5894–5907 *Nucleic Acids Research, 2009, Vol. 37, No. 17* *Published online 2 September 2009*
doi:10.1093/nar/gkp594

Two-dimensional combinatorial screening and the *RNA Privileged Space Predictor* program efficiently identify aminoglycoside–RNA hairpin loop interactions

Dustin J. Paul, Steven J. Seedhouse and Matthew D. Disney*

Department of Chemistry and Center of Excellence in Bioinformatics and Life Sciences, University at Buffalo, The State University of New York, 657 Natural Sciences Complex, Buffalo, NY 14260, USA

Received May 8, 2009; Revised June 27, 2009; Accepted June 29, 2009

ABSTRACT

Herein, we report the identification of RNA hairpin loops that bind derivatives of kanamycin A, tobramycin, neamine, and neomycin B via two-dimensional combinatorial screening, a method that screens chemical and RNA spaces simultaneously. An arrayed aminoglycoside library was probed for binding to a 6-nucleotide RNA hairpin loop library (4096 members). Members of the loop library that bound each aminoglycoside were excised from the array, amplified and sequenced. Sequences were analyzed with our newly developed *RNA Privileged Space Predictor* (RNA-PSP) program, which analyzes selected sequences to identify statistically significant trends. RNA-PSP identified the following unique trends: 5′UNNNC3′ loops for the kanamycin A derivative (where N is any nucleotide); 5′UNNC3′ loops for the tobramycin derivative; 5′UNC3′ loops for the neamine derivative; and 5′UNNG3′ loops for the neomycin B derivative. The affinities and selectivities of a subset of the ligand–hairpin loop interactions were determined. The selected interactions have K_d values ranging from 10 nM to 605 nM. Selectivities ranged from 0.4 to >200-fold. Interestingly, the results from RNA-PSP are able to qualitatively predict specificity based on overlap between the RNA sequences selected for the ligands. These studies expand the information available on small molecule–RNA motif interactions, which could be useful to design ligands targeting RNA.

INTRODUCTION

Because RNA folds into three-dimensional structures, it has many diverse and essential roles in biology beyond serving as a passive carrier of genetic information. For example, viral RNAs regulate viral replication (1) and microRNAs regulate gene expression (2). Their malfunction can cause diseases such as cancer (3). Riboswitches also regulate gene expression but do so in response to the concentration of metabolites in cells. Since many riboswitches have been identified in bacteria, they are also antibacterial drug targets (4,5). Additional functional roles for RNA are being identified through genomic sequencing efforts (6). These examples illustrate the interest in developing compounds that target RNA to either harness their therapeutic potential or to more thoroughly define RNA's cellular functions.

Despite this interest, there has been limited progress in finding compounds that target RNA with high affinity and specificity. This is best illustrated by the observation that high-throughput screening provides lower hit rates for RNA targets than their protein counterparts (7). Nonetheless, several ligands have been identified that bind RNA targets using these and other methods (8–12), and high throughput screening currently remains the most common approach used to identify ligands that bind RNA.

Identifying compounds that bind RNA targets could be hampered by low hit rates from high throughput screening because the wrong chemical space for the RNA or the wrong RNA space for the chemical library is being probed. At present, there is not enough information about the chemical space appropriate for RNA targeting endeavors and about the RNA scaffolds that are

*To whom correspondence should be addressed. Tel: +1 716 645 4242; Fax: +1 716 645 6963; Email: mddisney@buffalo.edu

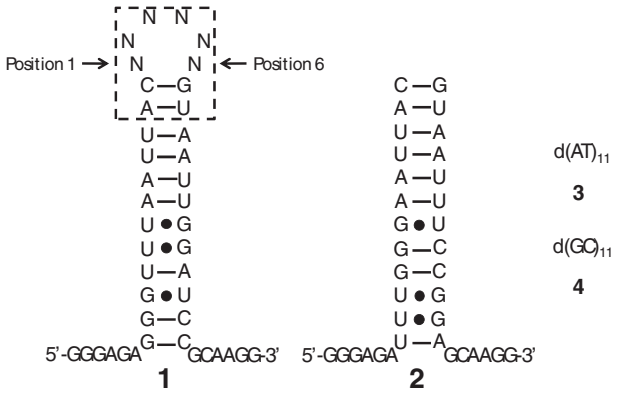

Figure 1. Secondary structures of the hairpin loop library (**1**, 4096 members) and competitor oligonucleotides **2–4**. The boxed nucleotides highlight the randomized region; for simplicity, only this region is shown for the selected hairpins (Figure 6). The competitor oligonucleotides **2–4** were used in 5000 times excess over **1** to ensure RNA–ligand interactions occur within the randomized region.

targetable in order to improve hit rates. If enough information of this type were available, it could serve as a platform to facilitate the rational and modular design of ligands targeting RNA, analogous to those used in polyamide targeting of DNA (13).

In order to gather information about RNA motif-small molecule interactions, we developed an approach termed two-dimensional combinatorial screening (2DCS) that probes RNA and chemical spaces in parallel (14). In 2DCS, a library of ligands is arrayed onto an agarose microarray surface and incubated with an RNA motif library (such as **1**, Figure 1) under conditions of high oligonucleotide stringency using competitor oligonucleotides (**2–4**, Figure 1). Once a ligand binds to members of the library, the bound RNA is excised from the array surface, amplified, cloned and sequenced. The output of 2DCS is a database of RNA motif-ligand partners. Previously, this approach was used to screen libraries of internal loops for binding aminoglycosides (14) and a 6-nucleotide hairpin loop library for binding 6′ acylated kanamycin and neamine derivatives (15). In our earlier report, the 6′ acylated kanamycin A and neamine derivatives were displayed in a manner that mimics modification of the aminoglycosides by 6′ aminoglycoside acetyltransferase (AAC) enzymes (16,17).

In this report, 2DCS was used to probe the interactions of a library of four arrayed aminoglycosides to a 6-nucleotide hairpin loop library. The aminoglycosides arrayed were: 6″-azido-kanamycin A (**5**), 6″-azido-tobramycin (**6**), 5-*O*-(2-azidoethyl) neamine (**7**) and 5″-azido neomycin B (**8**) (Figure 2A). In contrast to our earlier report on aminoglycoside–hairpin interactions, these aminoglycosides are displayed on the microarray in a manner that mimics their biological presentation to the bacterial rRNA A-site (18). That is, the positions in the aminoglycosides that were functionalized for immobilization do not make direct contacts with the A-site, at least for the kanamycin and tobramycin derivatives. In our previous studies, the aminoglycosides were immobilized through a 5-hexynoate linker that was installed by acylation of the aminoglycoside's 6′NH_2, which forms direct contacts with the A-site RNA.

In all previous studies using 2DCS, the RNA loop structure preferences were manually identified. To streamline trend identification, the *RNA Privileged Space Predictor* (RNA-PSP) computer program was developed to automate statistical analysis and is disclosed here for the first time. RNA-PSP determined that the following aminoglycoside–hairpin loop interactions have the highest statistical confidence: for **5**, 5′UNNNC3′ loops (where N represents any nucleotide, two-tailed p-value <0.00001); for **6**, 5′UNNC3′ loops (two-tailed p-value = 0.0006); for **7**, 5′UNC3′ loops (two-tailed p-value = 0.0001); and for **8**, 5′UNNG3′ loops (two-tailed p-value = 0.0071). By cross-analyzing the trends for all aminoglycosides, overlap between the RNA space identified for two or more ligands can be identified. If overlap is present, it qualitatively indicates that the RNAs would not be particularly specific for an aminoglycoside. In contrast, lack of overlap indicates that the RNAs could be specific for a particular aminoglycoside. These results contribute to the development of a database of RNA hairpin loop–ligand partners and further lay a foundation for the rational and modular design of small molecules that target RNA (19).

MATERIALS AND METHODS

Synthesis

All azido-aminoglycosides were synthesized from the corresponding parent aminoglycoside in the free base form according to previously published procedures (14,20,21).

*Syntheses of the dye-functionalized aminoglycosides (**5–8-TMR**).* Boc-protected azido-aminoglycosides were reacted with propargylamine via a Huisgen dipolar cycloaddition reaction (HDCR), a variant of 'click chemistry' (22), as previously described (14). The product of this reaction was dissolved in *N,N*-dimethylformamide (DMF) with 10% *N,N*-diisopropylethylamine (DIPEA) and reacted with 5-carboxytetramethylrhodamine, succinimidyl ester (5-TAMRA, SE, Invitrogen). Compounds were lyophilized and purified by preparative thin layer chromatography (TLC). Product bands were visualized with UV light and were excised from the plate. The product was extracted from the silica gel by vortexing in methanol for 10 min. Samples were lyophilized and deprotected by tumbling in 500 µl of a 1:1 trifluoroacetic acid (TFA): dichloromethane (DCM) mixture for 30 min at room temperature. Samples were then lyophilized and purified by HPLC. Complete synthetic details for each dye-labeled aminoglycoside are available in the Supplementary Data.

*Synthesis of the triazole-functionalized TAMRA, **TMR-triazole**.* An alkyne-functionalized rhodamine dye was synthesized by dissolving 5-TAMRA, SE (52.7 µg, 100 nmol) in DMF with 10% DIPEA and reacting it with propargylamine (12.8 µl, 200 nmol). After sonicating the reaction at room temperature for 3 h, it was lyophilized and the solid washed with water to remove unreacted

propargylamine. The product was dissolved in dimethylsulfoxide (DMSO) and reacted with 3-azidopropylamine (0.02 µl, 200 nmol) (23) for 3 h in the presence of 1 mM $CuSO_4$, 1 mM ascorbic acid and 100 µM tris(benzyltriazolylmethyl)amine (TBTA, dissolved in 4:1 butanol: DMSO) (24). The sample was lyophilized and purified by HPLC. Complete synthetic details are in the Supplementary Data.

Construction of azido-aminoglycoside microarrays

Azido-aminoglycosides were immobilized onto alkyne-funtionalized microarrays via a HDCR. Alkyne-functionalized microarray slides were constructed as previously described (20,21). Serial dilutions of the aminoglycoside derivatives were spotted in 10 mM tris(hydroxymethyl)aminomethane-hydrochloride (Tris–HCl) (pH 8.5), 100 µM TBTA, 1 mM $CuSO_4$, 1 mM ascorbic acid and 10% glycerol. The slides were placed into a humidity chamber for 3 h, washed with water and dried at room temperature.

General nucleic acids

All DNA oligonucleotides were purchased from Integrated DNA Technologies, Inc. and used without purification unless noted otherwise. The RNA competitor oligonucleotides were purchased from Dharmacon (Lafayette, Colorado) and deprotected according to the manufacturer's standard procedure. All aqueous solutions were made with diethylpyrocarbonate (DEPC)-treated NANOpure water.

RNA library and competitor oligonucleotides

The RNA 6-nucleotide hairpin loop library (**1**, Figure 1) (15,25) was synthesized by *in vitro* transcription from the corresponding DNA template using a Stratagene RNAMaxx High Yield transcription kit. This DNA template was synthesized using custom mixing at the randomized positions to ensure that there was equivalent representation of each nucleotide. Internally labeled hairpin loop library was synthesized and purified as previously described (15).

Competitor oligonucleotides were used in order to select interactions between the ligand and the randomized region of the hairpin RNA. Oligonucleotide **2** is a mimic of the stem; the sequence was changed so that it does not bind to RT–PCR primers but maintains similar nearest neighbors. Competitors **3** and **4** are DNA oligonucleotides.

RNA transcription and purification

Selected RNA hairpins were transcribed using a Stratagene RNAMaxx transcription kit according to the manufacturer's protocol using 10 µl of the amplified DNA from the PCR reaction of a synthetic DNA template. After transcription, 1 unit RQ1 RNase-free DNase I (Promega) was added, and the sample was incubated at 37°C for an additional 1 h. The transcribed RNAs were then purified by gel electrophoresis on a denaturing 15% polyacrylamide gel. The RNAs were visualized by UV-shadowing and extracted into 300 mM NaCl by tumbling overnight at 4°C. The resulting solution was concentrated with 2-butanol and ethanol precipitated. Concentration was determined by measuring the absorbance at 260 nm and by using the corresponding extinction coefficient. Extinction coefficients were determined using HyTher version 1.0 (Nicolas Peyret and John SantaLucia, Jr, Wayne State University) (26,27) and are based on the absorbances of nearest neighbors (28).

RNA selection

Azido-aminoglycoside microarrays were pre-equilibrated with 1× hybridization buffer (HB; 20 mM *N*-(2-hydroxyethyl)piperazine-*N*′-2-ethanesulfonic acid (Hepes), pH 7.5, 150 mM NaCl, 5 mM KCl, 1 mM $MgCl_2$) supplemented with 40 µg/ml BSA for 30 min at room temperature. The remainder of the experiment was completed as previously described (14,15) with 10 pmol of internally labeled hairpin loop library (**1**) and 50 nmoles of each competitor oligonucleotide (**2–4**, Figure 1).

RT–PCR amplification

RT–PCR reactions were completed as previously described (20). The RT–PCR product was then cloned into pGEM T Vector (Promega) according to the manufacturer's standard protocol. Sequencing was completed by Functional Biosciences (Madison, WI).

Determination of trends in selected hairpins

The program RNA-PSP (v. 1.0) was used to extract the selected sequences from the raw sequencing data file for both statistical analysis and RNA secondary structure prediction (29,30). The RNA-PSP program was developed to identify trends in selected sequences and determine their statistical significance. This was completed by comparing the occurrence of the trend within a selected mixture of sequences to the occurrence of that trend within the entire 6-nucleotide hairpin loop library (**1**, Figure 1). RNA-PSP ranks the most statistically significant trends by performing a *Z*-test using Equations (1) and (2) (31):

$$\Phi = \frac{n_1 p_1 + n_2 p_2}{n_1 + n_2} \quad \mathbf{1}$$

$$Z_{\text{obs}} = \frac{(p_1 - p_2)}{\sqrt{\Phi(1-\Phi)((1/n_1) + (1/n_2))}} \quad \mathbf{2}$$

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 trend, and p_2 is the observed proportion for Population 2 (entire library) displaying the trend. The *Z*-scores are then manually converted to the corresponding two-tailed *p*-value (Figure 4).

Fluorescence binding assays

Dissociation constants from direct assays were determined as previously described (14,20) using a Bio-Tek FLX-800 plate reader and 50 nM of the corresponding fluorescently labeled aminoglycoside (**5-TMR, 6-TMR, 7-TMR** or **8-TMR**, Figure 2A). The excitation and emission filters used were 530/25 and 590/35, respectively, and the

sensitivity was set to 70. An average of a 40% decrease in fluorescence intensity was observed. The resulting curves were fit to Equation 3 (32):

$$I = I_0 + 0.5\Delta\varepsilon([\mathrm{FL}]_0 - (([\mathrm{FL}]_0 + [\mathrm{RNA}]_0 K_t)^2 - 4[\mathrm{FL}]_0[\mathrm{RNA}]_0)^{0.5}) \quad \textbf{3}$$

where I is the observed fluorescence intensity, I_0 is the fluorescence intensity in the absence of RNA, $\Delta\varepsilon$ is the difference between the fluorescence intensity in the absence of RNA and in the presence of infinite RNA concentration and has units of M^{-1}, $[FL]_0$ is the concentration of the fluorescently labeled aminoglycoside, $[RNA]_0$ is the concentration of the selected hairpin loop(s) or control RNA, and K_t is the dissociation constant.

Control experiments were performed in the same manner for binding to **TMR-triazole**, which contains the dye and the triazole linkage but no aminoglycoside (Supplementary Data). No change in fluorescence is observed when up to 2 μM of the selected RNA hairpin mixture is added to 50 nM **TMR-triazole**. These results show that the change in fluorescence is due to binding of the aminoglycoside to the oligonucleotides and not the binding of the dye.

Competition experiments were also performed as previously described (14,15) using 1.1 μM, 0.4 μM, 1.0 μM and 0.15 μM of the library of RNAs selected for **5**, **6**, **7** and **8**, respectively. These are the concentrations at which the direct fluorescence binding curves reached saturation. The selected RNA libraries were incubated with 50 nM of **5-TMR, 6-TMR, 7-TMR** or **8-TMR**, and increasing concentrations of the free base form of the aminoglycoside. The expected increase in fluorescence was observed and the resulting curves were fit to Equation (4) (20):

$$\Theta = \frac{1}{2[\mathrm{FL}]_0}\left[K_t + \frac{K_t}{K_d}[C_t]_0 + [\mathrm{RNA}]_0 + [\mathrm{FL}]_0 - \sqrt{\left([K_t] + \frac{K_t}{K_d}[C_t]_0 + [\mathrm{RNA}]_0 + [\mathrm{FL}]_0\right)^2 - 4[\mathrm{FL}]_0[\mathrm{RNA}]_0}\right] + A \quad \textbf{4}$$

where Θ is the fraction of the fluorescently labeled aminoglycoside bound, K_t is the dissociation constant determined by direct binding assay, K_d is the dissociation constant of the competing unlabeled aminoglycoside, $[FL]_0$ is the total concentration of the fluorescently labeled aminoglycoside, $[C_t]_0$ is the total concentration of competing aminoglycoside, A is the fraction of fluorescently labeled aminoglycoside that is bound at infinite concentration of unlabeled competitor with values ranging between 0 and 0.1, and $[RNA]_0$ is the total concentration of the hairpin of interest.

RESULTS

Selection of aminoglycoside structures

We identified the hairpin loops that bind azido-aminoglycoside derivatives of kanamycin A (**5**), tobramycin (**6**), neamine (**7**) and neomycin B (**8**) (Figure 2A) simultaneously via microarray. These ligands are known to bind bacterial rRNA A-sites (33) and other RNAs (34). Thus, defining the RNA hairpin loops that bind these ligands could provide insights into the optimal RNAs that they should be used to target.

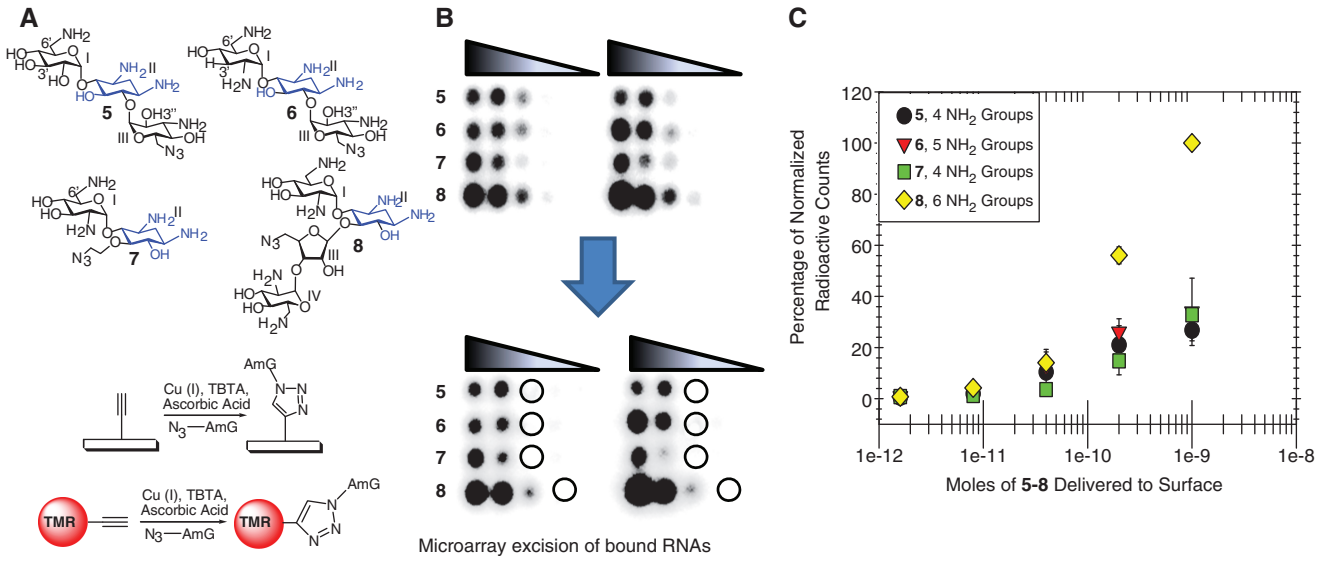

Figure 2. (**A**, top) Chemical structures of the azido-aminoglycosides used to study RNA–ligand interactions: **5** (kanamycin A derivative), **6** (tobramycin derivative), **7** (neamine derivative), **8** (neomycin B derivative). The 2-deoxystreptamine ring common to all four aminoglycosides is highlighted in blue. (**A**, bottom) Immobilization of **5–8** via 1, 3 Huisgen dipolar cycloaddition (14,21) on alkyne-displaying microarrays for 2DCS or conjugation to 5′-TAMRA (TMR, red ball) to study binding affinities. AmG refers to aminoglycoside. (**B**) Image of a microarray displaying compounds **5–8** that was hybridized with oligonucleotides from Figure 1 before (top) and after excision of the bound RNAs (bottom). Circles indicated the positions where the RNAs were excised. (**C**) Plot of the data for binding of ^{32}P-internally labeled **1** to array immobilized **5–8** in the presence of competitor oligonucleotides **2–4**. Plots were normalized to the highest signal for binding **8**.

Table 1. Binding affinities of the mixtures of selected RNAs and **1** & **2** to different aminoglycosides and their selectivities[a]

Aminoglycoside studied for binding	Aminoglycoside from which RNAs were harvested[b]				Other RNAs		
	5	**6**	**7**	**8**	**1**	**2**	A-site[c]
5-TMR	200 ± 46; —	130 ± 30; 0.7	> 2000[d]; > 10[e]	900 ± 157; 5	> 2000[d]	> 2500[d]	18 000
6-TMR	> 2000[d]; > 19[e]	110 ± 44; —	1020 ± 168; 10	105 ± 78; 1	> 1300[d]	> 2500[d]	1500
7-TMR	1000 ± 100; 3	130 ± 30; 0.4	350 ± 51; —	130 ± 67; 0.4	> 5000[d]	> 1200[d]	7800
8-TMR	680 ± 296; 34	100 ± 5; 5	1050 ± 264; 52	20 ± 4; —	> 1100[d]	> 2500[d]	19

All affinities are reported in nM.
[a]Selectivities (given after semicolon) were calculated by dividing the K_d for the other aminoglycoside by the K_d for the aminoglycoside for which the mixture was selected.
[b]These are the mixtures of RNAs that were selected to bind each aminoglycoside.
[c]Binding affinities of the bacterial rRNA A-site were measured using surface plasmon resonance (SPR) with the RNA immobilized on the SPR chip (39).
[d]These values are the lower limit of the dissociation constant because the binding curves did not reach saturation at these concentrations.
[e]These values represent the lower limit of specificity.

In order to implement 2DCS, a chemical handle on the ligands is required to allow precise immobilization onto a microarray surface. In the cases of **5–8**, an azide chemical handle was installed to anchor ligands onto alkyne-functionalized surfaces via a Huisgen dipolar cycloaddition reaction (HDCR) (35). The position of azido group functionalization in **5–8** was chosen based on the ease of modifying a primary hydroxyl group (21,34,36) and on the interactions of these ligands with mimics of the bacterial A-site. For example, crystal structures of a bacterial A-site mimic complexed with kanamycin A and tobramycin show that the 6″-OH, which is the position that was functionalized with our chemical handle, does not form direct contacts with the RNA (18,37). Neamine (**7**) and neomycin B (**8**) mimics were modified at the 5- and 5″-OH positions, respectively (18). The 5-OH of neamine forms an intramolecular hydrogen bond to the 2′NH_2 in a crystal structure with a mimic of the bacterial rRNA A-site. The 5″-OH in neomycin B forms an intramolecular hydrogen bond to the 2′NH_2 and an intermolecular hydrogen bond with N7 on G1491 (18). Thus, to varying degrees, the ligands represent therapeutically relevant forms of these drugs, at least for the A-site. The 6′-*N*-5-hexynoate kanamycin A and neamine ligands used in previous 2DCS studies with RNA hairpins did not as the 6′NH_2 forms direct contacts with the A-site (18).

2DCS of aminoglycoside–RNA hairpin loop interactions

Serial dilutions of **5–8** were spotted onto array surfaces and immobilized using a HDCR. For each aminoglycoside, five ligand loadings or concentrations were arrayed onto the surface (Figure 2B). The arrays were then probed for binding to the ^{32}P-internally labeled 6-nucleotide hairpin loop library **1** (Figure 1). The library contains 1536 predicted 4- and 2560 predicted 6-nucleotide hairpins, assuming that hairpins that can form canonical pairs at positions 1 and 6 (AU, GC, GU) are 4-nucleotide loops. To ensure that the interactions probed are to the random region in **1**, competitors **2–4** (Figure 1) were used in 5000 times excess over **1** and 5-fold excess over the total amount of ligand delivered to the array surface. Based on previous studies, ~10% of the ligand delivered (200 nl) at 5 mM concentration is actually immobilized (38). Since a 4096 member RNA library was probed for binding an array with five different loadings of four aminoglycosides, this 2DCS experiment probed 81 920 types of interactions in duplicate and in parallel on a single array surface.

The array was imaged and the amount of RNA bound to each ligand was normalized to the highest loading of the neomycin B derivative, **8** (Figure 2C). Not surprisingly, the aminoglycosides with the most amino groups, **6** and **8**, gave the highest signals. Signals above background were observed for delivery of as little as 20 pmol of ligand onto the surface. The lowest loading spot that gave signal above background for each aminoglycoside was excised from the array as indicated in Figure 2B. The RNAs at lower ligand loading were excised because previously it was shown that RNAs harvested from lower ligand loadings are higher affinity (20).

RNA mixtures are generally selective for the corresponding aminoglycoside

The binding affinities of the mixtures of RNA selected to bind each aminoglycoside were determined using a fluorescence-based assay with TAMRA-labeled derivatives of the aminoglycosides (**5-TMR, 6-TMR, 7-TMR** and **8-TMR**, Figure 2A) (20). Table 1 provides a summary of the binding affinities. The mixtures of RNAs selected by **5**, **6**, **7** and **8** bind to their TAMRA-labeled aminoglycosides with K_d values of 200, 110, 350 and 20 nM, respectively. In contrast, binding affinities of the entire library, **1**, to the TAMRA-labeled aminoglycosides are much weaker; the highest affinity interaction is between **1** and **8-TMR** with a K_d of > 1100 nM. Comparison of the affinities of the selected hairpin RNA mixtures to the affinities of a mimic of the bacterial rRNA A-site for the ligands shows that interactions are in general higher affinity for the selected RNAs than the A-site (39). The exception is neomycin B (**8**), which binds similarly to both the bacterial A-site mimic and the selected hairpins.

In order to determine if conjugation to TAMRA affects the binding affinity of each aminoglycoside derivative, two sets of experiments were completed—competition experiments with unlabeled aminoglycosides and direct binding

assays with a triazole-functionalized TAMRA, **TMR-triazole**. Competition assays were completed by titrating the corresponding parent aminoglycoside (which does not contain the azide handle) into solutions of the selected RNA library equilibrated with the fluorescently labeled aminoglycoside derivative. Results show that the TAMRA dye and triazole ring decrease affinity by about 5-fold for **5-TMR**, increase affinity about 3-fold for **6-TMR**, decrease affinity by about 5-fold for **7-TMR**, and has no effect on affinity for **8-TMR**. These values are consistent with a previous report that found that conjugation of TAMRA to a kanamycin and a neamine derivative affected binding affinity by about 4-fold (15). The affinity of the triazole-functionalized dye, **TMR-triazole**, for the selected mixtures of RNAs was also determined (Table 1). There was $\leq$5% change in fluorescence when up to 2 μM RNA is added to a solution of **TMR**-triazole, indicating that the aminoglycoside is required for binding.

The specificity of each hairpin mixture was determined by measuring the binding affinities for all of the arrayed ligands. Previous studies have shown that RNA internal loop–ligand interactions identified by 2DCS were selective (14). In these studies, the RNA hairpin loop–ligand interactions have varying levels of specificity (Table 1).

The hairpin mixture selected for **5** binds with a similar affinity to the tobramycin derivative, **6**. Structurally, **5** and **6** are similar in size and linkages between rings. However, **6** contains a hydrogen instead of a hydroxyl group at the 3′ position and an amino group instead of a hydroxyl group at the 2′ position. In contrast, the RNA mixture selected to bind **5** is >10-fold selective over the neamine derivative, **7**. This value represents a lower limit since no change in fluorescence was observed when up to 2 μM of the RNA mixture selected to bind **5** was added to **7-TMR**. The kanamycin hairpin mixture is also 5-fold selective over the neomycin derivative, **8**, despite the fact that **8** has two more amino groups than **5**. This suggests that binding of **5** to the hairpins is not purely electrostatic in nature.

Interestingly, the hairpin mixture selected to bind **6**, the tobramycin derivative, also displays varying levels of specificity that are different than the hairpin mixture selected to bind **5**. For example, the hairpins selected to bind **6** bind >19-fold more weakly to **5**, the kanamycin derivative. This is in contrast to the mixture selected to bind **5**, which binds to **5** and **6** similarly. The **6**-mixture is also about 10-fold more specific over the neamine derivative, **7**, while exhibiting no specificity over the neomycin derivative, **8**.

The RNAs selected to bind the neamine derivative, **7**, are the least specific. Although the mixture binds **7** 3-fold more tightly than **5**, the mixture binds ~3-fold more weakly to **6** and **8**. The RNA mixture selected to bind **8**, however, is the most selective. This mixture binds 34-fold more weakly to **5**, 52-fold more weakly to **7**, and 5-fold more weakly to **6**. Taken together, the binding data for the mixtures suggest overlapping RNA hairpin loop space for some of the aminoglycosides.

RNA-PSP

The RNA mixtures were then sequenced to identify the hairpin loops that bound **5–8**. The computer program RNA-PSP was developed to address the need for fast and accurate statistical analysis of selected RNAs. RNA-PSP, developed on a Microsoft Visual Basic 2008 platform, allows direct input of sequence files from any selection. The inputted sequence file is then analyzed to extract the sequences of the variable region for each selected library member. For the automated extraction of selected sequences, users specify the constant and the variable regions of the library, allowing RNA-PSP to sort through a sequencing file and identify embedded RNAs from the selection. An image of the user interface and a representative schematic of the fundamental algorithm are shown in Figures 3 and 4, respectively.

Once the selected structures are extracted by the program, it generates all possible combinations of sequences from the original library and stores the results. For example, in the 6-nucleotide hairpin library **1**, there are 4096 possible hairpins. The representation of a trend in selected hairpins is then compared to the representation of that trend in all library members. *Z*-scores [Equations (1) and (2)] are computed and converted into two-tailed *p*-values.

It is critical that *Z*-scores be determined for all trends within the selected sequence mixtures in order to define privileged RNA space. The corresponding two-tailed *p*-value quantitatively assesses whether the trend truly represents privileged RNA space. That is, if the following null hypothesis is invoked: 'Ligands tested are completely unbiased for the RNAs which they bind', then we expect to find a selected mixture of RNAs that precisely represents the entire library. The two-tailed *p*-value describes the probability that the actual sequence results obtained can occur if this hypothesis were true. For example, for trends with two-tailed *p*-values <0.05, we can reject the null hypothesis with >95% confidence. In the context of 2DCS, two-tailed *p*-values represent a direct assessment of the likelihood that an identified trend is due to a real preference for the ligand to bind RNAs that contain it.

As an initial test of the accuracy with which statistical analysis can be completed using RNA-PSP, the selected sequences from a previous 2DCS selection to identify the hairpins that bind 6′-*N*-5-hexynoate neamine (15) were uploaded into RNA-PSP and analyzed. The most significant trend identified by RNA-PSP was 5′GC3′ or 5′CG3′ steps, with a two-tailed *p*-value <0.0001. Comparison of this result to those determined by manual analysis and computation in the previous study showed that these two analyses are in agreement. Thus, RNA-PSP produced complete and accurate results from automated statistical analysis in seconds compared to the days required for manual computation. This program is available for download free of charge from the Disney Lab web site (http://www.nsm.buffalo.edu/Research/rna/).

The RNA-PSP program determined the following trends in hairpin loops to have the highest statistical confidence for binding each aminoglycoside: for **5**,

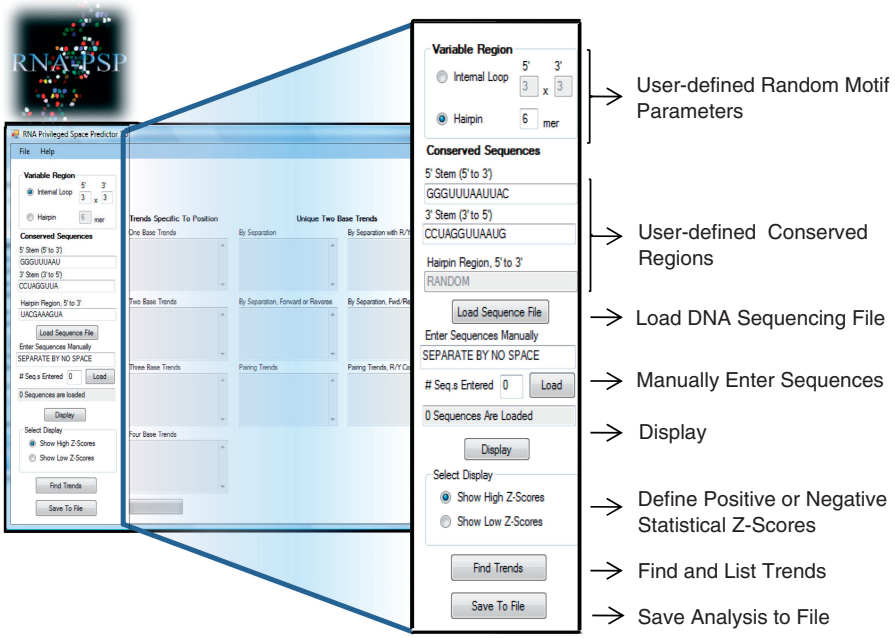

Figure 3. User interface for RNA-PSP. The control panel on the left-hand side of the screen allows control over search parameters including: (i) ability to define the random motif region by size and type; (ii) ability to define conserved regions in RNAs selected, providing a guide for automated searching and extracting of all random motif sequences from a file; (iii) automated or manual input of random motif sequences; (iv) ability to display both positive and negative *Z*-scores; and (v) ability to save the complete analysis to a text file.

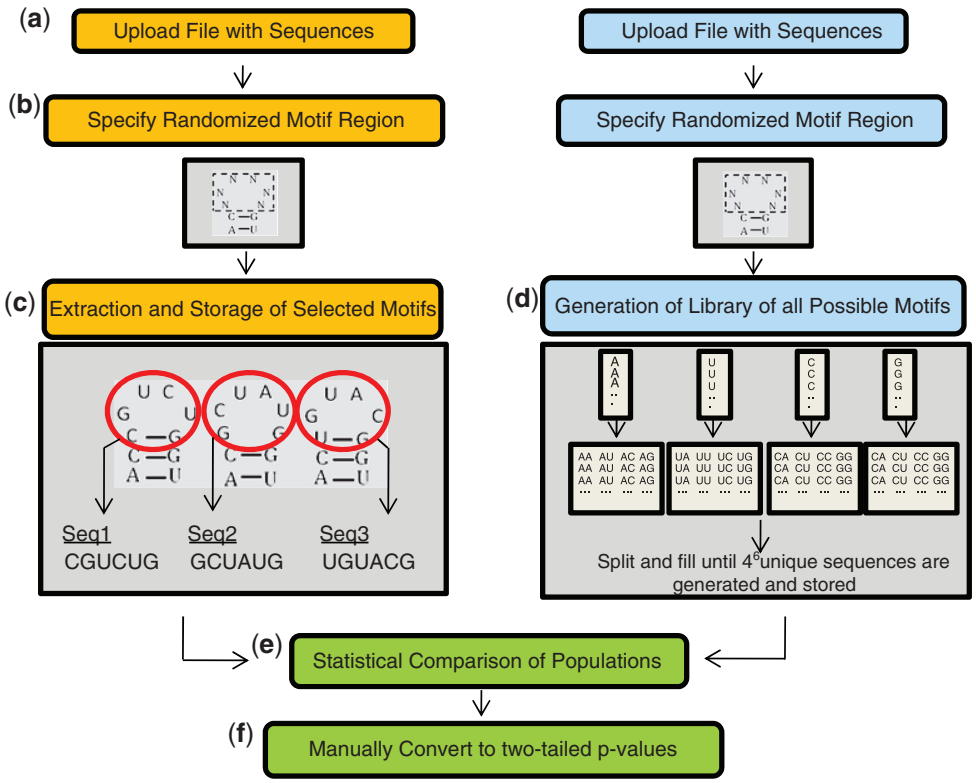

Figure 4. RNA-PSP program used to find and rank trends found within selected mixtures of sequences. (**a**) File containing complete DNA sequences with embedded selected hairpins is uploaded into RNA-PSP; (**b**) The randomized motif is identified within the RNA sequence; (**c**) Each random motif is located in the sequencing file, extracted, and stored as a member of Population 1. Each sequence is listed as a combination of A/U/G/C in order 5′ to 3′ in the randomized sequence (e.g. CUGGCA); (**d**) The entire library of all possible motifs is generated and stored as Population 2; (**e**) Populations are compared, and statistical *Z*-scores are calculated for all trends. Trends are then sorted and ranked according to *Z*-score; (**f**) *Z*-scores are manually converted to two-tailed *p*-values and statistically significant trends with $>95\%$ confidence level are displayed.

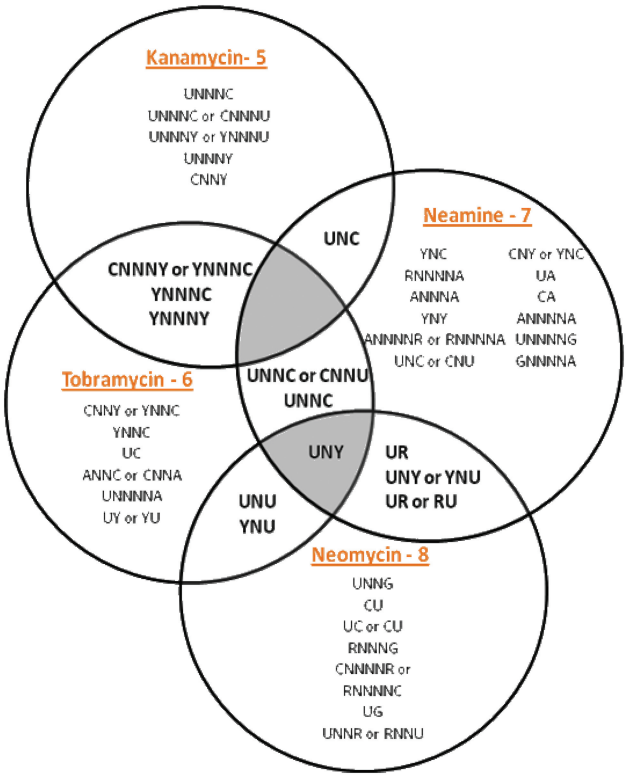

Figure 5. Venn diagram of the statistically significant trends indentified in the RNA sequence space selected to bind the four aminoglycoside derivatives. Overlapping trends are shown in bold. The most statistically signifcant trend for the kanamycin A derivative, **5**, is 5′UNNNC3′; for the tobramycin derivative, **6**, 5′UNNC3′; for the neamine derivative, **7**, 5′UNC3′; and for the neomcyin B derivative, **8**, 5′UNNG3′.

5′UNNNC3′ loops (two-tailed p-value <0.0001); for **6**, 5′UNNC3′ loops (two-tailed p-value = 0.0006); and for **7**, 5′UNC3′ loops (two-tailed p-value = 0.0001). The second highest statistical confidence was 5′UNNG3′ loops (two-tailed p-value = 0.0071) for **8**. In addition to this sequence analysis, the secondary structures of each selected RNA hairpin loop were predicted using the *RNAstructure* program (29,30). Coupling structure prediction with statistical analysis showed that there is no preference for predicted 4- or 6-nucleotide hairpin loops for any aminoglycoside. A list of all selected hairpins for each aminoglycoside and more details of the statistical analysis are available in the Supplementary Data.

RNA-PSP identifies overlap in RNA space that correlates with specificity

A cross-analysis was also completed to determine if there was overlap between statistically significant trends for all aminoglycosides (Figure 5). This analysis revealed that the trend 5′UNY3′ was statistically significant for three of the aminoglycosides: **6**, two-tailed p-value = 0.0324; **7**, two-tailed p-value = 0.0008; and **8**, two-tailed p-value = 0.0028. This trend is not significant for **5** (two-tailed p-value = 0.1425). There are additional trends that overlap as depicted in Figure 5. Two-tailed p-values for all statistically significant trends are available in the Supplementary Data. Comparison of the RNA motifs that bind compounds **5–7** show that they prefer hairpin loops that have a U 5′ to a C, but the numbers of bases separating the two nucleotides differ. Unique trends were identified by comparing the statistically significant trends for a selected aminoglycoside mixture with the statistically significant trends identified for the other aminoglycosides. RNA-PSP identified the following trends that were the most statistically significant and unique for binding **6**, **7** and **8**: for **6**, 5′CNNA3′ and 5′ANNC3′ loops (two-tailed p-value = 0.0444); for **7**, 5′RNNNNA3′ loops (where R is a purine; two-tailed p-value = 0.0285); for **8**, 5′UNNG3′ loops (two-tailed p-value = 0.0071). Although 5′UNNNC3′ is unique to **5**, there are many similar trends in **6**, for example YNNNY (two-tailed p-value = 0.0151) and YNNNC (two-tailed p-value = 0.0160).

Affinities of selected hairpin loop–aminoglycoside partners

Dissociation constants were measured for the individual loop–ligand partners that were predicted to bind tightly and specifically according to the statistical data. This prediction was based on the identification of loops displaying the most statistically significant trend that was also unique to the aminoglycoside of interest. In general, the K_d values observed for the individual hairpins selected for each aminoglycoside are consistent with those measured for the mixtures (Figure 6 and Tables 1 and 2).

For the kanamycin A derivative, **5**, binding affinities were determined for six hairpins that display the most statistically significant trend, 5′UNNNC3′ (two-tailed p-value < 0.0001). Dissociation constants range from 140 to 605 nM (Figure 6), in good agreement with the dissociation constant of 200 nM for the mixture. Four of the hairpin loops, Kan HP1, Kan HP3, Kan HP4 and Kan HP6 have similar affinities and bind mostly tightly to **5**. They also share sequence similarities. For example, in addition to displaying the 5′UNNNC3′ consensus trend, all have U as the first nucleotide in the hairpin loop (not necessarily the first nucleotide of the sequence). The loop sequences of Kan HP1 and Kan HP4 are even more similar: both have the sequence 5′UGCRCY3′ (where Y is a pyrimidine). The two loops with the lowest affinities, Kan HP2 and Kan HP5, have the 5′U of the 5′UNNNC3′ consensus trend as the second nucleotide in predicted 6-nucleotide hairpin loops. Binding of **5-TMR** to Kan HP2 and Kan HP5 is about 4-fold weaker than the hairpins that have a U as the first nucleotide in the loop.

Careful examination of Kan HP3, Kan HP4 and Kan HP6 revealed that they have the potential to form duplexes or kissing complexes (40). In order to determine if kissing complexes were formed under the conditions used in 2DCS and binding affinity measurements, optical melting experiments were completed. Previous studies have shown that optical melts of kissing complexes have two transitions. The transition at the lower temperature is concentration dependent and represents the dissociation

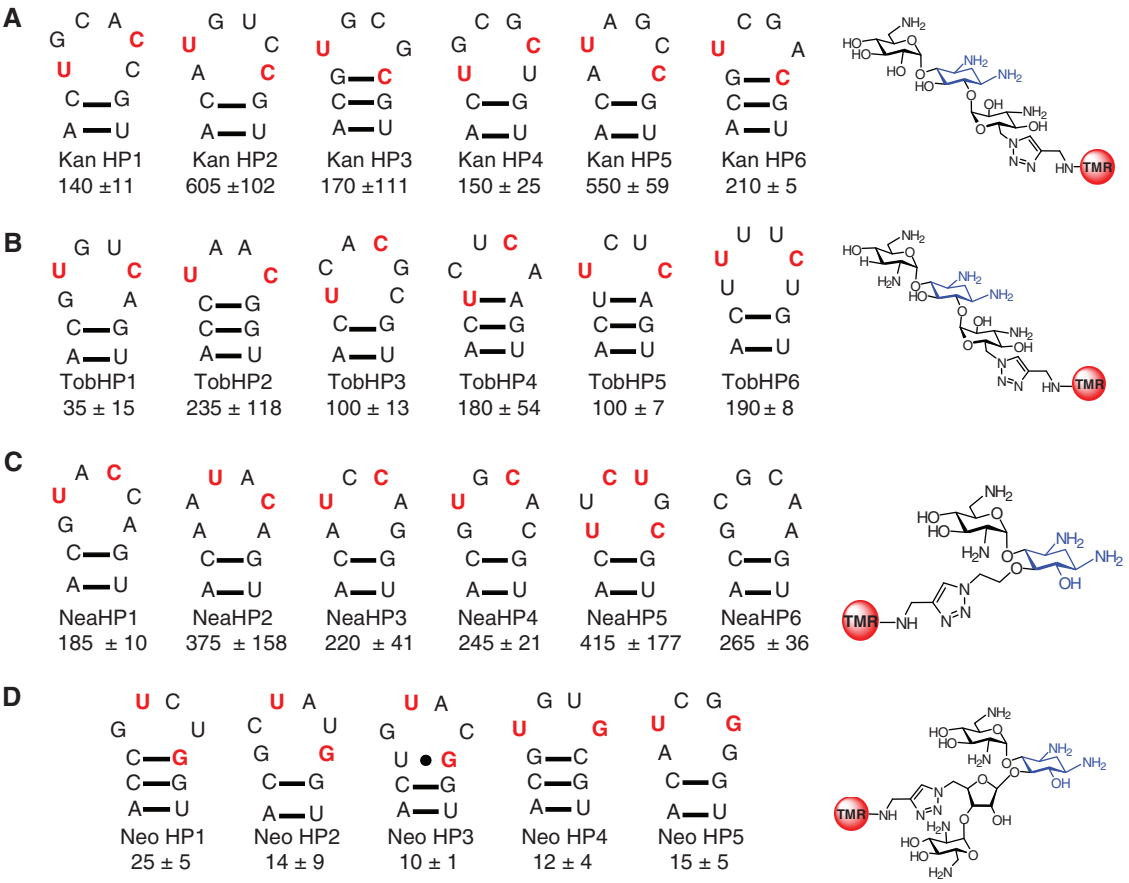

Figure 6. The secondary structures of a subset of the RNA hairpin loops that were selected to bind **5**, **6**, **7** and **8**. The nucleotides shown are derived from the boxed region in **1** (Figure 1). The affinities (in nM) for the respective aminoglycoside are shown below the hairpin structure. Statistical analysis shows that kanamycin A binds 5′UNNNC3′ loops (**A**), tobramycin binds 5′UNNC3′ loops (**B**), neamine binds 5′UNC3′ loops (**C**), and neomycin B binds 5′UNNG3′ loops (**D**). Nucleotides highlighted in red have the statistically most significant trend for that aminoglycoside.

Table 2. Binding affinities and selectivities of individual hairpin loops selected from **5**, **6**, **7** and **8**

	Dissociation constants (nM); selectivity[a]			
Hairpin loop	**5-TMR**	**6-TMR**	**7-TMR**	**8-TMR**
Selected from **5**				
Kan HP1	140 ± 11; —	210 ± 10; 1.5	> 1000[b]; > 7[c]	> 1000[b]; > 7[c]
Kan HP3	170 ± 111; —	275 ± 35; 1.6	790 ± 59; 5	100 ± 19; 0.6
Kan HP6	210 ± 5; —	130 ± 16; 0.6	> 1300[b]; > 7[c]	190 ± 57; 1
Selected from **6**				
Tob HP1	250 ± 102; 8	35 ± 15; —	635 ± 200; 19	> 2000[b]; > 59[c]
Tob HP2	> 2300[b]; > 10[c]	235 ± 118; —	> 2300[b]; > 10[c]	> 1500[b]; > 6[c]
Tob HP3	> 2000[b]; > 18[c]	100 ± 13; —	> 1500[b]; > 15[c]	200 ± 69; 2
Selected from **7**				
Nea HP1	400 ± 182; 2	260 ± 67; 1	185 ± 10; —	205 ± 27; 1
Nea HP2	220 ± 148; 1	140 ± 26; 0.4	375 ± 158; —	540 ± 127; 1.4
Nea HP6	390 ± 134; 1	110 ± 27; 0.4	265 ± 36; —	> 2000[b]; > 8[c]
Selected from **8**				
Neo HP3	> 2000[b]; > 200[c]	64 ± 15; 6	> 2000[b]; > 200[c]	10 ± 1; —
Neo HP4	> 700[b]; > 57[c]	180 ± 52; 15	> 900[b]; > 72[c]	12 ± 4; —
Neo HP5	> 1000[b]; > 67[c]	170 ± 61; 11	> 1200[b]; > 80[c]	15 ± 5; —

[a]Selectivities (given after semicolon) were calculated by dividing the K_d for the other aminoglycoside by the K_d for the aminoglycoside for which the mixture was selected.
[b]These values are the lower limit of the dissociation constant because the binding curves did not reach saturation at these concentrations.
[c]These values represent the lower limit of specificity.

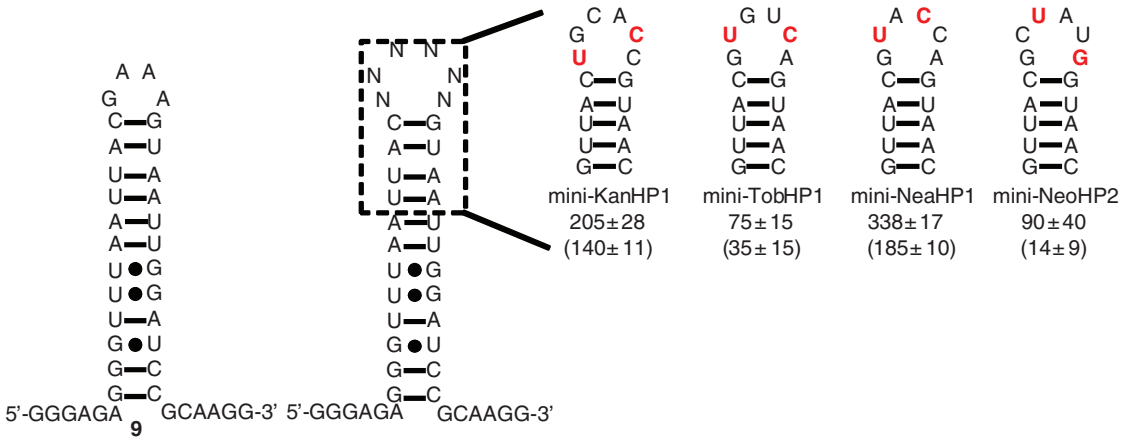

Figure 7. Secondary structures of the hairpin loops tested to determine if cassette nucleotides contribute to binding affinity. **9** is a mimic of the cassette in the hairpin loop library (**1**, Figure 1) with a GAAA tetraloop closed by a CG pair. The dissociation constants (nM) to the minimized hairpin loop are given below the secondary structure of the hairpin. The dissociation constants (nM) to the original cassette are shown in parentheses.

of the hairpin dimers. The transition at higher temperature is the melting of the hairpin and is concentration independent (40). For both hairpins tested (Kan HP3 and Kan HP4), there is only a single transition at all concentrations tested (0.5–8 μM, a 16-fold concentration range). The melting temperatures are also independent of concentration. Taken together, these results indicate that Kan HP3 and Kan HP4 fold into hairpin structures that do not form kissing complexes under conditions used for selections or determination of binding affinity. It should be noted that the concentration of **1** used in the 2DCS selection is 25 nM, or 6 pM of each unique hairpin. Data and experimental procedures for optical melting experiments are available in the Supplementary Data.

The most statistically significant trend identified for the hairpins that bind **6** is 5′UNNC3′ (two-tailed p-value = 0.0006). Therefore, six loops with this trend were further studied (Figure 6). As was also the case for the RNAs selected to bind **5**, the range of dissociation constants for the individual hairpins selected to bind **6** (35–235 nM) is in good agreement with the dissociation constant for the mixture (110 nM). Both predicted 4- and 6-nucleotide hairpins were identified (Figure 6), although there is no statistically significant preference for either when compared to the percentages of 4- and 6-nucleotide hairpins in the entire library. The sequence for Tob HP1 appeared twice in the sequencing data. Interestingly, it is also the highest affinity loop with a K_d of 35 nM. The other hairpins selected to bind **6-TMR** are 3- to 7-fold lower affinity than Tob HP1. In addition to the most statistically significant trend for **6**, 5′UNNC3′, some of the individual hairpins also contained the most statistically significant trend that is also unique for **6**, or 5′CNNA3′ and 5′ANNC3′ loops. These include Tob HP2, Tob HP3, Tob HP4 and Tob HP5.

Consistent with the data collected for **5** and **6**, the range of binding affinities for the individual hairpins selected to bind **7** (185–415 nM) mirrors the dissociation constant for the mixture (350 nM). The highest affinity loop for **7**, Nea HP1 (Figure 6), displays both the most statistically significant trend, 5′UNC3′, and the most statistically significant unique trend, 5′RNNNNA3′. This loop appeared twice in the sequencing data and has a K_d of 185 nM. There are several other hairpin loops, however, that have similar affinities including Nea HP3, Nea HP4, and Nea HP6. This is in contrast to the results with **6** in which the loop that appeared twice in the sequencing data was the highest affinity hairpin selected by 3-fold. Nea HP1, Nea HP3, and Nea HP4 display both the 5′UNC3′ and 5′RNNNNA3′ trend perhaps suggesting that both are required for binding. However, Nea HP6, which also binds with similar affinity, only contains the 5′RNNNNA3′trend.

The affinities of some of the individual hairpins selected to bind the neomycin B derivative, **8**, were also determined. All five hairpins display the most statistically significant unique trend, 5′UNNG3′ (Figure 6). Consistent with the mixture, the affinities of the loops selected to bind **8** are higher than the other hairpin–aminoglycoside interactions. The exception is Tob HP1, which has a similar affinity for **6**. As was the case for the other aminoglycosides, some of the sequences form predicted 4-nucleotide hairpins. However, there is no correlation between predicted size and affinity as expected from the statistical analysis. For the 4-nucleotide loops, there is also no difference in affinity between hairpins in which the 5′UNNG comprises only loop nucleotides and hairpins in which the 5′U or 3′G is predicted to form a base pair.

For comparison, binding affinities were determined for a hairpin loop not selected to bind any of the arrayed aminoglycosides (**9**; Figure 7). Hairpin loop **9** contains the same stem as the hairpin loop library (**1**) with a GAAA tetraloop closed by a CG base pair. As expected, **9** binds much more weakly than the selected hairpins to all of the aminoglycosides. No change in fluorescence was observed when up to 2.5 μM of **9** was added to compounds **5-TMR**, **6-TMR** and **8-TMR**. **7-TMR** did not reach saturable binding up to 2.5 μM of **9**, but the fluorescence did change by ~20%.

To investigate if cassette nucleotides contribute to the binding affinities of **5–8-TMR**, minimized constructs were designed in which half of the stem and the dangling ends were deleted (Figure 7). Optical melting experiments were completed to confirm that mini-Kan HP1, mini-Tob HP1, mini-Nea-HP1 and mini-Neo HP2 form hairpins. There is only a single transition at all concentrations tested (0.5–56 μM, 112-fold concentration range), and the melting temperatures of the four hairpin loops are independent of concentration. The dissociation constants for all minimized constructs are similar to the dissociation constants for the hairpins displayed in the original cassette (**1**) (Figure 7). The dissociation constant for mini-Neo HP2 had the largest difference from its parent hairpin with an ~5-fold higher dissociation constant. It has been shown in previous studies that cassette nucleotides do not significantly contribute to binding (14,15). Thus, the RNA motif–ligand pairs of this 2DCS are portable to other larger RNAs containing similar hairpins.

Selectivity of individual RNA hairpin–aminoglycoside interactions

As with the mixtures of selected structures, the binding affinities of some of the individual sequences for all four arrayed aminoglycosides were measured to determine specificity. The results are summarized in Table 2. In general, the selectivities of the individual loops reflect the specificities of the corresponding mixtures (Table 1).

For the hairpins selected to bind **5**, all of the hairpins bind **5** (the kanamycin A derivative) and **6** (the tobramycin derivative) similarly, which was also observed for the mixtures. In contrast, the hairpins selected to bind **5** discriminate against **7**. For example, Kan HP1 and Kan HP6 bind at least 7-fold more tightly to **5** than the neamine derivative, and Kan HP3 binds 5-fold more tightly. Only one of the three hairpins used to gain insight into selectivity, Kan HP1, binds to **5** more tightly than **8**. This is somewhat surprising since the mixture of hairpins selected for **5** bind 5-fold more weakly to **8**. However, the lack of specificity of Kan HP3 can be explained by an overlap in sequence space. It contains the trend for both **5** (5′UNNNC′) and **8** (5′UNNG3′). Kan HP3 is a member of the ultrastable UNCG tetraloop family, and binds all four aminoglycoside derivatives in this study with sub-micromolar affinities. The dissociation constants for **5**, **6**, **7** and **8** are 170, 275, 790 and 100 nM, respectively.

The individual hairpins that bind **6** are selective, even more so than the corresponding mixture. Tob HP1 binds **6** 8-fold more tightly than **5**; 19-fold more tightly than **7**; and >59-fold more tightly than **8**. Tob HP2 is also selective for **6** by >10-fold over **5** and **7** and by >6-fold over **8**. Tob HP3 shows at least 15-fold selectivity over **5** and **7** but only binds 2-fold more tightly to **6** than **8**. Taken together, the hairpins selected to bind **6** bind with high affinity and specificity over both **5** and **7**. The selectivity of the hairpins for **6** over **8** (the neomycin derivative) is variable, ranging from 2- to >59-fold. However, there was no difference in the affinity of the mixture selected for the tobramycin derivative between **6** and **8**. This suggests overlap between the RNA space that both ligands prefer that was present in the mixture but not in the individuals hairpins. Thus, determining the most statistically significant and unique trend in the sequencing data allows for computation to identify the most selective RNA motif–ligand interactions.

As was the case for the mixture of RNAs selected to bind **7**, the individual hairpins showed very little specificity for binding. Nea HP1, Nea HP2 and Nea HP6 bind all four aminoglycosides with the same affinity with one exception. Although Nea HP6 has the same approximate affinity for **5**, **6**, and **7**, it binds much more weakly to **8**, by at least 8-fold. This is likely because Nea HP6 only contains a trend unique to **7**, 5′RNNNNA3′ (Figure 5).

The hairpins selected to bind **8** overall have the highest levels of selectivity, ranging from 6-fold to >200-fold. The lowest specificities observed were between the **8**-selected hairpins and **6**, ranging from 6- to 15-fold. The hairpins discriminate between **5** and **7** by at least 57-fold. Neo HP3 is able to distinguish between **8** and both **5** and **7** by >200-fold. A similar trend is observed for Neo HP4 and Neo HP5 which bind >57-fold more weakly to **5** and >72-fold more weakly to **7**. The selectivities of the individual hairpins for **8** over **5** and **7** are similar to the selectivity of hairpin mixture (Table 1). However, the selectivity of the individual **8**-selected hairpins over the tobramycin derivative (**6**) is better than the selectivity observed for the mixture. The mixture binds **8** only ~5-fold more tightly than **6** while Neo HP4 and Neo HP5 bind at least 15- and 11-fold more tightly, respectively. Again, this can be explained by the lack of overlap of the individual sequences: Neo HP4 and Neo HP5 contain the trend unique for **8** (5′UNNG3′) and do not contain the most statistically significant trend for **7** (5′UNNC3′). Although 5′UNNC3′ is a statistically significant trend for **8**, it is not the most statistically significant one.

DISCUSSION

Comparison to previous aminoglycoside–RNA studies

Previous studies of aminoglycoside–RNA interactions have investigated validated RNA targets such as the prokaryotic A-site RNA (41–44), HIV trans-activation response element (TAR) RNA (45), rev-responsive element (RRE) RNA (46–48), and group I introns (49,50). Despite these efforts to understand RNA–aminoglycoside interactions, no general sequence preferences were identified. In order to identify optimal sites within a large RNA for binding aminoglycosides and aminoglycoside derivatives, we have probed an aminoglycoside library for binding libraries of small RNAs that contain internal loops and hairpin loops (14,15,20).

In this study, RNA space and chemical space were probed simultaneously to identify aminoglycoside derivative–RNA hairpin loop partners. In general, the hairpin loops and internal loops selected to bind **5** are similar in sequence and affinity. The internal loops selected to bind **5** are pyrimidine-rich (two-tailed *p*-value of 0.0373). The consensus hairpin loop sequence for **5** is 5′UNNNC3′. The internal loop mixture selected for **5** had a dissociation constant of 75 nM while the individual loops had K_d

values between 64 and 250 nM (14). However, the internal loops are more selective than the hairpins. Selectivities of the internal loops ranged from 4- to 35-fold with an average selectivity of 14-fold. A selection of RNAs that bind kanamycin completed by Goertx *et al*. also contains the hairpin consensus sequence selected for binding **5** (51).

Hairpin loops have also been selected to bind a different kanamycin derivative, 6′-*N*-5-hexynoate kanamycin A (15). Interestingly, the hairpins selected for this derivative bind with similar affinities but differ in the consensus trend—an adenine in position 1 separated by at least 2 nucleotides from a cytosine (two-tailed *p*-value of 0.0010). Thus, the site of immobilization changes the hairpin loop preference for kanamycin derivatives.

The most statistically significant trends identified for **5** and **6** differ by only one nucleotide: 5′UNNNC3′ and 5′UNNC3′, respectively. Despite the similarity, the mixture for **6** binds >19-fold more weakly to **5**. Therefore, **5** can accommodate the 5′UNNC3′ trend identified for **6**, but **6** cannot accommodate the 5′UNNNC3′ trend identified for **5**. This could be due to another trend also present in hairpins that bind **6**: 5′CNNA3′ and 5′ANNC3′.

A selection of internal loops that bind the tobramycin derivative **6** was also completed by 2DCS. The trend identified for the internal loops was different from the trend for hairpins—guanine across from guanine (two-tailed *p*-value = 0.0109). The mixture of selected internal loops binds **6** similarly with a dissociation constant of 50 nM. As in the case of the hairpin loops reported herein, the internal loops that bind **6** are most selective over **5** and **7** (at least 11-fold) and less selective over **8** (5-fold). Tobramycin aptamers have been identified and contain the most statistically significant trend identified in these studies (52–54). NMR studies of a tobramycin–RNA aptamer containing 5′UNNC3′ showed strong NOEs between all the rings of the tobramycin aminoglycoside and the uracil and cytosine nucleotides present in the hairpin loop (53).

The affinities for the internal loops and hairpins selected to bind **7** by 2DCS are similar—350 and 200 nM, respectively (14). A trend was identified for **7** and hairpin loops, 5′UNC3′; however, a trend for internal loops could not be identified. Many of the internal loops contain an adenine across from a guanine although it was not statistically significant. Interestingly, another statistically significant trend for the hairpins that bind **7** is 5′RNNNNA3′. Many of the hairpins selected to bind **7** contain a guanine in position 1 and an adenine in position 6 (Figures 1 and 6).

Hairpin loops that bind another neamine derivative, 6′-*N*-5-hexynoate neamine, were identified in a previous study; the privileged hairpin loops contained both 5′GC and 5′CG steps (15). In this study, Nea HP6 also contains this trend and binds with similar affinity to **7** (Figure 6) as the hairpins identified to bind 6′-*N*-5-hexynoate neamine. It does not, however, contain the 5′UNC3′ trend identified by RNA-PSP. Nea HP6 was also the most selective for **7** as it binds 2-fold more weakly to **6** and >8-fold more weakly to **8** (Table 2). This is perhaps not surprising based on the overlap in sequence space of 5′UNY3′ but not 5′RNNNA3′ for **7** and **8** (Figures 5 and 6). As observed with **5**, the site of functionalization, and hence immobilization, tunes the privileged sequence space for an aminoglycoside derivative.

The hairpin loops selected by **8** bind more tightly than the internal loops from our previous 2DCS experiment by ~8-fold (14). The privileged hairpin loops have the sequence 5′UNNG3′ while the privileged internal loops have a guanine across from adenine (two-tailed *p*-value = 0.0019). In general, the individual hairpin loops and internal loops are specific for **8** with the least specific interaction for a hairpin due to overlap with **6** and the least specific interaction for an internal loop due to overlap with **7**.

The structure of a neomycin B aptamer has been solved (55) and has a similar affinity as the hairpin loops reported herein. The aptamer forms a GAGA hairpin closed by a GC pair. Adjacent to the loop closing pair in the stem are three consecutive GU pairs. Neomycin B forms contacts with the GU pairs, the GC closing pair, and guanine and adenine residues in the hairpin loop. The hairpin loop from this study that is most similar to such an arrangement is Neo HP5.

The data reported herein and from our previous studies (15,20,56) suggest that it is likely that many loops present in biological RNAs should bind aminoglycosides with high affinity. In fact, they should bind preferentially over the A-site, their known therapeutic target (33). However, the A-site is not the only binding site for aminoglycosides even in the ribosome. It has been shown that binding of aminoglycosides to RNA helix 69 (H69) in the large subunit prevents ribosome recycling (57). Side effects associated with the clinical use of aminoglycoside antibiotics (58,59) also point to binding of other RNAs or biomolecules. It is likely that the ribosomal A-site is the most occupied target for aminoglycosides *in vivo* for many reasons: (i) the relative abundance of rRNAs compared to pre-mRNAs and other non-coding RNAs (60); (ii) the slower turnover rate of rRNAs compared to other RNAs (61); and (iii) the potential inaccessibility of some loops due to formation of tertiary contacts or interactions with protein.

CONCLUSIONS

These results have identified the privileged RNA hairpins loops that bind derivatives of four aminoglycosides (Figure 6). The RNA–ligand interactions bind with nanomolar affinities and are generally selective for the corresponding aminoglycoside. The combination of 2DCS and efficient statistical analysis by RNA-PSP streamlines the identification of RNA–ligand partners. The development of RNA-PSP has been essential for analysis of sequencing data to identify unique trends in the loops selected to bind each aminoglycoside. More importantly, it identified trends that overlap between aminoglycosides in order to predict or understand selectivity (Figure 5). RNA-PSP can be used to analyze any nucleic acid selection and is available free of charge at the Disney Lab web site (http://www.nsm.buffalo.edu/Research/rna/). These studies have the potential to elucidate new RNA targets for

aminoglycosides and further accelerate the rational design of drugs that target RNA.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

The authors thank Jessica Childs-Disney in the Department of Chemistry & Biochemistry at Canisius College for helpful suggestions, discussions and critical review. They also thank Alexei Pushechnikov for helpful advice about the syntheses of aminoglycoside derivatives.

FUNDING

University at Buffalo; the NYS Center of Excellence and Bioinformatics and Life Sciences; and the National Institutes of Health [Grant R01-GM079235]. Funding for open access charge: National Institutes of Health.

Conflict of interest statement. None declared.

REFERENCES

1. Vilfan,I.D., Candelli,A., Hage,S., Aalto,A.P., Poranen,M.M., Bamford,D.H. and Dekker,N.H. (2008) Reinitiated viral RNA-dependent RNA polymerase resumes replication at a reduced rate. *Nucleic Acids Res*., **36**, 7059–7067.
2. Matzke,M.A. and Birchler,J.A. (2005) RNAi-mediated pathways in the nucleus. *Nat. Rev. Genet*., **6**, 24–35.
3. Negrini,M., Ferracin,M., Sabbioni,S. and Croce,C.M. (2007) MicroRNAs in human cancer: from research to therapy. *J. Cell Sci*., **120**, 1833–1840.
4. Winkler,W.C. and Breaker,R.R. (2003) Genetic control by metabolite-binding riboswitches. *Chembiochem*, **4**, 1024–1032.
5. Blount,K.F., Wang,J.X., Lim,J., Sudarsan,N. and Breaker,R.R. (2007) Antibacterial lysine analogs that target lysine riboswitches. *Nat. Chem. Biol*., **3**, 44–49.
6. Guttman,M., Amit,I., Garber,M., French,C., Lin,M.F., Feldser,D., Huarte,M., Zuk,O., Carey,B.W., Cassady,J.P. *et al*. (2009) Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals. *Nature*, **458**, 223–227.
7. Joyce,G.F. (1994) *In vitro* evolution of nucleic acids. *Curr. Opin. Struct. Biol*., **4**, 331–336.
8. Swayze,E.E., Jefferson,E.A., Sannes-Lowery,K.A., Blyn,L.B., Risen,L.M., Arakawa,S., Osgood,S.A., Hofstadler,S.A. and Griffey,R.H. (2002) SAR by MS: a ligand based technique for drug lead discovery against structured RNA targets. *J. Med. Chem*., **45**, 3816–3819.
9. Yu,L., Oost,T.K., Schkeryantz,J.M., Yang,J., Janowick,D. and Fesik,S.W. (2003) Discovery of aminoglycoside mimetics by NMR-based screening of *Escherichia coli* A-site RNA. *J. Am. Chem. Soc*., **125**, 4444–4450.
10. 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. *et al*. (2005) SAR by MS: discovery of a new class of RNA-binding small molecules for the hepatitis C virus: internal ribosome entry site IIA subdomain. *J. Med. Chem*., **48**, 7099–7102.
11. Thomas,J.R. and Hergenrother,P.J. (2008) Targeting RNA with small molecules. *Chem. Rev*., **108**, 1171–1224.
12. Tor,Y. (2003) Targeting RNA with small molecules. *Chembiochem*, **4**, 998–1007.
13. Dervan,P.B. (2001) Molecular recognition of DNA by small molecules. *Bioorg. Med. Chem*., **9**, 2215–2235.
14. Disney,M.D., Labuda,L.P., Paul,D.J., Poplawski,S.G., Pushechnikov,A., Tran,T., Velagapudi,S.P., Wu,M. and Childs-Disney,J.L. (2008) Two-dimensional combinatorial screening identifies specific aminoglycoside-RNA internal loop partners. *J. Am. Chem. Soc*., **130**, 11185–11194.
15. Aminova,O., Paul,D.J., Childs-Disney,J.L. and Disney,M.D. (2008) Two-dimensional combinatorial screening identifies specific 6′-acylated kanamycin A- and 6′-acylated neamine-RNA hairpin interactions. *Biochemistry*, **47**, 12670–12679.
16. Llano-Sotelo,B., Azucena,E.F. Jr, Kotra,L.P., Mobashery,S. and Chow,C.S. (2002) Aminoglycosides modified by resistance enzymes display diminished binding to the bacterial ribosomal aminoacyl-tRNA site. *Chem. Biol*., **9**, 455–463.
17. Magnet,S. and Blanchard,J.S. (2005) Molecular insights into aminoglycoside action and resistance. *Chem. Rev*., **105**, 477–498.
18. Francois,B., Russell,R.J., Murray,J.B., Aboul-ela,F., Masquida,B., Vicens,Q. and Westhof,E. (2005) Crystal structures of complexes between aminoglycosides and decoding A site oligonucleotides: role of the number of rings and positive charges in the specific binding leading to miscoding. *Nucleic Acids Res*., **33**, 5677–5690.
19. Lee,M.M., Pushechnikov,A. and Disney,M.D. (2009) Rational and modular design of potent ligands targeting the RNA that causes myotonic dystrophy 2. *ACS Chem. Biol*., **4**, 345–355.
20. Childs-Disney,J.L., Wu,M., Pushechnikov,A., Aminova,O. and Disney,M.D. (2007) A small molecule microarray platform to select RNA internal loop-ligand interactions. *ACS Chem. Biol*., **2**, 745–754.
21. Disney,M.D. and Barrett,O.J. (2007) An aminoglycoside microarray platform for directly monitoring and studying antibiotic resistance. *Biochemistry*, **46**, 11223–11230.
22. Kolb,H.C., Finn,M.G. and Sharpless,K.B. (2001) Click chemistry: diverse chemical function from a few good reactions. *Angew. Chem. Int. Ed. Engl*., **40**, 2004–2021.
23. Carboni,B., Benalil,A. and Vaultier,M. (1993) Aliphatic amino azides as key building-blocks for efficient polyamine syntheses. *J. Org. Chem*., **58**, 3736–3741.
24. Chan,T.R., Hilgraf,R., Sharpless,K.B. and Fokin,V.V. (2004) Polytriazoles as copper(I)-stabilizing ligands in catalysis. *Org. Lett*., **6**, 2853–2855.
25. Bevilacqua,J.M. and Bevilacqua,P.C. (1998) Thermodynamic analysis of an RNA combinatorial library contained in a short hairpin. *Biochemistry*, **37**, 15877–15884.
26. Peyret,N., Seneviratne,P.A., Allawi,H.T. and SantaLucia,J. Jr. (1999) Nearest-neighbor thermodynamics and NMR of DNA sequences with internal A.A, C.C, G.G, and T.T mismatches. *Biochemistry*, **38**, 3468–3477.
27. SantaLucia,J. Jr. (1998) A unified view of polymer, dumbbell, and oligonucleotide DNA nearest-neighbor thermodynamics. *Proc. Natl. Acad. Sci. USA*, **95**, 1460–1465.
28. Puglisi,J.D. and Tinoco,I. Jr. (1989) Absorbance melting curves of RNA. *Methods Enzymol*., **180**, 304–325.
29. Mathews,D.H., Disney,M.D., Childs,J.L., Schroeder,S.J., Zuker,M. and Turner,D.H. (2004) Incorporating chemical modification constraints into a dynamic programming algorithm for prediction of RNA secondary structure. *Proc. Natl Acad. Sci. USA*, **101**, 7287–7292.
30. Mathews,D.H., Sabina,J., Zuker,M. and Turner,D.H. (1999) Expanded sequence dependence of thermodynamic parameters improves prediction of RNA secondary structure. *J. Mol. Biol*., **288**, 911–940.
31. Zou,K.H., Fielding,J.R., Silverman,S.G. and Tempany,C.M. (2003) Hypothesis testing I: proportions. *Radiology*, **226**, 609–613.
32. Wang,Y. and Rando,R.R. (1995) Specific binding of aminoglycoside antibiotics to RNA. *Chem. Biol*., **2**, 281–290.
33. Moazed,D. and Noller,H.F. (1987) Interaction of antibiotics with functional sites in 16S ribosomal-RNA. *Nature*, **327**, 389–394.
34. Michael,K., Wang,H. and Tor,Y. (1999) Enhanced RNA binding of dimerized aminoglycosides. *Bioorg. Med. Chem*., **7**, 1361–1371.
35. Rostovtsev,V.V., Green,L.G., Fokin,V.V. and Sharpless,K.B. (2002) A stepwise huisgen cycloaddition process: copper(I)-catalyzed regioselective ''ligation'' of azides and terminal alkynes. *Angew. Chem. Int. Ed. Engl*., **41**, 2596–2599.
36. Wang,H. and Tor,Y. (1998) RNA-aminoglycoside interactions: design, synthesis, and binding of 'amino-aminoglycosides' to RNA. *Angew. Chem. Int. Ed. Engl*., **37**, 109–111.

37. Vicens,Q. and Westhof,E. (2002) Crystal structure of a complex between the aminoglycoside tobramycin and an oligonucleotide containing the ribosomal decoding A site. *Chem. Biol.*, **9**, 747–755.
38. Barrett,O.J., Pushechnikov,A., Wu,M. and Disney,M.D. (2008) Studying aminoglycoside modification by the acetyltransferase class of resistance-causing enzymes via microarray. *Carbohydr. Res.*, **343**, 2924–2931.
39. Wong,C.H., Hendrix,M., Priestley,E.S. and Greenberg,W.A. (1998) Specificity of aminoglycoside antibiotics for the A-site of the decoding region of ribosomal RNA. *Chem. Biol.*, **5**, 397–406.
40. Gregorian,R.S. Jr. and Crothers,D.M. (1995) Determinants of RNA hairpin loop-loop complex stability. *J. Mol. Biol.*, **248**, 968–984.
41. Recht,M.I., Fourmy,D., Blanchard,S.C., Dahlquist,K.D. and Puglisi,J.D. (1996) RNA sequence determinants for aminoglycoside binding to an A-site rRNA model oligonucleotide. *J. Mol. Biol.*, **262**, 421–436.
42. Ryu,D.H. and Rando,R.R. (2001) Aminoglycoside binding to human and bacterial A-Site rRNA decoding region constructs. *Bioorg. Med. Chem.*, **9**, 2601–2608.
43. Kaul,M. and Pilch,D.S. (2002) Thermodynamics of aminoglycoside-rRNA recognition: the binding of neomycin-class aminoglycosides to the A site of 16S rRNA. *Biochemistry*, **41**, 7695–7706.
44. Kaul,M., Barbieri,C.M., Kerrigan,J.E. and Pilch,D.S. (2003) Coupling of drug protonation to the specific binding of aminoglycosides to the A site of 16 S rRNA: elucidation of the number of drug amino groups involved and their identities. *J. Mol. Biol.*, **326**, 1373–1387.
45. Krebs,A., Ludwig,V., Boden,O. and Gobel,M.W. (2003) Targeting the HIV trans-activation responsive region—approaches towards RNA-binding drugs. *Chembiochem*, **4**, 972–978.
46. Hendrix,M., Priestley,E.S., Joyce,G.F. and Wong,C.H. (1997) Direct observation of aminoglycoside-RNA interactions by surface plasmon resonance. *J. Am. Chem. Soc.*, **119**, 3641–3648.
47. Luedtke,N.W., Baker,T.J., Goodman,M. and Tor,Y. (2000) Guanidinoglycosides: a novel family of RNA ligands. *J. Am. Chem. Soc.*, **122**, 12035–12036.
48. Park,W.K.C., Auer,M., Jaksche,H. and Wong,C.H. (1996) Rapid combinatorial synthesis of aminoglycoside antibiotic mimetics: Use of a polyethylene glycol-linked amine and a neamine-derived aldehyde in multiple component condensation as a strategy for the discovery of new inhibitors of the HIV RNA Rev responsive element. *J. Am. Chem. Soc.*, **118**, 10150–10155.
49. von Ahsen,U., Davies,J. and Schroeder,R. (1991) Antibiotic inhibition of group I ribozyme function. *Nature*, **353**, 368–370.
50. Liu,Y., Tidwell,R.R. and Leibowitz,M.J. (1994) Inhibition of *in vitro* splicing of a group I intron of *Pneumocystis carinii*. *J. Eukaryot. Microbiol.*, **41**, 31–38.
51. Goertz,P.W., Cox,J.C. and Ellington,A.D. (2004) Automated selection of aminoglycoside aptamers. *J. Assoc. Lab. Automat.*, **9**, 150–154.
52. Cho,J., Hamasaki,K. and Rando,R.R. (1998) The binding site of a specific aminoglycoside binding RNA molecule. *Biochemistry*, **37**, 4985–4992.
53. Jiang,L., Suri,A.K., Fiala,R. and Patel,D.J. (1997) Saccharide-RNA recognition in an aminoglycoside antibiotic-RNA aptamer complex. *Chem. Biol.*, **4**, 35–50.
54. Verhelst,S.H., Michiels,P.J., van der Marel,G.A., van Boeckel,C.A. and van Boom,J.H. (2004) Surface plasmon resonance evaluation of various aminoglycoside-RNA hairpin interactions reveals low degree of selectivity. *Chembiochem*, **5**, 937–942.
55. Jiang,L., Majumdar,A., Hu,W., Jaishree,T.J., Xu,W. and Patel,D.J. (1999) Saccharide-RNA recognition in a complex formed between neomycin B and an RNA aptamer. *Structure*, **7**, 817–827.
56. Disney,M.D. and Childs-Disney,J.L. (2007) Using selection to identify and chemical microarray to study the RNA internal loops recognized by 6′-N-acylated kanamycin A. *Chembiochem*, **8**, 649–656.
57. Borovinskaya,M.A., Pai,R.D., Zhang,W., Schuwirth,B.S., Holton,J.M., Hirokawa,G., Kaji,H., Kaji,A. and Cate,J.H. (2007) Structural basis for aminoglycoside inhibition of bacterial ribosome recycling. *Nat. Struct. Mol. Biol.*, **14**, 727–732.
58. Hainrichson,M., Nudelman,I. and Baasov,T. (2008) Designer aminoglycosides: the race to develop improved antibiotics and compounds for the treatment of human genetic diseases. *Org. Biomol. Chem.*, **6**, 227–239.
59. Mingeot-Leclercq,M.P., Brasseur,R. and Schanck,A. (1995) Molecular parameters involved in aminoglycoside nephrotoxicity. *J. Toxicol. Environ. Health*, **44**, 263–300.
60. Berg,J.M., Tymoczko,J.L. and Stryer,L. (2007) *Biochemistry, Sixth Edition*. W.H. Freeman and Company, New York.
61. Deutscher,M.P. (2006) Degradation of RNA in bacteria: comparison of mRNA and stable RNA. *Nucleic Acids Res.*, **34**, 659–666.