Published in final edited form as:

Biochemistry. 2020 September 22; 59(37): 3473–3486. doi:10.1021/acs.biochem.0c00586.

Selection of high affinity RNA aptamers that distinguish between doxycycline and tetracycline

Zachary J. Tickner, Guocai Zhong, Kelly R. Sheptack, Michael Farzan*

Department of Immunology and Microbiology, The Scripps Research Institute, Jupiter, FL, 33458

Abstract

Oligonucleotide adapters are found in prokaryotes and eukaryotes, and they can be selected from large synthetic libraries to bind protein or small-molecule ligands with high affinities and specificities. Aptamers can function as biosensors, as protein recognition elements, and as components of riboswitches allowing ligand-dependent control of gene expression. One of the best studied laboratory selected aptamers binds the antibiotic tetracycline, but it binds with much lower affinity to the closely related but more bioavailable antibiotic doxycycline. Here we report enrichment of doxycycline-binding aptamers from a selectively-randomized library of tetracycline-aptamer variants over four selection rounds. Selected aptamers distinguish between doxycycline, which they bind with approximately 7 nanomolar dissociation constants, and tetracycline, which they bind undetectably. They thus function as orthogonal complements to the original tetracycline aptamer. Unexpectedly, doxycycline aptamers adopt a distinct conformation from the tetracycline aptamer and depend on constant regions originally installed as primer-binding sites. We show that doxycycline fluorescence emission intensity increases upon aptamer binding, permitting their use as biosensors. This new class of aptamers can be used in multiple contexts where doxycycline detection, or doxycycline-mediated regulation, is necessary.

INTRODUCTION

Aptamers are short oligonucleotides or peptides capable of binding to molecular targets with high affinity and specificity^{1, 2}. Aptamers were first reported in 1990, when Tuerk and Gold isolated RNA aptamers to T4 DNA polymerase³ and Ellington and Szostak developed RNA aptamers to several small organic dyes⁴. Both groups used in-vitro Darwinian selection techniques, typically referred to as SELEX (Systematic Evolution of Ligands by EXponential enrichment), to enrich aptamers from large pools of randomized RNA sequences. That same year, Robertson and Joyce used in-vitro selection to alter the catalytic function of a self-splicing intron to specifically cleave ssDNA⁵. The following three decades have seen in-vitro selected aptamers used as small-molecule biosensors^{6, 7}, targeting motifs for therapeutic delivery systems^{8, 9}, enzymatic and signal transduction

*Corresponding author: Michael Farzan – Department of Immunology and Microbiology, The Scripps Research Institute, Jupiter, Florida 33458, United States; Phone:561-228-2300; mfarzan@scripps.edu.

SUPPORTING INFORMATION

Library quality control data, predicted secondary structures for the tet aptamer library template with constant regions, data for enrichment of P2MC during dox aptamer selections, concentration-dependent folding studies, low-concentration (50 nM RNA) PAGE binding studies, control SPR experiments, locations of enriched mutations within the tet aptamer binding pocket.

inhibitors¹⁰, and ligand binding sites for synthetic riboswitches controlling transcription and translation¹¹⁻¹⁵.

Doxycycline (dox) is attractive as a SELEX target and as a regulatory ligand for RNA devices. Dox is a suitable RNA aptamer target due to its potential for π -stacking interactions and hydrogen bonding, properties which facilitate its binding to ribosomal RNA. Fluorescence emission of the dox structural isomer tetracycline (tet) is amplified upon binding to the tet aptamer, suggesting that dox aptamers may similarly serve as dox biosensors¹⁶. Several synthetic riboswitches controlled by tet association with an *in vitro*-selected aptamer have shown efficient regulation of gene expression in mammalian cells, but did not perform as well in mice. Dox is a preferred ligand for other tet-dependent gene expression control systems (e.g. Tet-On, Tet-Off) because of its increased bioavailability relative to tet. Unfortunately this tet aptamer binds dox too weakly to enable-dependent regulation of gene expression¹⁵⁻¹⁷. Therefore, substitution of a dox aptamer for the tet aptamer in synthetic riboswitches may allow increased performance in animal models.

A typical SELEX experiment attempts to enrich aptamers from libraries containing up to 10^{15} randomized RNA species¹⁸. Because the number of possible RNA sequences of length 'n' is given by 4^n , fully-randomized libraries of RNAs longer than 24 nucleotides sample a small fraction of all possible sequences, and the vast majority of these will include multiple deleterious mutations. Because of this, it is desirable to design starting libraries restricted to sections of "sequence space" containing greater concentrations of potential high-fitness RNA species. One option is to use an existing aptamer to a structurally-related ligand as a library template. For example, Famulok used 30% randomization of each position in an *in vitro* selected L-citrulline-binding RNA aptamer to allow selection of an L-arginine binder from the resulting "soft-randomized" library¹⁹. However, re-randomization of all positions within an existing aptamer may still generate many low-fitness RNA species by disrupting secondary structure elements required for proper aptamer folding. Additionally, changes in folding conditions may also explain why many aptamers selected *in vitro* fail to function when transferred into cells²⁰. To avoid these potential problems, Porter *et al.* developed SELEX libraries by partially randomizing "scaffolds," secondary structures derived from riboswitches and ribozymes known to function in the target cell environment²¹. By avoiding randomization of critical structural regions, libraries of RNAs with conserved folding patterns and a wide variety of potential binding sites were generated and specific aptamers to various neurotransmitter precursors enriched over seven selection rounds. This approach is especially useful for the goal of generating an aptamer to a closely related ligand. A previously described, well characterized tet aptamer binds detectably binds dox, albeit weakly¹⁶. This aptamer has been successfully used in various mammalian-cell riboswitches, indicating that it can fold appropriately and stably in these cells^{11, 15}. A high-resolution crystal structure of the tet aptamer bound to 7-chloro-tetracycline is also available, facilitating informed selection of randomized nucleotides²².

Here we report the selection and characterization of high-affinity, high-specificity RNA aptamers to dox from a library derived from selective randomization of the tet aptamer. These aptamers bind to dox with affinities as low as 3 nM and do not detectably bind tet. Dox aptamers also amplify fluorescence emission upon binding, permitting their use as dox

biosensors. These aptamers were found to require upstream and downstream constant regions installed for reamplification during selections, sequences that are not required for tet aptamer binding to tet. Structural characterization of dox aptamers revealed a partial reorganization of the tet aptamer secondary structure involving formation of a new stem-loop structure formed by base pairing between the 5' SELEX primer binding site and the 5' portion of the tet aptamer sequence. The orthogonal functionalities of the tet aptamer and these new dox aptamers may enable their tandem use in advanced gene-control systems.

MATERIALS AND METHODS

PAGE Purification of Oligos

For native PAGE preparation of DNA samples, 6X Orange G loading dye containing 0.4X TBE, 1.5 mg/mL Orange G (Acros Organics #416550100), and 60% (v/v) glycerol was added to resuspended DNA to a final dye concentration of 1X. Native PAGE gels were assembled with 1X TBE and 10% 19:1 acrylamide/bis-acrylamide and polymerized using 0.8 mg/mL ammonium persulfate and 0.08% (v/v) TEMED. Nucleic acid/dye mixes were loaded and separated via PAGE for 3h at 8W after a 30 minute pre-run. Following electrophoresis, gels were stained using 1X SYBR-Safe (ThermoFisher #S33102) in water for 10min and washed twice in water for 5min. Bands were visualized using a UV transilluminator, excised and crushed, suspended in 300 mM NaCl, and tumbled overnight at 4°C. Gel fragments were spun down and the supernatant transferred to new tubes. Fragments were washed twice in water, all washes were combined, and gel fragments removed using a 0.2 um SFCA filter (Thermo #723-2520) followed by precipitation using isopropanol.

For denaturing PAGE purification of RNA samples, RNA samples were mixed with 2X denaturing orange G loading dye containing 0.4X TBE, 0.5 mg/mL orange G, and 10 M urea to a final dye concentration of 1X. PAGE gels were assembled as described above, with the addition of 8 M urea. Gels were pre-run at 8W for 30 min. RNA samples were denatured at 65°C for 5 min prior to electrophoresis at 8W for 3h. Following separation, gels were wrapped in saran and bands visualized and marked using 254 nm UV shadowing. Bands were marked, excised, crushed, and tumbled in 300 mM NaCl overnight at 4°C. RNA was separated from gel fragments, concentrated, and quantified as described above.

Crystal Structure Analysis

The crystal structure of 7-chloro-tetracycline bound to the tet aptamer (PDB #3EGZ) was visualized and structure images generated using UCSF Chimera v1.14^{22, 23}. Interatomic distances were measured between 7-chloro-tetracycline and aptamer bases to select randomization sites for library design.

Structure Prediction

RNA secondary structure prediction was performed using RNAstructure 5.8²⁴. For minimum free-energy structure prediction, RNAs were folded as single sequences with a maximum energy difference of 10%²⁵. For Multilign analysis, P2MC was set as the index sequence and compared to the nine most-enriched sequences from round 3, round 4, or the most enriched sequences from each round²⁶. The maximum energy difference was set to

20%, the structure window size to 2, the alignment window size to 1, and the gap penalty to 0.4 for Multalign analysis, with single base-pair deletions allowed. For Turbofold analysis, P2MC was compared to 30 RNAs enriched in rounds 3 and 4²⁷. Maximum predicted accuracy analysis was performed with max % score difference set to 50.0, window size set to 5, and MEA gamma set to 1. Partition functions were generated for the tet aptamer, P2MC, and P1MC iv. to enable base pairing probability annotation.

In-Line Probing

In-line probing was performed according to methods described by Regulski and Breaker²⁸. 20 µg of gel-purified aptamer RNA was dephosphorylated in using calf intestinal phosphatase (NEB #M0290), purified using RNEasy mini spin columns (Qiagen #74104), and eluted in water. Dephosphorylated RNAs were 5' end-labeled using T4 polynucleotide kinase (NEB #M0201) and ~1 µCi α-³²P-ATP (PerkinElmer #BLU003H250UC), followed by denaturing PAGE purification. In-line probing reactions were set up using the following conditions: 1-100 nM RNA, 0-500 nM ligand, 50 mM Tris-HCl pH 8.3, 100 mM KCl, and 20 mM MgCl₂. Samples were denatured at 65°C for 5min, followed by incubation at room temperature for 40h. Control reactions were prepared from frozen stocks of radiolabeled aptamer RNA immediately prior to electrophoresis. Undigested RNA samples were prepared with 1-100 nM RNA in water. RNase T1 digests were prepared with 1-100 nM RNA, 25 mM sodium citrate pH 5.0, 7M urea, and 1 mM EDTA pH 8.0, then incubated at 55°C for 5 min. Alkaline digests were prepared with 1-100 nM RNA, 50 mM NaCO₃ pH 9.0, and 1 mM EDTA, and incubated at 90°C for 5 min. All samples were brought to 1X denaturing Orange G loading dye and separated on a 0.1 mm denaturing 20% polyacrylamide gel at 1W for 8h, following a 30 minute pre-run. Gels were imaged by autoradiography following a 24-48h exposure. Exposure data was analyzed using Quantity One software by Bio-Rad.

Selection Resin Coupling

Resin coupling to tet and dox was performed following methods used in tet selections^{29, 30}. Test-scale couplings were performed and showed an optimum concentration of 2.5-5.3 coupling site equivalents of ligand. For bulk couplings, 5g of Epoxy-Sepharose 6B (GE Healthcare #17-0480-01) was swelled in five 200 mL washes of distilled water over 1h. 2.5-5.3 coupling site equivalents of tetracycline hydrochloride (Alfa Aesar #B21408) or doxycycline hydrochloride (Alfa Aesar #J60422) were dissolved in water and brought to pH 11 using HCl and NaOH to generate coupling solutions. 50 mL coupling solutions were added to swelled and drained resins, protected from light, and shaken at room temperature for 8h. Following coupling, excess ligand was removed by six alternating 50 mL washes of 100 mM Tris-Cl pH 8.0 and 100 mM NaOAc pH 4.0, followed by a 50 mL wash, draining, and storage in 10 mM Tris-Cl pH 8.0. A control resin for library preselection was prepared by coupling 1 M ethanolamine (Alfa Aesar #A11697) using the same methods. Absorbance of dox-coupled, tet-coupled, EA-coupled, and uncoupled resins was measured from 400 – 800 nm and compared to standard curves of tet and dox to yield predicted densities of approximately 6×10^{17} molecules displayed per 200 uL selection column.

Library Synthesis

Library synthesis was performed following methods described by Groher and Suesse³⁰. All ssDNA selection library templates and primers were ordered from Integrated DNA Technologies (Supplementary File 1). Forward primers contained a 5' extension for installation of the upstream T7 transcription site onto PCR products³¹. 122.2 nmol of template DNA was resuspended to 100 μ M and primers were resuspended to 1 mM in water. Gradient PCR was performed to optimize annealing temperatures prior to large-scale reactions, with annealing temperatures of 60°C for Taq and 65°C for Q5 and Phusion polymerases showing high yields with low side product formation. 5×10^{14} template molecules were amplified over 8 PCR cycles using each polymerase with the following reaction conditions:

Reaction mixes were divided into 100 μ L aliquots in 96-well PCR plates and amplified over eight cycles using the above annealing temperatures and 30s extension times. PCR products were pooled, concentrated by extraction with 2-butanol, and precipitated overnight using isopropanol. Precipitates were pelleted, dried, and extracted with pH 6.5 phenol/chloroform/IAA (Fisher BioReagents #FL-06-0205). Extracted DNAs were PAGE purified on a 3mm 10% native gel as described above.

Approximately 1.5×10^{15} template DNA molecules, or 3-fold theoretical representation of library PCR templates, were transcribed using T7 RNA Pol transcription kits (Lucigen #30223). NTP mixes were made by resuspending ATP, CTP, GTP, and UTP (Chem Impex Intl #00015, 00095, 00348, 00311) in water, adjusting pH values to 7 using NaOH and HCl, diluting each to 40 mM, filtering through 0.2 μ m SFCA filters, and combining into a stock solution with 10 mM each NTP. 200 μ g template DNA was transcribed for 8 h in 16.66 mL according to the manufacturer's instructions. After 8h, 180 U of DNase I (NEB #0303) was added to each transcription and incubated at 37°C for 20min. Following DNase I treatment, transcription products were butanol concentrated, precipitated with isopropanol as described above, resuspended in water, extracted using acid phenol/chloroform/isoamyl alcohol (Ambion #9721G), and reprecipitated. Library RNA was further purified by denaturing PAGE, followed by quantitation by nanodrop and RT-qPCR using a Qiagen OneStep RT-qPCR kit (Qiagen #210212) with a standard curve composed of serial dilutions of gel-purified dead tet aptamer RNA.

Dox Aptamer SELEX

Dox aptamer SELEX conditions were adapted from several published methods and used RT-qPCR quantification of selection pools^{30, 32, 33}. For each selection cycle, three 200 μ L dox resin columns were assembled in sterile 2.5 mL fritted syringes (Torviq #SF-0250) and washed twice for 10 minutes with 1 mL selection buffer (50 mM Tris-HCl pH 7.6, 250 mM NaCl, 0.5% Tween-20, 5 mM $MgCl_2$). 10^{15} library RNAs were brought to 600 μ L in selection buffer without Tween-20 or $MgCl_2$, denatured at 90°C for 3 min, brought to 0.5% Tween-20 and 5 mM $MgCl_2$, and cooled at room temperature for 15 min. 10 mg samples of yeast tRNA (ThermoFisher #AM7119) were brought to 1 mL and folded in identical conditions. Dox affinity resin columns were blocked with yeast tRNA for 15 min, drained, incubated with library RNA for 15 min, and library RNA flow-through collected. Columns

were then subjected to three 1 mL selection buffer washes in round 1, six washes in round 2, six washes in round 3, and 10 washes in round 4. Following washes, RNA was eluted using successive washes of 1, 10, 100, and 1000 μM dox in selection buffer. All flow-through, wash, and eluent fractions were frozen with liquid nitrogen immediately following collection. Beginning in round 3, counterselection on EA-derivatized resin was performed. Selection pool RNAs were folded in 600 μL and incubated on EA-derivatized resin for 15 minutes prior to dox affinity resin binding. Flow-through was collected and preselection columns washed with 400 μL selection buffer. Preselection flow-through and wash fractions were then combined, introduced to the dox affinity resin column, and further steps performed as usual.

Fraction RNA contents were quantified using a Qiagen One Step RT-PCR kit (Qiagen #210212) according to the manufacturer's instructions, with 5 μM each selection primer and 1X EvaGreen dye (Biotium #31000). Relative quantification was performed using a standard curve of binding-deficient tet aptamer RNA with installed constant regions. Elution profiles were used to select which fractions would be pooled for the next round of selections. Selected elutions were pooled, pelleted, and reverse-transcribed using TGIRT-III reverse transcriptase according to the manufacturer's instructions. RT products were quantified by qPCR using each library's respective polymerase, with 5 μM each primer and 1X EvaGreen dye. The number of qPCR cycles required to reach the end of exponential amplification was determined and used as the cycle count for bulk PCR amplification of RT products. Remaining RT products were amplified in 900 μL PCR reactions with each library's respective polymerase and 5 μM each primer. PCR products were combined and purified using 4 PCR cleanup columns (Qiagen #28106). 12.7 μg PCR product DNA ($\sim 1 \times 10^{14}$ molecules) was transcribed in 1.5 mL using T7 RNA polymerase and templates removed by DNase I digestion. Transcription products were purified by phenol-chloroform extraction and purification on RNEasy columns (Qiagen #74106). Aliquots containing $\sim 1 \times 10^{15}$ RNAs were taken to serve as starting pools for the next selection round.

Next-Generation Sequencing

NGS was used for library quality control and to monitor selection progress³⁴. NGS samples were prepared from 100 ng (~ 1 pmol) aliquots of starting library transcription templates and RT-PCR products generated following each selection round. Adapter ligation PCR reactions were assembled in 50 μL using 2X Clontech HiFi PCR Premix (Takara Bio #639298) with 300 nM forward and reverse primers. 25 amplification cycles were performed with annealing at 60°C and 30s extension steps. A universal forward primer (P5) was used in all reactions, while reverse primers (P7 index 1-7) containing index sequences for sample identification and 8-nt, fully-randomized UMI sequences were assigned to specific samples³⁵. Primer sequences are available in Supplementary File 1. PCR products were separated on 2% agarose-TAE gels, ~ 200 bp bands excised, and samples purified using QIAquick Gel Extraction Kits (Qiagen #28706). Samples were submitted to the Scripps Research Institute Genomics Core for quality control by TapeStation 4200 (Agilent #G2991AA) and/or BioAnalyzer (Agilent #G2939BA) systems, following which samples were pooled and sequenced using a MiSeq V3 system (Illumina #SY-410-1003) with single-ended, 150 bp reads. Reads were quality-controlled for minimum read length and the presence of intact

constant regions, then ranked by the number of UMIs associated with each sequence. Data analysis software was written by Michael Farzan and is available upon request. Samples generated using Taq polymerase and containing high numbers of G-repeat expansions were submitted during previous dox aptamer selections. These samples showed abnormal, high-running bands during TapeStation quality control and poor sequencing results, reflecting difficulties encountered during PCR amplification of these constructs during selections.

Column Elutions

Dox aptamer column elution binding analysis was based on the method used by Lorsch and Szostak³². Dox aptamer, tet aptamer, and dead tet aptamer RNAs were transcribed from ssDNA templates from Integrated DNA Technologies and purified by denaturing PAGE. Folding, blocking, binding, wash, and elution steps were performed as in dox aptamer selections, without counterselection on EA resin. Preliminary experiments with tet and dead tet RNAs and using 21 washes showed a plateau in wash fraction RNA content following 6-8 washes (data not shown), and a wash count of 10 was chosen for specific elution experiments. Elutions were performed using 1 nM – 100 μ M dox or tet in order to detect elution in response to lower ligand concentrations. 1 nM tet did not produce significant elution of the tet aptamer despite a reported K_d of 770 pM¹⁶, possibly because the approximately 10^{12} tet molecules was too small relative to the amount of column-bound RNA present. Fractions were frozen using liquid nitrogen immediately after collection, thawed on ice, and quantified by RT-qPCR as during dox aptamer SELEX.

Fluorescence Titration Spectroscopy

FTS experiments were based on the method described by Müller *et al*¹⁶. Gel-purified aptamer RNA was brought to 10 μ M in binding buffer (20 mM Potassium phosphate buffer (pH 7.5), 100 mM NaCl, 10 mM MgCl₂), denatured at 65°C for 5 min, and cooled at room temperature for 45 min. RNAs were diluted to assay concentrations in 1 nM tet or 20 nM dox in binding buffer and incubated for 1 h prior to data collection. Tet and dox assay concentrations were determined by the minimum concentration at which fluorescence could be measured in the absence of ligand. Samples were transferred to 1 cm quartz cuvettes, excited at 370 nm, and spectra recorded from 450-600 nm in 1 nm increments using a Cary Eclipse spectrofluorometer (Agilent Technologies). Excitation slits were set to 5 nm and emission slits set to 20 nm, with integration time set to slow and PMT sensitivity to high. The instrument was zeroed using blank binding buffer prior to measurement of dox and dox/RNA samples. Folding and all subsequent steps were performed in triplicate, with data represented as the average of three values.

In-Gel Fluorescence Assays

FTS experiments were based on the method described by Müller *et al*¹⁶. Gels were prepared with 50 mM Tris-acetate (pH 7.5), 10 mM MgAc₂, and 10% (19:1) acrylamide/bis-acrylamide, then polymerized as described above. RNAs and tet/dox were combined and diluted to assay concentrations in binding buffer (20 mM Tris-HCl (pH 7.5), 100 mM NaCl, 10 mM MgCl₂), denatured at 65°C for 5 min, and cooled at room temperature for 45 minutes. Following incubations, 16.7% glycerol was added and vortexed to weight samples for gel loading. Gels were pre-run for 30 min at 1W, loaded, and samples separated for 6h at

1W. Following electrophoresis, gels were transferred to a ChemiDoc XRS gel imaging system (Bio-Rad #1708265). Tet/dox fluorescence was imaged using UV excitation and a 530/28 nm filter. Gels were then stained for 5 min using 0.2X SYBR-Gold (Thermo #S11494), washed, destained for 10 min, and imaged using SYBR-Gold filter settings. Composite gel images were generated using ImageJ.

Surface Plasmon Resonance

RNA was transcribed and PAGE-purified as described previously. Biotin-labeling was performed using pCp-biotin (Jena Bioscience #NU-1706-BIO) and T4 ssRNA ligase (NEB #M0437M) according to the manufacturer's instructions: 5 µg gel-purified RNA was incubated in 1X T4 RNA ligase reaction buffer with 10 µM pCp biotin and 50 U ligase overnight at 16°C and purified using an RNEasy cleanup column (Qiagen #74106). RNAs were brought to 20 µg/mL in selection buffer (50 mM Tris-HCl pH 7.6, 250 mM NaCl, 0.5% Tween-20, 5 mM MgCl₂) and folded as described previously. Experiments were performed using a Biacore X100 system (Cytiva Life Sciences). SPR chips were docked, equilibrated in buffer overnight, and two priming runs performed prior to RNA loading. For streptavidin (SA) chips (Cytiva Life Sciences #BR100032) biotinylated RNA was immobilized using 120s exposure and 300s stabilization, while chips were regenerated using 240s buffer contact and 600s stabilization between injection series. For CAP chips (Cytiva Life Sciences # 28920233), CAP reagent was exposed for 180s and stabilized for 300s, followed by exposure to biotinylated RNA for 180s and equilibration for 300s. CAP chip regeneration was performed using guanidine/NaOH according to the manufacturer's instructions. For binding assays, single-cycle kinetics series were set up with 60s exposure and 600s dissociation steps following each injection. Two- or ten-fold serial dilutions of ligand were assembled in selection buffer to span predicted dox aptamer dissociation constants. Two blank injection series were performed prior to dox/tet injections in order to obtain baseline data. Kinetic data were generated using the kinetics analysis tool in Biacore Insight Evaluation Software (Cytiva Life Sciences). Dissociation constants were also estimated using the one-site specific binding model in GraphPad Prism 8.4.0 (GraphPad Software).

RESULT

Library Design and Synthesis

Because of the structural similarity of dox to tet, dox aptamer selection libraries were based on partial randomization of the tet aptamer (Figure 1A-B) ²⁹. Randomization sites were chosen based on analysis of the tet aptamer crystal structure (Figure 1C) and in-line probing of the tet aptamer (Figure 1D) ^{22, 28}. The central junction of stems P1, P2, and P3, as well as the loop region of stem P3 were randomized based on proximity to 7-chloro-tetracycline in the crystal structure, while several additional bases were selected based on protection by tet binding during in-line probing. In total, 29 of 65 positions in the tet aptamer were selected for randomization. A base substitution rate of 30% (10% for each alternative nt) was chosen to yield a suitable distribution of mutants (Figure S1). 18-nucleotide 5' and 17-nucleotide 3' constant regions were installed to serve as primer binding sites during selection. These regions included restriction digest sites as well as a portion of the T7 RNA polymerase promotor, and were designed to allow proper folding of the tet aptamer (Figure. S2). Three 5

$\times 10^{14}$ -molecule aliquots of the DNA template library were amplified using three separate DNA polymerases and transcribed to yield the Taq (Taq DNA Polymerase), Q5 (Q5 Hot Start II DNA Polymerase), and Phu (Phusion Hot Start II) selection libraries. Libraries were quality-controlled for read length and randomization rates using next-generation sequencing (NGS, Figure S1). Unique molecular identifiers (UMIs) were used to assist in determining accurate copy numbers of library and selection pool RNAs; for quality-controlled sequences, 99.99% of UMIs were associated with unique sequences³⁵.

Dox Aptamer SELEX

Four selection rounds were performed for each library in parallel (Figure 2A). Stringency was increased by increasing the number of wash steps throughout the selection, counter-selecting with ethanolamine-derivatized epoxy Sepharose-6B columns in rounds 3 and 4, and excluding high-dox elutions from selection pool reamplifications in later rounds³⁰. Selections were monitored using two primary endpoints in order to detect aptamer enrichment. First, RT-qPCR of wash and elution fractions showed increases in overall retention during washes as well as specific elution of library RNAs by dox in round 4, suggesting specific binding to column-displayed dox (Figure 2B). Notably, 1 mM dox elutions contained much less RNA than the previous 1-100 μ M elutions.

Based on these results, only 1 – 100 μ M dox elutions were carried forward for all libraries following round 2, and only 1 – 10 μ M dox elutions of the Phu library were carried forward in round 2. Only 1 μ M elutions were submitted for NGS analysis following round 4. Second, selection pool composition was monitored using NGS. Significant enrichment of several sequences was observed in rounds 3 and 4, with some sequences comprising up to 10% of all UMIs (Figure 2C). Sequence analysis revealed common patterns of mutations in the most enriched constructs from all three selection libraries (Figure 2D). These similarities suggested convergent evolution, and a consensus sequence was also found to be enriched in later selection rounds (Figure S3A).

Dox aptamer binding studies began following round 3 based RT-qPCR evidence of specific elution and NGS results showing sequence enrichment. Each library's most enriched RNAs following round 3 (Taq 3-1, Q5 3-1, Phu 3-1), as well as the tet aptamer and a binding-deficient triple mutant (A21G A32G A40G, "dead tet aptamer") were synthesized. 10^{15} copies of control or aptamer RNA were bound to tet or dox affinity columns, washed, and eluted using tet or dox similarly to a selection round. When bound to dox resin, all three aptamer RNAs showed specific elution by dox, while the control tet and dead tet aptamers did not (Figure 3A, left panel). The RNA content of washes from dox resin was higher for dox aptamer RNAs than for either tet aptamer. Dox aptamer RNAs showed lower retention on tet resin and little to no specific elution by tet, in contrast to tet aptamers (Figure 3A, right panel). Specific elution of dead-tet aptamer RNA by tet suggests that some affinity is retained despite its mutations. Selections were stopped after round 4 after specific elution, sequence enrichment, and preliminary binding studies suggested enrichment of dox aptamers.

Fluorescence Binding Studies

Tet aptamer binding increases the fluorescence emission of tet, enabling the measurement of binding interactions using fluorescence titration spectroscopy (FTS) ¹⁶. We next determined whether the enriched aptamers would similarly increase dox fluorescence. For dox aptamer FTS, RNAs were folded at 10 μ M in the absence of dox, then incubated with 20 nM dox for one hour prior to fluorimetry. FTS revealed dose-dependent increases in dox fluorescence emission upon incubation with dox aptamer RNAs (Figure 3B). The consensus sequence aptamer P2MC produced greater increases in dox fluorescence emission beginning at significantly lower RNA concentrations compared to RNAs enriched in selections. Titrations of tet aptamer RNA into 1 nM tet produced significant increases in fluorescence emission, but titration of P2MC did not (Figure 3C), suggesting that each of these two RNAs exhibits a high degree of selectivity for its ligand.

To better understand the relatively long (~30 minute) equilibration times required for FTS, binding dynamics over time were studied using native PAGE ¹⁶. To accomplish this, aptamer RNA was folded in the absence of dox, then incubated with equimolar dox for 0 to 60 minutes prior to gel loading. Gels were imaged for dox fluorescence, stained using SYBR-Gold, and re-imaged to localize RNA (Figure 3D). At 500 nM RNA, two bands were observed for each dox aptamer: an upper band which did not colocalize with dox (orange arrow) and a lower band which did (green arrow). Longer incubations did not appear to affect the relative distribution of RNA between these two bands. Aptamer RNA was purified from a single band following denaturing PAGE prior to binding studies, which suggests that these banding patterns reflect differences in folding or multimerization. Aptamer RNA folded at higher concentrations shifted toward higher-running bands (Figure S3B). Very small amounts of dox were found to colocalize with P2MC upper bands following prolonged incubations. Folding dox aptamers in the presence of dox resulted in dox-dependent shifts of RNA staining intensity from upper to lower bands (Figure 3E). The ratios of the two bands and the degree of ligand-dependent upper band depletion varied between dox aptamers, with P2MC demonstrating the highest preference for lower band formation and Q5 3-1 showing the highest response to the presence of dox during folding. Native PAGE using 50 nM RNA was complicated by weak dox fluorescence emission, but showed aptamer binding at low nanomolar concentrations (Figure S3C). Together these results show tight, specific binding of the aptamer RNAs to dox.

Structural Analysis of Dox Aptamers

Based on its folding properties and performance during binding assays, we investigated the secondary structure of the P2MC consensus sequence aptamer (Figure 4A). Its eleven mutations were individually reverted to their tet aptamer base identities, revealing that 7 of these 11 mutations are required for binding to 500 nM dox (Figure. 4B). The majority of essential mutations were found within 5 nucleotides of the base of stem P1 in the tet aptamer structure. The constant regions installed for use as primer binding sites during dox aptamer selections did not interfere with tet aptamer folding or binding in preliminary studies (Figure 4C, upper left panel), and were intended to be removed following dox aptamer selections³⁶. To our surprise, deletion of these sequences completely abolished 500 nM dox binding in all dox aptamers tested (Figure 4C, upper right panel). Partial 5' and 3' truncations of these

sequences revealed that only the final two nucleotides of the 3' constant region could be removed without loss of binding (Figure 4C, lower panel). Mutagenesis of constant region bases produced variable effects on 500 nM dox binding (Figure 4D).

To help explain these results, predicted secondary structures were generated using the free energy minimization, Multilign and Turbofold programs included with the RNAstructure software package²⁴⁻²⁷. Most of these structures showed formation of tet aptamer stem P1, as well as a novel stem-loop involving bases 5 – 30 ("stem P4"). Native PAGE fluorescence experiments supported these results; transitions which converted between A-U, G-U, and G-C pairs within stem P4 were well-tolerated while transversions disrupting these pairs were not, while all mutations in the predicted P4 loop were well-tolerated. Exchanging all predicted base pairs from A6-U39 to U11-A24 did not disrupt 500 nM dox binding, differing from the effects of point mutations at positions 6-11 and providing strong evidence for stem P4 formation (data not shown). Mutagenesis suggests that base identity in positions 87-97 is important for dox binding, but predicted structures showed a variety of possible configurations for the 3' constant region.

Additional attempts were made to generate a minimized dox aptamer, validate structure predictions, and find locations where the aptamer could be ligated to a riboswitch. Stem P1 is dispensable for tet aptamer binding to tet^{15, 29}, and mutagenesis results suggested that the P2MC stem P4 loop is not directly involved in dox binding. Truncation of both stems and replacement of their loop regions with GAAA tetraloops resulted in a 74-nt aptamer with unimpaired binding to 500 nM dox (Figure 4E). In addition, attempts were made to circularize the aptamer. The parent tet aptamer can bind tet with its 5' and 3' termini localized to the base of stems P1 or P2, and P1 loop truncations do not prevent P2MC binding to dox. P1 dox aptamers (referring to 5' and 3' termini localized to stem P1) were generated by deletion of P2MC bases 54-63 and bridging of positions 1 and 97 using a variety of sequences (Figure 4F). The distal portion of stem P1 and two common RNA tetraloop sequences support dox binding when used as bridging sequences, with the CUUG tetraloop (P1MC iv.) construct showing the highest binding to 500 nM dox. Free energy minimization predicts short stems adjacent to these bridging sequences, with base pairing beginning at C34-G42 for the CUUG tetraloop and C34-G57 for the P1 stem-loop.

The mutability of stems P1 and P4, as well as the placement of essential mutations between them, suggested that dox binding interactions involve P2MC bases 30-45 and 72-98. In-line probing of P2MC and CUUG-bridged P1MC RNA was performed to observe dox-dependent changes to the structure of these regions, and to validate the presence of stems P1 and P4 (Figure 5A-B)²⁸. In-line probing exploits conformation-dependent rates of Mg²⁺-catalyzed RNA backbone cleavage to reveal structure and binding interactions, with base paired regions more protected from cleavage than unpaired regions. In brief, 50 nM 5' radio-labeled aptamer RNA was folded at room temperature in the presence of 20 mM Mg²⁺ and 0 to 1 mM dox and incubated for 40 hour before separation by denaturing PAGE and autoradiography. As expected, bands representing stems P1 and P4 were clearly visible in probing results from both aptamers, consistent with low cleavage rates in base-paired regions and high cleavage rates in loops. P2MC bases 80-84 and P1MC iv. bases 17-20 and 37-39 showed increased cleavage consistent with their predicted localization to loop regions.

However most positions between P1 and P4 showed intermediate cleavage, complicating structure determination. Dox-dependent reductions in band intensity were observed in corresponding regions of both aptamers (gray arrows). One base, P1MC iv. A23, showed a modest increase in cleavage upon treatment with dox (green arrow). Regions adjacent to the base of stem P1 and containing essential mutations for dox aptamer binding showed the greatest degree of protection upon incubation with dox, with 50 nM P2MC RNA showing reduced cleavage at these positions. Together these results show clear evidence of structural changes upon dox binding as well as the formation of stems P1 and P4.

The ability to truncate both stem P1 and the final two 3' bases of P2MC suggested that these sites could be modified to serve as immobilization sites for more precise binding studies such as SPR^{37, 38}. Biotinylated cytidine was ligated to P1MC, P2MC, library template, and dead tet RNA 3' termini without alteration of binding to 100 nM dox/tet in a gel fluorescence assay (data not shown). SPR results showed both dox aptamers to bind with $K_d \approx 7$ nM (Figure 6A-B). Dox aptamers exhibit fast dissociation kinetics compared to the tet aptamer library template (Figure 6C). The library template was found to bind tet with $K_d = 0.780 \pm 1.76$ nM, almost exactly the value determined by ITC for the tet aptamer without constant regions (770 ± 110 pM)¹⁶. Control experiments failed to show dox binding to the dead tet aptamer or tet binding to P2MC (Figure S4).

DISCUSSION

High-affinity RNA aptamers to dox were isolated following four selection rounds on dox affinity resin. More sensitive methods for detecting enrichment (RT-qPCR, NGS) have been shown to enable identification of aptamer species in fewer rounds compared to traditional methods; the use of scaffolded libraries likely also contributed to rapid dox aptamer enrichment^{18, 21, 33, 34}. Further optimizing SELEX by limiting randomization sites and rates ensured that the enriched dox aptamers were more highly represented in starting pools derived from tet aptamer-scaffolded libraries compared to libraries with less focused randomization. To minimize the emergence of G-repeat expansions observed with pilot efforts using Taq DNA polymerase, we amplified libraries with two additional polymerases, Q5 Hot Start II and Phusion Hot Start DNA polymerases. As we previously observed, some G-repeat expansions were observed in selection rounds 3 and 4 with Taq polymerase, and 5' poly-A repeats were enriched with both Taq and Phu libraries following round 4. Overall, we found that the Q5 Hot Start II DNA polymerase performed best during selections, and that scaffolded libraries outperformed more highly randomized libraries.

The most enriched dox aptamers following selection rounds 3 and 4 contained 12-14 substitutions or deletions, all located within the 29 randomized positions. The high-affinity P2MC consensus aptamer contains 11 mutations, 7 of which were found to be essential for 500 nM dox binding (Figures 2D and 4B). Apart from A37G none of these mutations are completely conserved within selected aptamers. In the tet aptamer template, A37 is localized to the P3 loop, shows strong protection upon tet binding, and interacts with A74, C75, and A76 to help form the tet binding site (Figure S5)²². Like A37C, the essential A32C mutation is also localized to stem P3 of the tet aptamer library template. A32 interacts with A22, encapsulating an Mg^{2+} ion to help form the tet aptamer binding site. Mutation of this

position to a smaller C nucleotide would alter or abolish this interaction, helping to facilitate structural rearrangement from the tet aptamer structure. Neighboring bases G39 and A40, along with U77, A78, and C79, contact tet directly. In P2MC the A37G, A74C, C75A, and A76G mutations are all essential, but in-line probing revealed dox-dependent protection of bases 74-76 only (Figures 5A-B). Positions 77-79 showed no protection upon dox binding, contain no essential mutations, and are localized to short stem-loops in several predicted structures, suggesting a reduced role in ligand interactions compared to the tet aptamer. Protection of P2MC positions 40-45 by dox during in-line probing, as well as the presence of essential mutations C42G and C45U, indicate that this region is also important for binding. Bases 42-45 are sequestered in the tet aptamer P3 stem region and show modest protection by tet during in-line probing; however stem P3 is not predicted to form in P2MC, possibly due in part to the disruptive effect of the C42G mutation. Protection upon ligand binding is observed at positions 40-41 in both tet and dox aptamers. Together these results suggest a rearrangement of the ligand binding site in P2MC such that dox interacts with bases 74-76 and 40-45, while interactions with former P3 loop nucleotides 31-39 are altered or abolished.

Structure probing revealed dramatic rearrangements outside of the ligand binding site, with loss of stem P3 and formation of stem P4 via base-pairing to the 5' constant region. Porter *et al.* found that varying the choice of reverse transcriptase led to different degrees of library scaffold rearrangement, as low-fidelity RT enzymes introduce mutations in non-randomized scaffolding regions. To circumvent this, dox aptamer selections were performed using the high-fidelity TGIRT-III reverse transcriptase^{21, 39} and mutations in non-randomized regions were rare in both starting libraries and selection pools. In light of this, rearrangements from the tet aptamer scaffold appear to be caused by interactions with the installed constant regions rather than mutations at non-randomized scaffold positions. Probing and mutagenesis studies clearly demonstrate both the incorporation of 5' constant region nucleotides into stem P4 and the preservation of stem P1. Structures predicted by free energy minimization, Multilign, and Turbofold all show several possible configurations for bases 30-45 and 72-98 in P2MC and enriched aptamers. However, deletion studies and structure prediction suggest that incorporation of both constant regions, as well as formation of stem P4, are necessary for binding of all dox aptamers studied here.

All dox aptamers formed several higher-running, non-binding bands when folded at high RNA concentrations, complicating binding studies and suggesting multimerization (Figure S3B). However, even at low RNA concentrations all dox aptamers formed a second, higher, non-dox-colocalizing band which is partially depleted by the presence of dox during folding (Figures 3E, S3C). Small amounts of dox colocalized with this band in native PAGE of 500 nM P2MC following prolonged binding times (Figure 3D). Comigration of this band with tet aptamer RNA suggests possible formation of tet aptamer-like secondary structure, although additional experiments are required to validate this hypothesis. The P2MC consensus sequence forms less of this higher-running, non-colocalizing structure relative to enriched sequences, inviting further mutagenesis experiments to characterize and potentially improve dox aptamer folding⁴⁰. P2MC's preference for lower band formation when folded in the absence of dox may explain its higher sensitivity during FTS, as more binding-competent RNAs would have been introduced during each titration (Figure 3C). Native PAGE and

probing experiments performed using 50 nM dox aptamer RNA suggest a low-nanomolar dissociation constant for dox binding to RNA in the lower band. Attempts to perform biolayer interferometry using 25 nM P2MC RNA showed no detectable binding to dox concentrations up to 1 μ M despite efficient RNA loading onto sensor heads⁴¹. This may be due to the small size of dox and/or limited conformational changes upon dox binding, a possibility supported by the results of in-line probing with and without dox.

SPR was able to detect dox aptamer binding, albeit with predictably small responses due to the size of dox compared to aptamer RNA^{37, 38}. Response magnitudes were low for dox aptamers compared to the tet aptamer despite similar RNA loading, possibly due to greater proportions of nonbinding RNA immobilized on sensor chips. If so this reflects the results from FTS binding studies, but still allowed affinity measurements for the binding-competent population. SPR results also showed that dox aptamers have fast off-rates compared to the parent tet aptamer. Use of affinity elutions selects strongly for specific binders, but also biases selection pools toward molecules with fast dissociation kinetics. The enriched dox aptamers were selected to adhere to dox columns for up to 100 min of buffer washes, providing at least a minimum of selective pressure for stable binding; adjusting the length and sequence of affinity elution steps and extended column binding times might have improved results further³⁰. Aptamers enriched in three parallel selection pools using affinity elution all showed obvious structural similarities; nonspecific elution of column-bound RNA using EDTA, urea, or heat may have permitted isolation of more diverse species with more stable binding to dox, but also may have enriched species which preferred column-immobilized to free dox.

Binding-induced fluorescence amplification makes dox aptamers immediately useful as biosensors. However, in-vitro selected aptamers often fail to perform properly in cellular environments due to differing folding conditions; the tet aptamer functions well in mammalian cells, but the structural rearrangement observed in these dox aptamers may lead to misfolding in this environment. We anticipate that determination of the 3D structure of a dox-aptamer complex will precisely define the nature of the partial structural rearrangement from the tet aptamer scaffold. Further work in characterizing dox aptamer folding and binding properties in and out of cellular environments may assist in stabilizing dox aptamer secondary structure. This information will provide information relevant to future selections from scaffolded libraries, as well as enabling development of dox-dependent riboswitches in conjunction with additional rounds of conformational change based SELEX.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGEMENTS

This work was supported by NIH grants NIH DP1 DA043912 and U19 AI149646 (M.F.).

REFERENCES

- [1]. Adachi T, and Nakamura Y (2019) Aptamers: A Review of Their Chemical Properties and Modifications for Therapeutic Application, *Molecules* 24, 4229–4242.
- [2]. Banerjee J, and Nilsen-Hamilton M (2013) Aptamers: multifunctional molecules for biomedical research, *Journal of molecular medicine (Berlin, Germany)* 91, 1333–1342.
- [3]. Tuerk C, and Gold L (1990) Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase, *Science* 249, 505–510. [PubMed: 2200121]
- [4]. Ellington AD, and Szostak JW (1990) In vitro selection of RNA molecules that bind specific ligands, *Nature* 346, 818–822. [PubMed: 1697402]
- [5]. Robertson DL, and Joyce GF (1990) Selection in vitro of an RNA enzyme that specifically cleaves single-stranded DNA, *Nature* 344, 467–468. [PubMed: 1690861]
- [6]. Karunanayake Mudiyanse A, Wu R, Leon-Duque MA, Ren K, and You M (2019) “Second-generation” fluorogenic RNA-based sensors, *Methods (San Diego, Calif.)* 161, 24–34.
- [7]. Pfeiffer F, and Mayer G (2016) Selection and Biosensor Application of Aptamers for Small Molecules, *Frontiers in Chemistry* 4, 1–21. [PubMed: 26835446]
- [8]. Chen C.-h. B., Dellamaggiore KR, Ouellette CP, Sedano CD, Lizadjohry M, Chernis, Gonzales M, Baltasar FE, Fan AL, Myerowitz R, and Neufeld EF (2008) Aptamer-based endocytosis of a lysosomal enzyme, *Proc Natl Acad Sci U S A* 105, 15908–15913. [PubMed: 18838694]
- [9]. Huang Y-F, Shangguan D, Liu H, Phillips JA, Zhang X, Chen Y, and Tan W (2009) Molecular Assembly of an Aptamer–Drug Conjugate for Targeted Drug Delivery to Tumor Cells, *ChemBioChem* 10, 862–868. [PubMed: 19253922]
- [10]. Nimjee SM, White RR, Becker RC, and Sullenger BA (2017) Aptamers as Therapeutics, *Annu Rev Pharmacol Toxicol* 57, 61–79. [PubMed: 28061688]
- [11]. Mou H, Zhong G, Gardner MR, Wang H, Wang YW, Cheng D, and Farzan M (2018) Conditional Regulation of Gene Expression by Ligand-Induced Occlusion of a MicroRNA Target Sequence, *Molecular therapy : the journal of the American Society of Gene Therapy* 26, 1277–1286. [PubMed: 29567311]
- [12]. Tang J, and Breaker RR (1997) Rational design of allosteric ribozymes, *Chemistry & Biology* 4, 453–459. [PubMed: 9224568]
- [13]. Weigand JE, and Suess B (2007) Tetracycline aptamer-controlled regulation of pre-mRNA splicing in yeast, *Nucleic acids research* 35, 4179–4185. [PubMed: 17567606]
- [14]. Win MN, and Smolke CD (2007) A modular and extensible RNA-based gene-regulatory platform for engineering cellular function, *Proceedings of the National Academy of Sciences* 104, 14283–14288.
- [15]. Zhong G, Wang H, Bailey CC, Gao G, and Farzan M (2016) Rational design of aptazyme riboswitches for efficient control of gene expression in mammalian cells, *eLife* 5, 1–17.
- [16]. Muller M, Weigand JE, Weichenrieder O, and Suess B (2006) Thermodynamic characterization of an engineered tetracycline-binding riboswitch, *Nucleic acids research* 34, 2607–2617. [PubMed: 16707663]
- [17]. Welling PG, Koch PA, Lau CC, and Craig WA (1977) Bioavailability of tetracycline and doxycycline in fasted and nonfasted subjects, *Antimicrobial agents and chemotherapy* 11, 462–469. [PubMed: 856000]
- [18]. Ozer A, Pagano JM, and Lis JT (2014) New Technologies Provide Quantum Changes in the Scale, Speed, and Success of SELEX Methods and Aptamer Characterization, *Molecular Therapy - Nucleic Acids* 3, e183. [PubMed: 25093707]
- [19]. Famulok M (1994) Molecular Recognition of Amino Acids by RNA-Aptamers: An L-Citrulline Binding RNA Motif and Its Evolution into an L-Arginine Binder, *Journal of the American Chemical Society* 116, 1698–1706.
- [20]. Groher F, and Suess B (2014) Synthetic riboswitches - A tool comes of age, *Biochimica et biophysica acta* 1839, 964–973. [PubMed: 24844178]

- [21]. Porter EB, Polaski JT, Morck MM, and Batey RT (2017) Recurrent RNA motifs as scaffolds for genetically encodable small-molecule biosensors, *Nat Chem Biol* 13, 295–301. [PubMed: 28092358]
- [22]. Xiao H, Edwards TE, and Ferré-D'Amaré AR (2008) Structural basis for specific, high-affinity tetracycline binding by an in vitro evolved aptamer and artificial riboswitch, *Chemistry & biology* 15, 1125–1137. [PubMed: 18940672]
- [23]. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, and Ferrin TE (2004) UCSF Chimera--a visualization system for exploratory research and analysis, *Journal of computational chemistry* 25, 1605–1612. [PubMed: 15264254]
- [24]. Bellaousov S, Reuter JS, Seetin MG, and Mathews DH (2013) RNAstructure: Web servers for RNA secondary structure prediction and analysis, *Nucleic acids research* 41, W471–W474. [PubMed: 23620284]
- [25]. Xu ZZ, and Mathews DH (2016) Secondary Structure Prediction of Single Sequences Using RNAstructure, *Methods in molecular biology* (Clifton, N.J.) 1490, 15–34.
- [26]. Xu Z, and Mathews DH (2011) Multilign: an algorithm to predict secondary structures conserved in multiple RNA sequences, *Bioinformatics* 27, 626–632. [PubMed: 21193521]
- [27]. Harmanci AO, Sharma G, and Mathews DH (2011) TurboFold: iterative probabilistic estimation of secondary structures for multiple RNA sequences, *BMC Bioinformatics* 12, 108–108. [PubMed: 21507242]
- [28]. Regulski EE, and Breaker RR (2008) In-line probing analysis of riboswitches, *Methods in molecular biology* (Clifton, N.J.) 419, 53–67.
- [29]. Berens C, Thain A, and Schroeder R (2001) A tetracycline-binding RNA aptamer, *Bioorganic & medicinal chemistry* 9, 2549–2556. [PubMed: 11557342]
- [30]. Groher F, and Suess B (2016) In vitro selection of antibiotic-binding aptamers, *Methods* (San Diego, Calif.) 106, 42–50.
- [31]. Bandwar RP, Jia Y, Stano NM, and Patel SS (2002) Kinetic and Thermodynamic Basis of Promoter Strength: Multiple Steps of Transcription Initiation by T7 RNA Polymerase Are Modulated by the Promoter Sequence, *Biochemistry* 41, 3586–3595. [PubMed: 11888274]
- [32]. Lorsch JR, and Szostak JW (1994) In vitro selection of RNA aptamers specific for cyanocobalamin, *Biochemistry* 33, 973–982. [PubMed: 7508262]
- [33]. Mencin N, Smuc T, Vranicar M, Mavri J, Hren M, Galesa K, Krkoc P, Ulrich H, and Solar B (2014) Optimization of SELEX: comparison of different methods for monitoring the progress of in vitro selection of aptamers, *Journal of pharmaceutical and biomedical analysis* 91, 151–159. [PubMed: 24463043]
- [34]. Schütze T, Wilhelm B, Greiner N, Braun H, Peter F, Mörl M, Erdmann VA, Lehrach H, Konthur Z, Menger M, Arndt PF, and Glöckler J (2011) Probing the SELEX Process with Next-Generation Sequencing, *PLOS ONE* 6, e29604. [PubMed: 22242135]
- [35]. Kivioja T, Vähärautio A, Karlsson K, Bonke M, Enge M, Linnarsson S, and Taipale J (2012) Counting absolute numbers of molecules using unique molecular identifiers, *Nature Methods* 9, 72–74.
- [36]. Cowperthwaite MC, and Ellington AD (2008) Bioinformatic analysis of the contribution of primer sequences to aptamer structures, *J Mol Evol* 67, 95–102. [PubMed: 18594898]
- [37]. Chang AL, McKeague M, and Smolke CD (2014) Facile characterization of aptamer kinetic and equilibrium binding properties using surface plasmon resonance, *Methods Enzymol* 549, 451–466. [PubMed: 25432760]
- [38]. Zhu Z, Feng M, Zuo L, Zhu Z, Wang F, Chen L, Li J, Shan G, and Luo S-Z (2015) An aptamer based surface plasmon resonance biosensor for the detection of ochratoxin A in wine and peanut oil, *Biosensors and Bioelectronics* 65, 320–326. [PubMed: 25461176]
- [39]. Mohr S, Ghanem E, Smith W, Sheeter D, Qin Y, King O, Polioudakis D, Iyer VR, Hunnicke-Smith S, Swamy S, Kuersten S, and Lambowitz AM (2013) Thermostable group II intron reverse transcriptase fusion proteins and their use in cDNA synthesis and next-generation RNA sequencing, *RNA* (New York, N.Y.) 19, 958–970.

- [40]. Carothers JM, Oestreich SC, and Szostak JW (2006) Aptamers selected for higher-affinity binding are not more specific for the target ligand, *J Am Chem Soc* 128, 7929–7937. [PubMed: 16771507]
- [41]. Lou X, Egli M, and Yang X (2016) Determining Functional Aptamer-Protein Interaction by Biolayer Interferometry, *Current Protocols in Nucleic Acid Chemistry* 67, 7.25.21–27.25.15.

Figure 1: Selection library design.

(A) Tetracycline (tet) and doxycycline (dox). Structural differences are highlighted in orange. (B) Secondary structure of the tet aptamer. Constant regions installed to enable reverse transcription and PCR amplification during SELEX are shown below in green. Bases are numbered according to both their positions within the tet aptamer (black), and also within the full-length library template incorporating constant regions (green). Cyan-fill circles indicate bases protected from RNase digestion as shown in panel D. (C) Crystal structure of the tet aptamer bound to 7-chloro-tet. Base colors correspond to Fig. 1B. A U1A cocrystallization protein bound to the P1 loop region has been omitted for clarity. Data from Xiao *et al.* [22]. (D) In-line probing of the tet aptamer. Band intensities correspond to RNA backbone cleavage rates after the indicated base. Cleavage rates are reduced in base-paired regions, or in regions protected upon ligand binding. Positions of G nucleotides revealed by RNase T1 digestion are labeled to the left. Randomization sites are highlighted in blue to the right, and the positions of stem loops P1, P2, and P3 are shown in black.

Figure 2: Dox aptamer SELEX.

(A) SELEX method. Selection columns containing dox- or ethanolamine (EA)-derivatized affinity resin are prepared and washed in selection buffer. Dox resin is blocked using 10 mg yeast tRNA to prevent nonspecific binding. 10^{15} library RNAs are incubated with EA resin to remove RNAs with affinities to the resin or linker domains. Preselected library RNA is then bound to the dox affinity resin. Nonbinding RNAs are removed in several wash steps, followed by elution with increasing concentrations of dox. Wash and eluent fractions are quantified by RT-qPCR. One or more elution fractions are reverse-transcribed, RT products quantified using qPCR, and the remaining RT products PCR amplified. PCR products are then transcribed to yield the selection pool for the subsequent round, with a sample taken for adapter ligation and NGS analysis. Stringency was increased throughout selections by increasing the number of wash steps from 3 to 10 and by introduction of counterselection on EA resin in rounds 3-4. (B) Elution profiles. RT-qPCR quantifications of final wash and elution fractions are shown. The RNA content of final washes increased throughout selections despite increased numbers of wash steps. Pronounced specific elution of RNA by 1 μ M dox was observed in all libraries during round 4. Quantification was not performed following round 1 to avoid loss of low-abundance RNA species. (C) Sequence enrichment. NGS analysis of starting libraries, as well as PCR products following each round, was

performed. Sets of 150,000 reads were quality-controlled for the presence of intact constant regions and minimum sequence length. Abundance was determined by the number of unique molecular identifiers (UMIs) associated with a given sequence to minimize artifacts associated with exponential amplification. Abundances of the most enriched sequences in round 3 (e.g. Taq 3-1) and round 4 (e.g. Taq 4-1) are shown. Phu 3-1 remained its library's most common sequence in both rounds 3 and 4. **(D)** Alignment of enriched sequences. Common patterns of mutations were observed in enriched RNAs, including total conservation of the A37G mutation. A consensus sequence, described as P2MC in the text, is also shown with conserved mutations highlighted in green.

Figure 3: Dox aptamer binding studies.

(A) Column elutions. Column binding, washes, and elutions were performed as in SELEX. Dox aptamers showed higher retention on dox resin than on tet resin, and showed specific elution in response to dox but not tet. The tet aptamer exhibited opposite behavior, while a binding-deficient tet aptamer control was retained poorly on both resins. Wash fractions of enriched sequences bound to dox resin contained higher amounts of RNA than wash fractions from the tet aptamer bound to tet resin. The tet aptamer included 5' and 3' constant regions in all binding experiments shown here. (B) Dox fluorescent titration spectroscopy (FTS). Aptamer RNA was folded at 10 μ M, diluted to various concentrations in 20 nM dox, and incubated for 1h prior to fluorimetry. Titration of dox with dox aptamer RNAs produced aptamer-dependent, dose-dependent increases in dox fluorescence emission, with the P2MC dox aptamer showing the highest sensitivity. (C) Tet FTS. The tet aptamer amplified fluorescence emission of 1 nM tet, but amplification of dox fluorescence was low and only occurred in the presence of high tet aptamer concentrations. All dox aptamers failed to amplify tet fluorescence. (D) Native PAGE of binding timecourses. 500 nM aptamer RNA was folded in the absence of dox, then incubated with 500 nM dox for the indicated times prior to native PAGE. All dox aptamers were observed to form two bands, with dox fluorescence colocalized to the lower band. Prolonged incubation did not alter the

distribution of RNA between bands. P2MC's upper band colocalized with small amounts of dox following prolonged incubation. P2MC showed the lowest rates of upper band formation out of all RNAs studied, while Q5 3-1 showed the highest. (E) Native PAGE of ligand-assisted folding experiments. 500 nM aptamer RNA was folded in the presence of increasing concentrations of dox. Incomplete depletion of the upper band was observed at high dox concentrations.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Figure 4: Dox aptamer structure determination.

(A) Predicted secondary structure for P2MC. Colors are as in Fig. 1, except where they reflect information from Fig. 4B and 4G-H. (B) Reversion of enriched mutations. Native PAGE binding experiments were performed as in Fig. 3E. 7 of 11 mutations were found to be essential for dox binding. Reversions also altered banding patterns. (C) Constant region truncation studies. Removal of constant regions installed as RT and/or PCR primer binding sites for SELEX completely abolished binding to 500 nM dox (upper panel). Formation of a single band running alongside truncated tet aptamer RNA was also observed in all truncated constructs except Q5 3-1. Partial truncations from 5' and 3' termini also prevented dox binding, except for deletion of positions 99-100 (lower panel). (D) Summary of constant region mutagenesis. Experiments were performed as in 4B. Mutations within the 5' constant region reflect deletion experiments and their predicted conformations in stem P4. (E) Minimized dox aptamer. Truncation and loop substitution at P2MC stem-loops P1 and P4 did not disrupt binding to 500 nM dox. (F) P1 dox aptamers. P1 dox aptamers were generated by deletion of the P1 loop and positions 99-100, followed by bridging of positions 1 and 98 with a variety of sequences (i. – vi.) Aptamers bridged with the P1 stem or two tetraloop sequences showed binding to 500 nM dox. The CUUG bridge construct showed binding comparable to aptamers enriched during selections (Fig. 3B).

Figure 5: In-line probing of dox aptamers.

(A) Full-length probing gels. Experiments were performed as in 1D. Stems P1 and P4 are clearly visible, with intermediate cleavage rates at other positions. Dox-dependent protection is observed at positions surrounding the base of stem P1. Bases exhibiting dox-dependent protection are indicated by gray arrows; a single base showing dox-dependent deprotection is indicated by a green arrow. (B) Detail view of probing gels from A. Contrast has been adjusted to allow better visualization of upper bands.

Figure 6: SPR Binding Studies.

(A) P2MC Dox Binding. P2MC RNA was 3' end-labeled using T4 ssRNA ligase and pCp-biotin, bound to an SA-coated sensor chip, and titrated with two series of successive 60s injections of 6.25 – 100 nM dox. Each injection was followed by a 600s dissociation step. A sensorgram is shown with signal from two prior, buffer-only injection series subtracted (left panel). A kinetic curve was fit using Biacore Insight Evaluation Software and used to provide the estimated binding constants, with $R_{\max} = 3.967 \text{ RU}$ and $X^2 = 0.0884 \text{ RU}^2$. Blank-subtracted sensorgram heights were recorded 4s following each injection (right panel). A binding curve was fit using GraphPad Prism's one-site specific binding method ($R^2 = 0.6945$) to yield the indicated dissociation constant. (B) P1MC Dox Binding. Experiments were performed as in (A), apart from substitution of a CAP chip and capture reagent for an SA chip. Kinetic curve fitting (left panel) had $R_{\max} = 10.01 \text{ RU}$ and $X^2 = 6.06 \text{ RU}^2$. Baseline decay following injections may explain the high error of this fit. Peak heights were analyzed as in (A), with $R^2 = 0.6544$. (C) Library Template Tet Binding. Library template RNA consisted of the tet aptamer with 5' and 3' SELEX constant regions. Experiments were performed as in (B) except injections used 6.25 – 100 nM tet. Kinetic curve fitting had $R_{\max} = 23.13 \text{ RU}$ and $X^2 = 1.10 \text{ RU}^2$. Peak height curve fitting had $R^2 = 0.9681$.