

# Selecting RNA aptamers for synthetic biology: investigating magnesium dependence and predicting binding affinity

James M. Carothers<sup>1</sup>, Jonathan A. Goler<sup>1,2</sup>, Yuvraaj Kapoor<sup>1</sup>, Lesley Lara<sup>1</sup> and Jay D. Keasling<sup>1,2,3,\*</sup>

<sup>1</sup>California Institute for Quantitative Biosciences and Berkeley Center for Synthetic Biology, University of California, Berkeley, CA 94720 and Joint BioEnergy Institute, Emeryville, CA 94608 <sup>2</sup>Synthetic Biology Engineering Research Center, University of California, Berkeley, CA 94720 and <sup>3</sup>Departments of Bioengineering and Chemical Engineering, University of California, Berkeley and Physical Biosciences Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA

Received November 19, 2009; Revised January 12, 2010; Accepted January 19, 2010

## ABSTRACT

The ability to generate RNA aptamers for synthetic biology using *in vitro* selection depends on the informational complexity (IC) needed to specify functional structures that bind target ligands with desired affinities in physiological concentrations of magnesium. We investigate how selection for high-affinity aptamers is constrained by chemical properties of the ligand and the need to bind in low magnesium. We select two sets of RNA aptamers that bind planar ligands with dissociation constants ( $K_d$ s) ranging from 65 nM to 100  $\mu$ M in physiological buffer conditions. Aptamers selected to bind the non-proteinogenic amino acid, *p*-amino phenylalanine (pAF), are larger and more informationally complex (i.e., rarer in a pool of random sequences) than aptamers selected to bind a larger fluorescent dye, tetramethylrhodamine (TMR). Interestingly, tighter binding aptamers show less dependence on magnesium than weaker-binding aptamers. Thus, selection for high-affinity binding may automatically lead to structures that are functional in physiological conditions (1–2.5 mM  $Mg^{2+}$ ). We hypothesize that selection for high-affinity binding in physiological conditions is primarily constrained by ligand characteristics such as molecular weight (MW) and the number of rotatable bonds. We suggest that it may be possible to estimate aptamer–ligand affinities and predict whether a particular aptamer-based design goal is achievable before performing the selection.

## INTRODUCTION

RNA structures have emerged as important tools for programming gene expression for metabolic engineering and synthetic biology (1–5). Dynamic genetic controls responsive to small molecule ligands can be constructed from RNA aptamers generated with *in vitro* selection (6–8). In principle, aptamer-based genetic controls could be used as *in situ* readouts of product formation or as feedback mechanisms for regulating metabolic flux through engineered pathways in response to any desired metabolite (9).

Several different mechanistic approaches have been devised to couple fluctuations in the concentration of a metabolite to changes in gene expression using RNA aptamers. For example, a synthetic riboswitch that anti-attenuates translation in *Bacillus subtilis* was made by incorporating an *in vitro* selected aptamer that binds theophylline (10) into the region just upstream of the Shine–Dalgarno (SD) sequence of a reporter gene (11,12). Adding ligand to the cell activates translation more than 3-fold because ligand-binding stabilizes an aptamer conformation that reduces the amount of secondary structure involving the SD sequence. A screen for synthetic gene-regulatory aptamers identified theophylline aptamer structures that employ a similar mechanism to activate translation up to 35-fold upon the addition of ligand (13). Gene expression has also been modulated by controlling mRNA stability and access to the ribosome binding site using synthetic aptamer-controlled self-cleaving RNA structures ('aptazymes') (14–18).

Regardless of the specific genetic control mechanism to be employed, it is impractical to generate or select high-affinity aptamers directly *in vivo*. This is because the frequency of functional ligand-binding structures in

\*To whom correspondence should be addressed. Tel: 510 4952626; Fax: 510 4952630; Email: keasling@berkeley.edu

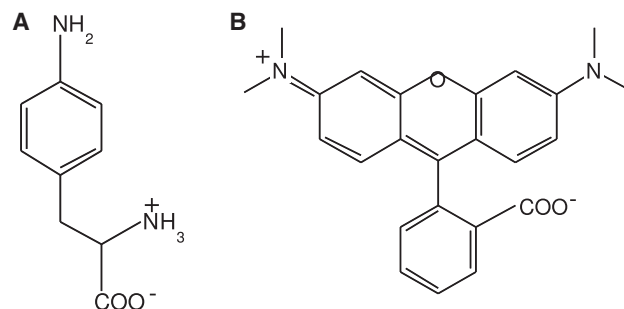
The authors wish it to be known that, in their opinion, the first two authors should be regarded as joint First Authors.

RNA sequence space is small (from  $\sim 1$  in  $10^9$  for low-affinity aptamers to 1 in  $10^{20}$  for very high-affinity aptamers (19)) compared to the size of the sequence library that can be directly searched due to the limit imposed by transformation efficiency,  $\sim 10^9$  in an organism such as *Escherichia coli* (20). However, *in vitro* affinity chromatography selections can be used to search libraries containing  $>10^{15}$  unique molecules, providing access to aptamers that bind target ligands with very high affinity.

To date, a large number of the examples (12,17,18,21,22) of synthetic aptamer-based control of living systems have used the theophylline aptamer, first selected by Jenison *et al.* (10) almost 15 years ago. Although that selection was not intended to produce structures for use *in vivo*, the theophylline aptamer has been repeatedly employed because it is small (15 nt binding site), discriminates between two closely related cell-permeable molecules (theophylline and caffeine), and binds with high affinity in physiological concentrations of magnesium (the aptamer–ligand dissociation constant ( $k_d$ ) is 400 nM in 1 mM  $Mg^{2+}$ ). The latter point is significant because RNA function tends to decrease at lower concentrations of magnesium (5), such as those typically found intracellularly (0.5 mM to 1.5 mM (23)). Since the RNA backbone is polyanionic, the negative–negative charge repulsion may inhibit formation of complex structures in the absence of counterions. Divalent cations like  $Mg^{2+}$  can help screen this charge and enable the backbone to fold closer to itself (24).

Many selections, generally carried out in 5–10 mM  $Mg^{2+}$ , have isolated aptamers that bind targets such as amino acids, proteins and nucleotides (8). While some of the resulting structures may bind in physiological conditions (25–28), we are interested in knowing whether lowering the magnesium concentration increases the informational complexities (ICs) (i.e. lowers the probability of finding a functional molecule in sequence space) of functional aptamers such that the probability of selecting active molecules becomes very low (19).

Here we investigated both how the need to bind in physiological magnesium concentrations affect selection and aptamer affinity and how ligand MW and chemical characteristics affect potential aptamer affinity. We addressed these questions by *in vitro* selecting aptamers that bind two different cell-permeable small molecules, *p*-aminophenylalanine (pAF) and tetramethylrhodamine (TMR), (Figure 1). pAF is a non-proteinogenic amino acid (MW = 162 g/mol) that can be produced heterologously *in vivo* in *E. coli* from *Streptomyces venezuelae* gene products and is a precursor for making the pristanamycin family of antibiotics (29). With a small, relatively simple structure, it is representative of other metabolic building blocks; it has four potential hydrogen bond-forming moieties and a single aromatic ring. TMR is a larger fluorogenic molecule (MW = 343 g/mol) with four aromatic rings. TMR is presumably an easier binding target for RNA than pAF and is representative of more complicated polycyclic organic compounds that might be found in the later stages of an engineered biosynthetic pathway (Figure 1).



**Figure 1.** Chemical structures of target ligands. (A) pAF; MW = 180 g/mol and (B) TMR; MW = 343 g/mol.

We conducted parallel selections for pAF aptamers with four different concentrations of magnesium (1, 2.5, 5 and 10 mM). After 10 rounds, all of the selections produced molecules that bound pAF, though there were fewer different sequences in the 1 and 2.5 mM  $Mg^{2+}$  selections than the 5 and 10 mM selections. Two of the pAF structures were isolated and reselected to generate IC and affinity profiles. We selected four different TMR aptamers in 1 mM  $Mg^{2+}$  and characterized them in the same way. We then compared the IC and binding affinities of these newly selected aptamers to structures in the literature. We found that aptamers selected in physiological concentrations of magnesium have ICs consistent with those selected in more permissive conditions and show that aptamer IC versus ligand affinity relationships can, to a first approximation, be grouped according to ligand molecular weight (MW). Our results suggest that it may be possible to predict aptamer affinities for a target ligand on the basis of MW and chemical properties before performing the selection. We discuss the implications of these findings for the application of *in vitro* selected RNA aptamers as genetic controls for synthetic biological systems.

## MATERIALS AND METHODS

### pAF and TMR aptamer *in vitro* selection

**Pool design.** Libraries of  $\sim 10^{14}$  sequences were chemically synthesized as DNA oligonucleotides (Keck Facility, New Haven, CT, USA) (Supplementary Figure 1). The pAF selection used a library containing a designed 12-base stem-tetraloop (30) flanked on each side by 28 random nucleotides (STR pool). The library pool for the TMR aptamer selection was comprised of two equal portions. The RAN2 portion of the pool had a straight run of 72 random nucleotides while the structured portion of pool (STR2) contained a designed 12 base stem-tetraloop flanked on each side by 30 nt of completely random sequence. All sequences contained 5' 40 nt T7 promoter/PCR primer binding sites and 3' Reverse Transcription and PCR (RT-PCR) primer binding sites (19).

Reselection pools were created from the sequences isolated in the initial selections, with new RT and PCR primers designed with  $T_m = 55^\circ C$ . The aptamer region was doped with a mixture of 79% original base, and the

other bases at a coupling-efficiency balanced ratio of 7% each.

**In vitro transcription.** Pools were transcribed with T7 RNA polymerase, in transcription buffer containing 20 mM each NTP, 25 mM  $Mg^{2+}$ , 40 mM Tris (pH 7.8), 2.5 mM spermidine, 10 mM DTT, 0.01% Triton X-100, 50 U T7 RNA Polymerase, 10 U/ $\mu$ l thermostable inorganic pyrophosphatase (TIPP, New England Biolabs, UK), 200 U/ $\mu$ l RNAsin (Promega) and, for the selection, trace  $\alpha^{32}P$ -ATP (Perkin Elmer). The RNA was purified on a PAGE gel, and ethanol (EtOH) precipitated with potassium chloride (KCl).

**Selection matrix preparation.** 200  $\mu$ l selection columns were made by coupling the carboxy terminus of Fmoc-p-aminophenylalanine-Boc (Bachem AG) to the amine-terminus of NovaBiochem Amino PEGA resin (amino polyethylene glycol) in dimethyl formamide (DMF). Unreacted amine groups were blocked by acetylation (31) and the Fmoc and Boc protecting groups were removed using established procedures (32) to produce  $\sim$ 0.5 mM pAF columns. TMR matrix was similarly prepared using 5-carboxy TMR (EMD Biosciences), but without the deprotection steps.

**In vitro selection.** Selection columns were washed with  $H_2O$ , capped, inverted 2–3 times, and then pre-soaked in 4 ml selection buffer (SB: 130 mM K-glutamate (pH 7.5), 15 mM NaCl, 10 mM DTT and  $MgCl_2$  at various levels) for 15–30 min. This step was important because the swelling characteristics of the PEGA resin differ between the DMF solvent and water-based solvents. The RNA was resuspended in 200  $\mu$ l nuclease-free water and incubated at 37°C for 5 min. 2 $\times$  SB and 4  $\mu$ l each 200  $\mu$ M acetylated tRNAs and 100 $\times$  BSA (10 mg/ml) were added to the RNA, and the mixture was agitated, applied to the column and incubated for 5–30 min.

Columns were then washed with 20 column volumes (CVs) SB, 4 CV at a time. For the pAF selections, elutions were performed in 2 CV fractions of 1 mM pAF in SB. After 6 CV, the columns were capped and 2 CV of 1 mM pAF in SB was incubated for 30–60 min, followed by 4 CV of further wash. For the TMR selections, elutions were performed in 2 CV fractions of 0.5 mM TMR in SB. As above, after 6 CV, the columns were capped and 2 CV of 0.5 mM TMR in SB was incubated for 30–60 min, followed by 4 CV of further wash. Fractions of  $^{32}P$ -labeled RNA were counted on a Beckman Coulter LS 6500 liquid scintillation counter without scintillation fluid. The column resin was also counted. All selections were conducted at room temperature. The elution fractions were pooled, and the RNA in these fractions was de-salted using NAP-25 Sephadex columns (GE Healthcare), EtOH precipitated with KCl and amplified by RT-PCR using Ready-To-Go beads (GE Healthcare). The resultant PCR product was either transcribed for the next round of selection or cloned via TOPO cloning (Invitrogen). At the end of the selection (see ‘Result section’), 60–160 TOPO-cloned samples from each pool were sequenced and aligned using ClustalW (33).

## Binding constants and IC

**Binding assays.** Apparent  $K_{d}$ s were determined with the spin-filter method (10) using at least three independent measurements taken across the linear range (40–60% binding). Briefly, aptamer RNA was arranged in dilution series in 100  $\mu$ l volumes. For pAF binding assays, samples were heated to 85°C for 5 min. 100  $\mu$ l of 2 $\times$  SB containing 0.5 nM  $^3H$ -labeled pAF (Moravek Biochemicals) and either 1–10 mM  $Mg^{2+}$  was added and samples equilibrated at room temperature for at least 2 h (sufficient to reach equilibrium, data not shown). Samples were transferred to Microcon YM-10 spin columns (Millipore), spun for 5 s at 10 000g to saturate the membrane. Collection tubes were replaced, and samples were again spun for 60 s. 15  $\mu$ l samples were collected from the top and bottom of the column, added to 4 ml of EcoScint fluid, and counted for  $^3H$  on a Beckman LS 6500 for 3 min to determine the fraction of pAF that was bound to the aptamer.

The TMR binding assays were performed in a similar manner, with the following modifications. The RNA was heated to 37°C for 5 min and the second spin was 2 min. The amount of TMR bound to the aptamer was determined by measuring the relative fluorescence of 25  $\mu$ l samples taken from the top and bottom of the Microcon YM-10 spin columns. Fluorescence was measured using a TECAN Safire with excitation wavelength of 550 nm and emission wavelength of 590 nm.

**Calculating ICs.** The IC of each aptamer was computed using the reselection sequence data and predicted secondary structures (19). The IC is given by the sum of the amount of information needed to specify each position in the recognition loops and Watson–Crick (W–C) base-paired stems. Positions in recognition loops that are invariant require two bits of information, two-base varying positions require one bit, and four-base varying positions require zero bits to specify their identities. For positions in stems, each nucleotide involved in a W–C base-pair is specified by  $\sim$ 1 bit (i.e. 2 bits for each W–C pair). The observed frequency,  $F_i$ , of each base was normalized and adjusted in order to calculate the aptamer ICs in terms of random sequence space (19).

**Correlation coefficients and regression.** Non-parametric tests of association were applied because these are relatively insensitive to data that deviates from a normal distribution (19). Spearman’s rank correlation test was carried out using the ‘pspearman’ package for R (R Foundation for Statistical Computing, Vienna, Austria). Confidence intervals of 95% were used for all rankings; significance was determined under the null hypothesis of no correlation with a one-tailed test.  $P$ -values were calculated from the test statistic as described (19). Non-parametric line-fits were performed using Kendall’s robust line fit provided in the ‘mblm’ package for R. Bootstrap confidence intervals were computed for the non-parametric line fits using the bias-corrected percentile method with 10 000 replicates (34).



**Estimating functional group binding energies.** The potential binding energy contributions of ligand functional groups were approximated using intrinsic binding energies derived through multiple regression analysis by Andrews *et al.* (35). The potential contribution of individual functional groups was obtained by multiplying the empirical ' $E_x$ ' value for each type of functional group at 298 K (35) by the number of moieties present in the ligand (see Supplementary Table 3). Functional group classifications were taken from The CAS Registry accessed via SciFinder Scholar (American Chemical Society, Columbus, OH, USA) (also see Supplementary Table S3).

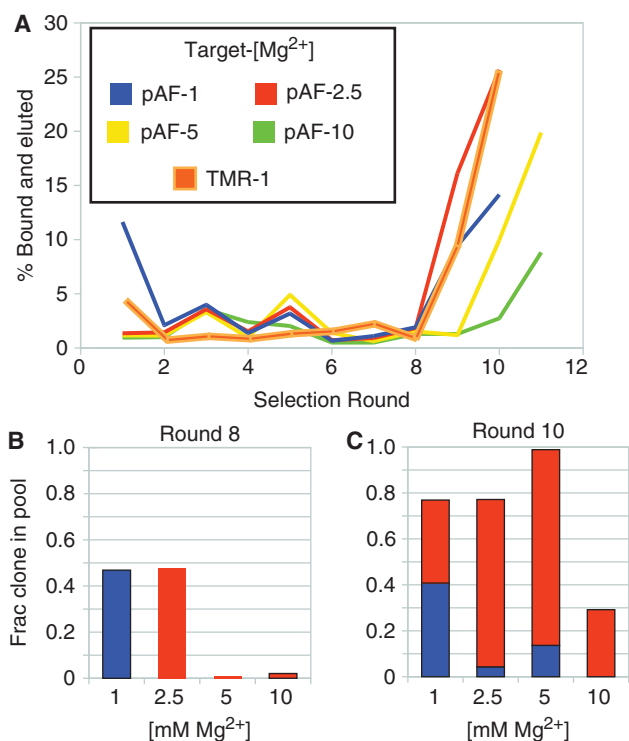
## RESULTS

### *In vitro* selection of pAF aptamers with varying magnesium concentrations

We wanted to test how much of a constraint the concentration of magnesium is on selecting high-affinity aptamers to small molecule ligands. We carried out parallel selections for pAF-binding RNA aptamers in four different concentrations of magnesium: 1, 2.5, 5 and 10 mM. 1 and 2.5 mM magnesium concentrations were chosen because these are consistent with the amount of free magnesium that is present *in vivo*, whereas 5 and 10 mM magnesium are typical *in vitro* selection conditions (31,36–38). An initial pool of RNA was prepared by transcribing  $\sim 3 \times 10^{14}$  unique DNA templates designed to contain an internal stem–tetraloop flanked by stretches of completely random sequence ('Materials and Methods' section and Supplementary Figure S1). The pool of RNA was then divided into four equal parts to provide the starting material for each of the parallel selections.

Before Round 8, <5% of the RNA in any of the four selections bound to a pAF column, survived 20 column volumes of wash, and specifically eluted with 0.5 mM pAF in solution. After that point, the fraction of the RNA pools that bound and specifically eluted rapidly increased (Figure 2). By Round 10, 25% of the total RNA in the 2.5 mM magnesium selection exhibited binding and elution with 0.5 mM pAF, while 14%, 9% and 2% of the total RNA in the 1, 5 and 10 mM magnesium selections bound the column and eluted with pAF, respectively. The four pools were sequenced after Rounds 8 and 10.

In Round 10, the 1 mM magnesium selection was dominated by two different sequences, pAF-R1-1 and a less abundant pAF-4Z1. The 2.5 mM magnesium sample was almost entirely pAF-4Z1. Both the 5 and 10 mM samples had significant pAF-4Z1 populations. The 10 mM magnesium sample still contained significant diversity though many of those sequences probably do not bind well to the pAF ligand, as indicated by the small fraction of the total pool RNA (2%) that bound and eluted in Round 10. Overall, in the four pools combined, copies of pAF-4Z1 accounted for 56% of the available sequences, while pAF-R1-1 represented another 15% of the sequences. Two more sequences with multiple copies, pAF-2B and -4B, made up  $\sim 2\%$  of the combined pools. Together these four clones represented more than 70% of the total sequences (Supplementary Figure S2).



**Figure 2.** Course of pAF and TMR aptamer selections. (A) Fraction of total RNA surviving 20 CV of wash, by round, for each of the five selections performed. The concentration of magnesium in the selection buffer is indicated by the number following the name of the target ligand, in mM. The selection pools were sequenced after Rounds 8 and 10. Fraction of the Round 8 (B) and Round 10 (C) pAF selection pool sequences comprised of clones pAF-4Z1 (red) and pAF-R1-1 (blue) as a function of magnesium concentration in the selection buffer.

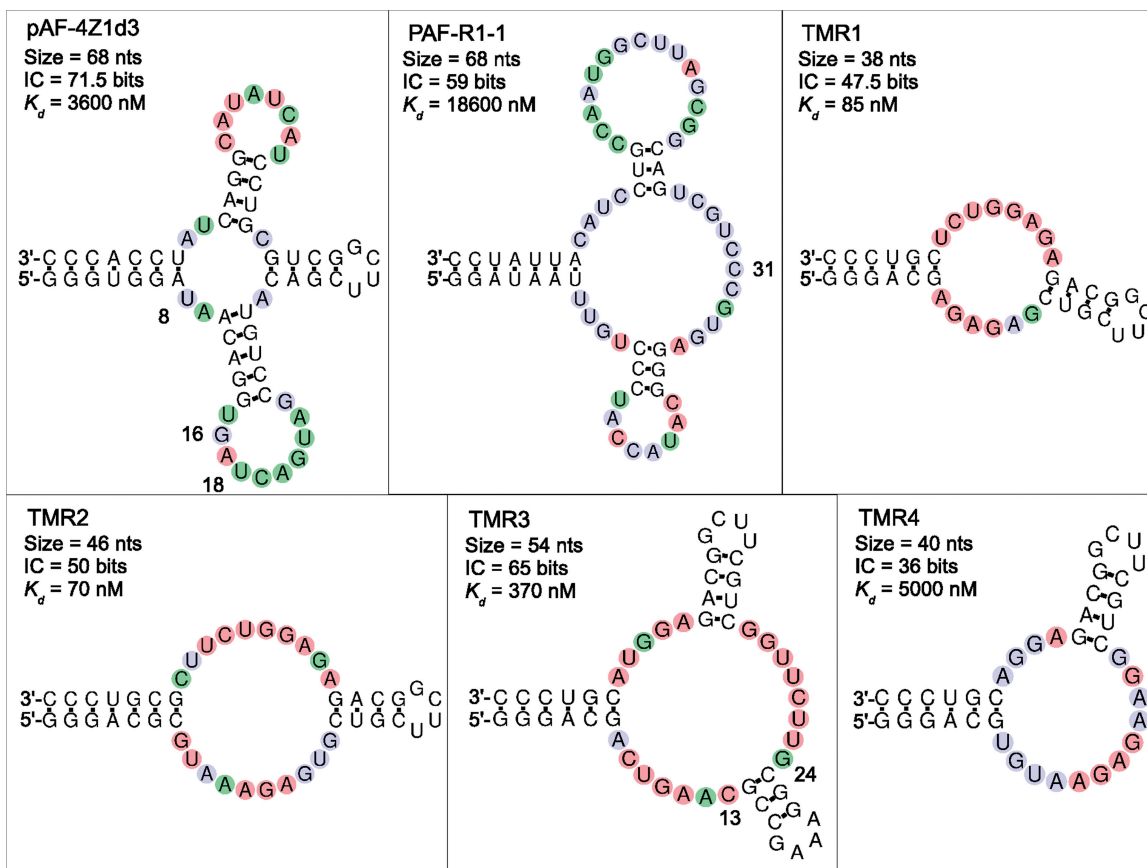
**Table 1.** ICs and affinities (apparent  $K_d$ s) of the pAF- and TMR-binding aptamers in 5 and 1 mM  $Mg^{2+}$

Aptamer	Size (nt)	IC (bits)	$K_d$ nM (5 mM $Mg^{2+}$ )	$K_d$ nM (1 mM $Mg^{2+}$ )
pAF-4Z1d3	68	71.5 $\pm$ 0.5	3600 $\pm$ 400	5400 $\pm$ 2600
pAF-R1-1	68	59.0 $\pm$ 1.5	18 600 $\pm$ 8400	103 000 $\pm$ 5800
TMR1	38	47.5 $\pm$ 1.0	85 $\pm$ 25	70 $\pm$ 15
TMR2	46	50.0 $\pm$ 1.0	70 $\pm$ 25	65 $\pm$ 20
TMR3	54	65.0 $\pm$ 2.5	370 $\pm$ 150	400 $\pm$ 230
TMR4	40	36.0 $\pm$ 1.0	5000 $\pm$ 1600	10 900 $\pm$ 200

The activities of the four most abundant RNAs isolated in Round 10 were measured in column-binding assays. Of the four clones, only pAF-R1-1 and -4Z1 exhibited significant binding. Approximately 29% of the pAF-R1-1 RNA and 19% of the pAF-4Z1 RNA bound to a pAF column, survived 20 column volumes of wash, and specifically eluted with pAF in solution. In contrast, <5% of the pAF-2B or pAF-4B RNA bound and specifically eluted (Supplementary Table S1). pAF-R1-1 and -4Z1 were chosen for further analysis (see below).

### *In vitro* selection of TMR aptamers in 1 mM $Mg^{2+}$

We anticipated that TMR would be a better binding target for RNA than pAF and that it should be possible to



**Figure 3.** Secondary structures of aptamers selected here. The size (in nucleotides), IC (in bits) and apparent  $K_d$  (in nanomolars) in 5 mM  $Mg^{2+}$  are shown for each aptamer. Red circles indicate positions that are invariant, green circles indicate positions that vary between two bases and blue circles indicate positions that can be any base. Watson-Crick (W-C) base-paired stems designed to flank the putative recognition bulges appear on the left in each secondary structure. Positions referenced in the text are numbered.

recover high-affinity TMR aptamers in 1 mM  $Mg^{2+}$  and possibly from an unstructured sequence pool. Therefore a single selection for TMR was conducted in 1 mM  $Mg^{2+}$  with a starting pool of RNA transcribed from an equal mixture of two populations totaling  $3 \times 10^{14}$  unique DNA templates. One portion contained a designed internal stem–tetraloop (STR2 library), as above, while the other contained a completely random region (RAN2 library) ('Materials and Methods' section).

After 10 rounds of selection, ~25% of the RNA bound to a column derivatized with 5-carboxy TMR, survived 20 column volumes of wash and specifically eluted with 0.5 mM TMR in solution (Figure 2A). At this point, the fraction of TMR aptamer pool RNA binding and eluting from the selection column was equal to or greater than the fractions of pAF aptamer pool RNAs that bound and eluted from the pAF selection columns. Despite similar column-binding profiles, sequencing showed that the output of the TMR aptamer Round 10 selection pool contained significantly more diversity than the pAF aptamer selection pools had contained at the same point. Specifically, 14 distinct clones comprised ~70% of the TMR aptamer selection pool compared to only 4 distinct clones that comprised ~70% of the combined pAF aptamer selection pools (Supplementary

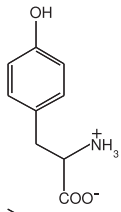
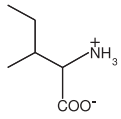
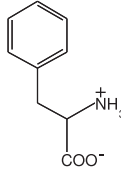
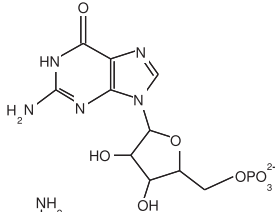
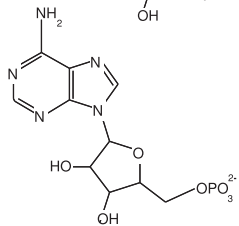
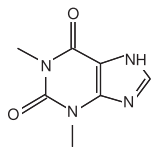
Figures S2 and S3). This is consistent with our expectation that there should be many more structural solutions to the problem of binding the higher MW TMR ligand than to the problem of binding the lower MW pAF ligand.

Ten of the 14 unique TMR aptamer selection pool sequences originated in the portion of the starting library containing the designed stem–tetraloop, as did all of the RNA sequences that showed substantial binding (>5% binding and specifically eluting with 0.5 mM TMR) in column-binding assays (Supplementary Figure S3 and Supplementary Table S2). Thus, the difference in the number of unique clones that bind TMR compared to the number of unique clones that bind pAF does not appear to be due to variations in the starting sequence libraries. Four of the TMR aptamers that exhibited substantial activity in the column-binding assays (TMR1, TMR2, TMR3 and TMR4) were arbitrarily chosen for further study (Figure 3).

#### Reselections to optimize binding affinity and measure IC

We performed reselections of the pAF and TMR-binding RNAs to identify functional sequence variants to optimize binding to the target ligand and to determine the amount of IC required to specify each of the structures.

**Table 2.** Ligand structures and ICs of aptamers in literature

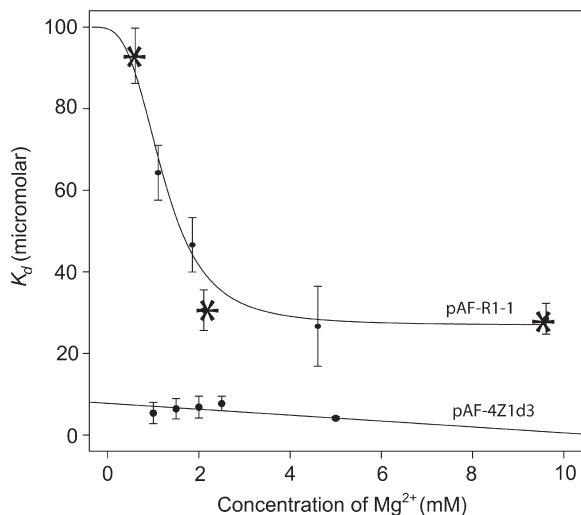
Ligand	Structure	MW (g/mol)	IC (bits)	$K_d$ nM (5 mM $Mg^{2+}$ )	No. of aptamers
Tyrosine (36)		181	75	23 000	1
Isoleucine (38,40)		131	33–38	600 000–1 200 000	2
Phenylalanine (31)		165	70	45 000	1
GMP (19)		363	32.5–67	9–8000	11
AMP (41)		363	35	1000	1
Theophylline (10)		180	37	320	1

Reselection pools were transcribed from chemically-synthesized oligonucleotide pools mutated at a rate of 21% per position. After five rounds of selection in 1 mM  $Mg^{2+}$  buffer, >20% of the RNA in each of the reselections bound to the column and specifically eluted with the respective ligand in solution. At that point, the reselection pools were cloned and sequenced.

Upon analyzing the reselection sequences, we found that there were three variations in the pAF-4Z1 alignment that occurred more often than predicted by chance (Supplementary Figure S4).  $A_8$  changed to U in 25% of the sequences (39 out of 153 sequenced clones),  $A_{16}$  changed to G in 22% of the sequences and  $C_{18}$  changed to U in 58% of the sequences. The latter two mutations appeared together in 11 out of 153 sequences and all three mutations were observed together once. We measured the

apparent  $K_d$  of a sequence containing all three mutations (aptamer pAF-4Z1d3) in a solution binding assay and found a 20-fold reduction in  $K_d$  compared to the original isolate (Table 1). One of the sequences recovered from the pAF-R1-1 reselection had a deletion in the designed internal stem-tetraloop. Replacing the 12 nt after  $C_{31}$  that comprise the designed stem-tetraloop with a short 3 nt sequence does not alter the binding affinity of pAF-R1-1 compared to the original isolate (Table 1 and Figure 4).

The region of TMR3 between  $C_{13}$  and  $G_{24}$  showed covariation indicative of a W-C base-paired stem. Replacing that portion of TMR3 with a designed stem-tetraloop yielded a structure that binds with a submicromolar  $K_d$  (Table 1). W-C base-paired stems, appearing on the left of each secondary structure in



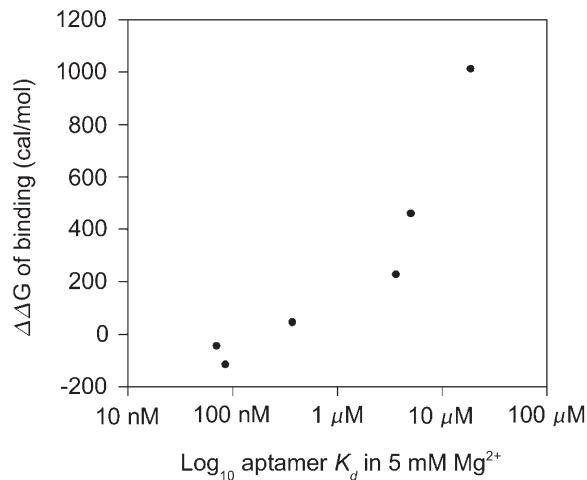
**Figure 4.** Magnesium dependence of pAF aptamers. pAF-R1-1 shows significant magnesium dependence, and fitting a parameterized curve to the data gives a Hill coefficient of 3.4, results for pAF-R1-1d (stem-tetraloop removed) are superimposed as asterisk symbols (\*). pAF-4Z1d3 maintains its  $K_d$  in a small band of  $\sim 3\text{--}7\ \mu\text{M}$  with no discernible trend.

Figure 4, were designed to flank the putative recognition loops in all of the aptamers. No other structural changes were made to TMR aptamers 1, 2 or 4.

We calculated the IC of each aptamer structure as a measure of the difficulty of binding the target ligands in specified buffer conditions (19). IC was computed as a sum of the information content needed to specify the loops of each structure, obtained from the per-position frequencies of each base identified in the reselections, and the information content needed to specify the W–C base-paired stems. The ICs for the pAF aptamers are substantially greater than the ICs for all of the TMR aptamers except TMR3 (Table 1). The fact that the pAF aptamers are longer than all of the TMR aptamers and are more informationally complex than three of the four TMR aptamers provides objective evidence that TMR is a better binding target for RNA than pAF.

### Magnesium dependence

We examined the extent to which the aptamers depend upon magnesium for activity by measuring the apparent  $K_d$ s as a function of  $\text{Mg}^{2+}$ . The pAF-4Z1d3, TMR1, TMR2 and TMR3 aptamers exhibit the same  $K_d$ s for their target ligands in a physiological concentration of magnesium (1 mM) as they do in more permissive conditions (5 mM) (Table 1). In contrast, the TMR4 aptamer  $K_d$  is 2-fold higher in 1 mM  $\text{Mg}^{2+}$  than in 5 mM  $\text{Mg}^{2+}$ . The pAF-R1-1 aptamer shows an even greater dependence on magnesium concentration: the pAF-R1-1  $K_d$  is more than 5-fold higher in 1 mM  $\text{Mg}^{2+}$  than in 5 mM  $\text{Mg}^{2+}$ . For illustrative purposes, the pAF aptamer  $K_d$ s are plotted as a function of  $\text{Mg}^{2+}$  concentration in Figure 4. The sigmoidal curve fit of the pAF-R1-1  $K_d$  versus  $\text{Mg}^{2+}$  concentration data has a Hill coefficient of 3.4, suggesting that



**Figure 5.** pAF and TMR aptamer magnesium dependence. The aptamer free energy of binding in 1 mM  $\text{Mg}^{2+}$  minus the free energy of binding in 5 mM  $\text{Mg}^{2+}$  ( $\Delta\Delta G$  of binding) is plotted against  $\log_{10}$  aptamer  $K_d$  in 5 mM  $\text{Mg}^{2+}$ .

there are specific metal binding sites that must be occupied to form the active structure, as compared to a more general need for counterions (39).

Figure 5 illustrates magnesium dependence as the difference in binding affinities in 1 mM  $\text{Mg}^{2+}$  and 5 mM  $\text{Mg}^{2+}$  [i.e. ( $\Delta G$  of binding in 1 mM  $\text{Mg}^{2+}$ ) - ( $\Delta G$  of binding in 5 mM  $\text{Mg}^{2+}$ ) =  $\Delta\Delta G$  of binding]. As aptamer affinity in 5 mM  $\text{Mg}^{2+}$  decreases (higher  $K_d$ ), there is a larger difference between 1 and 5 mM  $\text{Mg}^{2+}$  binding affinities (Spearman correlation coefficient,  $r_s = 0.94$ ,  $P = 0.008$ ). Put another way, aptamers that bind their ligands more tightly (lower  $K_d$ ) exhibit less dependence on magnesium than aptamers with higher  $K_d$ s, implying that the same structural mechanisms that confer high-affinity binding may also obviate the need for high levels of magnesium.

### DISCUSSION

*In vitro* selection produces the simplest (i.e. the least informationally complex) structural solutions to a given problem (19). It is timely to revisit the general question of RNA aptamer IC and functional activity, first explored by Carothers *et al.* and later by others (42–48), in light of new interest in generating RNA aptamers to bind small molecule targets *in vivo*. Beyond the need for the *in vitro* selected aptamers to bind in physiological conditions, the nature of the target ligand may place further restrictions on the ability to generate functional structures. In practice, there may be limitations on the kinds of ligands to which RNA aptamers can be *in vitro* selected to bind in physiological conditions with affinities that are sufficient for metabolic engineering and synthetic biology applications.

Structural and biochemical analysis of aptamer-ligand complexes has shown that steric complementarity, planar stacking and intercalation and hydrogen bonding are the

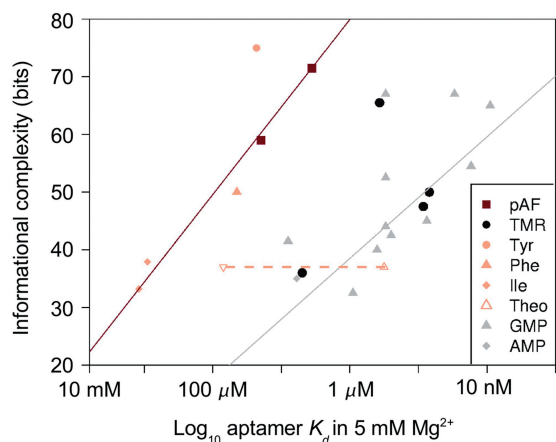


primary modes of interaction (49,50). Because of this, intrinsic chemical properties of the target molecule, such as MW, the number of rotatable bonds and the number of polar groups potentially serving as hydrogen bond donors and acceptors are likely to be important for determining the affinity of aptamers that can be isolated using *in vitro* selection. As a consequence, it is reasonable to expect that aptamers that bind small, flexible, hydrophobic molecules may be less abundant in a pool of random sequences than aptamers that bind with equal affinity to larger, more planar, and more polar ligands (19).

### Generalizing to other ligands

Through a literature search we collected *in vitro* selected aptamers that bind six different small molecules whose ICs were either previously presented or could be estimated from published data (Table 2). These included four aptamers that bind amino acids (one binds tyrosine, two isoleucine and one phenylalanine), 11 aptamers that bind guanosine mono/triphosphate (GMP), one aptamer that binds adenosine mono/triphosphate (AMP) and one aptamer that binds the xanthine alkaloid theophylline. We plot the ICs and binding affinities of aptamers from the literature alongside data for the two pAF and four TMR aptamers selected and characterized here in Figure 6. All of the aptamers from the literature were selected in 5 mM  $Mg^{2+}$ . For consistency, all of the  $K_d$ s in Figure 6 are for binding in 5 mM  $Mg^{2+}$ .

Interestingly, the amino acid-binding aptamers selected elsewhere and the pAF aptamers selected here fall into a region of the IC versus  $K_d$  plot that is distinct from the region of the plot where the aptamers that bind the larger ligands TMR, GMP and AMP appear. Without



**Figure 6.** Aptamer IC versus binding affinity. Aptamer IC in bits is plotted against  $\log_{10}$  apparent  $K_d$  in 5 mM  $Mg^{2+}$ . Line fit of the GMP aptamers (19) is grey; line fit of the amino acid-binding aptamers is dark red. The Spearman rank correlation coefficient ( $r_s$ ) for informational complexity and binding affinity for the GMP aptamers is 0.65,  $P < 0.025$ ,  $r_s = 0.67$ ,  $P < 0.0025$  for the GMP, AMP and TMR aptamers and  $r_s = 0.83$ ,  $P < 0.030$  for the amino acid aptamers. The orange dotted line connects positions on the graph between the actual theophylline aptamer IC versus  $K_d$  relationship (triangle pointed up) and the theophylline aptamer IC versus  $K_d$  relationship adjusted for the number of rotatable bonds relative to the set of amino acids (triangle pointed down).

exception, aptamers that bind the amino acids (ligand MWs = 131–181 g/mol) are more informationally complex and/or have lower affinities for their ligands than the aptamers that bind TMR, GMP and AMP (ligand MWs = 343–363 g/mol). Thus, to a first approximation, the IC versus affinity relationship appears to be dictated by ligand MW, a finding that may be consistent with the weak correlations between affinity and ligand mass observed in other studies (50).

The red and grey lines in Figure 6 are line fits of the data for the set of amino acid-binding aptamers and the extensively characterized GMP aptamers (19), respectively. For the set of amino acid binding aptamers, the informational cost of each 10-fold increase in binding affinity is  $15 \pm 8$  bits. This value agrees with the informational cost of a 10-fold improvement in GMP aptamer binding affinity ( $10 \pm 5$  bits, equal to a 1000-fold decrease in abundance in a pool of random sequences) and the informational cost of a 10-fold improvement in binding affinity for the entire set of aptamers that bind the larger ligands ( $8 \pm 2$  bits). Therefore, we take the horizontal distance between the lines at a given IC to represent the average difference in potential binding affinity that can be ascribed to the difference in ligand MWs. Specifically, the average affinities of aptamers that bind the smaller MW ligands (MW = 131–181 g/mol) are 1000-fold worse than the the average affinities of aptamers that bind the larger MW ligands (MW = 343–363 g/mol). Equivalently, 40 bits of additional IC are needed, on average, to specify an aptamer that binds an amino acid-sized ligand with the same affinity as an aptamer that binds a larger ligand.

The Jenison (10) theophylline aptamer is a significant outlier on this plot since it has an affinity for the target that is several orders of magnitude higher than predicted based solely on MW (open triangle pointed up in Figure 6). It has an IC versus affinity relationship that resembles the aptamers that bind the larger ligands, even though the theophylline MW is only 180 g/mol. We can begin to understand this apparent discrepancy by considering other characteristics of the target ligands.

Table 3 shows estimates for the potential free energy of binding (35) intrinsic to the theophylline ligand functional groups compared to the average potential free energy of binding for the functional groups in the amino acids and larger ligands (also see Supplementary Table S3). Employing values for the intrinsic binding potential of each type of functional group calculated by Andrews *et al.* (35), we find that the potential energy contributions of the carbons and polar groups in theophylline are similar to the average potential energy contributions of the carbons and polar groups in the amino acids.

In contrast, we note that theophylline has fewer rotatable bonds than the ligands in either set (Supplementary Table S3). Fixing a rotatable bond in a ligand, which reduces the internal degrees of freedom (DOF), has been determined to contribute as much as 0.6–1.0 kcal/mol to the free energy of binding (35,51). Based on the Andrews *et al.* functional group intrinsic binding energies, the fixed bonds in theophylline could contribute, on average,



**Table 3.** Intrinsic ligand functional group binding energies

Ligand set	MW (g/mol)	Carbons (sp <sup>2</sup> , sp <sup>3</sup> ) (kcal/mol)	Polar (N <sup>+</sup> , N, OH, O, S) (kcal/mol)	DOF (kcal/mol)
pAF, Tyr, Ile, Phe	164	5.4	12.4	-3.2
TMR, GMP, AMP	366	10.3	13.8	-3.7
Theophylline	180	5.1	11.6	0

Intrinsic ligand binding energies were calculated using the functional group binding energies ( $E_x$ ) presented by Andrews *et al.* (35). The mean intrinsic binding energies (in kcal/mol) for the functional groups within each set of ligands is given. Carbons (Csp<sup>2</sup>, Csp<sup>3</sup>) corresponds to the potential binding energy attributable to planar and aliphatic carbons. The potential binding energy penalty ascribed to rotatable bonds is listed under the DOF heading. Supplementary Table S3 enumerates the functional groups for individual ligands and the  $E_x$  values employed for these calculations.

3.2 kcal/mol of favorable binding energy to the affinity of the theophylline aptamer relative to the amino acids (Table 3). Interestingly, subtracting 3.2 kcal/mol from the binding affinity of the theophylline aptamer moves that data point toward a region of the graph consistent with the IC versus  $K_d$  relationship predicted on the basis of MW alone (orange triangle pointed down in Figure 6).

Caution should be exercised when using this approach to make predictions because of our limited dataset and the uncertainty in ascribing values for the intrinsic binding energy of individual ligand functional groups in RNA-ligand interactions (35,51). With these important caveats, we suggest that this treatment may be employed as a rough guide for estimating potential binding affinities relative to other aptamer–ligand  $K_d$ s. Extending this analysis to aptamers that bind other ligands differing in MWs and with a variety of functional groups should enable aptamer binding affinity predictions to be further refined.

### Estimating the maximal $K_d$

Ultimately, we would like to develop a mathematical framework for predicting the potential IC/ $K_d$  ratio because this is useful for determining what affinities are accessible with a particular sequence pool length, diversity and design. We have shown that aptamer IC versus  $K_d$  relationships can be used to begin understanding how ligand MW and the number of rotatable bonds affect potential binding affinities. For aptamers with an internal stem–tetraloop and ICs <55 bits, the prior probability of occurring in a pool with a sequence design and diversity ( $10^{14}$  unique sequences) similar to STR and STR2 is approximately one (19). This implies that it should always be possible to isolate aptamers from these pools to bind a ligand with a given affinity if <55 bits of IC are needed to specify those structures.

In contrast, the prior probabilities for the most informationally complex aptamers are much less than one and are, therefore, expected to be among the highest affinity structures that can be selected from libraries with the lengths, diversities and designs of STR and STR2. For

example, the prior probability that pAF-4Z1d3, whose  $K_d$  for pAF is 3.6  $\mu$ M, will be present in the STR/STR2 pool is <0.0001 (Supplementary Table S4). The prior probabilities of the most complex and highest affinity GMP aptamers ( $K_d$ s for GMP = 10–30 nM) in structured pools were previously shown to be <0.0035 (19).

We suggest that the lowest  $K_d$ s attainable for aptamers selected from pools like these to bind ligands similar to the amino acids are in the low micromolar range and that the lowest  $K_d$ s attainable for aptamers that bind molecules similar to the larger targets (i.e. planar moieties with MW  $\sim$ 350 g/mol) are in the low nanomolar range. It may be possible to generate high-affinity aptamers that bind targets smaller than pAF and TMR or molecules with more internal DOF, but this will likely require structurally-complex sequence libraries with additional designed elements (47) that ‘pre-pay’ more of the IC (19).

### Implications of magnesium and IC relationship

Two lines of evidence suggest that there is a general relationship between IC,  $K_d$ , and magnesium sensitivity (illustrated by  $\Delta\Delta G$  of binding between 1 and 5 mM  $Mg^{2+}$ ) in aptamer structures. First, we observe that the  $K_d$ s of the highest affinity aptamers are not sensitive to the concentration of  $Mg^{2+}$  (Figure 5). For instance, the  $K_d$ s of the higher affinity pAF aptamer and the three highest affinity TMR aptamers are the same in 1 and 5 mM  $Mg^{2+}$ . In contrast, the aptamer with the highest  $K_d$  in both sets exhibits sensitivity to the concentration of  $Mg^{2+}$  (Figure 4).

We also note that the pAF-4Z1 and pAF-R1-1 aptamers dominated the selection pools in all of the  $Mg^{2+}$  conditions (Figure 2). Despite retaining significant sequence diversity in the last rounds, even the most permissive selection (in 10 mM  $Mg^{2+}$ ) had a pool where nearly 30% of the clones were pAF-4Z1. If there were other high-affinity aptamers in the library pool that can only bind in permissive magnesium conditions, then they should have been recovered from the 5 and 10 mM  $Mg^{2+}$  selections. The fact that this did not happen provides further evidence that selection in stringent magnesium conditions simultaneously selected for tighter-binding aptamers.

Khvorova *et al.* (52) were amongst the first to observe that the concentrations of  $Mg^{2+}$  required for high rates of hammerhead ribozyme self-cleavage are greatly reduced by structural elements (interaction loops) outside the minimally-active core that stabilize the tertiary fold (53). Penedo *et al.* (54) elucidated the magnesium sensitivity of minimal hammerhead ribozymes along with those harboring the interaction loops: the presence of two 3–8 bp loop structures rescues functionality in low (1 mM) magnesium relative to structures that do not contain the loop structures. Notably, the structural stability engendered by the interaction loops not only reduces the need for magnesium, but also increases the overall cleavage rates. Interaction loops increase the cleavage rate ( $k_{obs}$ ) of minimal hammerhead ribozymes 35–70 fold in 0.1 mM  $Mg^{2+}$  (52) and 4-fold in 10 mM  $Mg^{2+}$  (55). Interestingly, the minimal hammerhead ribozyme and the pAF-R1-1 aptamer have similar degrees of magnesium sensitivity, as indicated by

the fact that the half-maximum activity point is  $\sim 2$  mM  $Mg^{2+}$  for both structures [Figure 4 (54)].

We can approximate the IC needed to specify the hammerhead ribozyme interaction loops based on the published secondary and tertiary structures (52,53). There are about 11 bases within the loops themselves that are two-base varying (0.5 bits/nt = 5.5 bits) and two longer stems (2–3 bp each = 8–12 bits), totaling 13.5–17.5 bits of additional IC compared to the minimal hammerhead ribozyme. To put this value into perspective, 10 more bits of IC are needed to increase aptamer binding affinity by a factor of 10 and 40 more bits are required to specify an aptamer with the same affinity for an amino acid as one of the larger ligands (see above).

Taken together, we find that for a set of aptamers to a particular ligand, additional information (which in turn defines structure) produces more robust structures; i.e. structures that function in physiological magnesium concentrations. In the case of aptamers, this additional IC appears to result in not only greater affinity but also less reliance on magnesium for functionality. Perhaps additional structural elements within the aptamers improve the stability of the active conformations, resulting in higher binding affinity and less need for magnesium in the same way that the interaction loops stabilize the hammerhead ribozyme structure, increasing  $k_{obs}$  and allowing for cleavage in physiological magnesium conditions. These findings are notable in light of the observation that high concentrations of magnesium are often necessary for high levels of RNA functionality (5,52,54–58). The evidence discussed here suggests that dependence on magnesium for functionality can be readily overcome by additional structure (information). Although additional study is clearly warranted, we hypothesize that the goals of high-affinity binding and *in vivo* functionality are attainable in concert rather than in competition.

### Integrating aptamers into control devices

There are several immediate applications in synthetic biology and metabolic engineering for aptamer-based control devices: for sensing specific molecules in the environment; as *in situ* readouts for detecting product formation, useful for screening or evolving high-yield enzymes and pathways; and for creating small molecule-responsive feedback or feedforward controls in engineered systems (9). For example, an engineered metabolic pathway may have toxic intermediates whose concentrations could be dynamically regulated by controlling the expression of enzymes that produce them.

Knowledge about the concentration at which an intermediate is toxic could be used to estimate the IC needed to select an aptamer that binds in that regime (Figure 6). In some cases, this analysis may indicate that aptamers cannot be selected to bind a potential target with sufficient affinity. In others, there may be several metabolic intermediates that could serve as signals for aptamer-based pathway regulation. Choosing the largest intermediates with the fewest number of rotatable bonds should improve the chances of a successful aptamer selection. By recognizing which molecular characteristics contribute

positively to potential RNA binding, one can, at the least, estimate the affinity range (mM,  $\mu$ M, nM) obtainable for a given ligand and design sequence libraries and selection protocols accordingly.

Once isolated, aptamers should be reselected as above, in order to improve  $K_{ds}$ , and to obtain structural data. These data, such as which areas are stems, which bases can vary (blue and green bases in Figure 3) and which are unnecessary (such as the structured stem-tetraloop in pAF-R1-1 that was removed), can aid the integration of the aptamer into a biologically active context. For example, by combining pAF-R1-1 with a hammerhead ribozyme we have evolved pAF-responsive aptazymes that function *in vitro* and control gene expression in *E. coli* (Goler *et al.*, manuscript in preparation).

### CONCLUSION

The ability to rapidly select aptamers to target small molecules enables their use for *in vivo* sensing and control in engineered biological systems. The evidence presented here suggests that selection for high-affinity binding produces aptamers that do not require high concentrations of magnesium for activity. This may mean that selection for high-affinity aptamers to function in *in vivo* is not significantly constrained by the need to bind in physiological levels of magnesium.

We have shown that it may be possible to estimate the IC required to specify a structure that binds a ligand with a particular affinity before a selection is performed. To a first approximation, the potential aptamer binding affinity for a target can be predicted on the basis of the MW of the ligand. Predictions of achievable aptamer–ligand affinities are further improved by considering ligand properties such as the numbers of rotatable bonds and potential hydrogen bond donors and acceptors. Because IC is directly related to the probability that a structure will be present in a combinatorial sequence pool, we hypothesize that this analysis can be used to design libraries and selections that are more likely to yield aptamers of the desired affinity. Aggregating more comprehensive data on aptamers – their targets, affinities and ICs – will test this hypothesis and feasibly enable more accurate predictions of obtainable aptamer–ligand affinities.

### SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

### ACKNOWLEDGEMENTS

Special thanks to Adrienne McKee, Karen Wong and Nathan Hillson for comments helpful in preparing this manuscript.

### FUNDING

This work was supported by the Joint BioEnergy Institute through a contract between Lawrence Berkeley National Laboratory and the US Department of Energy

(DE-AC02-05CH11231); and the Synthetic Biology Engineering Research Center through a grant from the National Science Foundation [BES-0439124]. J.M.C. was supported in part by a Jane Coffin Childs Memorial Fund Postdoctoral Fellowship. Funding for open access charge: National Science Foundation [BES-0439124].

*Conflict of interest statement.* None declared.

## REFERENCES

- Isaacs, F.J., Dwyer, D.J. and Collins, J.J. (2006) RNA synthetic biology. *Nat. Biotechnol.*, **24**, 545–554.
- Pfleger, B.F., Pitera, D.J., Smolke, C.D. and Keasling, J.D. (2006) Combinatorial engineering of intergenic regions in operons tunes expression of multiple genes. *Nat. Biotechnol.*, **24**, 1027–1032.
- Davidson, E.A. and Ellington, A.D. (2007) Synthetic RNA circuits. *Nat. Chem. Biol.*, **3**, 23–28.
- Keasling, J.D. (2008) Synthetic biology for synthetic chemistry. *ACS Chem. Biol.*, **3**, 64.
- Saito, H. and Inoue, T. (2009) Synthetic biology with RNA motifs. *Int. J. Biochem. Cell. Biol.*, **41**, 398–404.
- Werstuck, G. and Green, M.R. (1998) Controlling gene expression in living cells through small molecule–RNA interactions. *Science*, **282**, 296.
- Grate, D. and Wilson, C. (2001) Inducible regulation of the *S. cerevisiae* cell cycle mediated by an RNA aptamer–ligand complex. *Bioorg. Med. Chem.*, **9**, 2565–2570.
- Carothers, J.M. and Szostak, J.W. (2006) In vitro selection of functional oligonucleotides and the origins of biochemical activity. In Klussmann, S. (ed.), *The aptamer handbook: functional oligonucleotides and their applications*. Springer-Verlag Press, Berlin, pp. 3–28.
- Carothers, J.M., Goler, J.A. and Keasling, J.D. (2009) Chemical synthesis using synthetic biology. *Curr. Opin. Biotechnol.*, **20**, 498–503.
- Jenison, R., Gill, S., Pardi, A. and Polisky, B. (1994) High-resolution molecular discrimination by RNA. *Science*, **263**, 1425–9.
- Suess, B. (2005) Engineered riboswitches control gene expression by small molecules. *Biochem. Soc. Trans.*, **33**, 474–476.
- Suess, B., Fink, B., Berens, C., Stentz, R. and Hillen, W. (2004) A theophylline responsive riboswitch based on helix slipping controls gene expression in vivo. *Nucleic Acids Res.*, **32**, 1610–1614.
- Lynch, S., Desai, S., Sajja, H. and Gallivan, J. (2007) A high-throughput screen for synthetic riboswitches reveals mechanistic insights into their function. *Chem. Biol.*, **14**, 173–184.
- Soukup, G.A. and Breaker, R.R. (1999) Engineering precision RNA molecular switches. *Proc. Natl Acad. Sci. USA*, **96**, 3584–3589.
- Thompson, K., Syrett, H., Knudsen, S. and Ellington, A. (2002) Group I aptazymes as genetic regulatory switches. *BMC Biotechnol.*, **2**, 21.
- Cho, S., Kim, J., Lee, B., Kim, J. and Kim, B. (2005) Bis-aptazyme sensors for Hepatitis C virus replicase and helicase without blank signal. *Nucleic Acids Res.*, **33**.
- Win, M.N. and Smolke, C.D. (2007) A modular and extensible RNA-based generegulatory platform for engineering cellular function. *Proc. Natl Acad. Sci. USA*, **104**, 14283.
- Ogawa, A. and Maeda, M. (2008) An artificial aptazyme-based riboswitch and its cascading system in *E. coli*. *Chem. Bio. Chem.*, **9**, 206–209.
- Carothers, J.M., Oestreich, S.C., Davis, J.H. and Szostak, J.W. (2004) Informational complexity and functional activity of RNA structures. *J. Am. Chem. Soc.*, **126**, 5130–5137.
- Dower, W. and Cwirla, S. (1992) *Guide to electroporation and electrofusion*. Academic, San Diego.
- Desai, S. and Gallivan, J. (2004) Genetic screens and selections for small molecules based on a synthetic riboswitch that activates protein translation. *J. Am. Chem. Soc.*, **126**, 13247–13254.
- Bayer, T. and Smolke, C. (2005) Programmable ligand-controlled riboregulators of eukaryotic gene expression. *Nat. Biotechnol.*, **23**, 337–343.
- Burke, D.H. and Greathouse, S.T. (2005) Low-magnesium, trans-cleavage activity by type III, tertiary stabilized hammerhead ribozymes with stem I discontinuities. *BMC Biochem.*, **6**, 14.
- Pyle, A. (2002) Metal ions in the structure and function of RNA. *J. Biol. Chem.*, **7**, 679–690.
- Wallis, M., von Ahsen, U., Schroeder, R. and Famulok, M. (1995) A novel RNA motif for neomycin recognition. *Chem. Biol.*, **2**, 543–552.
- Kiga, D., Futamura, Y., Sakamoto, K. and Yokoyama, S. (1998) An RNA aptamer to the xanthine/guanine base with a distinctive mode of purine recognition. *Nucleic Acids Res.*, **26**, 1755.
- Berens, C., Thain, A. and Schroeder, R. (2001) A tetracycline-binding RNA aptamer. *Bioorg. Med. Chem.*, **9**, 2549–2556.
- Famulok, M., Mayer, G. and Blind, M. (2000) Nucleic acid aptamers—from selection in vitro to applications in vivo. *Acc. Chem. Res.*, **33**, 591–599.
- Mehl, R.A., Anderson, J.C., Santoro, S.W., Wang, L., Martin, A.B., King, D.S., Horn, D.M. and Schultz, P.G. (2003) Generation of a bacterium with a 21 amino acid genetic code. *J. Am. Chem. Soc.*, **125**, 935–939.
- Davis, J. and Szostak, J. (2002) Isolation of high-affinity GTP aptamers from partially structured RNA libraries. *Proc. Natl Acad. Sci. USA*, **99**, 11616–11621.
- Illangasekare, M. and Yarus, M. (2002) Phenylalanine-binding RNAs and genetic code evolution. *J. Mol. Evol.*, **54**, 298–311.
- Amblard, M., Fehrentz, J., Martinez, J. and Subra, G. (2006) Methods and protocols of modern solid phase peptide synthesis. *Mol. Biotechnol.*, **33**, 239–254.
- Thompson, J., Higgins, D. and Gibson, T. (1994) CLUSTALW. *Nucleic Acids Res.*, **22**, 80.
- Efron, B. and Gong, G. (1983) A leisurely look at the bootstrap, the jackknife and cross-validation. *Am. Stat.*, **37**, 36–48.
- Andrews, P., Craik, D. and Martin, J. (1984) Functional group contributions to drug-receptor interactions. *J. Med. Chem.*, **27**, 1648–1657.
- Mannironi, C., Scerch, C., Fruscoloni, P. and Tocchini-Valentini, G.P. (2000) Molecular recognition of amino acids by RNA aptamers: the evolution into an L-tyrosine binder of a dopamine-binding RNA motif. *RNA*, **6**, 520–527.
- Mannironi, C., Nardo, A.D., Fruscoloni, P. and Tocchini-Valentini, G.P. (1997) In vitro selection of dopamine RNA ligands. *Biochemistry*, **36**, 9726–9734.
- Lozupone, C., Changayil, S., Majerfeld, I. and Yarus, M. (2003) Selection of the simplest RNA that binds isoleucine. *RNA*, **9**, 1315–1322.
- Bassi, G., Møllegaard, N., Murchie, A., von Kitzing, E. and Lilley, D. (1995) Ionic interactions and the global conformations of the hammerhead ribozyme. *Nat. Struct. Biol.*, **2**, 45–55.
- Legiewicz, M. and Yarus, M. (2005) A more complex isoleucine aptamer with a cognate triplet. *J. Biol. Chem.*, **280**, 19815.
- Sassanfar, M. and Szostak, J. (1993) An RNA motif that binds ATP. *Nature*, **364**.
- Legiewicz, M., Lozupone, C., Knight, R. and Yarus, M. (2005) Size, constant sequences, and optimal selection. *RNA*, **11**, 1701–1709.
- Gevertz, J., Schlick, T. and Gan, H.H. (2005) In vitro RNA random pools are not structurally diverse: a computational analysis. *RNA*, **11**, 853–863.
- Knight, R., De Sterck, H., Markel, R., Smit, S., Oshmyansky, A. and Yarus, M. (2005) Abundance of correctly folded RNA motifs in sequence space, calculated on computational grids. *Nucleic Acids Res.*, **33**, 5924–5935.
- Carothers, J.M., Davis, J.H., Chou, J.J. and Szostak, J.W. (2006) Solution structure of an informationally complex high-affinity RNA. *RNA*, **12**, 567–579.
- Carothers, J.M., Oestreich, S.C. and Szostak, J.W. (2006) Aptamers selected for higher-affinity binding are not more specific for the target ligand. *J. Am. Chem. Soc.*, **128**, 7929–7937.
- Kim, N., Gan, H.H. and Schlick, T. (2007) A computational proposal for designing structured RNA pools for in vitro selection of RNAs. *RNA*, **13**, 478–492.

48. Hazen, R.M., Griffin, P.L., Carothers, J.M. and Szostak, J.W. (2007) Functional information and the emergence of biocomplexity. *Proc. Natl Acad. Sci. USA*, **104**, 8574–8581.
49. Hermann, T. and Patel, D. (2000) Adaptive Recognition by Nucleic Acid Aptamers. *Science*, **287**, 820–825.
50. Pfeffer, P. and Gohlke, H. (2007) DrugScoreRNA-knowledge-based scoring function to predict RNA-Ligand interactions. *J. Chem. Inf. Model.*, **47**, 1868–1876.
51. Tirado-Rives, J. and Jorgensen, W.L. (2006) Contribution of conformer focusing to the uncertainty in predicting free energies for protein–ligand binding. *J. Med. Chem.*, **49**, 5880–5884.
52. Khvorova, A., Lescoute, A., Westhof, E. and Jayasena, S. (2003) Sequence elements outside the hammerhead ribozyme catalytic core enable intracellular activity. *Nat. Struct. Biol.*, **10**, 708–712.
53. Martick, M. and Scott, W. (2006) Tertiary contacts distant from the active site prime a ribozyme for catalysis. *Cell*, **126**, 309–320.
54. Penedo, J., Wilson, T., Jayasena, S., Khvorova, A. and Lilley, D. (2004) Folding of the natural hammerhead ribozyme is enhanced by interaction of auxiliary elements. *RNA*, **10**, 880.
55. Weinberg, M. and Rossi, J. (2005) Comparative single-turnover kinetic analyses of trans-cleaving hammerhead ribozymes with naturally derived nonconserved sequence motifs. *FEBS Lett.*, **579**, 1619–1624.
56. Koculi, E., Lee, N., Thirumalai, D. and Woodson, S. (2004) Folding of the Tetrahymena ribozyme by polyamines: importance of counterion valence and size. *J. Mol. Biol.*, **341**, 27–36.
57. Misra, V. and Draper, D. (2001) A thermodynamic framework for Mg<sup>2+</sup> binding to RNA. *Proc. Natl Acad. Sci. USA*, **98**, 12456.
58. Woodson, S. (2005) Metal ions and RNA folding: a highly charged topic with a dynamic future. *Curr. Opin. Chem. Biol.*, **9**, 104–109.