Identification and characterization of RNA binding sites for (p)ppGpp using RNA-DRaCALA

Jonathan Jagodnik [1,*]**, Brian Tjaden** [2]**, Wilma Ross** [1] **and Richard L. Gourse** [1,*]

[1]Department of Bacteriology, University of Wisconsin–Madison, Madison, WI 53706, USA and [2]Department of Computer Science, Wellesley College, Wellesley, MA 02481, USA

ABSTRACT

Ligand-binding RNAs (RNA aptamers) are widespread in the three domains of life, serving as sensors of metabolites and other small molecules. When aptamers are embedded within RNA transcripts as components of riboswitches, they can regulate gene expression upon binding their ligands. Previous methods for biochemical validation of computationally predicted aptamers are not well-suited for rapid screening of large numbers of RNA aptamers. Therefore, we utilized DRaCALA (Differential Radial Capillary Action of Ligand Assay), a technique designed originally to study protein-ligand interactions, to examine RNA-ligand binding, permitting rapid screening of dozens of RNA aptamer candidates concurrently. Using this method, which we call RNA-DRaCALA, we screened 30 *ykkC* family subtype 2a RNA aptamers that were computationally predicted to bind (p)ppGpp. Most of the aptamers bound both ppGpp and pppGpp, but some strongly favored only ppGpp or pppGpp, and some bound neither. Expansion of the number of biochemically verified sites allowed construction of more accurate secondary structure models and prediction of key features in the aptamers that distinguish a ppGpp from a pppGpp binding site. To demonstrate that the method works with other ligands, we also used RNA DRaCALA to analyze aptamer binding by thiamine pyrophosphate.

GRAPHICAL ABSTRACT
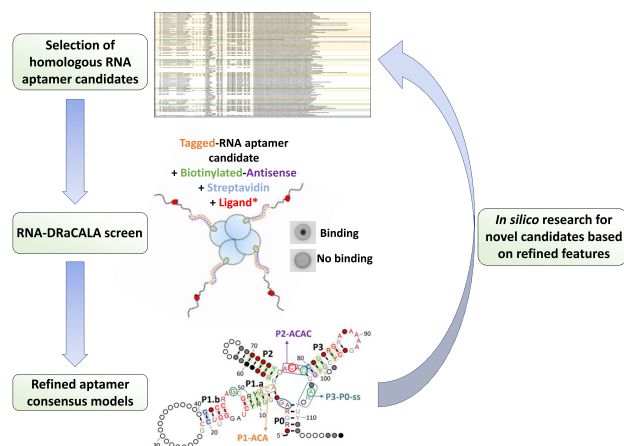
INTRODUCTION

RNA aptamers are RNA regions capable of binding specific ligands with high affinity and specificity (1–3). Often these ligand binding regions belong to larger RNA domains called riboswitches that function to regulate gene expression (2–7). Riboswitches are typically found within mRNAs and are among the most widespread regulators of RNA in biology. For example, they can restrict the expression of genes or operons in a metabolic pathway to a condition in which a precursor molecule in the pathway is present or the final product of the pathway is absent (2,3,6). This regulation can ensure that all the enzymes involved in the pathway are not translated when the synthesis of the final product is unnecessary. The aptameric portion of a riboswitch plays a key role by sensing the specific ligand level and driving a switch in RNA structure that provokes the regulatory event.

The first riboswitches were identified almost two decades ago (4–7), following the identification of riboregulation by tRNA binding to mRNAs for control of amino acyl tRNA synthetase expression (8). Numerous other riboswitches

*To whom correspondence should be addressed. Tel: +1 608 262 9813; Fax: +1 608 262 9865; Email: rgourse@bact.wisc.edu
Correspondence may also be addressed to Jonathan Jagodnik. Tel: +33 158415149; Fax: +33 158415025; Email: jagodnik@ibpc.fr
Present address: Jonathan Jagodnik, UMR8261, CNRS, Université de Paris, Institut de Biologie Physico-Chimique, 75005 Paris, France.

have also been identified based on the function of the regulated gene, such as the *thiM* and *thiC* genes involved in thiamine biosynthesis that are regulated by thiamine pyrophosphate (TPP) (9). Some, including the TPP-binding riboswitches found in the three domains of life, are of ancient origin (10–13). It is therefore hypothesized that aptamers first appeared during the RNA-world stage of evolution (10,14–15), where they served as the binding domains of ribozymes controlling biological enzymatic reactions.

The search for RNA aptamers proceeded along two different paths. Most potential aptamers were identified *in silico* by searching for evolutionarily conserved motifs preceding coding regions, followed by analysis of the RNAs for changes in secondary structure *in vitro* in the presence versus absence of specific ligands (6,16–18). More recently, analysis of RNA structures *in vivo* under different physiological conditions or in the presence of high concentrations of a given ligand *in vitro* revealed aptamers and riboswitches that were not identified by *in silico* approaches (19–21). In either case, demonstration of ligand binding to more than a small number of candidate aptamers at a time, often using probing with nucleases or spontaneous cleavage (in-line probing), has proven challenging.

The ligand binding specificities of aptamers are very diverse, including nucleotides and their derivatives, ions, and a wide range of other small molecules (10,14). For example, the *ykkC* aptamer family, found in multiple Gram positive and Gram negative bacterial species, was originally discovered in homologs of the *ykkC* and *yxkD* genes that are involved in multidrug resistance and transport (6). Although these aptamers remained orphans for a decade without known ligands, in recent years various ligands have been described that bind to them. For example, *ykkC* subtype 1 aptamers bind the toxic metabolic byproduct guanidine (22), subtype 2b aptamers bind the purine precursor phosphoribosylpyrophosphate (PRPP) (23), and subtype 2c aptamers bind the nucleoside diphosphates ADP or CDP, as well as their deoxyribose forms, dADP and dCDP (24).

ykkC subtype 2a RNA aptamers function in riboswitches in *cis* with genes involved in amino acid synthesis and transport (25) by binding guanosine 3′,5′-bisphosphate (ppGpp) and/or guanosine 3′-bisphosphate 5′-triphosphate (pppGpp). ppGpp and pppGpp are highly conserved stress signaling molecules in bacteria, chloroplasts, and mitochondria (26–28). In response to amino-acid depletion and other stresses, ppGpp and pppGpp, sometimes referred to as (p)ppGpp, are induced and bind to multiple protein targets involved in functions as diverse as transcription, translation, nucleotide metabolism, and other metabolic processes (29–32). Because one of those targets in proteobacteria is RNA polymerase (RNAP) (33,34), (p)ppGpp can regulate the expression of hundreds of genes whose promoters have the appropriate kinetic properties (35). In firmicutes like the *Bacilli*, however, (p)ppGpp does not bind directly to RNAP (27). Instead, these bacteria rely on many other mechanisms for control of gene expression by (p)ppGpp, including (p)ppGpp binding to enzymes critical for nucleotide biosynthesis and to *ykkC* subtype 2a RNA riboswitches (25).

More than 100 (p)ppGpp binding aptamers have been predicted bioinformatically within or upstream of more

than 20 different genes in about 50 different firmicute species (22,25). Binding to five of these predicted aptamers has been confirmed biochemically or genetically, one each in the Clostridia species *Thermosediminibacter oceani* and *Desulfitobacterium hafniense*, one each in the Incertae sedis species *Sulfobacillus acidophilus* and the unclassified Clostridium species BL8 (NZ_AUPA01000230.1, *Cbl8*), and one in the Negativicute species *Pelosinus fermentans*. These aptamers are located upstream of genes predicted to code either for the *ilvE* branched-chain-amino-acid transaminase, the *natA* ATP-binding cassette (ABC) domain transporter, or the *glt*/*glxB* methylglyoxalase, respectively. There are strong sequence similarities and predicted secondary structure models for the >100 proposed (p)ppGpp binding aptamers, but the secondary structure models generated from these proposed aptamers were created under the assumption that all bind both pppGpp and ppGpp with similar affinities. This was verified only for the *ilvE* aptamer from *T. oceani* (25).

Experimental validation and analysis of ligand binding to RNA aptamers have relied heavily on the in-line probing assay (36). This method has been used to measure the affinity of RNA aptamers for their ligands, and the assay also provides some information about the secondary structure of the aptamer. However, in-line probing often requires incubation of RNAs for extended periods of time (days to weeks) in the presence of the ligand of interest, which makes it unsuitable for testing the binding of an unstable ligand, as it will decay over the course of the experiment. Furthermore, in-line probing relies on spontaneous cleavage within unpaired flexible RNA regions, and a cleavage can potentially alter binding to the aptamer or otherwise affect interactions with the ligand during the experiment. While reliable for testing the binding of an RNA to multiple ligands, in-line probing is usually too labor-intensive for investigation of many RNAs at the same time.

Here we use a method that we call RNA-DRaCALA to screen and characterize the interaction of predicted RNA aptamers with their proposed ligands. We modified DRaCALA (<u>D</u>ifferential <u>R</u>adial <u>C</u>apillary <u>A</u>ction of <u>L</u>igand <u>A</u>ssay), which was originally designed to study binding of ligands to proteins (37) but has also been used for nucleic acids (38), to make it more suitable for screening multiple RNAs. We characterized previously identified (p)ppGpp aptamers and then screened a larger set of predicted *ykkC* subtype 2a aptamers for binding to ppGpp and/or pppGpp. A large majority of these aptamers bound both ppGpp and pppGpp, a few did not bind either ligand, and some were specific for one or the other. Using these experimentally validated aptamer subsets, we generated distinct models for ppGpp- and pppGpp-binding aptamers that recapitulate all previously reported features for binding (p)ppGpp, as well as new features identified here, and then we utilized these models to predict previously unreported aptamers.

MATERIALS AND METHODS

Production of RNA aptamers

RNA aptamers were prepared by transcription *in vitro* using T7 RNA polymerase and PCR-generated DNA templates. Primers for PCR amplification are listed in Supplementary

Table S3. For *Escherichia coli thiC* and *thiM* TPP-binding aptamers, the DNA templates were PCR amplified from genomic DNA from *E. coli* strain MG1655. For the *T. oceani ilvE (To ilvE)* and the *D. hafniense ilvE* (*Dh ilvE*) (p)ppGpp-binding aptamers, as well as the mutant *To ilvE* M2 aptamer that does not bind (p)ppGpp, DNA templates were PCR amplified from DNA generously provided by the Breaker laboratory. Primers for amplifying the *thiC*, *thiM* and *ilvE* DNA templates were designed with a T7 promoter adjacent to the DNA coding for the 5′ end of the aptamer and a 3′ tag where indicated (Supplementary Table S3). For (p)ppGpp-binding aptamer candidates (sequences from (25), Supplementary Tables S1 and S3) and for TPP-binding aptamer candidates (sequences from (19), Supplementary Tables S1 and S3), DNA templates for PCR amplification were purchased as double stranded DNA gene blocks from Integrated DNA Technologies. The gene blocks contained a T7 promoter adjacent to the DNA coding for the 5′ end of the aptamer. Primers for PCR amplification from the gene blocks included a 3′ tag (Supplementary Table S3). Up to three G nucleotides were added to the 5′end of the aptamer-encoding sequence at the transcription start sites to facilitate transcription by T7 RNAP, depending on the number of G's already present in the original aptamer sequence. The extra G nucleotides and/or 3′ tags were the only nucleotides absent from the aptamers described by Breaker and colleagues (25) (Supplementary Table S1). The amplified DNA templates were transcribed *in vitro* using the T7 Megascript kit (Invitrogen), following the manufacturer's instructions. RNAs were Trizol extracted using the Direct-zol RNA Microprep kit (Zymo Research) following the manufacturer's instructions and then resuspended in 20 µl DEPC-treated water. Purified RNAs were quantified by Nanodrop and analyzed on 7 M urea 8% PAGE to confirm their size and quality.

Radiolabeled ppGpp and pppGpp production and purification

^{32}P-ppGpp and ^{32}P-pppGpp were synthesized in separate reactions and purified as described previously (33). Briefly, a 50 µl reaction containing 10 µl of laboratory purified crude ribosomes, 25 mM Tris-acetate pH 8.0, 7.5 mM Mg-acetate, 30 mM K-acetate, 14 mM NH$_4$-acetate, 0.5 mM DTT, 0.1 mM EDTA, 1 mM unlabeled ATP, 0.5 mM GDP or GTP for ppGpp or pppGpp synthesis, respectively, and 100 µCi of γ^{32}P-ATP was incubated for 1 h at 37°C. ^{32}P-ppGpp or ^{32}P-pppGpp were then separated from other radioactive species by ascending PEI-cellulose chromatography and eluted from an excised strip of the plastic-backed PEI plate in 5 ml 4 M LiCl for 5 min. The eluate was then precipitated with 10–20 µl of a 10x dilution of NH$_4$OH, pelleted by centrifugation for 5 min at 4°C at 14 000 rpm, washed with 100% ethanol three times, and stored at 80°C until used. For daily use, the pellets were centrifuged, dried, resuspended in 0.5 M Na-formate pH 3.4, and neutralized with 1 M Tris-Cl pH 8.0. ppGpp or pppGpp concentrations were measured by Nanodrop.

ThiL protein production and purification

An N-terminally-10His-tagged version of ThiL (His10-ThiL) was purified by cloning the *Escherichia coli thiL* cod-

ing sequence into a pET28 plasmid, downstream from the vector-encoded T7 promoter, ribosome binding site, 10 His codons, and a thrombin cleavage site (sequence in Supplementary Table S3), to form plasmid pJJ01. pJJ01 was introduced into *E. coli* BL21DE3 cells (Novagen) by electroporation. This strain carries T7 RNA polymerase under the control of the IPTG inducible *lac*UV5 promoter. Cells were grown at 37°C in 500 ml LB supplemented with 30 µg/ml kanamycin to maintain the plasmid to an OD$_{600}$ of 0.45, and then His10-ThiL production was induced for 3 h with 1 mM IPTG. Cells were pelleted by centrifugation and lysed by sonication in a lysis buffer containing 10 mM Tris–HCl at pH 8.0, 5% glycerol, 300 mM NaCl, 2 mM β-mercaptoethanol, 0.14% sodium deoxycholate, a protease inhibitor cocktail and 0.4 mM phenylmethylsulfonyl fluoride, following the manufacturer's instructions (Thermo Fisher 78425). The lysate was cleared by centrifugation at 15 krpm for 30 min at 4°C, and the supernatant was purified on a nickel-agarose affinity column (Sigma). Eluted fractions were analyzed by SDS-PAGE with Coomassie staining, and the fractions containing His10-ThiL were further dialyzed into a buffer containing 20 mM Tris–HCl pH 8.3, 150 mM NaCl and 5 mM CaCl$_2$. The protein was Nanodrop- and Bradford assay-quantified and digested with 1 U of thrombin per mg of protein for 3 h at 37°C. The mixture was then purified again on a nickel-affinity column (Sigma) and dialyzed into a storage buffer containing 50% glycerol, 300 mM NaCl, 50 mM Na$_2$HPO$_4$ and 100 µM DTT. Proteins were analyzed by SDS-PAGE and Coomassie staining to confirm proper final size and purity and kept at –20°C for routine use or at -80°C for long-term storage.

Radiolabeled TPP production and purification

Purified ThiL was tested for TMP kinase activity in 3 µl reactions containing ThiL (either 10 µg or 20 µg, i.e. 3.33 or 6.67 µg/µl), TMP (varying from 2 to 20 mM), γ^{32}P-ATP (either 0.5 or 1 µCi), either 10 mM unlabeled ATP or no unlabeled ATP, and a 1x reaction buffer composed of 3 mM MgCl, 150 mM KCl and 1 mM Tris–HCl pH 8.0. Reactions were incubated for 1 h at 37°C and analyzed by ascending chromatography on PEI-cellulose alongside a control containing γ^{32}P-ATP. A 100 micron PEI-cellulose TLC plate (Select Scientific) was prewashed twice in water and once in methanol by ascending chromatography. ^{32}P-TPP was then separated from residual γ^{32}P-ATP and TMP by ascending chromatography on the PEI-cellulose plate in diethanolamine–methanol–formic acid–67 mM dibasic sodium phosphate at 1:15:1.5:5 proportions for 3h30, as described previously for TPP and TMP separation by ascending chromatography (39,40). Migration profiles of unlabeled thiamine pyrophosphate (TPP) and monophosphate (TMP), observed by UV-extinction, were used as standards. TPP migrated much closer to the origin than TMP, and radiolabeled TPP co-migrated with the unlabeled TPP standard. Under some test assay conditions, a second compound with slower mobility was also produced, possibly thiamine triphosphate (TTP), as has been observed in *E. coli*. TTP synthesis by ThiL has not been reported previously. Under conditions used for preparative ^{32}P-TPP

synthesis, no other radiolabeled reaction products were observed.

Purified ^{32}P-TPP for use in the RNA-DRaCALA assays was produced in 30 µl reactions containing 200 µg (i.e. 6.67 µg/ul) purified ThiL, 2 mM TMP, 10 µCi of γ^{32}P-ATP and 1× reaction buffer, and incubated for 1 h at 37°C. The reaction was then applied to the origin of a PEI-cellulose plate, and ^{32}P-TPP was separated from residual γ^{32}P-ATP and TMP by ascending chromatography as described above. After excision from the PEI plate and elution in 2 ml 4 M LiCl for 5 min, the ^{32}P-TPP eluate was precipitated with 8 ml 100% ethanol, pelleted by centrifugation for 5 min at 4°C at 14 000 rpm, washed with 100% ethanol three times, and stored at 80°C until used. For immediate use, the pellets were centrifuged, dried, and resuspended in water.

RNA-DRaCALA assays

In method A (Figure 1A), the 3′ends of *in vitro* transcribed RNAs were ligated with 3′-biotinylated cytosine bis(phosphate) at 16°C for 16 h using the RNA 3′end biotinylation kit (Pierce) following the manufacturer's instructions. The biotinylated RNAs were then extracted with chloroform:isoamyl alcohol, ethanol precipitated, washed, and resuspended in water. To ensure complete biotinylation of RNAs, we verified their migration profile with an electrophoretic mobility shift assay (EMSA) in the presence of a saturating concentration of streptavidin, followed by a Northern blot with aptamer-specific probes. In method B (Figure 1B), RNAs were transcribed with an extra aaaaag ccaccgccaccgccaccgccaccgccaccgcc 3′ tag, and these RNAs were tethered to streptavidin with an antisense biotinylated DNA (B-AS) (5′biot-ggcggtggcggtggcggtggcgg tggcggtggcggtggc). Samples containing 0.25 nM to 2 µM biotinylated or 0.25 nM to 1 µM tagged RNAs with 1 µM B-AS were denatured for 3 min at 80°C, flash frozen for 1 min on dry ice, and thawed on ice. They were then incubated for 5 min at 37°C with 1 nM up to 20 nM of ^{32}P-ppGpp or ^{32}P-pppGpp, or ^{32}P-TPP (<10 nM) in a buffer containing 10 mM MgCl$_2$ and 20 mM Tris–HCl, pH 7.5. Samples were incubated further for 5 min at 37°C after addition of streptavidin to a final concentration of 1 µM. 4 µl of these reactions were then spotted onto duplicate PRO-TRAN BA85 nitrocellulose filters (GE Healthcare Life Sciences). The dried filters were exposed to phosphor screens and analyzed subsequently with a Typhoon phosphorimager (GE Healthcare Life Sciences). Experimental error was calculated from the standard deviation, σ, of n separate replicates (at least three for each experiment) and expressed as 95% Confidence Intervals (CI), following the equation $(x - 1.96 * \frac{\sigma}{\sqrt{n}}, \ x + 1.96 * \frac{\sigma}{\sqrt{n}})$ with x as the mean value for the considered replicates. Results were fitted to a binding curve using SigmaPlot Software (Systat Software, San Jose, CA) according to the following equation at low levels of ligand : $Y = \frac{Bmax * X}{Kd + X}$ where Y is the fraction of ligand bound and X the aptamer concentration, and the K_d was determined from this curve. When binding saturation was not obtained, it was predicted using the SigmaPlot software's dynamic fit to a single binding site model, and the K_d was inferred from this prediction.

Bioinformatic model construction

RNA sequences from a family of aptamers were simultaneously folded and aligned to generate a consensus structure together with a multiple alignment using the LocaRNA software with default parameters (41). Visualization of the consensus RNA secondary structure was performed using the R2R program with the following parameters: *identity-levels* of 97%, 90% and 75% were used to indicate the degree of conservation of a nt identity, *present-levels* of 97%, 90%, 75% and 50% were used to indicate how frequently a nt is present, and *max-non-canon* of 10% was used to indicate the maximum allowable frequency of non-canonical base pairs in the structure (42). A covariance model was built using Infernal's *cmbuild* tool, and the resulting covariance model was calibrated using Infernal's *cmcalibrate* tool with default parameters (17). To assess how well an RNA fits one of the generated covariance models, Infernal's *cmsearch* tool was used to determine an E-value representing the statistical significance of the fit (17). Using the above-described approach, two models were constructed, one from the 24 aptamers binding ppGpp and one from the 22 aptamers binding pppGpp, as determined from the RNA-DRaCALA screen.

Hypergeometric *P*-value calculation

We used a hypergeometric test to calculate for a sample of $Z = 7$ aptamers, the probability H(n) of randomly selecting $n = 7$ aptamers belonging to a category of X = 14 aptamers (with a 5′-NAGG P3-P0-ss) out of a total population of Y = 27 aptamers, also called hypergeometric *P*-value, following the equation $H (n) = \frac{C(X, \ n)*C(Y-X, \ Z - n)}{C(Y, \ Z)} = \frac{(\frac{X!}{n!(X-n)!})*(\frac{(Y-X)!}{(Z-n)!((Y-X)-(Z-n))!})}{(\frac{Y!}{Z!(Y-Z)!})}$.

Bioinformatic homolog search

As a first approach, using the refined covariance models for ppGpp and pppGpp, 5242 replicons from ∼2700 bacterial genomes from all domains of life were explored. Infernal's *cmsearch* tool (17) was used to search these replicons for matches to the two covariance models and E-value were calculated indicating the statistical significance of each match (Supplementary Table S2). As a second approach, for select genes of interest, putative homologs of the gene were searched for candidate aptamers. Potential homologs were identified by conducting a TBLASTN search on a gene's protein sequence and retaining candidates with an E-value less than 1E–10 (43). A homolog's corresponding nucleotide sequence was extracted together with regions of 500 nucleotides upstream and 300 nucleotides downstream of the homolog's coding sequence. These nucleotide sequences corresponding to a putative homolog were then searched for fits to ppGpp and pppGpp covariance models using Infernal's *cmsearch* tool in order to
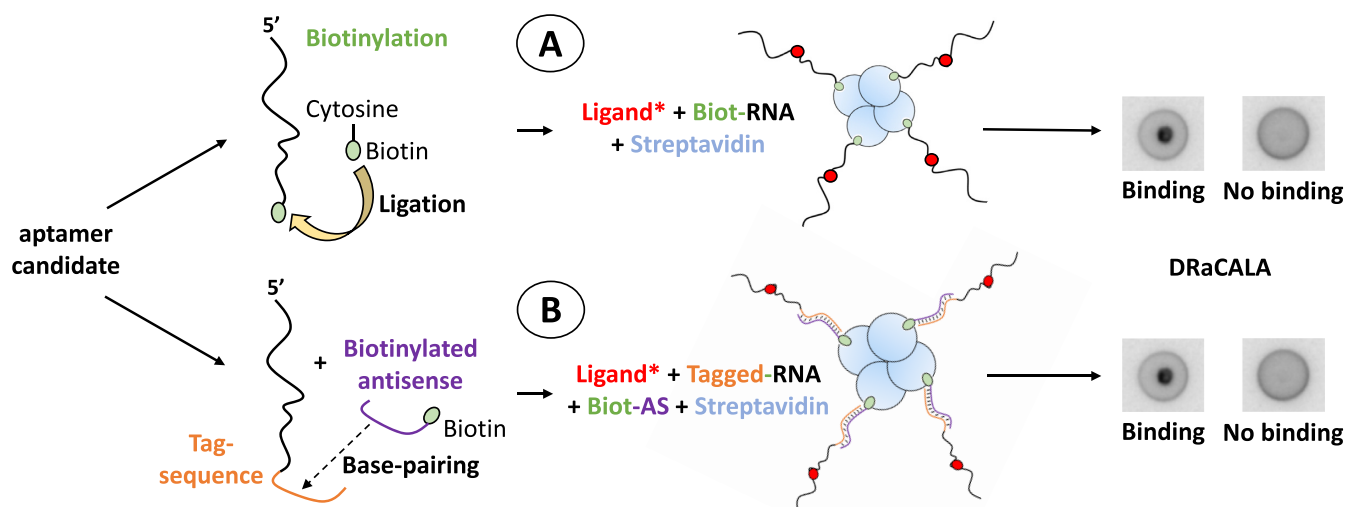
Figure 1. Outline of RNA-DRaCALA. Schematic representation of RNA-DRaCALA by two methods. In method (**A**), the RNA is directly biotinylated via the 3′-end ligation of a biotinylated cytosine. In method (**B**), the RNA is tagged with an additional 3′ sequence, targeted by a biotinylated antisense DNA oligonucleotide. In both cases, Streptavidin binds to the biotinylated RNA or RNA-antisense complex and anchors it to the nitrocellulose filter at the site of deposition. Radiolabeled ligands (Ligand*) bound to immobilized aptamer RNAs at the center of the filter are detectable by phosphorimaging.

identify candidate aptamers binding ppGpp or pppGpp, respectively (17).

RESULTS

Analysis of aptamer–ligand binding by RNA-DRaCALA

DRaCALA relies on binding of the ligand to its target in solution, followed by analysis of the diffusion pattern on a nitrocellulose membrane. Small ligand molecules diffuse across the membrane with the solvent front due to capillary action, while proteins and any ligand bound to them stick to the nitrocellulose and do not diffuse away from the original site of deposition (37,38). We used a modified version of DRaCALA in which RNA with an attached biotin moiety is immobilized on nitrocellulose through interaction with streptavidin (38) to test (p)ppGpp binding to two aptamers previously shown to interact with (p)ppGpp (25).

RNAs containing the known ppGpp-binding *ilvE* aptamer from *T. oceani* (*To ilvE*) (25) were prepared for RNA-DRaCALA, either by ligation to biotinylated cytosine (*To-B*) (Figure 1A) or by creation of aptamer RNAs with a 3′-tag (*To-T*) complementary to a pre-biotinylated antisense DNA oligonucleotide (B-AS) (Figure 1B). The *To ilvE* RNA aptamer was previously shown to bind ppGpp tightly (10 nM K_d) by in-line probing (25). The biotinylated *ilvE* RNAs were incubated with ^{32}P-ppGpp followed by addition of saturating amounts of streptavidin, and the reaction was deposited on a nitrocellulose membrane (Figure 2A). ^{32}P-ppGpp alone, or ^{32}P-ppGpp with streptavidin, diffused on the nitrocellulose, indicating that ^{32}P-ppGpp does not bind to the membrane or to the non-diffusing streptavidin. In the absence of streptavidin, the ^{32}P-ppGpp-RNA complexes also diffused on the nitrocellulose. However, in the presence of both streptavidin and the *To-B* RNA or the *To-T* RNA annealed to the antisense oligonucleotide, ^{32}P-ppGpp was concentrated in a spot at the center of the ni-

trocellulose filter, corresponding to the site of deposition, demonstrating that ^{32}P-ppGpp bound to the immobilized RNA aptamer-streptavidin complex. Slight differences in the percent bound with the two *To ilvE* RNA streptavidin tethering methods potentially could result from differences in ppGpp-binding properties of the aptamer with the 45 nt-long tag versus the 3′ end-ligated biotinylated cytosine.

A tagged version of the *To ilvE M2* mutant (*M2-T*) did not bind ^{32}P-ppGpp when measured by RNA-DRaCALA (Figure 2A), consistent with results obtained by in-line probing (25), indicating that the assay is specific. To demonstrate the specificity for ppGpp, the *To-B* aptamer was assayed in the presence of 10 nM ^{32}P-ppGpp and a molar excess (1 mM) of either unlabeled ppGpp or unlabeled GTP (Figure 2B). Unlabeled ppGpp competed with ^{32}P-ppGpp binding, whereas GTP did not, even in 10^5-fold excess. This is the only example we found where the in-line probing and RNA-DRaCALA approaches resulted in different conclusions. Sherlock *et al.* (25) found changes in the RNA cleavage profile by in-line probing of the *To ilvE* aptamer at high GTP concentrations. In contrast, GTP, even at high concentrations, did not compete with ppGpp for binding to the same aptamer when measured by RNA-DRaCALA. One possible explanation for this is the extended incubation time used for in line probing, which could result in cleavages caused by very low affinity GTP binding to the aptamer, compared to the short time used for RNA-DRaCALA.

The *ilvE* aptamer from *D. hafniense* was also tagged (*Dh-T*) and produced a clear binding signal although with lower yield than *To-T* RNA (Figure 2A), consistent with its previously reported higher K_d (6 μM for *Dh ilvE* versus 10 nM for *To ilvE* (25)). The difference in percent binding for the aptamers from the two species tethered by biotinylation using the same tagging method (*To-T* and *Dh-T*; Figure 2A) was consistent with the previous conclusion that the binding affinities of the two aptamers differed.
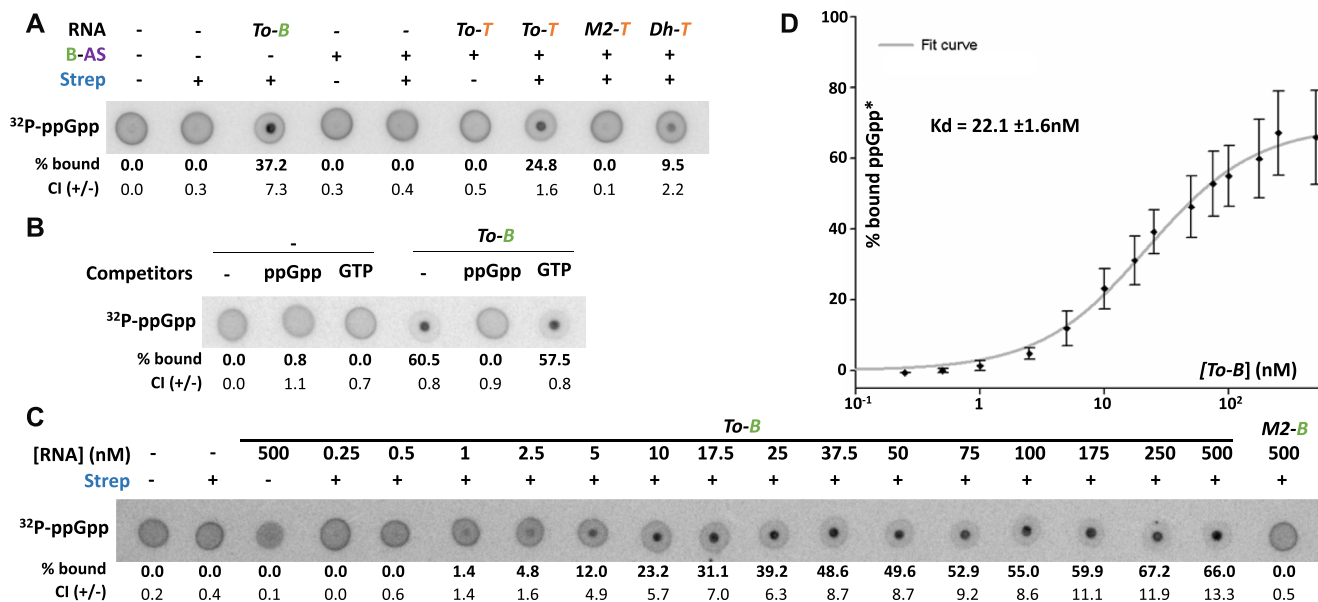
Figure 2. The RNA-DRaCALA method is specific and permits precise K_d measurement of ligand-aptamer binding. (**A**) RNA-DRaCALA assay on 0.5μM of biotinylated *ilvE* from *T. oceani* (*To-B*) or 3′-tagged *T.oceani ilvE* (*To-T*), M2 mutant *T. oceani ilvE* (*M2-T*), and *D. hafniense ilvE* (*Dh-T*) with 70 nM radiolabeled-ppGpp (^{32}P-ppGpp) in the presence or absence of 1 μM Streptavidin and/or 1 μM biotinylated antisense (B-AS). RNA-B: biotinylated RNAs; RNA-T: tagged RNAs; B-AS: biotinylated antisense; Strep: Streptavidin. All RNA-DRaCALA were reproduced in at least three independent replicates. Average values are represented under each DRaCALA sample as a percentage of ligand bound (radioactive signal at the center of the filter) divided by the total amount of signal. CI: confidence intervals for the quantified values. (**B**) RNA-DRaCALA competition assay on 1 μM of *To-B* in the presence of 1 μM Streptavidin, 20 nM ^{32}P-ppGpp, and 1 mM unlabeled ppGpp or GTP. Radiolabeled and unlabeled ligands were added simultaneously. (**C**) RNA-DRaCALA titration assay on concentrations of *To-B* ranging from 0.25 to 500 nM with 1 nM of ^{32}P-ppGpp and 1 μM Streptavidin, compared to the biotinylated M2 mutant of *T. oceani ilvE* (*M2-B*), the negative control. (**D**) The results from (C) were quantified and fitted to a saturation binding curve, resulting in a K_d of 22.1 nM for this experiment. Error bars represent standard error obtained from three biological replicates of the experiment.

Determination of K_d for ppGpp binding to the *ilvE* aptamer using RNA-DRaCALA

A more quantitative estimate of the affinity of the *To ilvE* RNA aptamer for ppGpp was also determined using the RNA-DRaCALA assay (Figure 2C, D). The concentration of the unlabeled reagent, the RNA aptamer, was varied and the concentration of the radiolabeled ligand was maintained at a constant relatively low concentration (1 nM), as described for ligand titration by DRaCALA (37,38). Increasing concentrations of *To-B* RNA were combined with a fixed low concentration of ^{32}P-ppGpp ligand (1 nM) (Figure 2C). Percentages of bound ligand measured for each *To-B* concentration were fit to a binding curve (Figure 2D), resulting in a K_d of 22.1 ± 1.6 nM, a value close to the 10 nM apparent K_d determined previously by in-line probing of *To ilvE* (25). The small difference could be attributable to the difference in RNA aptamer sequence, since in this case a biotinylated cytosine at the 3′ end of the *To ilvE* RNA was used for RNA-DRaCALA.

RNA-DRaCALA shows specific binding of ^{32}P-TPP to previously reported TPP aptamers

The RNA-DRaCALA method was also tested on two previously characterized aptamer-ligand interactions found in *E. coli* thiamine pyrophosphate (TPP) riboswitches (9). A strong binding signal for ^{32}P-TPP was observed with the *thiC-B* and *thiM-B* aptamers but not with a control, the

To ilvE aptamer, and ^{32}P-TPP binding was outcompeted by unlabeled TPP but not by unlabeled ATP, indicating that binding of TPP to the *thiC* and *thiM* aptamers was specific (Supplementary Figure S1A, B). The results with previously identified ppGpp- and TPP-binding aptamers, as well as the relative ease of application of RNA-DRaCALA to multiple samples, suggested that this method could be used to screen the binding of these or other ligands to large numbers of aptamers. However, TPP aptamer candidates identified in a recent differential structuromics study (19) did not bind ^{32}P-TPP by RNA-DRaCALA, possibly because of differences in the lengths of the RNA sequences and the solution conditions, such as the much higher ligand concentrations used in the structuromics study (Supplementary Figure S1C–E). We focused our further efforts on the analysis of aptamers from the *ykkC* subtype 2a homologous family of ppGpp binding aptamers.

RNA-DRaCALA screen of candidate (p)ppGpp aptamers reveals a range of binding affinities and specificities for pppGpp and ppGpp

We also screened additional aptamers for (p)ppGpp binding. Five homologs of *ykkC* subtype 2a aptamers were described in detail previously (25,44), *To ilvE*, *Dh ilvE*, *P. fermentans glxB* and Clostridium species BL8 (*Cbl8*) *natA*, as well as an homolog from *S. acidophilus*. Each of these aptamers bound ppGpp, and *To ilvE* bound pppGpp and ppGpp with the same K_d. However, the ppGpp binding

properties of most of the 105 *ykkC* subtype 2a aptamer candidates (25) were not tested directly. (RNAs are referred to as 'C' (for 'candidate') followed by a number and are listed with their adjacent gene and species of origin in Supplementary Table S1). Therefore, we screened 30 of the candidates by RNA-DRaCALA (Figure 3). Individual candidates were chosen for analysis (a) because they originated from bacterial species not shown experimentally to have a (p)ppGpp-binding riboswitch; (b) because the gene family in *cis* to the aptamer had not been tested experimentally for ppGpp binding and/or (c) because the aptamer had sequence or structural characteristics different from those identified in the *ykkC* subtype 2a aptamer consensus model (25,45).

Only two of the 30 tested candidates failed to bind either ^{32}P-ppGpp or ^{32}P-pppGpp, and another three bound very poorly to either ligand (2–6% of input ligand bound); the remaining 25 candidates bound at least one ligand and often both, although with different efficiencies (Figure 3; Supplementary Figure S2). For ppGpp, 24 candidates bound well (>10% of input bound), and 4 candidates bound poorly (3–10% of input bound). For pppGpp, 22 candidates bound well, and 3 bound poorly. Overall, our results strongly support the *in silico* predictions for (p)ppGpp binding from Breaker and colleagues (25), but they also highlight differences in binding by ppGpp and pppGpp.

The fact that most of the candidates bound ppGpp and pppGpp similarly (within 2-fold) suggests that binding by ppGpp and pppGpp relies on most of the same determinants (Figure 3). However, seven candidates preferred ppGpp. Four of these (C27, C38, C42 and C48) bound ppGpp two to five-fold better than pppGpp, and the other three (C21, C81 and C90) bound ppGpp more than five-fold better than pppGpp. Most strikingly, one candidate (C101) had a strong preference for pppGpp, binding 34% of pppGpp input versus only 3% of ppGpp input. Supplementary Figure S2 illustrates the diversity of binding by each ligand for all of the tested candidates.

We next performed RNA-DRaCALA with four representative aptamers at a range of RNA aptamer concentrations (rather than at a single concentration as in Figure 3) to determine K_ds for binding ppGpp and pppGpp (Figure 4). The four chosen aptamers included two with apparent specificity for ppGpp (C21 and C90)*, one that was pppGpp-specific (C101), and one that bound both ppGpp and pppGpp with very high apparent affinity (C16). The rank orders of the K_d measurements matched the rank orders from the single concentration binding measurements shown in Figure 3.

C16 had a very high affinity for ppGpp (K_d 33 ± 3.7 nM), similar to the estimated affinity for the *To ilvE* aptamer measured either by DRaCALA (22 nM in Figure 2D) or, as previously reported, by in-line probing (10 nM) (25). The affinity of C16 for pppGpp was also quite strong (K_d 137 ± 17 nM), correlating with the strong binding to the two ligands observed in the single concentration screen (65% for ppGpp and 57% for pppGpp; Figure 3A). In contrast, binding by ppGpp to the C21 and C90 aptamers was somewhat weaker than for C16 or *To ilvE* (estimated K_ds for ppGpp of 654 ± 121 nM for C21, Figure 4D; and 537 ± 154 nM for C90, Figure 4C), and was very poor for pppGpp

($K_{ds} > 5 \mu M$ for C21 and C90). These results correlate with the somewhat weaker ppGpp binding and the very poor pppGpp binding to C21 and C90 in the screen (Figure 3A; 32% and 28% ppGpp binding, respectively, compared to 5% pppGpp binding to each) and support the conclusion that these two aptamers display a definite preference for binding ppGpp versus pppGpp.

The C101 aptamer had a relatively high affinity for pppGpp (K_d of 119 ± 31 nM; Figure 4D), while displaying very little ppGpp binding even at the highest RNA concentration tested (1 μM), consistent with results from the screen in Figure 3. C101 is the only aptamer in the set of 30 candidates analyzed that had a higher affinity for pppGpp than ppGpp, the first example of a pppGpp-specific aptamer. In summary, the results of Figures 3 and 4 demonstrate that very closely related aptamers can bind two very similar ligands selectively, i.e. they can bind both, only one or the other, or neither ligand.

Refinement of the ppGpp-aptamer consensus model and creation of a pppGpp-aptamer model

Historically, sequence and secondary structure consensus models of aptamers have been based primarily on homologs close enough in sequence and predicted secondary structure to be part of the same family or subtype (22–25). However, such models are only as accurate as the aptamer datasets used to create them. Here we used aptamers with experimentally determined ligand binding properties to build secondary structure models, reasoning that these models could provide novel details about these RNAs.

As described above, not all of the tested aptamers bound similarly to both ppGpp and pppGpp. Therefore, we created two aptamer models. One model was based on all aptamers that bound ppGpp above a threshold of 10% ligand bound (i.e. the top 24 ppGpp-binding aptamers), and the other was based on all aptamers that bound pppGpp above the same threshold (i.e. the top 22 pppGpp-binding aptamers; Figure 3A). We hypothesized that differences between the two models might identify aptamer features important for distinguishing between the two ligands. A consensus secondary structure was generated for each aptamer set by multiple sequence alignment and covariance analysis (17,41,42) (Figure 5A,B). Secondary structure models were also created for each individual aptamer based on mFold predictions (46) and are shown for four representative aptamers in Figure 6 and five additional non-binding aptamers in Supplementary Figure S3. In addition to secondary structure models in Figure 5, these features are also represented in images adapted from the 3D crystal structure model of the *S. acidophilus* ppGpp aptamer (Figure 7; PDB ID 6dmc) (44).

Features that contribute to binding by both ppGpp and pppGpp

Not surprisingly, the secondary structure models show considerable similarity to previous models for *ykkC* subtype 2a (p)ppGpp aptamers, including three prominent stem-loop structures, P1, P2 and P3 (23,25). In addition, our secondary structure models include a fourth highly conserved
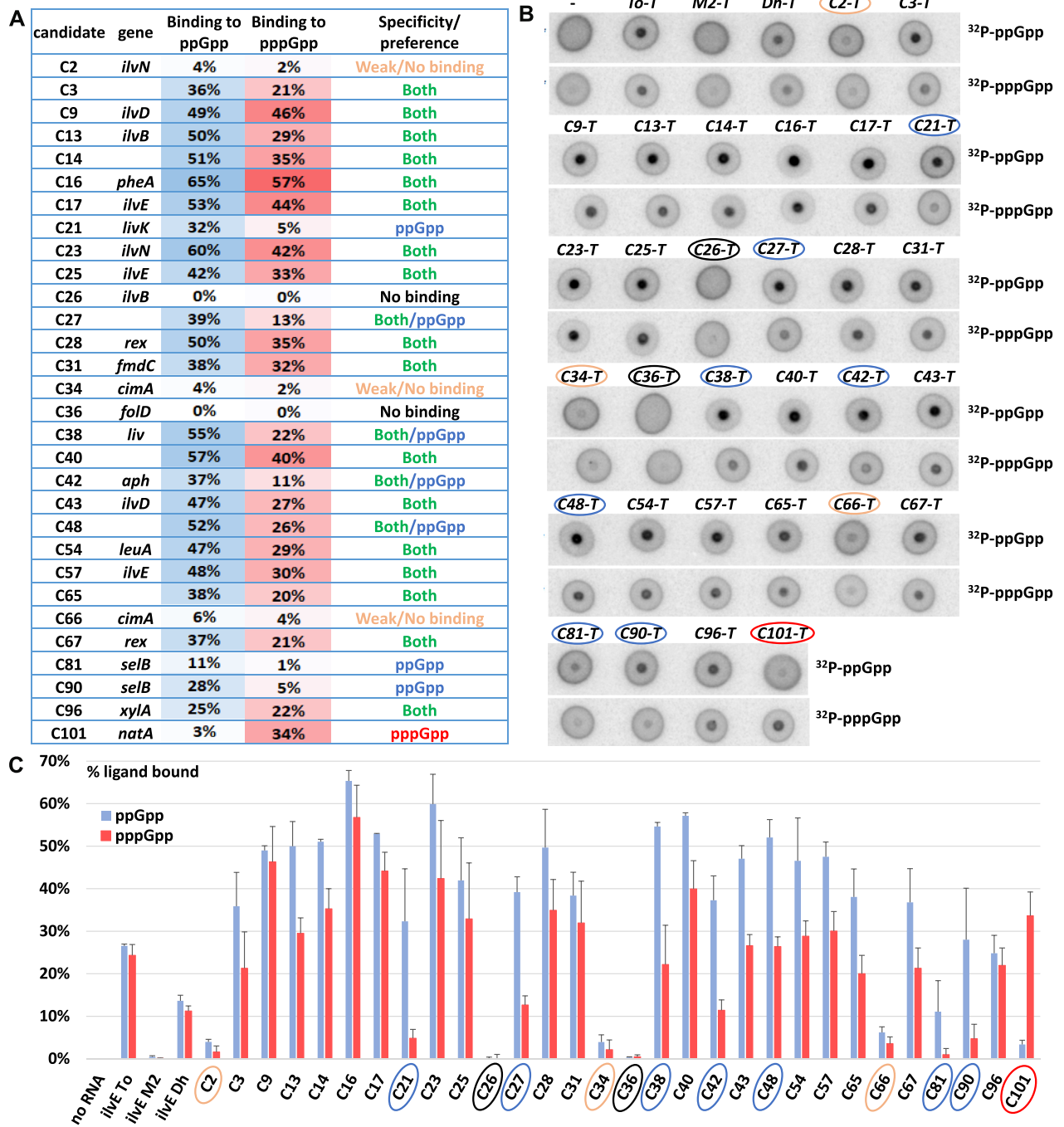
Figure 3. RNA-DRaCALA screen unveils a wide range of binding preferences to ppGpp and pppGpp in *ykkC* subtype 2a aptamers. (**A**) Subset of *ykkC* subtype 2a RNA-aptamer candidates selected for this screen. (**B**) RNA-DRaCALA screen of 0.5 μM 3′-tagged RNA-aptamer candidates in the presence of 1 μM biotinylated antisense DNAs, 1.5 μM Streptavidin, and 20 nM ^{32}P-ppGpp or ^{32}P-pppGpp. *To-T* and *Dh-T* were used as positive controls for the binding to both ligands, and *M2-T* was included as a negative control. (**C**) Quantitation of three independent biological replicates of the experiment shown in (**B**), represented as the fraction of radiolabeled ligand signal bound to the RNA aptamer candidate divided by the total radioactive signal, normalized to the control without an RNA candidate to account for background. Error bars correspond to 95% confidence intervals.
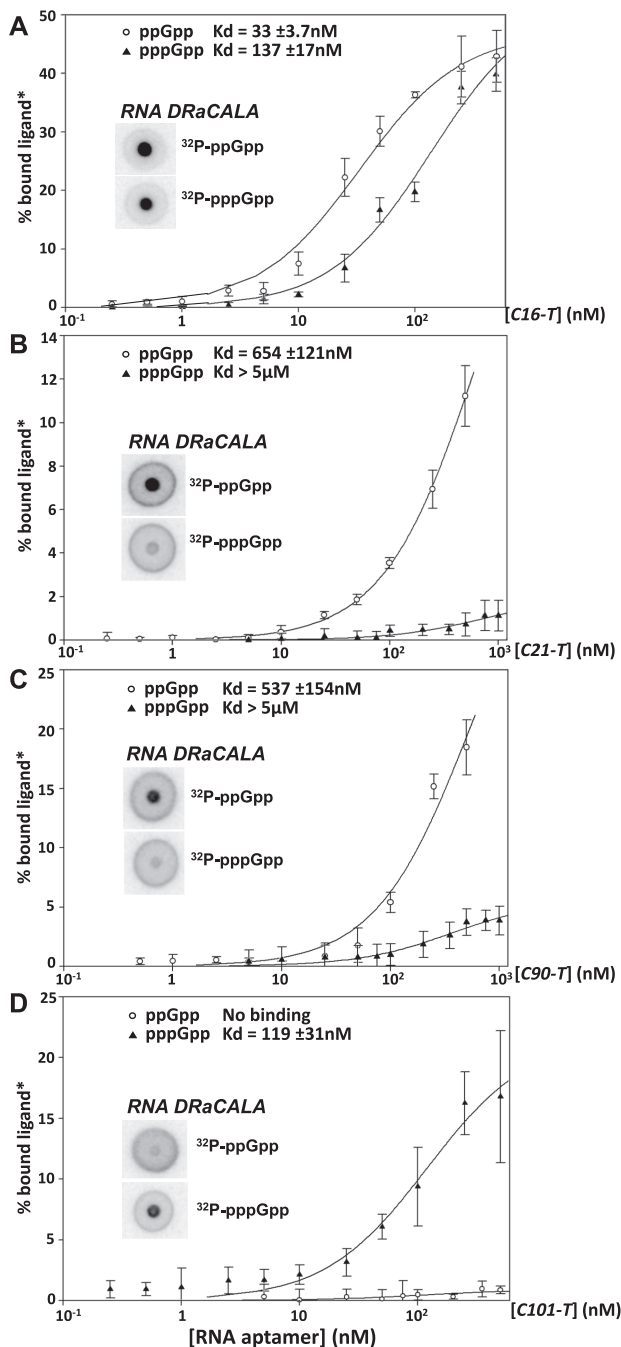
Figure 4. Determination of binding constants for representative *ykkC* subtype 2a aptamers. Binding curves and RNA-DRaCALA images for the C16 (**A**), C21 (**B**), C90 (**C**) and C101 (**D**) RNA aptamers. Titrations were performed with RNA concentrations ranging from 25 nM up to 500 nM or 1 μM with 10 nM ^{32}P-ppGpp or ^{32}P-pppGpp. Percent bound values shown are from three independent repeats. Results were fitted to a binding curve using the equation and described for Figure 1C. Binding constants (K_ds) were inferred from these curves where appropriate.

stem, referred to here as P0. P0 is formed by the 5′ and 3′ ends of the aptamer and was absent in the original model (23,25) but present in an updated model based on crystal structures (44,45,47). The junction of the four stem-loops in

these new models creates the pocket that contains the ligand binding site (Figures 5A, B and 6) (44,47). We conclude that models based on experimentally validated aptamer binding result in greater precision and extra detail that was obscured by inclusion of even small numbers of non-binding aptamers.

The ppGpp and pppGpp consensus models (Figure 5A, B) also emphasize the importance of several previously described features of (p)ppGpp-binding aptamers and reveal novel features that correlate with binding efficiency or specificity. Previously described features seen in these models include highly conserved nucleotides that contact the 5′ and 3′ phosphates and the guanine base of ppGpp in crystal structures (44,47) (see circled positions in Figures 5–7, and Supplementary Figures S3 and S4). Our new models include (a) the P0 stem (44,47); (b) the conserved residue A-5 in the P0 stem (A-4 in Figure 7), that binds the 5′ phosphates of (p)ppGpp (44) (circled in blue in Figures 5-7); (c) the conserved unpaired residue G-6 (G-5 in Figure 7) adjacent to A-5 of the P0 stem, that makes H$_2$O and Mg^{2+} -mediated contacts the with 5′ phosphates of (p)ppGpp, stacks with guanine of (p)ppGpp (Figure 7C) and forms a non-canonical base pair (Figure 7E) with the first A of a highly conserved P2-ACAC sequence in the region between the P2 and P3 stems (25,44) (Figure 5A, B; A-69 in Figure 7), contributing to stabilization of the ligand binding pocket (44,47); (d) the first C of P2-ACAC, a residue involved in base paring with the guanine base of (p)ppGpp (circled in red, Figure 5A, B, C-70 in Figure 7); (e) the second A of the P2-ACAC motif (A-71 in Figure 7), that hydrogen bonds with the conserved G nt in the bulge between P1.a and P1.b; this conserved G is circled in green in Figure 5A,B; G-43 in Figure 7; (f) the second C of the P2-ACAC motif (C-72 in Figure 7) contacts the 3′ phosphates of (p)ppGpp (25,44,47); (g) A-96 (in Figure 7) in the single stranded region between the P3 and P0 stems (P3-P0-ss) also contacts the 3′ phosphates of (p)ppGpp. In addition, each of the conserved purines in the P3-P0-ss (A-96, A-97, G-98 in Figure 7) stack and wrap around the ligand binding pocket, and G-98 also stacks with the second C of the P2-ACAC motif (C-72 in Figure 7C, D).

The importance of these features is consistent with their absence in aptamer candidates that bound the ligands weakly or not at all in our assays. These aptamers lacked at least one of the following features (a) a 4–5 nt unpaired region between the P3 and P0 stems (very poor binders C2, C34 and C66 have only 3 nt; Supplementary Figure S3 A,B,C); (b) the highly conserved P2-ACAC (non-binder C26 lacks the first A; Supplementary Figure S3D); (c) the highly conserved G at position 6, the first unpaired nucleotide 3′ to the P0 helix (very poor binder C2 and non-binder C36 have an A at this position; Supplementary Figure S3A,E); or (d) the highly conserved C between P1.a and P1.b, stacking with and immediately 3′ to the G (G-43 and C-44 in Figure 7E) that contacts the 3′ phosphates of (p)ppGpp (25,44,47) (non-binding aptamer C36 lacks this C, Supplementary Figure S3E). Of note, this C is mutated to A in the *To ilvE M2* mutant that has lost its ability to bind the ligand (25) (Figure 2A).
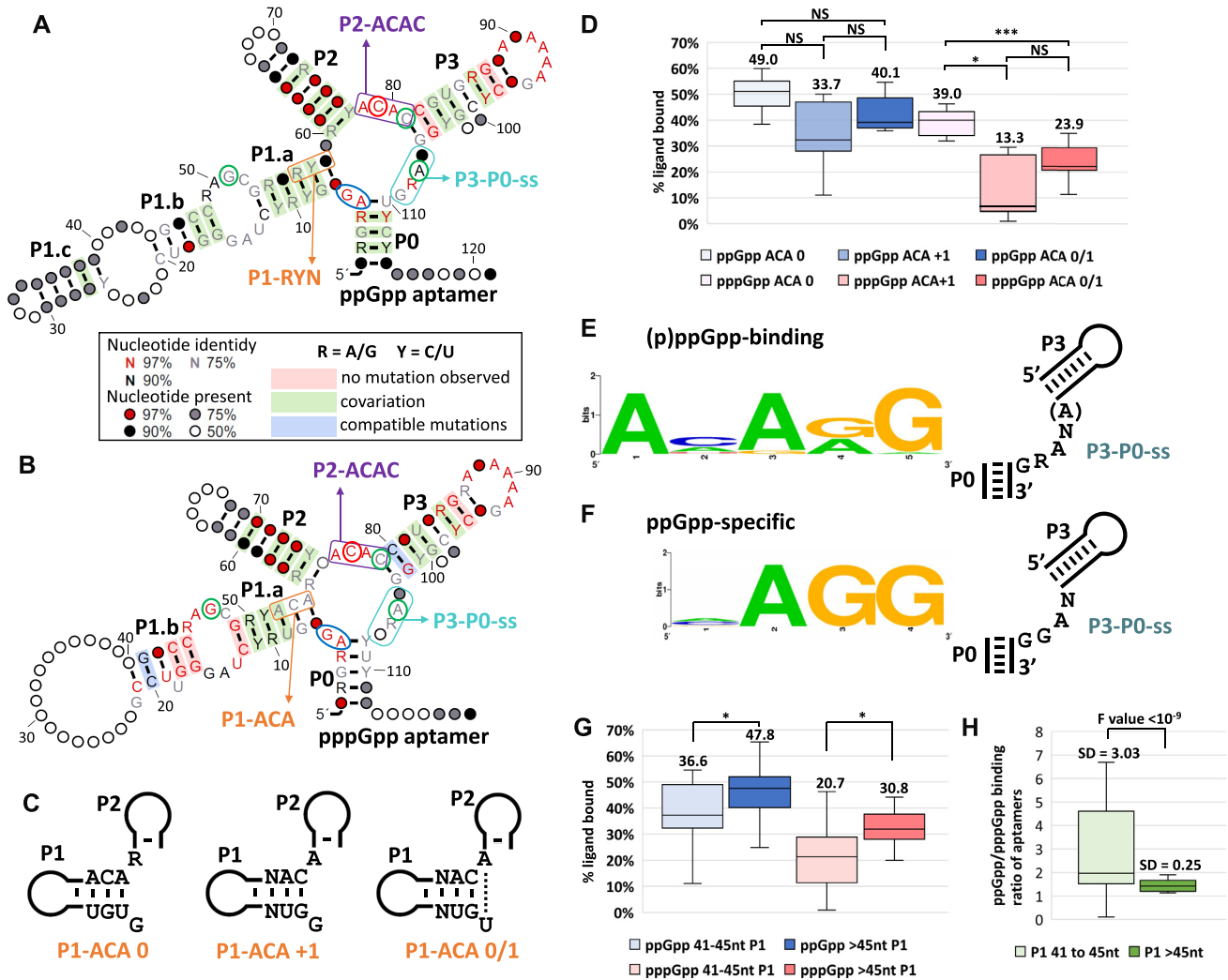
Figure 5. Refined ppGpp- and pppGpp-binding RNA aptamer models. Consensus RNA aptamer sequences and secondary structures derived from the top 24 ppGpp binding aptamers (**A**) and the top 22 pppGpp binding aptamers (**B**), respectively, determined from the RNA-DRaCALA screen in Figure 3. Stem structures are annotated from P0 to P3; nucleotides involved in contacts with the ligand in the binding pocket, as determined from the crystal structure of an RNA aptamer-ppGpp complex (44), are circled in green (contacts with 3′ phosphates), blue (contacts with 5′ phosphates, direct or via a Mg²⁺ ion), or red (base pairing with (p)ppGpp). Compatible mutations are U-to-C or C-to-U substitutions in C-G/U-G base pairs, or A-to-G or G-to-A substitutions in A-U/G-U base pairs. Relevant motifs, also described in the text, are indicated, and circled: P1-ACA in orange (P1-RYN in panel A), P2-ACAC in purple and P3-P0-ss in teal. (**C**) Representation of the 3 most frequent configurations in which the P1-ACA motif was found in aptamers. P1-ACA was either embedded at the 3′-end of the P1 stem-loop (P1-ACA 0), shifted 1 nt downstream (P1-ACA + 1), or formed an extra A–U base pair represented by a dotted line, leaving no single stranded nt between P1 and P2 (P1-ACA 0/1). (**D**). Box plot analysis of % ppGpp or pppGpp bound to aptamers harboring the motif P1-ACA 0 (7 aptamers), P1-ACA + 1 (5 aptamers), or P1-ACA 0/1 (11 aptamers). Using P3-P0-ss sequences (Supplementary Table S1) and results presented in Figure 3 and Supplementary Figure S4 (for C94, C99 and C100), we defined consensus sequences for the P3-P0-ss motif for aptamers that bind to both ligands (**E**) or that prefer ppGpp (**F**). Sequence Logos were designed using the WebLogo server (https://weblogo.berkeley.edu/logo.cgi (58,59). Consensus sequences are also represented in their proposed single stranded conformation between P3 and P0 stems. The consensus sequences are 5′-(A)NARG or NAGG for (p)ppGpp binding or ppGpp-specificity, respectively, where R is a purine, and N is any nucleotide. The first position, shown in parentheses, corresponds to the position adjacent to the P3 stem in the five candidates with a 5 nt P3-P0-ss. Note that the secondary structure consensus models in (A) and (B) do not illustrate the variation in length of the P3-P0-ss sequences among the aptamers. (**G**) Box plot analysis of % ppGpp or pppGpp bound to aptamers with P1 stem-loops ranging from 41 to 45 nt long (13 aptamers) or with P1 stem loops longer than 45 nt (12 aptamers). (**H**) Box plot analysis of the ratio of ppGpp/pppGpp bound to each aptamer for aptamers with P1 stem loops ranging from 41 to 45 nt long and for P1 stem loops longer than 45 nt. Standard deviations (SD) are indicated, as well as the result of a Fisher test for differences in variance. For (D), (G) and (H), the inclusive median method was used to identify quartiles. For (D) and (G), average values are indicated for each category, as well as the results of a two tailed Student's *t*-test, with *P*-values < 0.05 (*) or < 0.0005 (***) or not significant (NS).
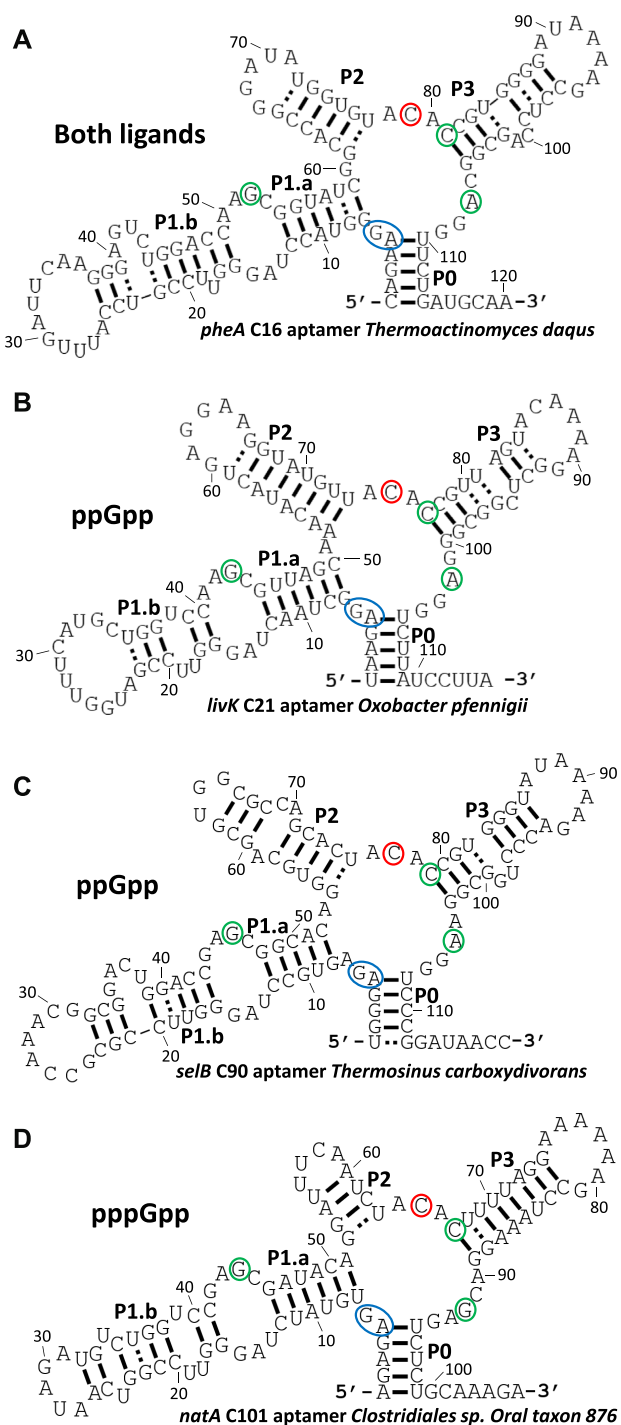
Figure 6. Proposed secondary structures for individual aptamers. Sequences and proposed secondary structure for the representative *ykkC* motif subtype 2a aptamers for which binding titrations were determined in Figure 4. These aptamers bind to both ppGpp and pppGpp (**A**) or bind specifically to either ppGpp (**B, C**) or pppGpp. (**D**) Aptamer candidate numbers and associated gene and species names are shown. Stem-loop numbers are indicated, and positions of interest are circled as in Figure 5.

Features that contribute to specificity for ppGpp over pppGpp

An ACA motif located at the 3′ end of the P1.a stem is present in > 75% of the aptamers that bind well to pppGpp (a motif distinct from the above-mentioned P2-ACAC motif) (Figure 5B). This ACA motif (hereafter referred to as the P1-ACA motif) is not as well conserved in aptamers that bind ppGpp, where the less specific sequence RYN is found instead (Figure 5A and previous ppGpp binding models (25,45)). This suggests that an ACA motif at this location may have a specific role in binding pppGpp versus ppGpp. The corollary is that aptamers without an adequate P1-ACA motif may not bind pppGpp and could tend to be ppGpp-specific. The exact position of the P1-ACA motif with respect to the 3′ end of the P1 stem varies slightly among aptamers, giving rise to three groups (Figure 5C). In group P1-ACA 0 (7 aptamers, Supplementary Table S1) the third position of the ACA is the last paired position in the P1 stem. In group P1-ACA + 1 (5 aptamers, Supplementary Table S1) there is an unpaired A at the third position of the P1-ACA motif. Finally, in Group P1-ACA 0/1 (11 aptamers, Supplementary Table S1) the third position in the ACA motif pairs with a U at the 5′ end of the P1 stem (A-50 and U6 in Figure 7A,B), leaving no single stranded nt between P1 and P2.

To quantify whether the three P1-ACA motif groups (0, +1 and 0/1) had statistically different binding properties for ppGpp or pppGpp, we performed a box-plot analysis of the percentage of ligand bound for each group (Figure 5D). The 5 aptamers that bound both ligands poorly or not at all and the 2 aptamers that lacked the P1-ACA were excluded. The position of the ACA motif did not significantly affect the percentage of ppGpp bound (33.7–49.0% bound in the three ACA groups), but there were statistically significant differences for the three groups in pppGpp binding. pppGpp binding by the P1-ACA 0 group (39%) was much better than by the P1-ACA + 1 group (13.3%) or the P1-ACA 0/1 group (23.9%), further suggesting that the ACA at the end of the P1 stem plays a role in binding pppGpp. In addition, we note that all the aptamers that favor ppGpp (C21, C27, C38, C42, C48, C81 and C90; Supplementary Table S1) belong to the ACA + 1 or ACA 0/1 groups, consistent with the interpretation that this feature contributes to ligand specificity, and that a P1-ACA 0 is generally more permissive for pppGpp-binding.

While the P1-ACA motif is not in close proximity to the ligand binding pocket as shown in the crystal structure (Figure 7), this motif could indirectly constrain P1 stem-loop folding, specifically at the bulge containing the single stranded G that contacts the 3′ phosphates of (p)ppGpp (G-43 in Figure 7). It could also impact the positioning of the two nt that contact the 5′ phosphates of (p)ppGpp (A-4 and G-5 in Figure 7), for instance in aptamers with P1-ACA 0/1 motifs (Figure 5C), with only one single stranded nt between P0 and P1 stem-loops and therefore less flexibility. Overall, variations in the binding pocket that result from formation of the A-50 and U-6 base-pair at the end of the P1 stem in P1-ACA 0/1 aptamers may affect accommodation of the extra 5′ phosphate of pppGpp in the binding pocket (Figure 5D).
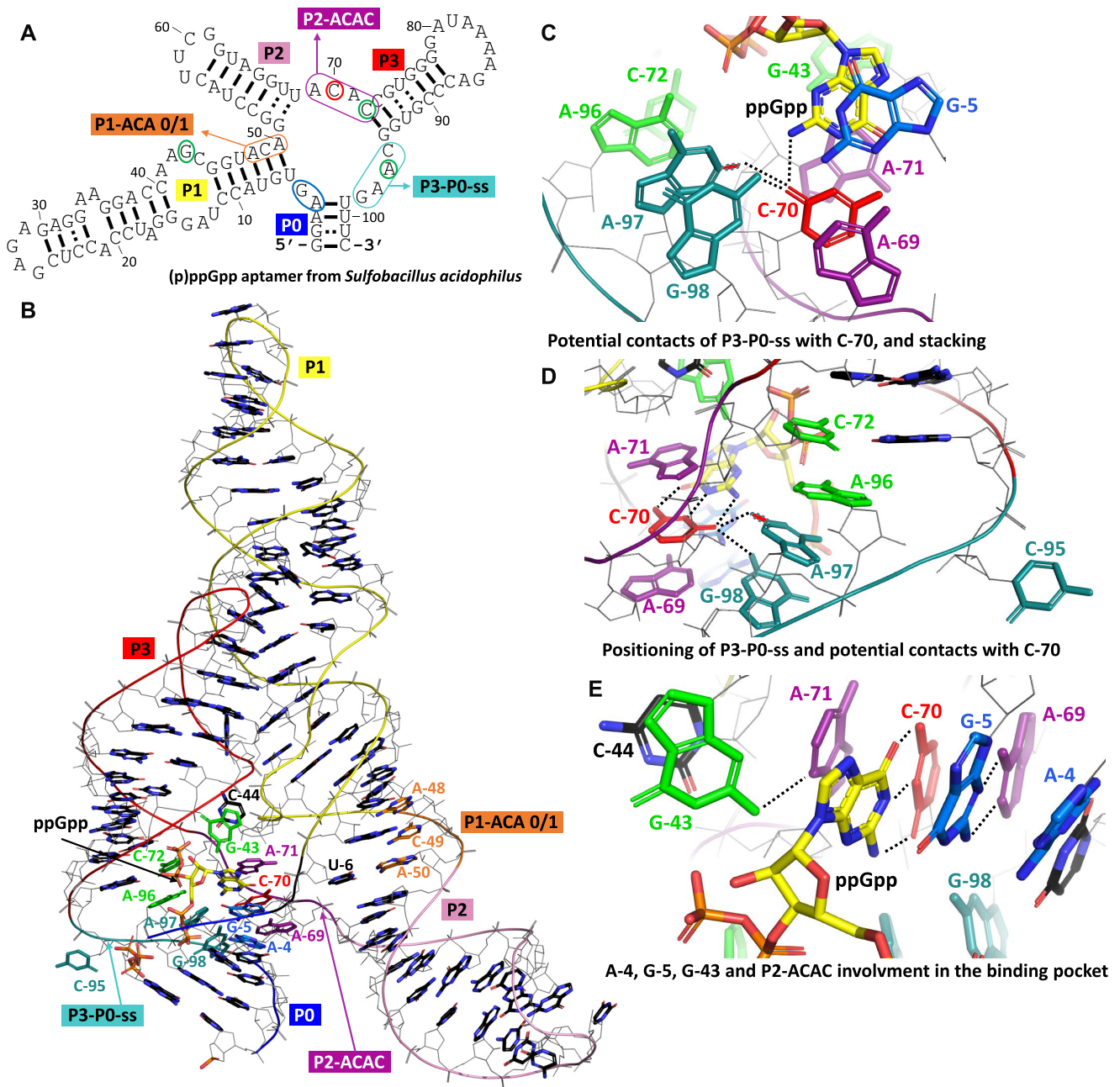
Figure 7. Structural perspective on the contribution of various features to ligand specificity. (**A**) Secondary structure and sequence of the *S. acidophilus* ppGpp-aptamer described in (44), PDB ID 6dmc. Positions of interest are circled as in Figure 5. Zoomed-in views of the features P3-P0-ss (**B–D**), P2-ACAC (**B–E**) and P1-ACA (**B**) with RNA backbone shown in ribbon form respectively represented in blue for P0, yellow for P1, pink for P2, red for P3, teal for P3-P0-ss, purple for P2-ACAC and orange for P1-ACA. Bases and ppGpp are depicted in stick representation, and those of interest are represented in blue when they contact ppGpp 5′-phosphates directly (A-4) or via Mg^{2+} interactions (G-5), in green when they contact ppGpp 3′-phosphates (G-43, C-72 and A-96), in red for the C-70 base-pairing with ppGpp, and otherwise in the color of their respective region (P3-P0-ss in teal, P2-ACAC in purple and P1-ACA in orange, respectively). ppGpp is depicted in yellow and atomic colors. Dotted black lines represent distances equal to or shorter than 3.3 Å between two bases where they can form hydrogen bonds. In (C) and (D), the NH_2 group of a hypothetical G-97 guanine was depicted in red and teal dotted lines, to show its proximity and potential interaction with C-70. Crystal structure figures were prepared using PyMol (http://pymol.org/).

Unpaired region between P3 and P0 stems

Our data also suggest that the sequence of the 4–5 nt unpaired region between the P3 and P0 stems (P3-P0-ss) may be one of the features that contributes to the preference for ppGpp over pppGpp. A consensus sequence for P3-P0-ss, 5′-(A)NARG (Figure 5E; R corresponding to a purine) was derived from all (p)ppGpp binding aptamers (25 binders, plus *To ilvE* and *Dh ilvE*, but not the non-binders) (Figure 5A, B, E, and Supplementary Table S1). However, the seven aptamers that preferentially bind ppGpp all have the 4 nt P3-P0-ss sequence 5′-NAGG (Figure 5F) and constitute half of the aptamers with the 5′-NAGG sequence. The probability that this distribution occurred randomly is 0,00386 (see Materials and Methods), suggesting that the 5′-NAGG motif is very favorable (but not sufficient) for ppGpp-specificity. The 5 nt P3-P0-ss sequence for the C101 aptamer that preferentially binds pppGpp is 5′-ACGAG, but it is unclear whether the difference in the P3-P0-ss length and sequence contributes to the preference for pppGpp, since the sample size of those that prefer pppGpp is too small for statistical comparison. Nevertheless, these results suggest that the sequence of the P3-P0-ss motif could influence ligand specificity, perhaps by accommodating the different 5′ phosphate moiety in the two ligands.

Crystal structure data of the *S. acidophilus* ppGpp aptamer, that has a 5′-CAAG P3-P0-ss, shows that the first C (C-95) bends away from the binding pocket (Figure 7B,D), probably explaining the low conservation of this nt position (Figure 5A, B). However, the last three positions of the P3-P0-ss region (A-96, A-97, G-98) stack together (Figure 7C,D), and position G-98 is in close proximity to and could form a hydrogen bond with the first C of P2-ACAC (C-70), the base that pairs with (p)ppGpp. We speculate that potential hydrogen bonding between C-70 and the G residues that correspond to both positions 98 and 97 in all of the aptamers that preferentially bind ppGpp (Figure 5F) could alter the position of C-70, leading to subtle changes in the binding pocket that favor the binding of ppGpp rather than pppGpp (Figure 7 C,D).

The length of the P1 stem-loop contributes to binding efficiency and to specificity for ppGpp or pppGpp

Each set of candidates used to generate the consensus models for ppGpp-binding or pppGpp-binding aptamers showed considerable variability in the overall length of the P1 stem (varying from 41 to 119 nt) (Figure 5A, B). To analyze the effect of P1 stem length, we analyzed aptamers in two groups, one group with P1 stems from 41–45 nt in length (shorter stems; 13 aptamers; see Supplementary Table S1) and one group with P1 stems longer than 45 nt (longer stems; 12 aptamers; Supplementary Table S1). Aptamers with weak to no binding of both ligands were excluded from the analysis. Statistically significant differences in percent binding were observed between the short and long groups for each of the ligands. An average of 47.8% of input ppGpp bound to aptamers with longer P1 stems versus an average of 36.6% binding to aptamers with shorter stems (Figure 5G). Similarly, for pppGpp a higher percentage of binding was observed for aptamers with longer P1

stems than for aptamers with shorter P1 stems (30.8% versus 20.7%). Although these effects are relatively small, they suggest that a longer P1 stem could result in a greater number of secondary structure folding options and therefore a higher probability that the optimal binding conformation is present in solution.

However, the range of binding preferences for the two ligands in aptamers in the short P1 group (represented as ppGpp/pppGpp ratios) differed substantially from the range exhibited by those in the long P1 group (Figure 5H). The ratio of ppGpp-bound to pppGpp-bound for each of the aptamers in the short P1 stem group varied widely (average value of 3.29 with a standard deviation of 3.03), indicating a large variation in the preference for binding to one ligand or the other. This group included all of the aptamers with strong preference for one ligand or the other (e.g. a ratio of ppGpp/pppGpp bound of 0.1 for the aptamer with strong pppGpp preference to ∼6.4 for an aptamer with a strong preference for ppGpp). In contrast, the aptamers in the longer P1 stem group all displayed similar binding to each of the two ligands (an average ratio of 1.43, with a standard deviation of 0.25). These observations are consistent with a model in which the shorter P1 stem affords less conformational variability, reducing the occurrence of conformations that are able to bind both ligands and thereby promoting binding to one over the other. Of note, the seven aptamers with a short P1 stem that also have a 5′-NAGG P3-P0-ss motif are the same seven aptamers that prefer ppGpp. This suggests that the combination of those features is sufficient to drive ppGpp-specificity.

Ligand-specificity and the *natA* and *selB* genes

We showed above that only one aptamer, C101, was strictly specific for pppGpp rather than ppGpp (Figures 3, 4D, 6D). This aptamer, from the *natA* gene of Clostridiales bacterium oral taxon 876, Strain F0540 (*Cbot*), has been reported to regulate ion transport (48)). We used RNA-DRaCALA to test whether pppGpp-binding specificity is also a property of *natA* aptamers from four other species from the list of 105 potential ppGpp-binding candidates reported by Breaker and colleagues (25) (Supplementary Table S1; Supplementary Figure S4A). C94 was previously shown to bind ppGpp with a K_d of ∼400 nM, but it was not tested for pppGpp binding (25). We found that C94 preferentially bound pppGpp (19% pppGpp bound versus 8% ppGpp bound), although it was not strictly specific for pppGpp (Supplementary Figure S4). Another aptamer, C105, was unable to bind either ligand, and the remaining two of the four candidates, C99 and C100, bound ppGpp and pppGpp similarly. For comparison in the same experiment, the *To ilvE* aptamer, described above, also bound equally to ppGpp and pppGpp (Supplementary Figure S4B,C).

Although C94 was not quite as pppGpp-specific as C101 (Supplementary Figure S4B,C), the two aptamers have ∼90% sequence identity (Supplementary Figure S4D,E). C94 and C101 both contain a P1 stem-loop shorter than 45 nt, i.e. the length more favorable for ligand specificity, as described above. They also have the same P3-P0-ss, a 5nt-long sequence that lacks the two G nt next to P0 typical of aptamers that display a preference for ppGpp (Supple-

mentary Figure S4E and Figure 7D). Most differences between C94 and C101 correspond to covariations, compatible mutations, or mutations in the single-stranded loops of P1 or P2, i.e. far from the binding pocket. However, they differ in the location of a feature identified above as a contributor to pppGpp binding, the P1-ACA motif, perhaps accounting for C101's greater degree of pppGpp specificity (Supplementary Figure S4). That is, they both share the P1-ACA 0/1 positioning, but in C101, this motif is the canonical ACA sequence (Figure 6D), whereas in C94, it is GCA (Supplementary Figure S4D,E).

C94 and C101 are found in the Clostridiales taxon while the other 3 *natA* aptamers that were tested for (p)ppGpp binding, C99, C100, C105, are from other bacterial taxa (Cohnella and Paenibacillus; see Figure 8; Supplementary Table S1). Thus, preferential binding of pppGpp may have evolved and/or propagated by horizonal transfer in the *Clostridiales* taxon. It seems likely that additional pppGpp-specific aptamers occur in *cis* with *natA* genes in other species in this taxon.

At least one other aptamer-containing gene may be associated with ligand-specific binding, the gene for selenocysteine-specific translation elongation factor *selB*, typically found in the Selenomonadales. Indeed, two *selB* aptamers, C81 and C90, were specific for ppGpp (Figures 3 and 6C; see also section below). Potential evolutionary relationships between species having (p)ppGpp-aptamers are depicted in Figure 8.

Additional *ykkC* subtype 2a aptamers

To expand the family of *ykkC* subtype 2a aptamers, two *in silico* approaches were used to search genomes for more homologs. First, 2700 bacterial genomes from all domains were searched for any matches to either of our refined models for ppGpp-binding or pppGpp-binding aptamers. 101 candidates were obtained from this search and a fitness score for each to the ppGpp and pppGpp models, as well as to the original *ykkC* subtype 2a (p)ppGpp aptamer model (25) was calculated (Supplementary Table S2). In addition, fitness scores to several other *ykkC* subtype models, including the *ykkC* subtype 2b PRPP aptamer model (23), the *ykkC* subtype 2c (d)ADP/(d)CDP aptamer model (24), the *ykkC* subtype 2d orphan aptamer model (45,49), and the *ykkC* subtype 1 guanidine I aptamer model (22) were determined (Supplementary Table S2). Each of these models has been described extensively and reviewed recently (49).

Twenty-two of the 101 candidates fit best to either the ppGpp or the pppGpp models, 19 of which were identified bioinformatically by Sherlock and colleagues (25) (Supplementary Tables S1 and S2). Forty-nine aptamers fit best to the consensus model for binding PRPP, and 30 fit best to other *ykkC* subtypes. The high incidence of 'best fits' to models other than those for (p)ppGpp results from the high similarity among *ykkC* aptamer subtypes, especially (p)ppGpp-binding subtype 2a and PRPP-binding subtype 2b, which can differ by as little as one nucleotide in stem-loop P3 (23,25,47). The bacterial class and order for each of the new aptamers identified in this search is shown in Figure 8. These results suggest potential evolutionary relationships among the experimentally confirmed (p)ppGpp

binding aptamers, the predicted PRPP binding aptamers, the predicted (d)ADP/CTP binding aptamers, and the predicted guanidine binding aptamers.

The second approach used a bioinformatic analysis of preselected genes from our experimental validation of (p)ppGpp aptamers (*natA,* an ABC transporter; *selB,* a selenocysteine-specific translation elongation factor; and *rex,* a redox-sensitive transcriptional repressor) (Figure 3). In a two-step procedure, a search for homologs of these proteins (annotated as well as unannotated) was followed by a search for *ykkC* aptamer homologs adjacent to the coding sequences. Among hundreds of identified protein coding homologs, twenty-three *ykkC* aptamers associated with *natA* genes were identified, almost exclusively in *Paenibacillaceae* bacteria (numbered *natA1* to *natA23*). One of the 23 *natA*-associated aptamers was identified previously (C99; Supplementary Table S1, Figure 8 and Supplementary Figure S4). Two additional aptamers were found in *selB* homologs within the order *Selenomonadales* (named *selB1* and *selB2*), and two in *rex* homologs in *Peptococcaceae Desulfosporosinus* bacteria (*rex1* and *rex2;* Figure 8). In contrast with the homologs discovered with the first method described above, most of these aptamers fit best to either of our new (p)ppGpp models (Figure 8 and Supplementary Table S2). More broadly, these results support the conclusion from Breaker and colleagues that aptamers and the ligands that bind them are not distributed randomly: they are often conserved with the genes they regulate.

DISCUSSION

Use of RNA-DRaCALA for aptamer identification and analysis

We show here that RNA-DRaCALA provides a rapid and accurate method for testing ligand binding to multiple RNA aptamers and for determining binding affinities *in vitro*. The affinities of the previously characterized *ilvE T. oceani* ppGpp aptamer, determined by the traditional in-line probing method (25) and by RNA-DRaCALA (Figure 2), were very similar. In addition, the specificity of RNA-DRaCALA was shown by the inability of similar molecules to compete with the radiolabeled (p)ppGpp or TPP ligand for binding to the known aptamers. Based on results of our screen of a large set of ppGpp aptamer candidates, we suggest that RNA-DRaCALA is a useful alternative to in-line probing for identifying aptamers that bind a specific ligand for which a radiolabeled form is available. Other recently identified modified nucleotides could be close enough to (p)ppGpp to be able to bind a subset of the *ykkC* subtype 2a aptamer homologs (32). For instance, the smaller pGpp might be able to bind aptamers with 3 nt P3-P0-ss that did not allow (p)ppGpp binding, such as C2, C33 and/or C66 aptamers (Supplementary Figure S3A,B,C). For many modified nucleotides, there is no corresponding aptamer family identified yet. One could select candidate aptamers from multiple orphan aptamer families and screen them with RNA-DRaCALA for binding newly described ligands.

Our analysis of 30 *in-silico* predicted ppGpp-binding aptamers revealed that most of them did bind to ppGpp, though with varying affinities. In addition, a large subset of
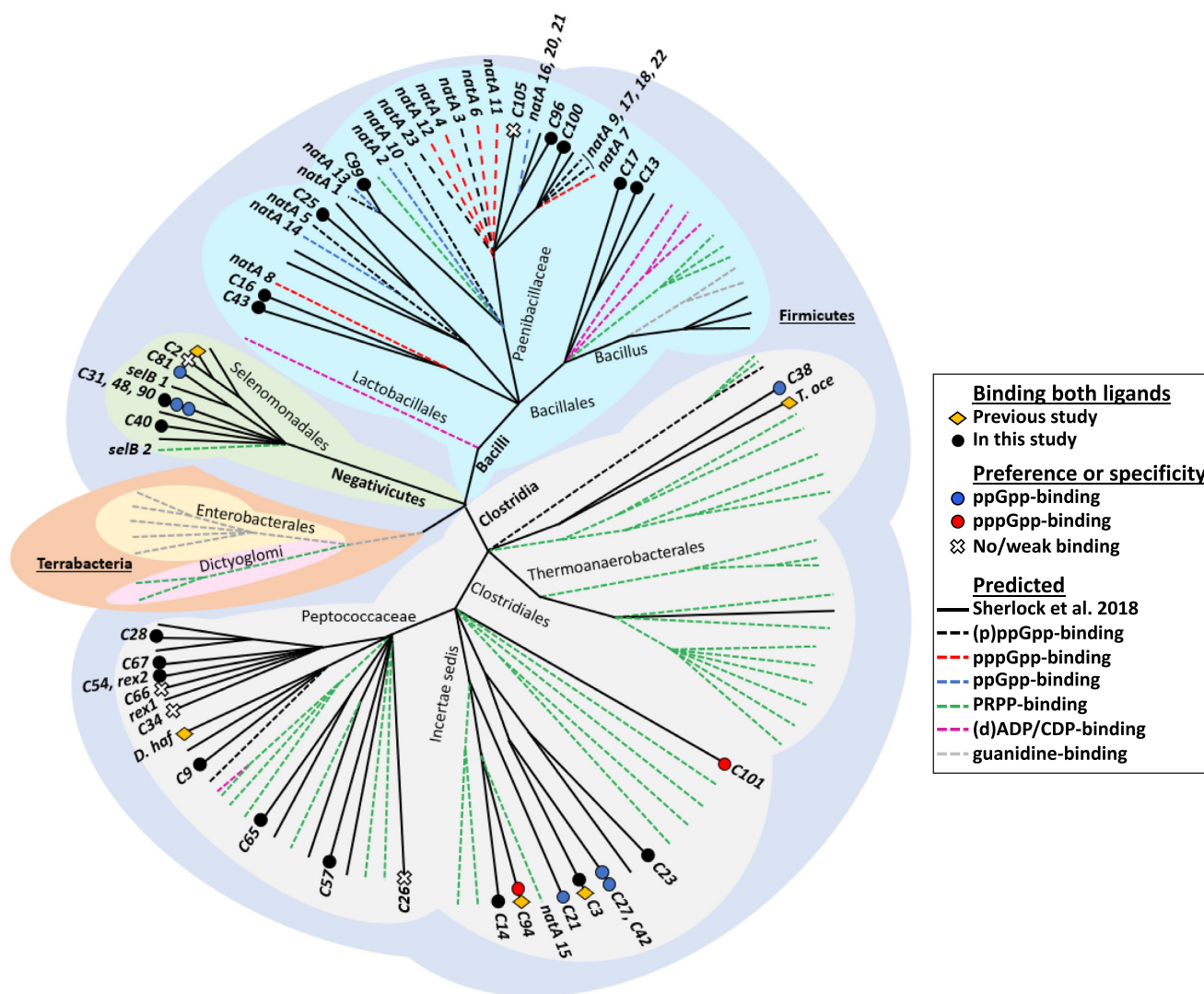
Figure 8. Diversity and conservation of *ykkC* aptamers homologs. Phylogenetic tree of the *ykkC* family of aptamers analyzed *in vitro* and/or *in silico* in this study using the PhyloT software (http://phylot.biobyte.de/). Candidates from the list of 105 candidates (25) that were tested *in vitro* appear as solid black lines with C numbers. New homologs found using a two-step *in silico* protein homology/aptamer search appear with a gene name tag and dotted lines in the color corresponding to their best fitness score (see legends), or by a solid black line if tested *in vitro*. Other homologs found using a one-step RNA-blast appear as dotted lines. A line can refer to multiple candidates in the case they are found in the same species. Binding preferences for ppGpp or for pppGpp or for both ligands are indicated by blue or red circles or by an X for weak or no binding.

the ppGpp-binders also bound to a second form of the ligand containing an extra 5′ phosphate, pppGpp, although in most cases the aptamers bound ppGpp better than pppGpp. There is a very high degree of ligand specificity among different *ykkC* aptamer subfamilies (e.g. for PRPP, guanidine, and (p)ppGpp), as noted previously (22,23,25). It is possible that some other aptamer subfamilies may also contain binding pockets capable of recognizing multiple very closely related compounds. There are also riboswitches containing two tandem aptamers, but these need not be closely related ligands.

Complexity of factors governing affinity and specificity

Not surprisingly, we found that elimination of even a small number of non-binding aptamers from bioinformatically-derived data sets used to generate secondary structure models improved the accuracy of the models. The consensus models determined from the set of experimentally validated *ykk2a* aptamers highlighted regions critical for ligand binding and are in good agreement with the features important for binding in the two crystal structures of ppGpp-bound aptamers.

There are also effects of stem length and motif position that influence ligand affinity but are not easily represented in the secondary structure models, suggesting additional complexity and/or context effects. For example, P1 stem length is highly variable among the confirmed (p)ppGpp aptamers. Analysis of the aptamer binding properties indicated that longer P1 stem lengths correlate with higher binding affinities for both ppGpp and pppGpp, whereas shorter P1 stem lengths correlate with specificity for one

ligand or the other. We suggest that longer stem length could increase variation in RNA conformation that could increase the probability of a conformation optimal for ligand binding. In contrast, shorter stem length could restrict conformational variation and thereby restrict binding by one of the two ligands. A second feature not apparent in the secondary structure models in Figure 5A and B is the variable position of the P1-ACA motif among the family of (p)ppGpp binding aptamers. Figure 5C illustrates the variable positions of the motif, one of which was preferred in pppGpp-binding aptamers.

We also note that (p)ppGpp ligands themselves can take on different conformations. In current protein-(p)ppGpp X-ray structures, about half contain (p)ppGpp in a ring-like form and half contain (p)ppGpp in an elongated form (31,50). It is possible that an RNA aptamer could have evolved to prefer one conformation of (p)ppGpp or the other. Additional mutational and biochemical analysis and crystal structures of ligand bound aptamers, such as those performed to compare PRPP- and ppGpp-binding (44,47), or PRPP- and guanidine binding (51), will be required to further identify sequence and structure determinants important for ligand binding and the specificity for similar but distinct ligands such as ppGpp and pppGpp.

Ligand specificity and gene homology

Bacteria can contain different relative amounts of ppGpp and pppGpp. pppGpp has been reported to be the predominant form in at least some firmicutes, and ppGpp appears to be the predominant form in many proteobacteria (32–34), suggesting that the relative affinities of an aptamer for the two ligand forms could reflect their abundance in the particular bacterial species. Alternatively, preference for one form could reflect a specific regulatory response to varying abundance of that form. We identified examples in which the function of the gene adjacent to the (p)ppGpp-sensing aptamer appeared to correlate with the aptamer's preference for ppGpp or pppGpp, and this connection was restricted to certain branches of the phylogenetic tree (Figure 8). The two tested *selB* aptamers were ppGpp-specific, whereas the only pppGpp-specific aptamers that we identified were associated with *natA* homologs. This suggests a connection between the aptamer's ligand specificity and the function of the regulated gene. Of note, the protein SelB has recently been identified as a potential ppGpp-binding protein by a capture-compound mass spectrometry approach (30), suggesting that ppGpp can affect both the activity and the synthesis of the SelB protein in various species.

Two *natA* aptamers from the same bacterial genus (*natA Cbot* and *natA* from *Clostridium sp. BL8*) preferentially bound pppGpp versus ppGpp, while *natA* aptamers from other genera did not display this preference. Thus, the aptamer's specificity might correlate with the phylum, genus, or species in which it is found, and the specificity of the aptamer might be predictive of the ligand that is dominant in a particular taxon. The basis for the apparent dominance of pppGpp for regulation of *natA* genes in *Clostridiales* taxons is unknown of course, but it could reflect the absence of the enzyme responsible for converting the pentaphosphate to the tetraphosphate or simply from the relative instability

of the tetraphosphate. If (p)ppGpp-binding RNA aptamers evolved before the protein targets that bind (p)ppGpp, the ligand binding specificity of the aptamer could predict the form of the ligand used in the current day stringent response. However, why the tested *selB* genes respond primarily to ppGpp, and why some *natA* genes respond primarily to pppGpp remains to be determined.

Improvements to the RNA DRaCALA approach

To adapt the DRaCALA technique for identification of RNA aptamers, it was necessary to immobilize RNA on the nitrocellulose filters. Instead of biotinylating the RNAs themselves, as was reported previously (38), we used biotinylated antisense oligonucleotides that hybridized to 3′-tags added to the aptamer RNAs (Figure 1). This antisense approach made the RNA-DRaCALA technique more suitable for screening large numbers of aptamers, because it was both more cost effective and more time and labor efficient. SELEX (systematic evolution of ligands by exponential enrichment) of streptavidin-binding RNA aptamers (52) has recently been optimized (53), and in theory this could be employed to immobilize a predicted aptamer on a filter without the need for biotinylation of the RNA. Alternatively, an MS2 coat protein binding sequence linked to the RNA could be used instead of biotin, replacing streptavidin with MS2 coat protein to anchor the RNA to a filter (54). Such an aptameric tag could greatly facilitate direct analysis of RNAs made in cells.

Although many small ligands are available in radiolabeled form (e.g. nucleotides, amino acids, sugars, urea, etc), radiolabeling is not the only option for identifying ligand binding by DRaCALA. For example, Cimdins-Ahne and colleagues describe a version of DRaCALA using mass spectrometry to determine ligand levels in the bound and unbound fractions of the diffusion assay (55). Alternative methods can also be imagined, such as fluorescent labelling (although with the risk of modifying the binding properties of the ligand), or the use of specific staining reagents to detect the ligand on the filter.

Relevance of dissociation constants obtained *in vitro* by RNA DRaCALA to binding of ligands *in vivo*

The K_ds for (p)ppGpp-aptamer interactions obtained *in vitro* by RNA-DRaCALA or by nuclease or in-line probing are sometimes much lower than predicted from the concentrations likely to be present in vivo. One possible explanation could be that regulation of transcription by these ligands is kinetically driven, as has been shown for the FMN riboswitch, where higher concentrations of the ligand than predicted by the K_d are needed to bind and drive the conformational change in the aptamer fast enough to occur in the narrow time window preceding passage of RNA polymerase through the transcription termination signal (56,57).

Another consideration possibly relevant to the high apparent affinity of (p)ppGpp for the *To ilvE* aptamer is that the solution conditions are quite different *in vivo* versus *in vitro*. For example, high salt and the presence of other potential ligands might reduce the affinity of (p)ppGpp for the aptamer in cells. Although we did not observe any competi-

tion between GTP and (p)ppGpp under our *in vitro* conditions (Figure 2), there are likely to be many potential ligands *in vivo*, including GTP, that bind weakly to the aptamer. At times these potential ligands are likely to be at much higher concentrations than (p)ppGpp. If the affinity of (p)ppGpp for the aptamer were not as high as it is, it is possible that the large number of weakly binding ligands could effectively lock the aptamer into its ligand bound conformation, preventing regulation of aptamer conformation throughout the full (p)ppGpp concentration range.

Alternatively, in some cases, the low K_d could be an artifact of measurement at a non-physiological temperature. For example, *T. oceani* grows optimally at 68°C, much higher than the temperature at which the affinity of ppGpp for the *To ilvE* aptamer was measured *in vitro* (37°C). Because the stability of aptamer RNA folding is a key for ligand binding, and RNA folding would be more stable at the lower temperature, ligand affinity might appear greater at the lower temperature. Similarly, a low K_d was also observed for the *pheA* aptamer from *Thermoactinomyces daqus* (Figure 4A), a bacterium typically grown at 55°C, consistent with the hypothesis that aptamers found in thermophilic species might have higher apparent affinities for ligands when measured at 37°C.

Potential uses for newly described ppGpp or pppGpp-specific aptamers

Testing predictions about whether ppGpp versus pppGpp is dominant in any particular evolutionary lineage requires experimental analysis. We suggest that aptamers primarily responsive to changing concentrations of only one of the ligands, i.e. ppGpp or pppGpp, could be used as biosensors *in vivo*. A riboswitch fused to a reporter could be designed to monitor the levels of the ligand without the need for radioactive labeling, harsh extraction techniques, or analysis by high performance liquid chromatography-mass spectrometry. Using key aptamer features identified in this study could help improve the design of more specific, and perhaps more sensitive, aptamers to use as biosensors for (p)ppGpp.

DATA AVAILABILITY

Synthetic oligonucleotide sequences are in the Supplementary Tables. All strains and plasmids are available on request.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

REFERENCES

1. Ellington,A.D. and Szostak,J.W. (1992) Selection in vitro of single-stranded DNA molecules that fold into specific ligand-binding structures. *Nature*, **355**, 850–852.
2. Winkler,W., Nahvi,A. and Breaker,R.R. (2002) Thiamine derivatives bind messenger RNAs directly to regulate bacterial gene expression. *Nature*, **419**, 952–956.
3. Winkler,W.C., Nahvi,A., Roth,A., Collins,J.A. and Breaker,R.R. (2004) Control of gene expression by a natural metabolite-responsive ribozyme. *Nature*, **428**, 281–286.
4. Nahvi,A., Sudarsan,N., Ebert,M.S., Zou,X., Brown,K.L. and Breaker,R.R. (2002) Genetic control by a metabolite binding mRNA. *Chem. Biol.*, **9**, 1043–1049.
5. Mironov,A.S., Gusarov,I., Rafikov,R., Lopez,L.E., Shatalin,K., Kreneva,R.A., Perumov,D.A. and Nudler,E. (2002) Sensing small molecules by nascent RNA: a mechanism to control transcription in bacteria. *Cell*, **111**, 747–756.
6. Barrick,J.E., Corbino,K.A., Winkler,W.C., Nahvi,A., Mandal,M., Collins,J., Lee,M., Roth,A., Sudarsan,N., Jona,I. *et al.* (2004) New RNA motifs suggest an expanded scope for riboswitches in bacterial genetic control. *Proc. Natl. Acad. Sci. U.S.A.*, **101**, 6421–6426.
7. McDaniel,B.A.M., Grundy,F.J., Artsimovitch,I. and Henkin,T.M. (2003) Transcription termination control of the S box system: direct measurement of S-adenosylmethionine by the leader RNA. *Proc. Natl. Acad. Sci. U.S.A.*, **100**, 3083–3088.
8. Grundy,F.J. and Henkin,T.M. (1993) tRNA as a positive regulator of transcription antitermination in *B. subtilis*. *Cell*, **74**, 475–482.
9. Winkler,W., Nahvi,A. and Breaker,R.R. (2002) Thiamine derivatives bind messenger RNAs directly to regulate bacterial gene expression. *Nature*, **419**, 952–956.
10. McCown,P.J., Corbino,K.A., Stav,S., Sherlock,M.E. and Breaker,R.R. (2017) Riboswitch diversity and distribution. *RNA*, **23**, 995–1011.
11. Bocobza,S., Adato,A., Mandel,T., Shapira,M., Nudler,E. and Aharoni,A. (2007) Riboswitch-dependent gene regulation and its evolution in the plant kingdom. *Genes Dev.*, **21**, 2874–2879.
12. Mukherjee,S., Retwitzer,M.D., Barash,D. and Sengupta,S. (2018) Phylogenomic and comparative analysis of the distribution and regulatory patterns of TPP riboswitches in fungi. *Sci. Rep.*, **8**, 5563.
13. Swati,D. (2017) Exploring riboswitches in archaeal metagenomes. *J RNAi Gene Silencing*, **13**, 536–543.
14. Serganov,A. and Nudler,E. (2013) A decade of riboswitches. *Cell*, **152**, 17–24.
15. Breaker,R.R. (2012) Riboswitches and the RNA world. *Cold Spring Harb. Perspect. Biol.*, **4**, a003566.
16. Barrick,J.E. and Breaker,R.R. (2007) The distributions, mechanisms, and structures of metabolite-binding riboswitches. *Genome Biol.*, **8**, R239.
17. Nawrocki,E.P. and Eddy,S.R. (2013) Infernal 1.1: 100-fold faster RNA homology searches. *Bioinformatics*, **29**, 2933–2935.
18. Sherwood,A.V. and Henkin,T.M. (2016) Riboswitch-Mediated gene regulation: novel RNA architectures dictate gene expression responses. *Annu. Rev. Microbiol.*, **70**, 361–374.
19. Tapsin,S., Sun,M., Shen,Y., Zhang,H., Lim,X.N., Susanto,T.T., Yang,S.L., Zeng,G.S., Lee,J., Lezhava,A. *et al.* (2018) Genome-wide identification of natural RNA aptamers in prokaryotes and eukaryotes. *Nat. Commun.*, **9**, 1289.
20. Ritchey,L.E., Tack,D.C., Yakhnin,H., Jolley,E.A., Assmann,S.M., Bevilacqua,P.C. and Babitzke,P. (2020) Structure-seq2 probing of RNA structure upon amino acid starvation reveals both known and novel RNA switches in *Bacillus subtilis*. *RNA*, **10**, 1441–1447.
21. Dar,D., Shamir,M., Mellin,J.R., Koutero,M., Stern-Ginossar,N., Cossart,P. and Sorek,R. (2016) Term-seq reveals abundant

ribo-regulation of antibiotics resistance in bacteria. *Science*, **352**, aad9822.

22. Nelson,J.W., Atilho,R.M., Sherlock,M.E., Stockbridge,R.B. and Breaker,R.R. (2017) Metabolism of free guanidine in bacteria is regulated by a widespread riboswitch class. *Mol. Cell*, **65**, 220–230.

23. Sherlock,M.E., Sudarsan,N., Stav,S. and Breaker,R.R. (2018) Tandem riboswitches form a natural boolean logic gate to control purine metabolism in bacteria. *Elife*, **7**, e33908.

24. Sherlock,M.E., Sadeeshkumar,H. and Breaker,R.R. (2019) Variant bacterial riboswitches associated with nucleotide hydrolase genes sense nucleoside diphosphates. *Biochemistry*, **58**, 401–410.

25. Sherlock,M.E., Sudarsan,N. and Breaker,R.R. (2018) Riboswitches for the alarmone ppGpp expand the collection of RNA-based signaling systems. *Proc. Natl. Acad. Sci. U.S.A.*, **115**, 6052–6057.

26. Cashel,M. (1975) Regulation of bacterial ppGpp and pppGpp. *Annu. Rev. Microbiol.*, **29**, 301–318.

27. Gourse,R.L., Chen,A.Y., Gopalkrishnan,S., Sanchez-Vazquez,P., Myers,A. and Ross,W. (2018) Transcriptional responses to ppGpp and DksA. *Annu. Rev. Microbiol.*, **72**, 163–184.

28. Atkinson,G.C., Tenson,T. and Hauryliuk,V. (2011) The RelA/SpoT homolog (RSH) superfamily: distribution and functional evolution of ppGpp synthetases and hydrolases across the tree of life. *PLoS One*, **6**, e23479.

29. Zhang,Y., Burkhardt,D.H., Rouskin,S., Li,G.-W., Weissman,J.S. and Gross,C.A. (2018) A stress response that monitors and regulates mRNA structure is Central to cold shock adaptation. *Mol. Cell*, **70**, 274–286.

30. Wang,B., Dai,P., Ding,D., Del Rosario,A., Grant,R.A., Pentelute,B.L. and Laub,M.T. (2019) Affinity-based capture and identification of protein effectors of the growth regulator ppGpp. *Nat. Chem. Biol.*, **15**, 141–150.

31. Irving,S.E., Choudhury,N.R. and Corrigan,R.M. (2021) The stringent response and physiological roles of (pp)pGpp in bacteria. *Nat. Rev. Microbiol.*, **19**, 256–271.

32. Anderson,B.W., Fung,D.K. and Wang,J.D. (2021) Regulatory themes and variations by the stress-Signaling nucleotide alarmones (p) ppGpp in bacteria. *Annu. Rev. Genet.*, **55**, 115–133.

33. Ross,W., Vrentas,C.E., Sanchez-Vazquez,P., Gaal,T. and Gourse,R.L. (2013) The magic spot: a ppGpp binding site on *E. coli* RNA polymerase responsible for regulation of transcription initiation. *Mol. Cell*, **50**, 420–429.

34. Ross,W., Sanchez-Vazquez,P., Chen,A.Y., Lee,J.-H., Burgos,H.L. and Gourse,R.L. (2016) ppGpp binding to a site at the RNAP-DksA interface accounts for its dramatic effects on transcription initiation during the stringent response. *Mol. Cell*, **62**, 811–823.

35. Sanchez-Vazquez,P., Dewey,C.N., Kitten,N., Ross,W. and Gourse,R.L. (2019) Genome-wide effects on *Escherichia coli* transcription from ppGpp binding to its two sites on RNA polymerase. *Proc. Natl. Acad. Sci. U.S.A.*, **116**, 8310–8319.

36. Regulski,E.E. and Breaker,R.R. (2008) In-Line probing analysis of riboswitches. In: Wilusz,J. (ed). *Post-Transcriptional Gene Regulation, Methods In Molecular Biology^{TM}*. Humana Press, Totowa, NJ, Vol. **419**, pp. 53–67.

37. Roelofs,K.G., Wang,J., Sintim,H.O. and Lee,V.T. (2011) Differential radial capillary action of ligand assay for high-throughput detection of protein-metabolite interactions. *Proc. Natl. Acad. Sci. U.S.A.*, **108**, 15528–15533.

38. Donaldson,G.P., Roelofs,K.G., Luo,Y., Sintim,H.O. and Lee,V.T. (2012) A rapid assay for affinity and kinetics of molecular interactions with nucleic acids. *Nucleic Acids Res.*, **40**, e48.

39. Levorato,C. and Cima,L. (1968) Thin-layer chromatography and determination of thiamine salts, phosphoric esters, disulphides, and their respective thiochromes. *J. Chromatogr. A*, **32**, 771–773.

40. Zhao,R., Gao,F. and Goldman,I.D. (2002) Reduced folate carrier transports thiamine monophosphate: an alternative route for thiamine delivery into mammalian cells. *Am. J. Physiol.-Cell Physiol.*, **282**, C1512–C1517.

41. Will,S., Joshi,T., Hofacker,I.L., Stadler,P.F. and Backofen,R. (2012) LocARNA-P: accurate boundary prediction and improved detection of structural RNAs. *RNA*, **18**, 900–914.

42. Weinberg,Z. and Breaker,R.R. (2011) R2R - software to speed the depiction of aesthetic consensus RNA secondary structures. *BMC Bioinformatics*, **12**, 3.

43. Johnson,M., Zaretskaya,I., Raytselis,Y., Merezhuk,Y., McGinnis,S. and Madden,T.L. (2008) NCBI BLAST: a better web interface. *Nucleic Acids Res.*, **36**, W5–W9.

44. Peselis,A. and Serganov,A. (2019) *ykkC* riboswitches employ an add-on helix to adjust specificity for polyanionic ligands. *Nat. Chem. Biol.*, **14**, 887–894.

45. Breaker,R.R. (2020) Imaginary ribozymes. *ACS Chem. Biol.*, **15**, 2020–2030.

46. Zuker,M. (2003) Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res.*, **31**, 3406–3415.

47. Knappenberger,A.J., Reiss,C.W. and Strobel,S.A. (2018) Structures of two aptamers with differing ligand specificity reveal ruggedness in the functional landscape of RNA. *Elife*, **7**, e36381.

48. Cheng,J., Guffanti,A.A. and Krulwich,T.A. (1997) A two-gene ABC-type transport system that extrudes Na+ in *Bacillus subtilis* is induced by ethanol or protonophore. *Mol. Microbiol.*, **23**, 1107–1120.

49. Sherlock,M.E. and Breaker,R.R. (2020) Former orphan riboswitches reveal unexplored areas of bacterial metabolism, signaling, and gene control processes. *RNA*, **26**, 675–693.

50. Steinchen,W. and Bange,G. (2016) The magic dance of the alarmones (p) ppGpp: the structural biology of the alarmones (p)ppGpp. *Mol. Microbiol.*, **101**, 531–544.

51. Knappenberger,A.J., Reiss,C.W., Focht,C.M. and Strobel,S.A. (2020) A modular RNA domain that confers differential ligand specificity. *Biochemistry*, **59**, 1361–1366.

52. Srisawat,C. and Engelke,D.R. (2001) Streptavidin aptamers: affinity tags for the study of RNAs and ribonucleoproteins. *RNA*, **7**, 632–641.

53. Leppek,K. and Stoecklin,G. (2014) An optimized streptavidin-binding RNA aptamer for purification of ribonucleoprotein complexes identifies novel ARE-binding proteins. *Nucleic Acids Res.*, **42**, e13.

54. Slobodin,B. and Gerst,J.E. (2011) RaPID: an aptamer-Based mRNA affinity purification technique for the identification of RNA and protein factors present in ribonucleoprotein complexes. In: Gerst,J.E. (ed). *RNA Detection and Visualization, Methods in Molecular Biology*. Humana Press, Totowa, NJ, Vol. **714**, pp. 387–406.

55. Cimdins-Ahne,A., Chernobrovkin,A., Kim,S., Lee,V.T., Zubarev,R.A. and Römling,U. (2022) A mass spectrometry-based non-radioactive capillary action of ligand assay (DRaCALA) to assess ligand binding to proteins. *J. Mass Spectrom.*, **57**, e4822.

56. Wickiser,J.K., Winkler,W.C., Breaker,R.R. and Crothers,D.M. (2005) The speed of RNA transcription and metabolite binding kinetics operate an FMN riboswitch. *Mol. Cell*, **18**, 49–60.

57. Grundy,F.J. and Henkin,T.M. (2004) Kinetic analysis of tRNA-Directed transcription antitermination of the *Bacillus subtilis glyQS* gene *in vitro*. *J. Bacteriol.*, **186**, 5392–5399.

58. Schneider,T.D. and Stephens,R.M. (1990) Sequence logos: a new way to display consensus sequences. *Nucleic Acids Res.*, **18**, 6097–6100.

59. Crooks,G.E., Hon,G., Chandonia,J.-M. and Brenner,S.E. (2004) WebLogo: a sequence logo generator. *Genome Res.*, **14**, 1188–1190.