Supporting Information

Bioinformatic analysis of riboswitch structures uncovers variant classes with altered ligand specificity

Zasha Weinberg[a,1,2], James W. Nelson[b,1,3], Christina E. Lünse[b,4], Madeline E. Sherlock[c], Ronald R. Breaker[a,b,c,5]

[a]Howard Hughes Medical Institute, [b]Department of Molecular, Cellular and Developmental Biology, [c]Department of Molecular Biophysics and Biochemistry, Yale University, P.O. Box 208103, New Haven, CT 06520-8103, USA.

[1]Co-first authors.

[2]Present address:Zasha Weinberg, Bioinformatics Group, Department of Computer Science and Interdisciplinary Center for Bioinformatics, Universität Leipzig, Härtelstraße 16-18, D-04107 Leipzig, Germany.

[3]Present address: James Nelson, Department of Chemistry and Chemical Biology, Harvard University, Cambridge, MA USA 02139; Howard Hughes Medical Institute.

[4]Present address: Christina E. Lünse, Institute of Biochemistry, Universität Leipzig, Brüderstraße 34, D-04103 Leipzig, Germany.

[5]To whom correspondence should be addressed. Email: ronald.breaker@yale.edu.

www.pnas.org/cgi/doi/10.1073/pnas.1619581114

SI Text.

Design of RNA Constructs to Examine Glycine Riboswitch D,G,A Variants by In-line Probing. We designed RNA constructs containing D,G,A group sequences for in-line probing as follows. First, we designed independent aptamer I and II constructs similar to RNAs tested previously (1, 2), where aptamer I and II refer to the 5´- and 3´-most aptamer domains, respectively. Even though glycine binding for aptamer I of the glycine riboswitch was never observed (2, 3), we decided to test this construct for the variant, because it is possible that aptamer II-independent ligand binding would be possible in this case. Aptamer II constructs were tested as concise (just containing nucleotides in the aptamer II region) or extended RNAs (containing part of aptamer I). The latter D,G,A variant construct is comparable to the VC-II construct examined previously (1). It has been previously shown for glycine riboswitches that part of aptamer I is necessary for aptamer II to bind (1-3). The complete set of RNA molecules tested for the D,G,A group is given in **Fig. S5*C*** and the primers used to construct those molecules are found in **Table S4**.

Variants of Ni^{2+}/Co^{2+}-binding Riboswitches Retain Ligand Specificity of the Parent Aptamer. Our analysis also uncovered novel variant RNAs related to the Ni^{2+}/Co^{2+}-binding riboswitch class (4). Parent 'Ni/Co' riboswitches are most often located upstream of genes encoding proteins with COG0053 domains, which are predicted to function as transporters of Co^{2+}, Zn^{2+} or Cd^{2+}. Since Ni/Co riboswitches bind up to four divalent metal ion ligands in a cooperative manner, these multiple binding sites involve greater number of nucleotides that make direct contact to ligands. Therefore, the key nucleotides for parent Ni/Co riboswitches consist of A,G,G,G,G,G at positions 14, 45, 46, 47, 87 and 88, respectively, following the numbering system described previously (4).

We investigated Ni/Co riboswitch variants with the key nucleotides A,G,G,G,A,G and A,C,G,G,A,G (**Fig. S7**), which share a G87A modification, and are both commonly associated with genes coding for the TIGR01525 protein domain. This protein domain is not associated with canonical Ni/Co riboswitches, but is annotated as a heavy-metal transporter. Although the heavy-metal transporting domain shares BLAST homology with other domains commonly associated with Ni/Co riboswitches, the distinct domain classification might suggest an altered substrate specificity of the proteins, and therefore a corresponding change to the riboswitches.

However, in-line probing data indicate that a representative of the A,G,G,G,A,G group of variants retains the ability to bind both Ni^{2+} and Co^{2+} with binding curves that are characteristic of cooperativity (**Fig. S7**). These results indicate that this single mutation of a key nucleotide in this particular example

does not alter the multiple binding pockets of Ni/Co riboswitches in a manner that disrupts its original metal ion specificity. Moreover, these results are consistent with the view that the strongest candidates for ligand changes can be found by identifying variants that are functionally and structurally distinct from genes regulated by the parent class. In the current case, the TIGR01525 domains associated with the A,G,G,G,A,G group of variants presumably have very similar biochemical function as genes associated with known Ni/Co riboswitches, and might be only modestly diverged in sequence and structure.

Another factor that might explain the retention of ligand specificity is that Ni/Co riboswitches bind 3 to 4 ions cooperatively (4). Therefore, they would need to change several binding sites simultaneously to completely change their ligand specificity. Perhaps they have altered the number of cooperative metal binding sites present, or have altered the specificity of only some sites.

Table S1. This table describes search parameters for each riboswitch class. "Riboswitch": the name of the riboswitch. "Rfam accession": the accession of the model used in the Rfam database. Some models were made by us to provide better coverage in sequence-space around the representative sequence that was crystallized. In comparison to other riboswitch structures, RNAs representing Mn^{2+} and $preQ_1$-I classes have less conserved sequence or structure that can be exploited by homology search algorithms to discriminate between true and false homologs. Current search technologies fail to find certain examples of these riboswitch classes unless multiple search models are used. Also, the Ni/Co and $preQ_1$-III riboswitch classes were not yet present in Rfam. Therefore, we made additional models, by applying iterative searches with Infernal. The accessions of these models do not begin with the letters 'RF' and are available in **Dataset S2** at http://breaker.yale.edu/variants. "PDB accession (chain)": the accession of the crystallized riboswitch. In parentheses, the chain we used is given. "Ligand name": is the name of the ligand molecule used to find key nucleotides. It is usually a HETATM in PDB, but in the case of the group I intron, we used the G nucleotide within the RNA chain. "Notes": exceptions to the general procedure. "Score" refers to a special score threshold used in the cmsearch program that was used when searching PDB in order to be able to find the relevant entry. "No GSC" means that the GSC algorithm was not used to weight sequences, because this process took too long. Each of the following distances was used independently to analyze which nucleotides are close to the ligand (distances in Å): 2, 2.25, 2.5, 2.75, 3, 3.25, 3.5, 3.75, 4, 4.5 and 5.

Riboswitch	Rfam accession	PDB accession (chain)	Ligand name in PDB (chain, if different from riboswitch chain)	Notes
FMN	RF00050	3fq2 (X)	FMN	
TPP	RF00059	2hoj (A)	TPP	
	RF00059	3d2g (A)	TPP	
	RF00059	2gdi (X)	TPP	
SAM-I	RF00162	4kqy (A)	SAM	
	RF00162	2gis (A)	SAM	
	RF00162	3gx5 (A)	SAM	
Lysine	RF00168	3dil (A)	LYS	
glmS	RF00234	2nz4 (P)	GLP	
	RF00234	2z75 (B)	GLP (A)	
Glycine	RF00504	3oww (A)	GLY	
	RF00504	3p49 (A)	GLY	
SAM-II	RF00521	2qwy (A)	SAM	Score: 27
$preQ_1$-I	RF00522	3k1v (A)	PRF	Score: 26
	preQ1-I-3q50	3q50 (A)	PRF	
$preQ_1$-II	RF01054	2miy (A)	PRF	Score: 29
	RF01054	4jf2 (A)	PRF	Score: 29
SAH	RF01057	3npq (A)	SAH	Score: 34
2´-deoxyG	RF01510	3ski (A)	GNG	
SAM-I/IV variant	RF01725	4oqu (A)	SAM	
SAM-III	RF01767	3e5c (A)	SAM	Score: 42
THF	RF01831	3sux (X)	THF	
	RF01831	4lvv (A)	FFO	
Fluoride	RF01734	4enb (A)	F	

C-di-AMP	RF00379	4qlm (A)	2BA	
	RF00379	4w90 (C)	2BA	
Mg^{2+}-I (*ykoK*)	RF00380	2qbz (X)	MG	
Ni/Co	NiCo	4rum (A)	CO	
AdoCbl (B_{12})	RF00174	4gma (Z)	B1Z	
	RF00174	4gxy (A)	B1Z	
	RF01482	4gxy (A)	B1Z	
AqCbl (B_{12})	RF01689	4frn (A)	I2A	
C-di-GMP-II	RF01786	3q3z (V)	C2E	
Guanine	RF00167	4fe5 (B)	HPA	
Adenine	RF00167	1y26 (X)	ADE	
C-di-GMP-I / C-AMP-GMP	RF01051	3irw (R)	C2E	No GSC
	RF01051	4yaz (A)	4BW	No GSC
$preQ_1$-III	preQ1-III	4rzd (A)	PRF	
ZTP	RF01750	4znp (A)	AMZ	
	RF01750	5btp (A)	AMZ	
	RF01750	4xwf (A)	AMZ	
Mn^{2+}	RF00080	4y1i (A)	MN	
	yybP-4Y1I	4y1i (A)	MN	
Glutamine	RF01739	5ddp (A)	GLN	

Table S2. Genes associated with various groups of c-di-GMP-I riboswitches that might have altered ligand specificity. At a distance of 2.75 Å, three key nucleotides are found in c-di-GMP-I riboswitches: G20, A47 and C92. Genes regulated by riboswitches in each candidate group are given with their frequencies (out of 1). For comparison, frequencies are also given for riboswitches in the G,A,C group, which correspond to c-di-GMP-I riboswitches whose key nucleotides are identical to those in the crystallized RNA, as well as the A,A,C group, which corresponds to c-AMP-GMP riboswitches. Blank entries mean the frequency is zero.

Group (key nucleotides)	Frequency in candidate group	Frequency in G,A,C group	Frequency in A,A,C group	Gene encodes
G,G,C	0.51	0.024	0.015	SCP-like extracellular domain
G,G,C	0.34	0.0025	0.001	PulF (Type-II secretory pathway)
G,G,C	0.17	0.002		Uncharacterized protein, YkwD (might be SCP-like)
U,A,C	0.44	0.0025	0.0031	Na^+/H^+ antiporter
U,A,C	0.11	0.00047	0.00069	Collagen-binding protein B
U,A,C	0.11			Urease, gamma subunit
U,A,C	0.055	0.0035	0.0016	ATPase
U,A,C	0.055			Stage V sporulation protein K
U,A,C	0.037	0.0006	0.0005	c-di-GMP synthase
U,A,C	0.037	0.0012	0.001	Methyl-accepting chemotaxis domain
U,A,C	0.037	0.0077	0.0035	Histidine kinase
U,A,C	0.037	0.0048	0.00086	Cache domain
U,A,C	0.037	0.0098		EAL domain (c-di-GMP phosphodiesterase)
U,A,C	0.037	0.00016		Putative c-di-GMP phosphodiesterase
U,A,U	0.5			DUF3179, unknown function
U,A,U	0.5	0.0028		Spherulation-specific family 4 (involved in sporulation)

Table S3. Genes of groups of c-di-GMP-II riboswitches that are candidates for changes in ligand specificity. Genes regulated by each candidate group of riboswitches are given with their frequencies (out of 1). For comparison, riboswitches in the better-characterized G,A,A group is given, as this group contains the crystallized c-di-GMP-II riboswitch. Because class II riboswitches are rarer than class I riboswitches, frequencies for class I riboswitches probably reflect gene associations better. Therefore, the cyclic-di-GMP-I class G,A,C group (presumably binding c-di-GMP) and A,A,C group (presumably binding c-AMP-GMP) are also represented in the table. Blank entries mean the frequency is zero. Many genes are associated with the G,A,G-group, so only the most common are listed. The complete list of genes is in **Dataset S1** at http://breaker.yale.edu/variants.

Group (key nucleotide)	Frequency in G,A,G	Frequency in G,A,A group	Frequency in G,A,C group of c-di-GMP-I riboswitches	Frequency in A,A,C group of c-di-GMP-I riboswitches	Gene encodes
G,A,G	0.11	0.011	0.01	0.012	S8/S53 peptidase (S8 is regulated by c-di-AMP riboswitches)
G,A,G	0.099		0.00035		COG4179 (unknown function)
G,A,G	0.06		0.007		Discoidin domain
G,A,G	0.056		0.00047		Tol transport system, periplasmic component
G,A,G	0.049		0.00035		N-acetylglucosaminephosphatidylinositol de-N-acetylase (GPI synthesis)
G,A,G	0.048	0.044	0.00096		Cellulose synthase
G,A,G	0.043	0.039	0.0025	0.016	COG4932, outer-membrane protein
G,A,G	0.041	0.0015	0.0022	0.036	Bacterial Ig-like
G,A,G	0.04	0.0096	0.00071	0.0062	Cna protein, B-type
G,A,G	0.034		0.001		Pentapeptide repeats
	Rarer genes not listed for G,A,G group.				
A,A,U	0.25				Glucose/galactose-binding protein (involved in chemotaxis)
A,A,U	0.23				Flagellar motor switch
A,A,U	0.24	0.079	0.02	0.027	Methyl-accepting chemotaxis
A,A,U	0.12	0.0038	0.0025		Alpha-tubulin suppressor
A,A,U	0.074	0.025	0.019	0.0092	HAMP domain (signaling)
A,A,U	0.056	0.00099	0.0048	0.0021	Cache domain
A,A,U	0.051				Large tegument protein UL36
G,C,G	0.25		0.0039		GYD domain (unknown function)
G,C,G	0.19				DNA methylase
G,C,G	0.16		0.0011		Acetyletransferase/hydrolase
G,C,G	0.12		0.00047		DUF3607 (unknown function)
G,C,G	0.074		0.0012	0.001	S53 peptidase
G,C,G	0.061	0.0065			SPRBCC ligand-binding domain

A,A,A	0.36		0.00018		YhcR endonuclease
A,A,A	0.36				NOP14-like (ribosome maturation)
A,A,A	0.31	0.0077	0.0083	0.01	PAS domain
A,C,U	0.31		0.003	0.033	Ig-like
A,C,U	0.22	0.038	0.021		Signal receiver domain
A,C,U	0.2	0.0048	0.0047	0.0039	Histidine-kinase-like ATPases
A,C,U	0.2	0.0036	0.0004		Signal transduction histidine kinase
A,C,U	0.074	0.0022	0.0038		Response regulator receiver domain

Table S4. List of oligonucleotides used in this study.

Name of Oligonucleotide	**Sequence**	**Purpose**
UCC5	TAATACGACTCACTATAGG GTAACCTTGTAATTCGGG GGATTGGCCCCGGAGTAT CCACGACCGG	Forward primer for the construction of a template for transcription of the 71 *env-23* RNA.
UCC6	TTTCCCTGGATGCACCGG ATTATGGTCCGGTCGTGG ATACTCCGG	Reverse primer for the construction of a template for transcription of the 71 *env-23* RNA.
UCC7	TAATACGACTCACTATAGG GGGACCTTGTAGTCCCCA GGATTGGCTGGGGAGCAT CCACCTCCGG	Forward primer for the construction of a template for transcription of the 71 *env-16* RNA from *G. tropica DSM 19535*.
UCC8	CTCCCCTGGCGGCTCCGG ATTAAGGTCCGGAGGTGG ATGCTCCCCAGC	Reverse primer for the construction of a template for transcription of the 71 *env-16* RNA from *G. tropica DSM 19535*.
UGC1	TAATACGACTCACTATAGG GTTCTTGTATAAGCTCAAT AATATGGTTTGAGCGTCTC GACCAGGC	Forward primer for the construction of a template for transcription of the 72 *coxL* RNA from *D. meridiei*.
UGC2	TTCATGTAGCCTGGGCAA TTTATGGTTGCCTGGTCGA GACGCTC	Reverse primer for the construction of a template for transcription of the 72 *coxL* RNA from *D. meridiei*.
CEL0371	TCGTAATACGACTCACTAT AGGTCGGATGAGGGACG TCCTAGACACCGCAAGGG GCCGAAGAAGCAAAACGT AAGGTAACATCTTCCTTTC GTGGAAACTGACAGGCAA AAAGGAAGGCGACCTGAC GAATCTTTGGAGAGTGTG CAAAGAAAGCACCACCTA CGAGGCTAGAACCAACTG GTAAAAGGACAGAGAA	gBlock containing the 175 *sdh* RNA from *C. comes* ATCC 27758 shown in Fig. S4. Designed for use as a PCR template using the primers below.

CEL0372	TCGTAATACGACTCACTATAGGTCGGATGAGGG	Forward primer for the construction of a template for transcription of the 175 *sdh* RNA from *C. comes* ATCC 27758 shown in Fig. S4.
CEL0373	TTCTCTGTCCTTTTACCAGTTGGTTCTAGCCTCG	Reverse primer for the construction of a template for transcription of the full length 175 *sdh* RNA from *C. comes* ATCC 27758 shown in Fig. S4.
CEL0405	TCGTCAGGTCGCCTTCCTTTTTGCC	Reverse primer for the construction of a template for transcription of the aptamer I 107 *sdh* RNA from *C. comes* ATCC 27758.
CEL0422	TCGTAATACGACTCACTATAGGAAGGCGACCTGACGAATCTTTGGAGAGTGTGCAAAGAAAGC	Forward primer for the construction of a template for transcription of the 85 *sdh* RNA from *C. comes* ATCC 27758.
CEL0423	TTCTCTGTCCTTTTACCAGTTGGTTCTAGCCTCGTAGGTGGTGCTTTCTTTGCACACTC	Reverse primer for the construction of a template for transcription of the 85 *sdh* RNA from *C. comes* ATCC 27758.
CEL0414	TCGTAATACGACTCACTATAGGAAAATGTAAACCGTATGATAACAGTCTGAAACATTCTTTATATAAAGAAGACCGAAGCGGAAAGCTAAAGTACTAACATCACTTTAGTGAAACTGACAGGTAAAGCAAATGACTGTTTGACGGAACCTTGGAAAGACTATCTTAGAGATAGAGCCAATGAGGCAATAACCAACTGGCAAAAAGACAGGGAAAGTTTATTAT	gBlock containing the 201 *potE* RNA from *Clostridium* sp. ASF356. Designed for use as a PCR template using the primers below.

CEL0415	TCGTAATACGACTCACTATAGGAAAATGTAAACCGTATGATAACAG	Forward primer for the construction of a template for transcription of the full length 201 *potE* RNA from *Clostridium sp.* ASF356.
CEL0416	ATAATAAACTTTCCCTGTCTTTTTGCCAGTTGGTTATTGCC	Reverse primer for the construction of a template for transcription of the full length201 *potE* RNA from *Clostridium sp.* ASF356.
CEL0417	TCGTAATACGACTCACTATAGGCCGTATGATAACAGTCTGAAACATTCTTTATATAAAGAAGACCG	Forward primer for the construction of a template for transcription of the concise full 181 *potE* RNA from *Clostridium sp.* ASF356.
CEL0409	TTCCCTGTCTTTTTGCCAGTTGGTTATTGCCTCATTGGCTCTATC	Reverse primer for the construction of a template for transcription of the concise full length181 *potE* RNA from *Clostridium sp.* ASF356 and 67 *potE* RNA from *Clostridium sp.* ASF356.
CEL0406	TCGTAATACGACTCACTATAGGCCGTATGATAACAGTCTGAAACATTCTTTATATAAAGAAGACCGAAGCGGAAAGC	Forward oligo for the construction of a template through primer extension for transcription of 115 *potE* RNA from *Clostridium sp.* ASF356.
CEL0407	TTCCGTCAAACAGTCATTTGCTTTACCTGTCAGTTTCACTAAAGTGATGTTAGTACTTTAGCTTTCCGCTTCGGTCTTC	Reverse oligo for the construction of a template through primer extension for transcription of 115 *potE* RNA from *Clostridium sp.* ASF356.
CEL0408	TCGTAATACGACTCACTATAGGACCTTGGAAAGACTATCTTAGAGATAGAGCCAATGAGGC	Forward oligo for the construction of a template through primer extension for transcription of 67 *potE* RNA from *Clostridium sp.* ASF356.

CEL0424	TCGTAATACGACTCACTATAGGCAAATGACTGTTTGACGGAACCTTGGAAAGACTATCTTAGAGATAGAGCC	Forward primer for the construction of a template through primer extension for transcription of 98 *potE* RNA from *Clostridium sp.* ASF356.
CEL0425	ATAATAAACTTTCCCTGTCTTTTTGCCAGTTGGTTATTGCCTCATTGGCTCTATCTCTAAGATAGTC	Reverse primer for the construction of a template through primer extension for transcription of 98 *potE* RNA from *Clostridium sp.* ASF356.
CEL0418	TCGTAATACGACTCACTATAGGTCAAATATGTTCGGATGAGGACTGCATTAGAGACTTTTTAAATGAAGCGCCGAAGAAGCAAAACAGAAATTTCATCTTGTTTCTGTGGAAACTGACAGGCAAAAGGAGGGCAGTCTGACGATACTTTGGAGAGAGTGATTTAATTCACCGCCTACGAGGATATAACCAACTGGCAAAAGGACAGAGATATAGATGCAC	gBlock containing the 198 *sdh* RNA from *Clostridium hiranonis* DSM 13275. Designed for use as a PCR template using the primers below.
CEL0419	TCGTAATACGACTCACTATAGGTCAAATATGTTCGGATGAGG	Forward primer for the construction of a template for transcription of 198 *sdh*RNA from *C. hiranonis* DSM 13275.
CEL0420	GTGCATCTATATCTCTGTCCTTTTGCCAGTTGG	Reverse primer for the construction of a template for transcription of 198 *sdh*RNA from *C. hiranonis* DSM 13275.
CEL0421	TCGTAATACGACTCACTATAGGTCGGATGAGGACTGC	Forward primer for the construction of a template for transcription of 178 *sdh*RNA from *C. hiranonis* DSM

	ATTAGAGAC	13275.
CEL0413	ATCTCTGTCCTTTTGCCAG TTGGTTATATCCTCGTAGG CGGTG	Reverse primer for the construction of a template for transcription of 178 *sdh*RNA and66 *sdh*RNA from *C. hiranonis* DSM 13275.
CEL0410	TCGTAATACGACTCACTAT AGGTCGGATGAGGACTGC ATTAGAGACTTTTTAAATG AAGCGCCGAAGAAGCAAA ACAG	Forward oligo for the construction of a template through primer extension for transcription of 113 *sdh* RNA from *C. hiranonis* DSM 13275.
CEL0411	TATCGTCAGACTGCCCTCC TTTTGCCTGTCAGTTTCCA CAGAAACAAGATGAAATT TCTGTTTTGCTTCTTCGGC GC	Reverse oligo for the construction of a template through primer extension for transcription of 113 *sdh* RNA from *C. hiranonis* DSM 13275.
CEL0412	TCGTAATACGACTCACTAT AGGACTTTGGAGAGAGTG ATTTAATTCACCGCCTACG	Forward oligo for the construction of a template through primer extension for transcription of 66 *sdh* RNA from *C. hiranonis* DSM 13275.
CEL0426	TCGTAATACGACTCACTAT AGGAGGGCAGTCTGACGA TACTTTGGAGAGAGTGAT TTAATTCACCG	Forward oligo for the construction of a template through primer extension for transcription of 93 *sdh* RNA from *C. hiranonis* DSM 13275.
CEL0427	GTGCATCTATATCTCTGTC CTTTTGCCAGTTGGTTATA TCCTCGTAGGCGGTGAAT TAAATCACTCTCTC	Reverse oligo for the construction of a template through primer extension for transcription of 93 *sdh* RNA from *C. hiranonis* DSM 13275.

NiCo1	TCGTAATACGACTCACTATAGGTCAAATAAGATAAATCTGAACAAGCGGTAAATTTATTTACGCAGCTGGGTTAC	Forward primer for the construction of a template for transcription of the 102 ATPase RNA from *F. hominis*.
NiCo2	TTAATAACTGTCTCACAGTGCGCTCTGGCTCTGCTGTTACATTTTTCATGTAACCCAGCTGCGTAAATA	Reverse primer for the construction of a template for transcription of the 102 ATPase RNA from *F. hominis*.

Fig. S1. PAGE analysis of in-line probing reactions of 5′ ^{32}P-labeled 71 *env-23* in the presence of increasing concentrations of 2′-dG. Concentrations range from 320 pM to 100 µM in half-log increments, i.e., the ratio between the concentration of each pair of consecutive concentrations is $10^{0.5}$. Annotations are as described for **Fig. 3C**.

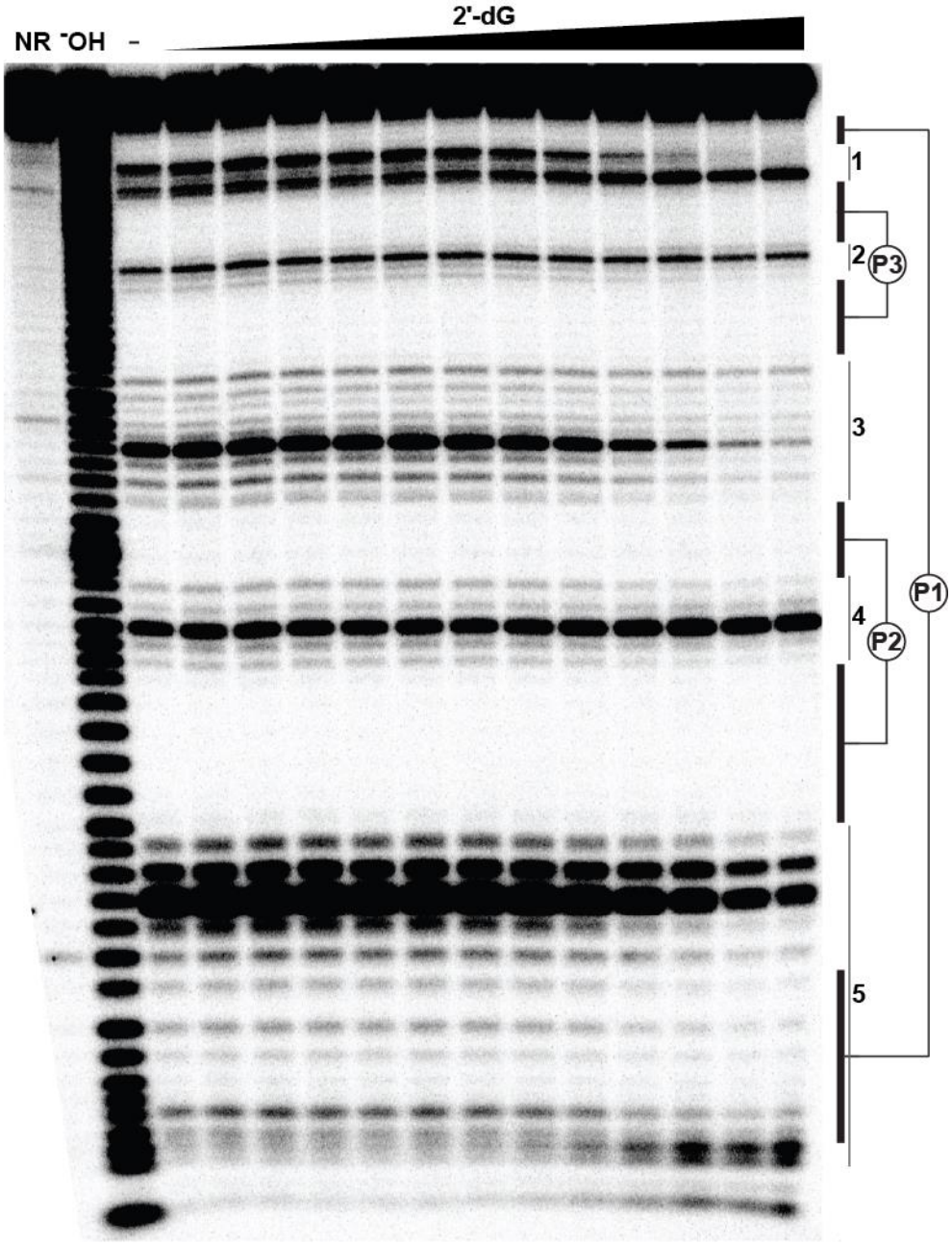


Fig. S2. Recognition of 2′-deoxyguanosine by a variant guanine riboswitch aptamer. (*A*) Sequence and secondary structure of the 71 *env-16* RNA. Labels are as in Fig. 3*B*. (*B*) PAGE analysis of in-line probing reactions of 5′ ^{32}P-labeled 71 *env-16* in the presence of increasing 2′-deoxyguanosine (2′-dG; 320 pM to 100 µM in half-log units). Annotations are as described for **Fig. 3C**. (*C*) Plot of the fraction of RNA bound to ligand versus the logarithm of the molar concentration (*c*) of 2′-deoxyguanosine. Data are derived from that presented in *B*. The theoretical binding curve for a one-to-one interaction between ligand and RNA is included.

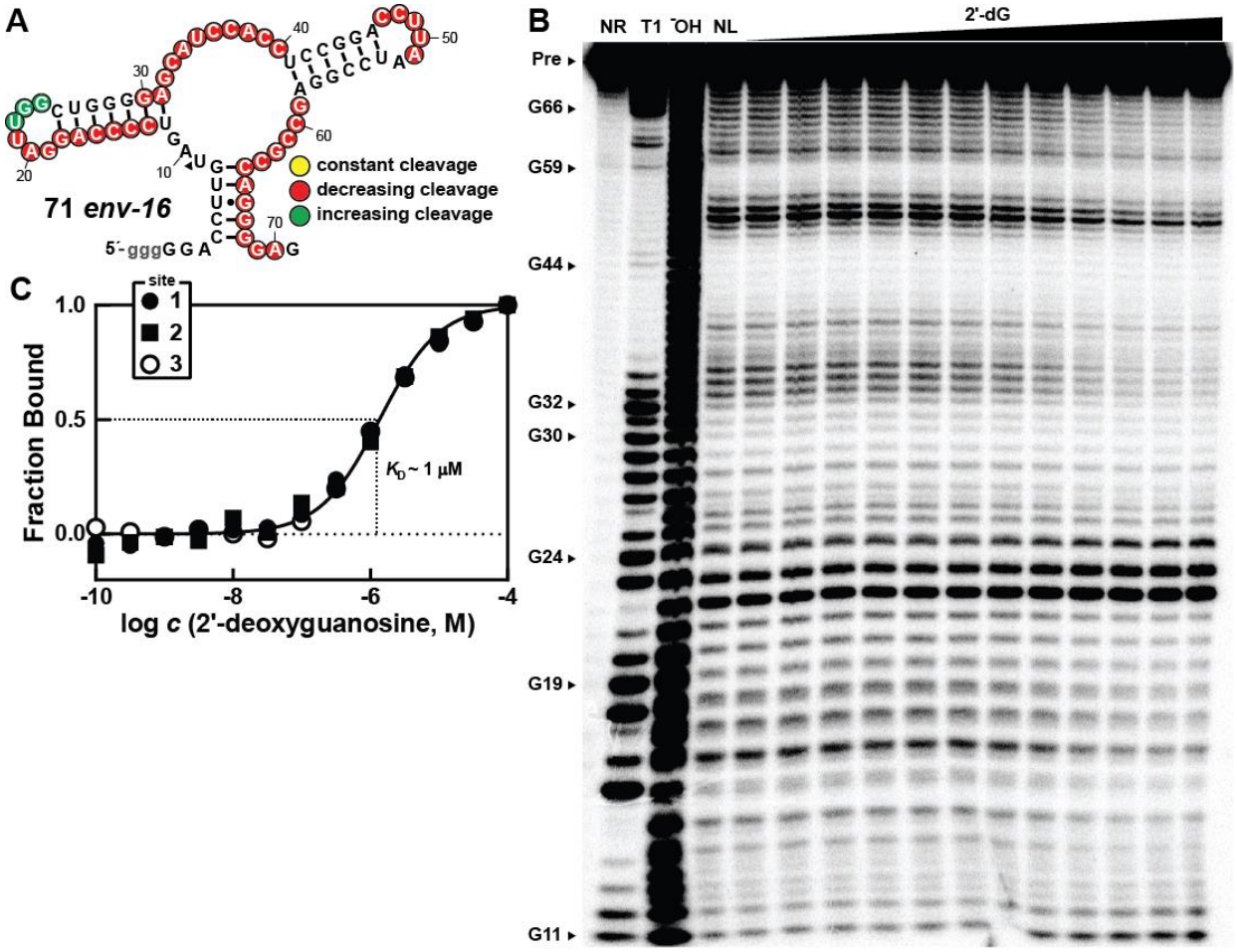


Fig. S3. Comparison of the consensus sequences and secondary structure models for (*A*) guanine, (*B*) 2´-dG-I, and (*C*) 2´-dG-II riboswitch aptamers (corresponds to the newly found U,C,C group). Annotations are as described for **Fig. 2*A***. Predicted secondary structures depicted in *B* and *C* are based on the data presented in *A*.

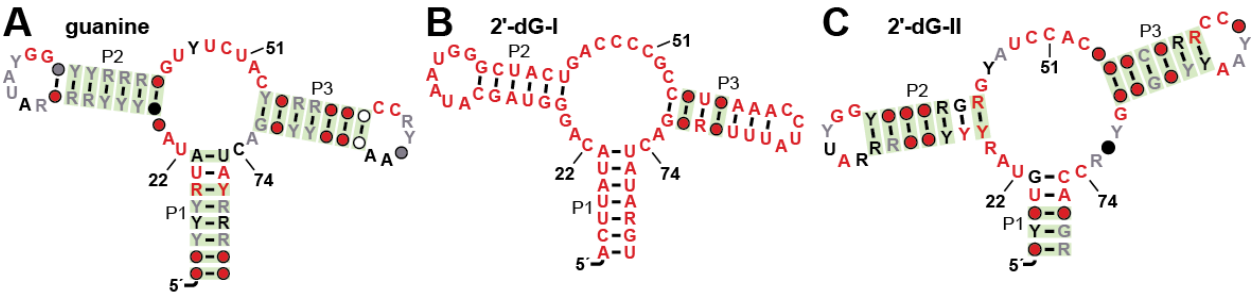


Fig. S4. No modulation with tested compounds against a representative U,G,C group variant of guanine riboswitch aptamers. (*A*) Sequence and secondary structure of the 72 *coxL* from *Desulfosporosinusmeridiei* DSM 13257. Annotations are as described for **Fig. 3*B***. (*B*) PAGE analysis of in-line probing reactions of 5′ ^{32}P-labeled 72 *coxL* in the presence of various ligand candidates. Annotations are as described for **Fig. 3*C***.

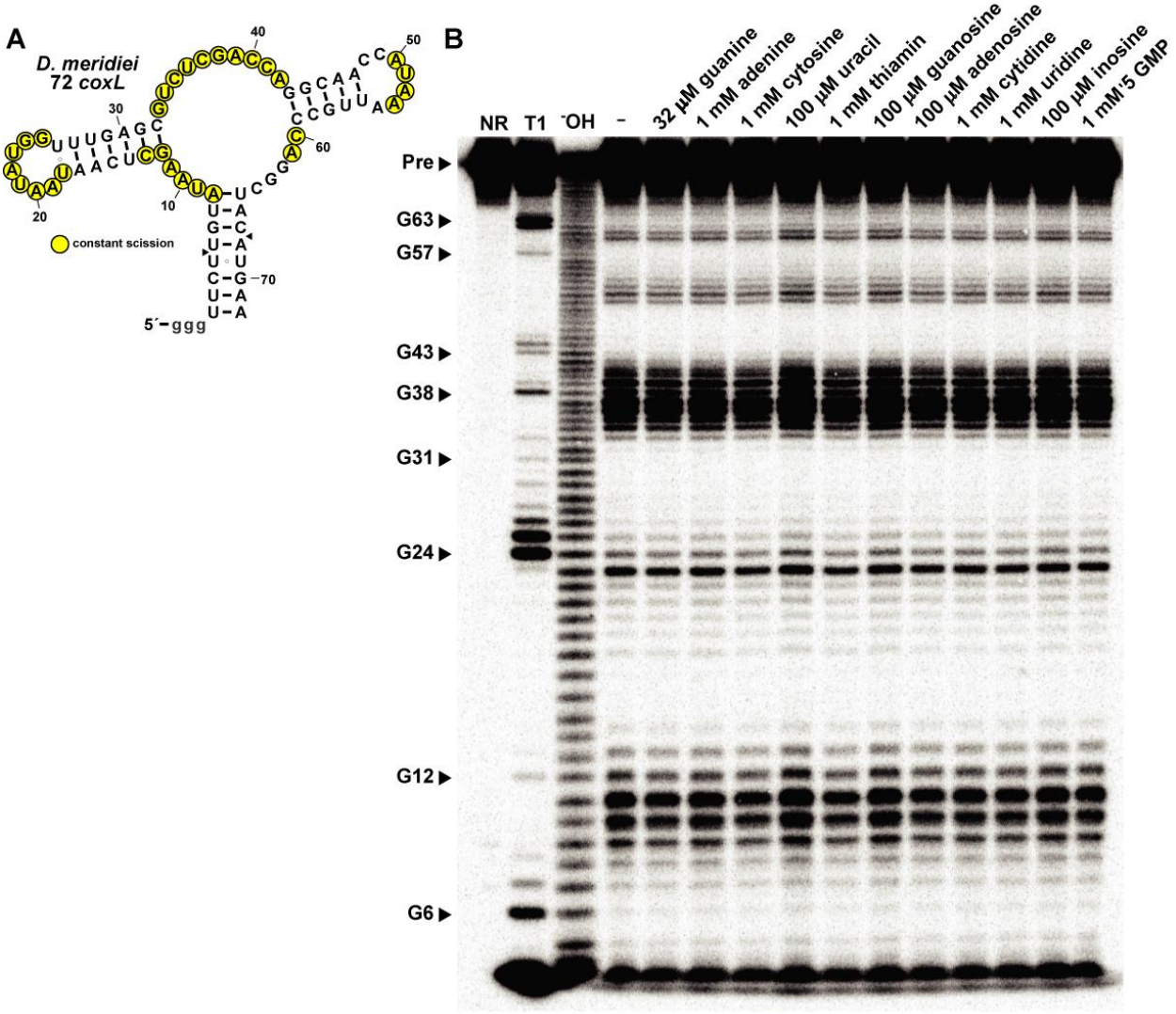
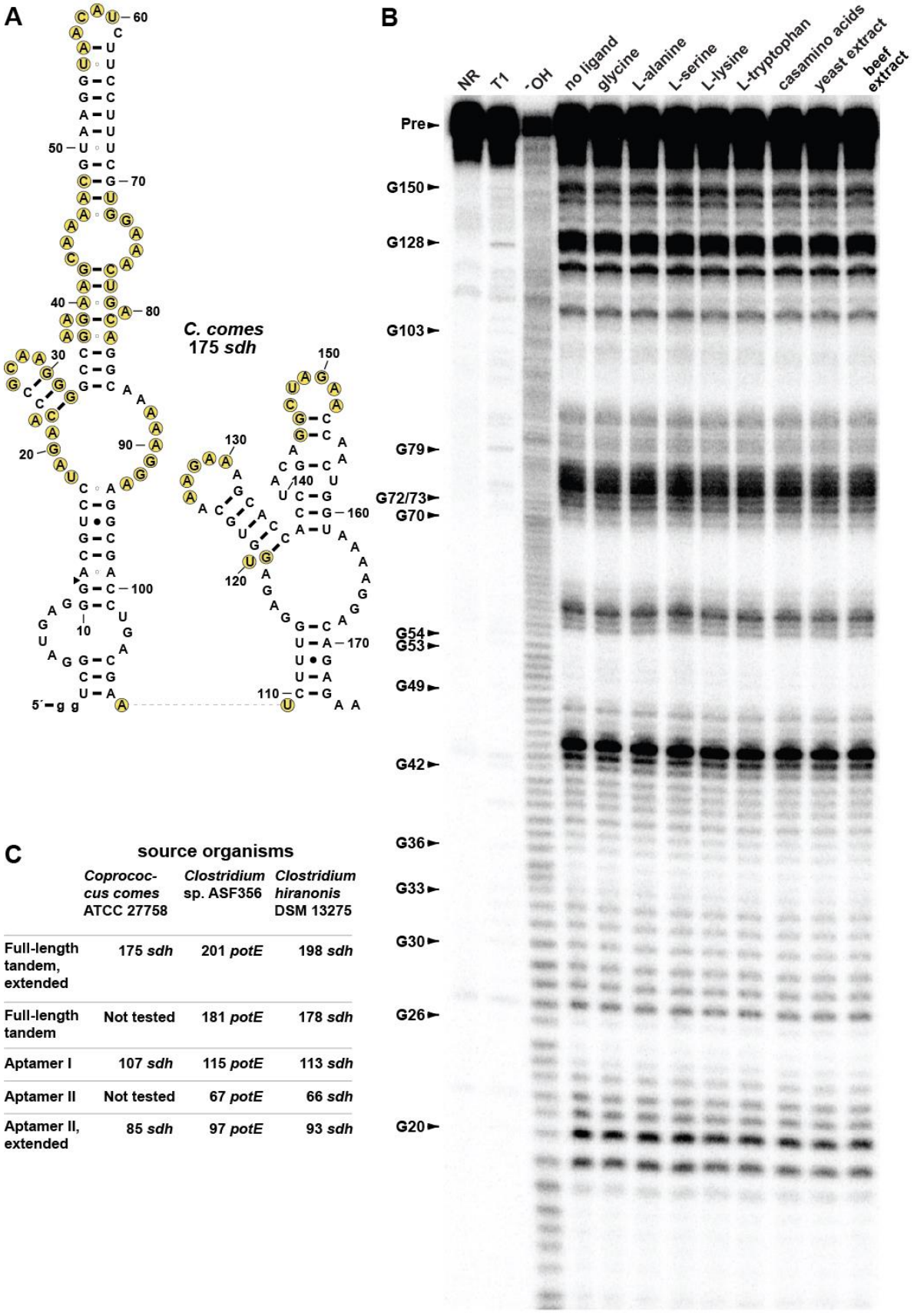


Fig. S5. (Next page) Examination of a candidate based on glycine riboswitch variants of the D,G,A group. (*A*) Sequence and secondary structure model of a D,G,A variant from *Coprococcus comes* ATCC 27758. Yellow circles indicate nucleotides whose 3′ linkage exhibits cleavage during in-line probing reactions. These nucleotides are likely to be relatively unstructured, and the data is largely consistent with the proposed secondary structure model. The arrowhead indicates the start of data on the gel. These data are typical of other D,G,A group constructs we tested. (*B*) PAGE analysis of in-line probing reactions of the molecule from part A in the presence of various compounds. Annotations are as described for **Fig. 3*C***. The compounds or mixtures listed were tested, and did not yield evidence of inducing structure modulation. By contrast, the VC I-II construct (1) exhibits structural modulation by glycine, casamino acids, and yeast and beef extract. The following compounds or mixtures were tested on all constructs (see **Table S4**), but did not show indications of binding: tryptophan, glycine, L-alanine, L-asparatate, L-cysteine, L-histidine, L -isoleucine, L-serine, L-threonine, L-lysine, glutamine, L-glutamate, L-valine, L-methionine, L-saccharopine, L-2-aminoadipic acid, 2-oxoadipic acid, 5-hydroxy-L-lysine, α-ketoglutarate, glycine methyl ester, glycinetert-butylester, glyoxylic acid monohydrate, glycolic acid, glycinamide, glycinehydroxamate, GABA, mercaptoacetic acid, sarcosine, putrescine, cadaverine, spermidine, spermine, c-di-AMP, c-di-GMP, c-AMP-GMP, casamino acids, and yeast and beef extract. (*C*) Names of D,G,A group variants of glycine riboswitch constructs that were tested. Constructs are sorted based on the organism from which they are derived (columns) and the length of the construct (row). A “tandem” construct is made up of Aptamer I (the 5′ aptamer) and Aptamer II (the 3′ aptamer). “Extended” tandem molecules add nucleotides on the 5′ and 3′ ends beyond the known conserved glycine riboswitch structure. “Extended” Aptamer II molecules are extended on their 5′ ends to include a 3′ part of Aptamer I. The oligonucleotides used to create the various molecules are listed in **Table S4**.

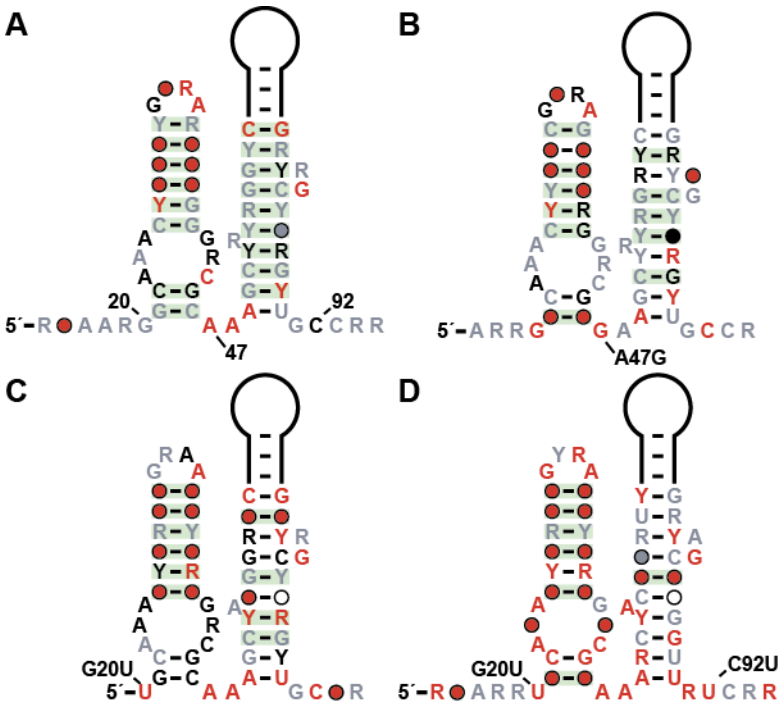


C

	source organisms		
	Coprococcus comes ATCC 27758	*Clostridium* sp. ASF356	*Clostridium hiranonis* DSM 13275
Full-length tandem, extended	175 *sdh*	201 *potE*	198 *sdh*
Full-length tandem	Not tested	181 *potE*	178 *sdh*
Aptamer I	107 *sdh*	115 *potE*	113 *sdh*
Aptamer II	Not tested	67 *potE*	66 *sdh*
Aptamer II, extended	85 *sdh*	97 *potE*	93 *sdh*

Fig. S6. Candidates based on c-di-GMP-I and c-di-GMP-II riboswitches. (*A*) Consensus diagram of established c-di-GMP-I riboswitches, with key nucleotides G,A,C labeled. Annotations are as described for **Fig. 2*A***. The enclosing P1 stem of c-di-GMP-I riboswitches is not depicted, because it is difficult to align, and not present in Rfam's alignments. The riboswitch sequences in **Dataset S1** are extended on their 5′ and 3′ ends so as to ensure that the nucleotides that participate in the P1 stem will be present. However, doing so likely includes several nucleotides that do not contribute to forming the binding pocket in **Dataset S1** at http://breaker.yale.edu/variants. (*B*) Consensus of the G,G,C variant group, (*C*) U,A,C group and (*D*) U,A,U group.

(Next page) (*E*) Consensus diagram of established c-di-GMP-II riboswitches, with key nucleotides A,A,G labeled. (*F*) Consensus of the G,A,G variant group, (*G*) A,A,U group, (*H*) G,C,G group, (*I*) A,A,A group and (*J*) A,C,U group.



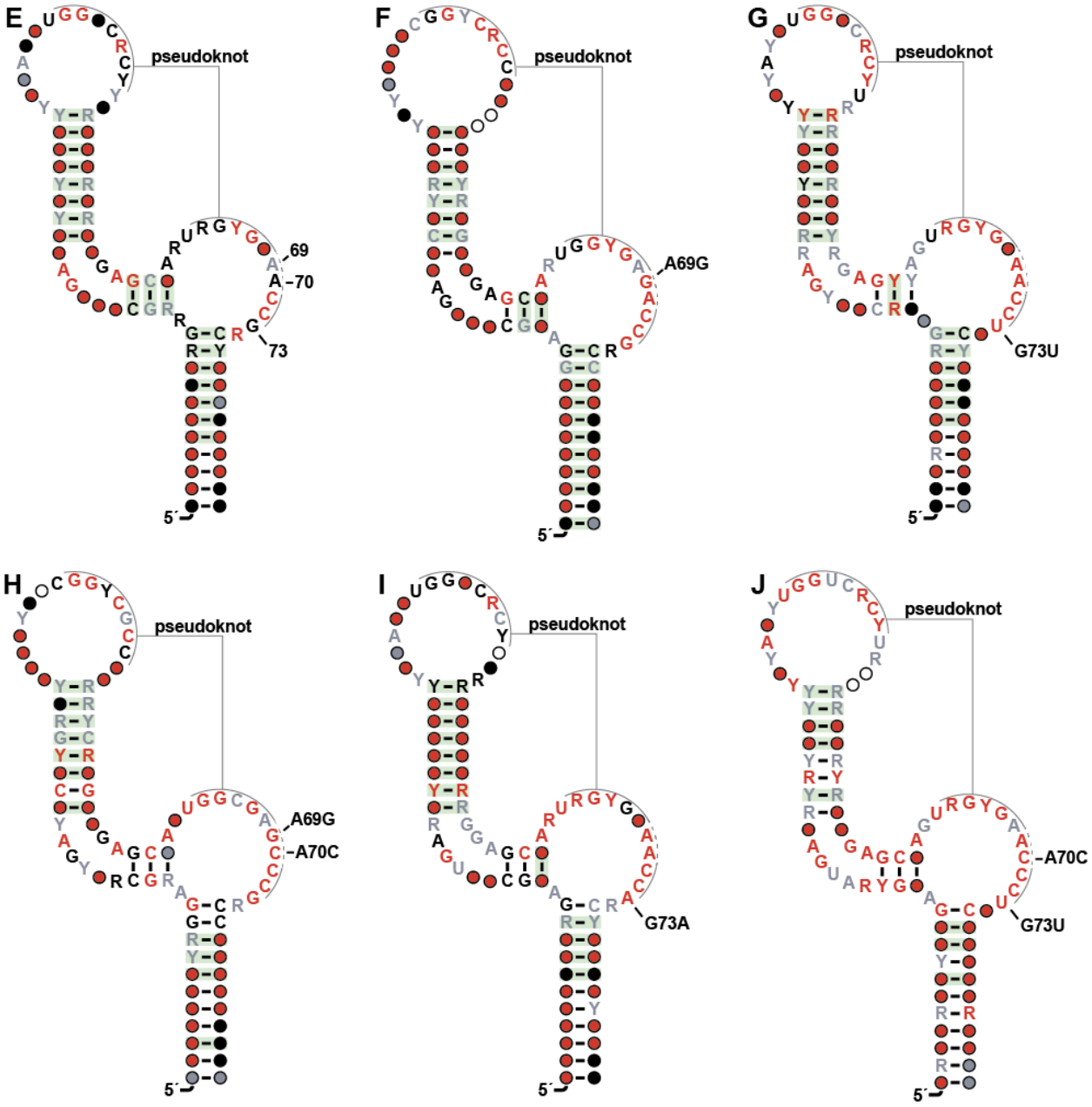

E
pseudoknot
69
70
73
5´
F
pseudoknot
A69G
5´
G
pseudoknot
G73U
5´
H
pseudoknot
A69G
A70C
5´
I
pseudoknot
G73A
5´
J
pseudoknot
A70C
G73U
5´

Fig. S7. Expanded consensus sequence and secondary structure of variant FMN riboswitches similar to an example published previously (5, 6). Annotations are as described for **Fig. 2*A***. Predicted secondary structure is based on bioinformatic and structural data (7-9).

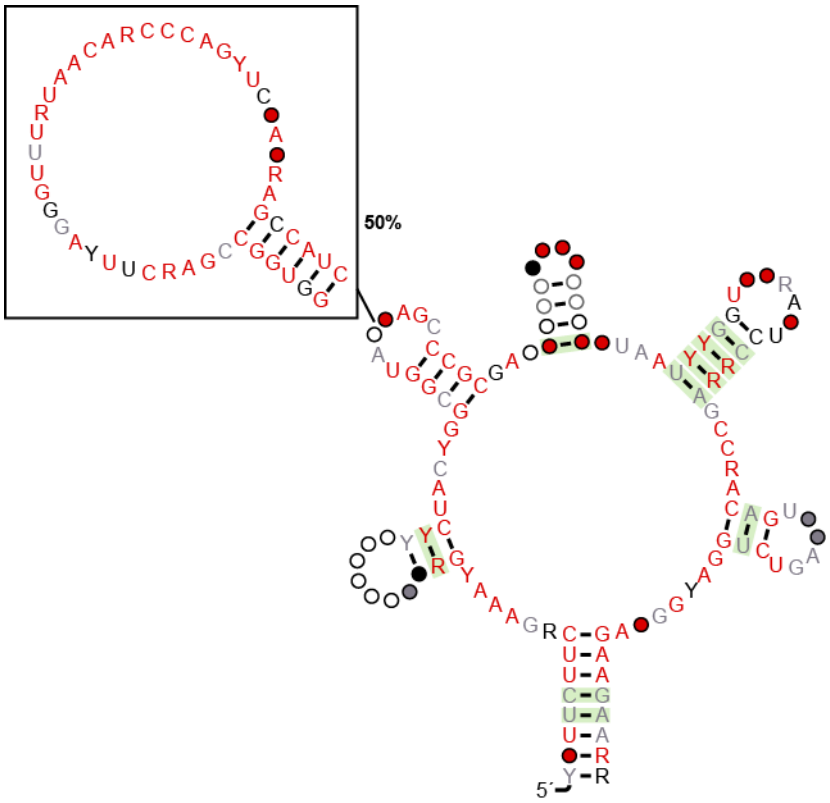
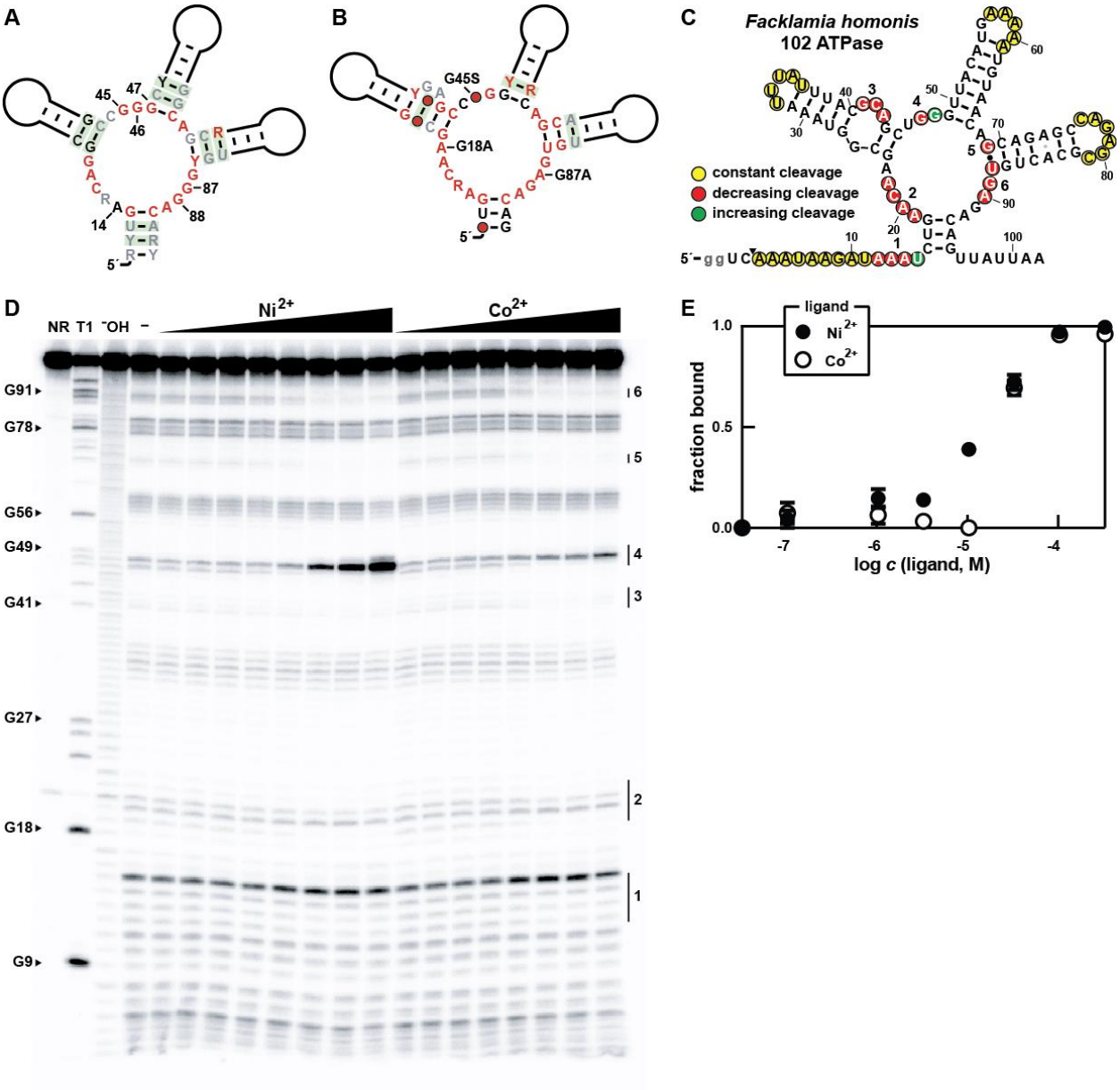


Fig. S8. A candidate based on Ni/Co riboswitches. (*A*) Consensus Ni/Co riboswitch aptamer, with key nucleotides labeled. Nucleotide depictions are as in **Fig. 2*A***. S represents either "G" or "C" nucleotides. (*B*) Consensus of variant riboswitches with G87→A mutation. Changes relative to the crystalized group are labeled in black. Predicted secondary structure is based on data in *A*. (*C*) Sequence and secondary structure of the 102 ATPase RNA from *Facklamiahomonis*. Labels are as in **Fig. 3B**. (*D*) PAGE analysis of in-line probing reactions of 5' ^{32}P-labeled 102 ATPase in the presence of increasing Ni^{2+} or Co^{2+} (100 nm to 32 µM in half-log units). Labels are as in **Fig. 3C**. (*E*) Plot of the fraction of RNA bound to ligand versus the log of the molar concentration *c* of added ligand. Data are derived from part *D*.



SI References

1. Mandal M*, et al.* (2004) A glycine-dependent riboswitch that uses cooperative binding to control gene expression. *Science* 306(5694):275-279.
2. Erion TV & Strobel SA (2011) Identification of a tertiary interaction important for cooperative ligand binding by the glycine riboswitch. *RNA* 17(1):74-84.
3. Sherman EM, Esquiaqui J, Elsayed G, & Ye JD (2012) An energetically beneficial leader-linker interaction abolishes ligand-binding cooperativity in glycine riboswitches. *RNA* 18(3):496-507.
4. Furukawa K, et al. (2015) Bacterial riboswitches cooperatively bind Ni^{2+} or Co^{2+} ions and control expression of heavy metal transporters. *Mol Cell* 57(6):1088-1098.
5. Blount KF (2013) *Methods for treating or inhibiting infection by Clostridium difficile. US Patent Appl No.* 13/576,989.
6. Blount KF*, et al.* (2012) *Flavin derivatives. US Patent Appl No.* 13/381,809.
7. Winkler, WC, Cohen-Chalamish, S, Breaker, RR (2002) An mRNA structure that controls gene expression by binding FMN. *Proc. Natl Acad Sci USA* 99(25):15908-15913.
8. Mironov, AA, Gusarov, I, Rafikov, R., Lopez, LE, Shatalin, K, Kreneva, RA, Perumov, DA, Nudler, E (2002) Sensing small molecules by nascent RNA: a mechanism to control transcription in bacteria. *Cell* 111(5):747-756.
9. Serganov, A, Huang L, Patel DJ (2009) Coenzyme recognition and gene regulation by a flavin mononucleotide riboswitch. *Nature* 458(7235):233-237.