Molecular Cell, Volume *29*

Supplemental Data

Riboswitches that Sense *S*-adenosylhomocysteine and Activate Genes Involved in Coenzyme Recycling

Joy Xin Wang, Elaine R. Lee, Dianali Rivera Morales, Jinsoo Lim, and Ronald R. Breaker

Supplemental Experimental Procedures

Identification and Unification of SAH Motif Structural Models

Although a previous bioinformatics search revealed the first representatives of SAH riboswitch candidates (Weinberg et al., 2007), a subsequent examination of bioinformatics hits (unpublished observations) uncovered a second putative RNA motif with gene associations similar to those for the SAH RNA motif described previously. Further analyses revealed that the predicted secondary structure for the original SAH motif could be reorganized to include a pseudoknot. The second collection of SAH RNA candidates also fit the revised structural model, thus unifying all hits into a single consensus sequence and structural class. This revised consensus pattern was used to search for more matches (see **Experimental Procedures**).

Bioinformatics

The initial seed alignments for the SAH aptamer were discovered using the CMfinder comparative genomics pipeline (Yao et al., 2007; Weinberg et al., 2007). After manual editing of the alignment, a probabilistic covariance model (CM) was derived and searched against microbial genomes and environmental sequences deposited in the NCBI RefSeq database (Pruitt et al., 2007), using the RAVENNA extension (Weinberg and Ruzzo, 2006) to the Infernal software package (Eddy, 2003).

After each search iteration, non-duplicative new matches were manually aligned and examined. Only those that exhibit a reasonable genomic context and that could be fully aligned to known SAH aptamer examples were added to the seed alignment for

subsequent searching. Genomic context examination employed the COG database (Tatusov et al., 2003) to predict functions of ORFs. Due to limitations of CM searches, the pseudoknot structure was not enforced in the search, but rather checked afterwards during the manual examination process. To find more-divergent matches, local mode CM searches derived from seed alignments were also performed against a database comprising sequences in front of ORFs that are homologous to ORFs immediately downstream of known SAH aptamer representatives. Sequences that were high-scoring but partial matches were expanded to cover the whole conserved domain and were subjected to the same manual alignment/examination process as described for global searches.

Sequences in the final alignment were weighted according to their similarity using the Infernal software package´s internal implementation of the Gerstein-Sonnhammer-Chothia algorithm (Gerstein et al., 1994) to reduce the bias caused by the existence of multiple strains of the same bacterial species that are present in the genomic sequence database. The conservation statistics and consensus were then calculated with consideration of the sequence weight. Separate weights and conservation statistics were calculated from subsets of the final alignment for regions with alternative structures.

Figure S1

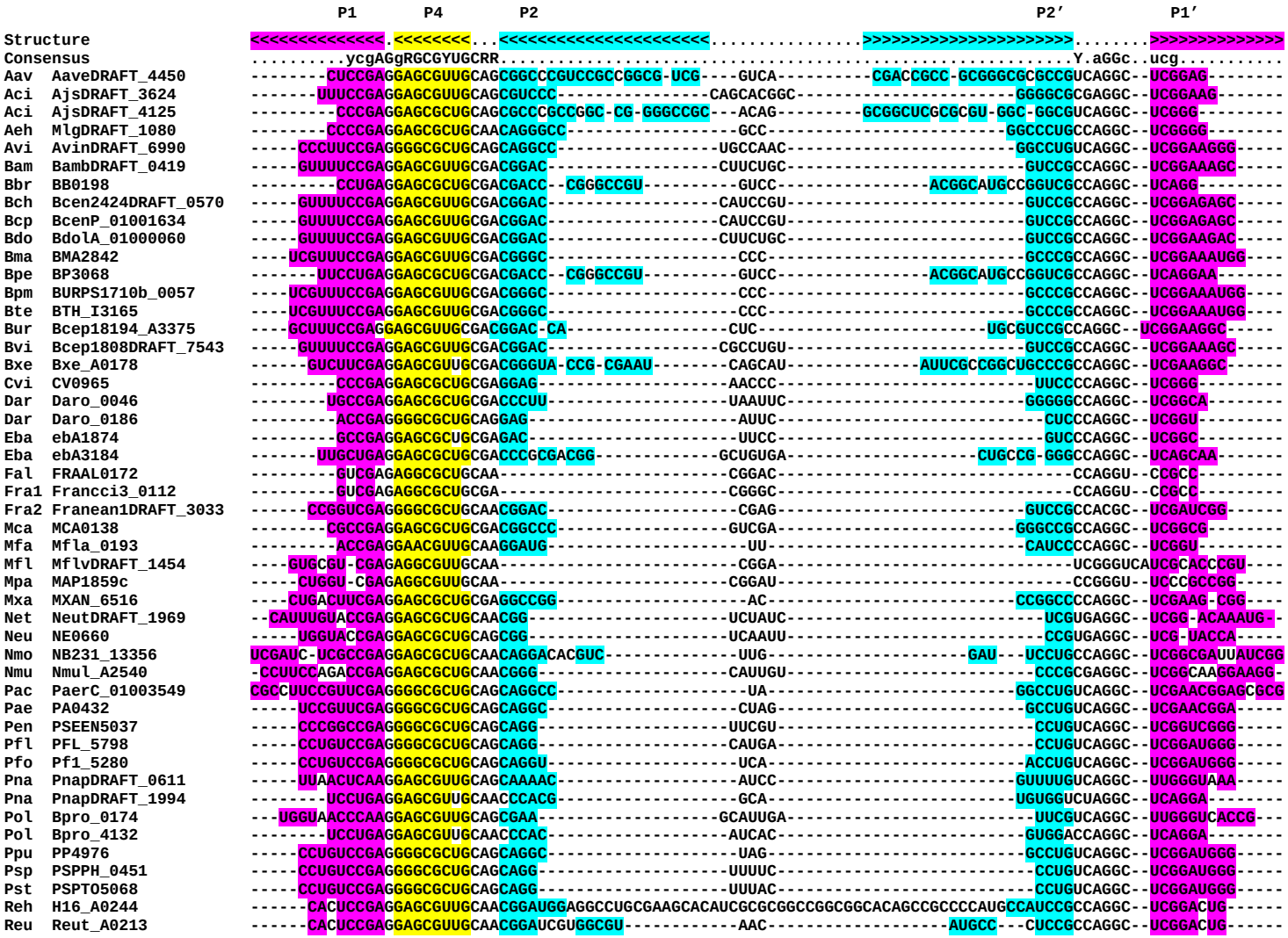
P1
P4
P2
P2′
P1′
Structure
Consensus
Aav AaveDRAFT_4450
Aci AjsDRAFT_3624
Aci AjsDRAFT_4125
Aeh MlgDRAFT_1080
Avi AvinDRAFT_6990
Bam BambDRAFT_0419
Bbr BB0198
Bch Bcen2424DRAFT_0570
Bcp BcenP_01001634
Bdo BdolA_01000060
Bma BMA2842
Bpe BP3068
Bpm BURPS1710b_0057
Bte BTH_I3165
Bur Bcep18194_A3375
Bvi Bcep1808DRAFT_7543
Bxe Bxe_A0178
Cvi CV0965
Dar Daro_0046
Dar Daro_0186
Eba ebA1874
Eba ebA3184
Fal FRAAL0172
Fra1 Francci3_0112
Fra2 Franean1DRAFT_3033
Mca MCA0138
Mfa Mfla_0193
Mfl MflvDRAFT_1454
Mpa MAP1859c
Mxa MXAN_6516
Net NeutDRAFT_1969
Neu NE0660
Nmo NB231_13356
Nmu Nmul_A2540
Pac PaerC_01003549
Pae PA0432
Pen PSEEN5037
Pfl PFL_5798
Pfo Pf1_5280
Pna PnapDRAFT_0611
Pna PnapDRAFT_1994
Pol Bpro_0174
Pol Bpro_4132
Ppu PP4976
Psp PSPPH_0451
Pst PSPTO5068
Reh H16_A0244
Reu Reut_A0213

Figure S1, continued

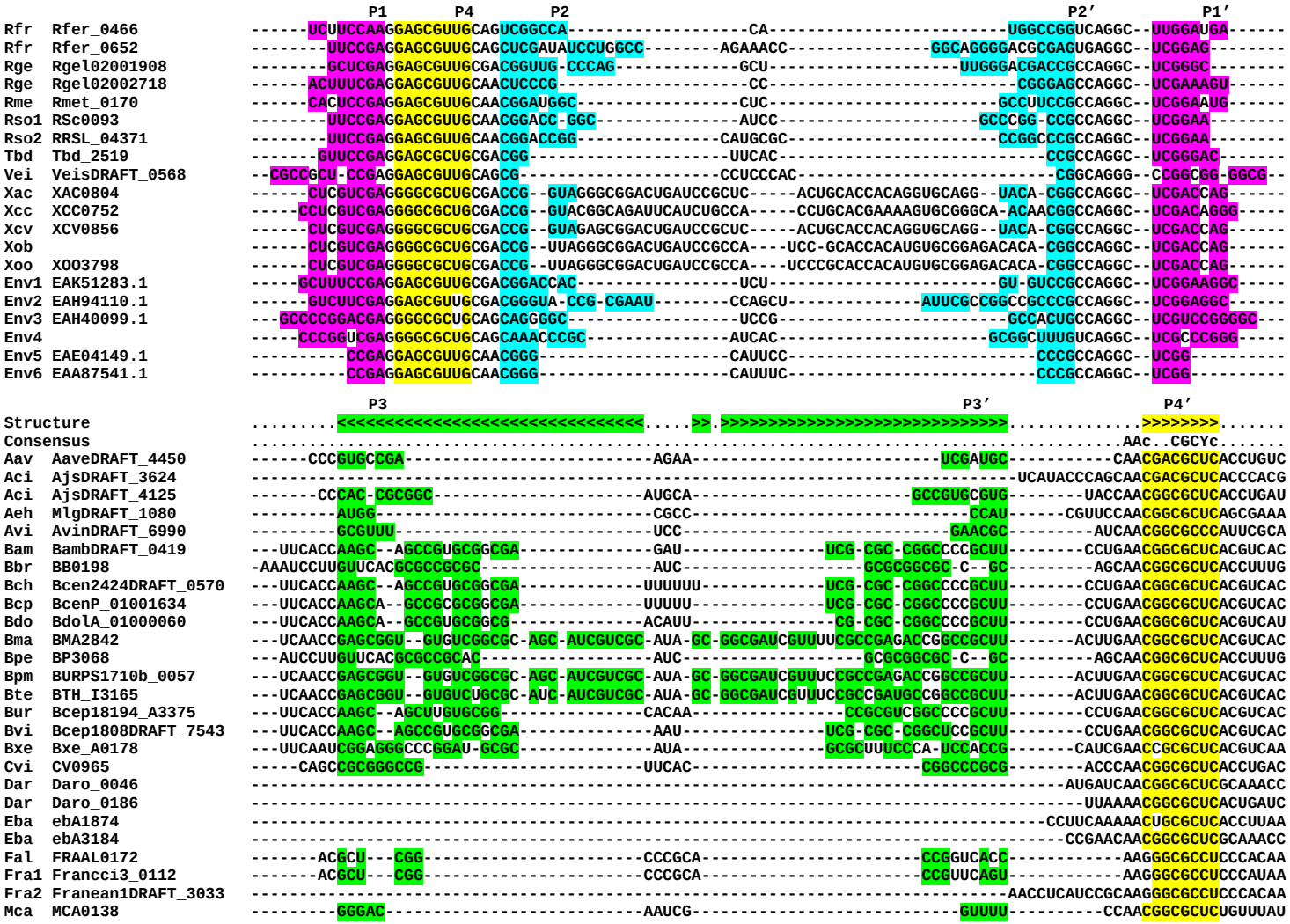

Figure S1, continued

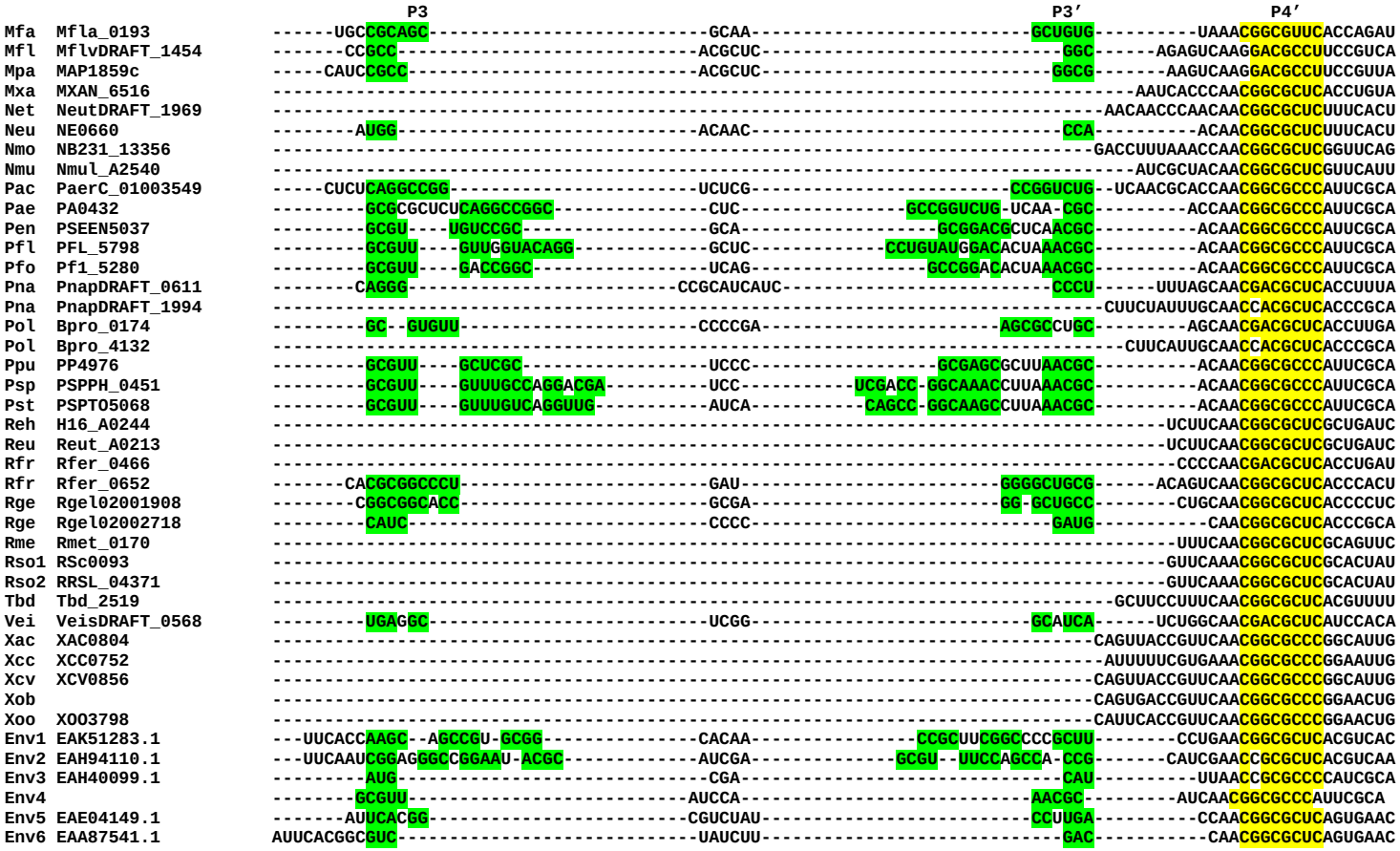

Figure S1. **Sequence Alignment of Nonduplicative SAH RNAs**

Predicted secondary structure elements are marked by brackets in the Structure row and are individually colored as follows: P1 (red), P4 pseudoknot (yellow), P2 (blue), and the optional P3 (green), where P represents "paired". Conserved nucleotides are labeled in the consensus row as follows: >97% (uppercase) and >90% (lowercase), where R = A or G, and Y = C or U. A three-letter organism abbreviation (see **Figure S2** for the bacterial strain) is followed by the downstream gene's locus tag from the original sequence file.

Figure S2 (next three pages)

Abb	*Organism*	Accession/Start-End
Aav	Acidovorax avenae subsp. citrulli AAC00-1	NZ_AASX01000001.1/552959-552847
Aci	Acidovorax sp. JS42	NZ_AASD01000006.1/141640-141560
Aci	Acidovorax sp. JS42	NZ_AASD01000001.1/236118-236236
Aeh	Alkalilimnicola ehrlichei MLHE-1	NZ_AALK01000011.1/76195-76276
Avi	Azotobacter vinelandii AvOP	NZ_AAAU03000001.1/1647268-1647357
Bam	Burkholderia ambifaria AMMD	NZ_AAJL01000019.1/45329-45213
Bbr	Bordetella bronchiseptica RB50	NC_002927.3/203883-204000
Bch	Burkholderia cenocepacia HI2424	NZ_AAHL01000093.1/16213-16094
Bcp	Burkholderia cenocepacia PC184	NZ_AAKX01000047.1/2738-2856
Bdo	*Burkholderia dolosa AUO158*	NZ_AAKY01000006.1/6273-6157
Bma	*Burkholderia mallei ATCC 23344*	NC_006348.1/2934894-2934753
Bpe	Bordetella pertussis Tohama I	NC_002929.2/3271246-3271127
Bpm	*Burkholderia pseudomallei 1710b*	NC_007434.1/62267-62126
Bte	Burkholderia thailandensis E264	NC_007651.1/3610419-3610278
Bur	Burkholderia sp. 383	NC_007510.1/233787-233904
Bvi	Burkholderia vietnamiensis G4	NZ_AAEH02000001.1/229831-229715
Bxe	Burkholderia xenovorans LB400	NC_007951.1/4657558-4657422
Cvi	Chromobacterium violaceum ATCC 12472	NC_005085.1/998460-998548
Dar	Dechloromonas aromatica RCB	NC_007298.1/57503-57435
Dar	Dechloromonas aromatica RCB	NC_007298.1/205030-204972
Eba	Azoarcus sp. EbN1	NC_006513.1/1088386-1088448
Eba	Azoarcus sp. EbN1	NC_006513.1/1900923-1901003
Fal	Frankia alni ACN14a	NC_008278.1/171653-171581
Fra1	Frankia sp. CcI3	NC_007777.1/139089-139017
Fra2	Frankia sp. EAN1pec	NZ_AAII01000057.1/13981-14057
Mca	Methylococcus capsulatus str. Bath	NC_002977.6/144279-144199
Mfa	Methylobacillus flagellatus KT	NC_007947.1/205625-205548
Mfl	Mycobacterium flavescens PYR-GCK	NZ_AAPA01000012.1/130354-130275
Mpa	Mycobacterium avium subsp. paratuberculosis K-10	NC_002944.2/2041607-2041528
Mxa	Myxococcus xanthus DK 1622	NC_008095.1/8035797-8035872
Net	Nitrosomonas eutropha C71	NZ_AAJE01000001.1/19337-19257
Neu	Nitrosomonas europaea ATCC 19718	NC_004757.1/715035-715112
Nmo	Nitrococcus mobilis Nb-231	NZ_AAOF01000009.1/53447-53352
Nmu	Nitrosospira multiformis ATCC 25196	NC_007614.1/2909970-2909888
Pac	Pseudomonas aeruginosa C3719	NZ_AAKV01000077.1/22341-22452
Pae	Pseudomonas aeruginosa PAO1	NC_002516.2/484244-484138
Pen	Pseudomonas entomophila L48	NC_008027.1/5348241-5348337
Pfl	Pseudomonas fluorescens Pf-5	NC_004129.6/6604863-6604970
Pfo	Pseudomonas fluorescens PfO-1	NC_007492.1/5940641-5940740
Pna	Polaromonas naphthalenivorans CJ2	NZ_AANM01000024.1/33980-33887
Pna	Polaromonas naphthalenivorans CJ2	NZ_AANM01000014.1/106477-106405
Pol	Polaromonas sp. JS666	NC_007948.1/188517-188613
Pol	Polaromonas sp. JS666	NC_007948.1/4339041-4338971
Ppu	Pseudomonas putida KT2440	NC_002947.3/5667867-5667964
Psp	Pseudomonas syringae pv. phaseolicola 1448A	NC_005773.3/515372-515261
Pst	Pseudomonas syringae pv. tomato str. DC3000	NC_004578.1/5771305-5771415
Reh	Ralstonia eutropha H16	NC_008313.1/256130-256245
Reu	Ralstonia eutropha JMP134	NC_007347.1/239471-239552

Rfr	Rhodoferax ferrireducens T118	NC_007908.1/475493-475563
Rfr	Rhodoferax ferrireducens T118	NC_007908.1/675989-676101
Rge	Rubrivivax gelatinosus PM1	NZ_AAEM01000003.1/306484-306386
Rge	Rubrivivax gelatinosus PM1	NZ_AAEM01000006.1/168890-168967
Rme	Ralstonia metallidurans CH34	NC_007973.1/177418-177491
Rso1	Ralstonia solanacearum GMI1000	NC_003295.1/107655-107582
Rso2	Ralstonia solanacearum UW551	NZ_AAKL01000003.1/110952-111026
Tbd	Thiobacillus denitrificans ATCC 25259	NC_007404.1/2591666-2591597
Vei	Verminephrobacter eiseniae EF01-2	NZ_AASQ01000026.1/18922-19012
Xac	Xanthomonas axonopodis pv. citri str. 306	NC_003919.1/955568-955456
Xcc	Xanthomonas campestris pv. campestris str. ATCC 33913	NC_003902.1/900295-900180
Xcv	Xanthomonas campestris pv. vesicatoria str. 85-10	NC_007508.1/977023-976911
Xob	Xanthomonas oryzae pv. oryzicola BLS256	NZ_AAQN01000001.1/878107-877993
Xoo	Xanthomonas oryzae pv. oryzae KACC10331	NC_006834.1/4051725-4051840
Env1	Environmental sequence	AACY01007480.1/48341-48457
Env2	Environmental sequence	AACY01119115.1/4057-4193
Env3	Environmental sequence	AACY01168624.1/789-874
Env4	Environmental sequence	AACY01215322.1/604-697
Env5	Environmental sequence	AACY01482824.1/508-429
Env6	Environmental sequence	AACY01802862.1/608-686

Abb	*Organism*	Operon										
Aav	Acidovorax avenae subsp. citrulli AAC00-1	AaveDRAFT_4450	COG1410									
Aci	Acidovorax sp. JS42	AjsDRAFT_3624	COG0499	AjsDRAFT_3623	COG1189	AjsDRAFT_3622	COG0685	AjsDRAFT_3621	N/A			
Aci	Acidovorax sp. JS42	AjsDRAFT_4125	COG1410	AjsDRAFT_4126	COG2989	AjsDRAFT_4127	COG3108					
Aeh	Alkalilimnicola ehrlichei MLHE-1	MlgDRAFT_1080	COG0499	MlgDRAFT_1081	COG0685	MlgDRAFT_1082	COG4067	MlgDRAFT_1083	COG0189			
Avi	Azotobacter vinelandii AvOP	AvinDRAFT_6990	COG0499	AvinDRAFT_6991	COG0685							
Bam	Burkholderia ambifaria AMMD	BambDRAFT_0419	COG0499	BambDRAFT_0418	COG1950	BambDRAFT_0417	COG0685					
Bbr	Bordetella bronchiseptica RB50	BB0198	COG0499	BB0199	N/A	BB0200	COG0685					
Bch	Burkholderia cenocepacia HI2424	Bcen2424DRAFT_0570	COG0499	Bcen2424DRAFT_0569	COG1950	Bcen2424DRAFT_0568	COG0685					
Bcp	Burkholderia cenocepacia PC184	BcenP_01001634	COG0499	BcenP_01001635	COG1950	BcenP_01001636	COG0685					
Bdo	*Burkholderia dolosa AUO158*	BdolA_01000060	COG0499	BdolA_01000059	COG1950	BdolA_01000058	COG0685					
Bma	*Burkholderia mallei ATCC 23344*	BMA2842	COG0499	BMA2841	COG1950	BMA2840	COG0685					
Bpe	Bordetella pertussis Tohama I	BP3068	COG0499	BP3067	N/A	BP3066	COG0685	BP3065	COG3415			
Bpm	*Burkholderia pseudomallei 1710b*	BURPS1710b_0057	COG0499	BURPS1710b_0056	COG1950	BURPS1710b_0055	COG0685					
Bte	Burkholderia thailandensis E264	BTH_I3165	COG0499	BTH_I3164	COG1950	BTH_I3163	COG0685					
Bur	Burkholderia sp. 383	Bcep18194_A3375	COG0499	Bcep18194_A3376	COG1950	Bcep18194_A3377	COG0685					
Bvi	Burkholderia vietnamiensis G4	Bcep1808DRAFT_7543	COG0499	Bcep1808DRAFT_7542	COG1950	Bcep1808DRAFT_7541	COG0685					
Bxe	Burkholderia xenovorans LB400	Bxe_A0178	COG0499	Bxe_A0179	COG1950	Bxe_A0180	COG0685					
Cvi	Chromobacterium violaceum ATCC 12472	CV0965	COG0499	CV0966	COG0685	CV0967	COG2862					
Dar	Dechloromonas aromatica RCB	Daro_0046	COG1410									
Dar	Dechloromonas aromatica RCB	Daro_0186	COG0499	Daro_0185	COG1950	Daro_0184	COG1189	Daro_0183	COG0685			
Eba	Azoarcus sp. EbN1	ebA1874	COG0499	ebA1875	COG0685	ebA1878	COG4454	ebA1879	N/A			
Eba	Azoarcus sp. EbN1	ebA3184	COG1410									
Fal	Frankia alni ACN14a	FRAAL0172	COG1410									
Fra1	Frankia sp. CcI3	Francci3_0112	COG1410	Francci3_0111	N/A							
Fra2	Frankia sp. EAN1pec	Franean1DRAFT_3033	COG1410									
Mca	Methylococcus capsulatus str. Bath	MCA0138	COG0499	MCA0137	COG0685							
Mfa	Methylobacillus flagellatus KT	Mfla_0193	COG0499	Mfla_0192	N/A	Mfla_0191	COG0685					
Mfl	Mycobacterium flavescens PYR-GCK	MflvDRAFT_1454	COG1410	MflvDRAFT_1453	COG0637							
Mpa	Mycobacterium avium subsp. paratuberculosis K-10	MAP1859c	COG1410									
Mxa	Myxococcus xanthus DK 1622	MXAN_6516	COG0499									
Net	Nitrosomonas eutropha C71	NeutDRAFT_1969	COG0499	NeutDRAFT_1968	COG0685							
Neu	Nitrosomonas europaea ATCC 19718	NE0660	COG0499	NE0661	COG0685							
Nmo	Nitrococcus mobilis Nb-231	NB231_13356	COG0499									
Nmu	Nitrosospira multiformis ATCC 25196	Nmul_A2540	COG0499	Nmul_A2539	COG0685							
Pac	Pseudomonas aeruginosa C3719	PaerC_01003549	COG0499	PaerC_01003550	COG0317	PaerC_01003551	COG0685	PaerC_01003552	N/A			
Pae	Pseudomonas aeruginosa PAO1	PA0432	COG0499	PA0431	COG0317	PA0430	COG0685	PA0429	N/A			
Pen	Pseudomonas entomophila L48	PSEEN5037	COG0499	PSEEN5038	COG0685	PSEEN5039	N/A	PSEEN5040	COG0834			
Pfl	Pseudomonas fluorescens Pf-5	PFL_5798	COG0499	PFL_5799	COG0685							
Pfo	Pseudomonas fluorescens PfO-1	Pf1_5280	COG0499	Pf1_5281	COG0685							
Pna	Polaromonas naphthalenivorans CJ2	PnapDRAFT_0611	COG1410									
Pna	Polaromonas naphthalenivorans CJ2	PnapDRAFT_1994	COG0499	PnapDRAFT_1993	COG1189	PnapDRAFT_1992	COG0685	PnapDRAFT_1991	N/A			
Pol	Polaromonas sp. JS666	Bpro_0174	COG1410									
Pol	Polaromonas sp. JS666	Bpro_4132	COG0499	Bpro_4131	COG1189	Bpro_4130	COG0685	Bpro_4129	N/A	Bpro_4128	COG0346	
Ppu	Pseudomonas putida KT2440	PP4976	COG0499	PP4977	COG0685	PP4978	N/A					
Psp	Pseudomonas syringae pv. phaseolicola 1448A	PSPPH_0451	COG0499	PSPPH_0450	COG0685							
Pst	Pseudomonas syringae pv. tomato str. DC3000	PSPTO5068	COG0499	PSPTO5069	COG0685							
Reh	Ralstonia eutropha H16	H16_A0244	COG0499	H16_A0245	COG1950	H16_A0246	COG0685	H16_A0247	COG3070	H16_A0248	COG1656	

Reu	Ralstonia eutropha JMP134	Reut_A0213	COG0499	Reut_A0214	COG1950	Reut_A0215	COG0685	Reut_A0216	COG3070	Reut_A0217	COG1656
Rfr	Rhodoferax ferrireducens T118	Rfer_0466	COG1410								
Rfr	Rhodoferax ferrireducens T118	Rfer_0652	COG0499	Rfer_0653	COG1189	Rfer_0654	COG0685				
Rge	Rubrivivax gelatinosus PM1	Rgel02001908	COG1410								
Rge	Rubrivivax gelatinosus PM1	Rgel02002718	COG0499	Rgel02002719	COG1189	Rgel02002720	COG0685				
Rme	Ralstonia metallidurans CH34	Rmet_0170	COG0499	Rmet_0171	COG1950	Rmet_0172	COG0685	Rmet_0173	COG3070	Rmet_0174	COG1656
Rso1	Ralstonia solanacearum GMI1000	RSc0093	COG0499	RSc0092	COG1950	RSc0091	COG0685	RSc0090	COG3070		
Rso2	Ralstonia solanacearum UW551	RRSL_04371	COG0499	RRSL_04372	COG1950	RRSL_04373	COG0685	RRSL_04374	COG3070		
Tbd	Thiobacillus denitrificans ATCC 25259	Tbd_2519	COG0499	Tbd_2518	N/A	Tbd_2517	COG0685				
Vei	Verminephrobacter eiseniae EF01-2	VeisDRAFT_0568	COG0499	VeisDRAFT_0569	COG1189	VeisDRAFT_0570	COG0685	VeisDRAFT_0592	N/A		
Xac	Xanthomonas axonopodis pv. citri str. 306	XAC0804	COG0499								
Xcc	Xanthomonas campestris pv. campestris str. ATCC 33913	XCC0752	COG0499								
Xcv	Xanthomonas campestris pv. vesicatoria str. 85-10	XCV0856	COG0499								
Xob	Xanthomonas oryzae pv. oryzicola BLS256										
Xoo	Xanthomonas oryzae pv. oryzae KACC10331	XOO3798	COG0499								
Env1	Environmental sequence	EAK51283.1	COG0499	EAK51284.1	COG1950	EAK51285.1	COG0685				
Env2	Environmental sequence	EAH94110.1	COG0499	EAH94111.1	COG1950	EAH94112.1	COG0685				
Env3	Environmental sequence	EAH40099.1									
Env4	Environmental sequence										
Env5	Environmental sequence	EAE04149.1	COG0499								
Env6	Environmental sequence	EAA87541.1	COG0499								

COG	Gene	Description
COG0499	***sahH, ahcY, acyH***	S-adenosylhomocysteine hydrolase [Coenzyme metabolism]
COG1410	***metH***	5-methyltetrahydrofolate–homocysteine S-methyltransferase (COG: Methionine synthase I, cobalamin-binding domain [Amino acid transport and metabolism])
COG0685	***metF***	5,10-methylenetetrahydrofolate reductase [Amino acid transport and metabolism]
COG1950		Predicted membrane protein [Function unknown]
COG1189		Predicted rRNA methylase [Translation, ribosomal structure and biogenesis]
COG3070	***tfoX***	Regulator of competence-specific genes [Transcription]
COG1656		Uncharacterized conserved protein [Function unknown]
COG0317	***spoT***	Guanosine polyphosphate pyrophosphohydrolases/synthetases [Signal transduction mechanisms / Transcription]

Figure S2. SAH Element Locations and Associated Genes

GenBank accession numbers and nucleotide positions are indicated for each representative from **Figure S1**. Operons likely to be controlled by riboswitches are on the same strand downstream of an SAH element, and have a predicted SAH recycling function or are separated by fewer than 100 nucleotides. Gene locus tags and COG database assignments are provided for up to five genes for each putative operon. N/A: no conserved domain was found within the predicted ORF when searching against COG database. COGs appearing more than once are individually colored.

Figure S3

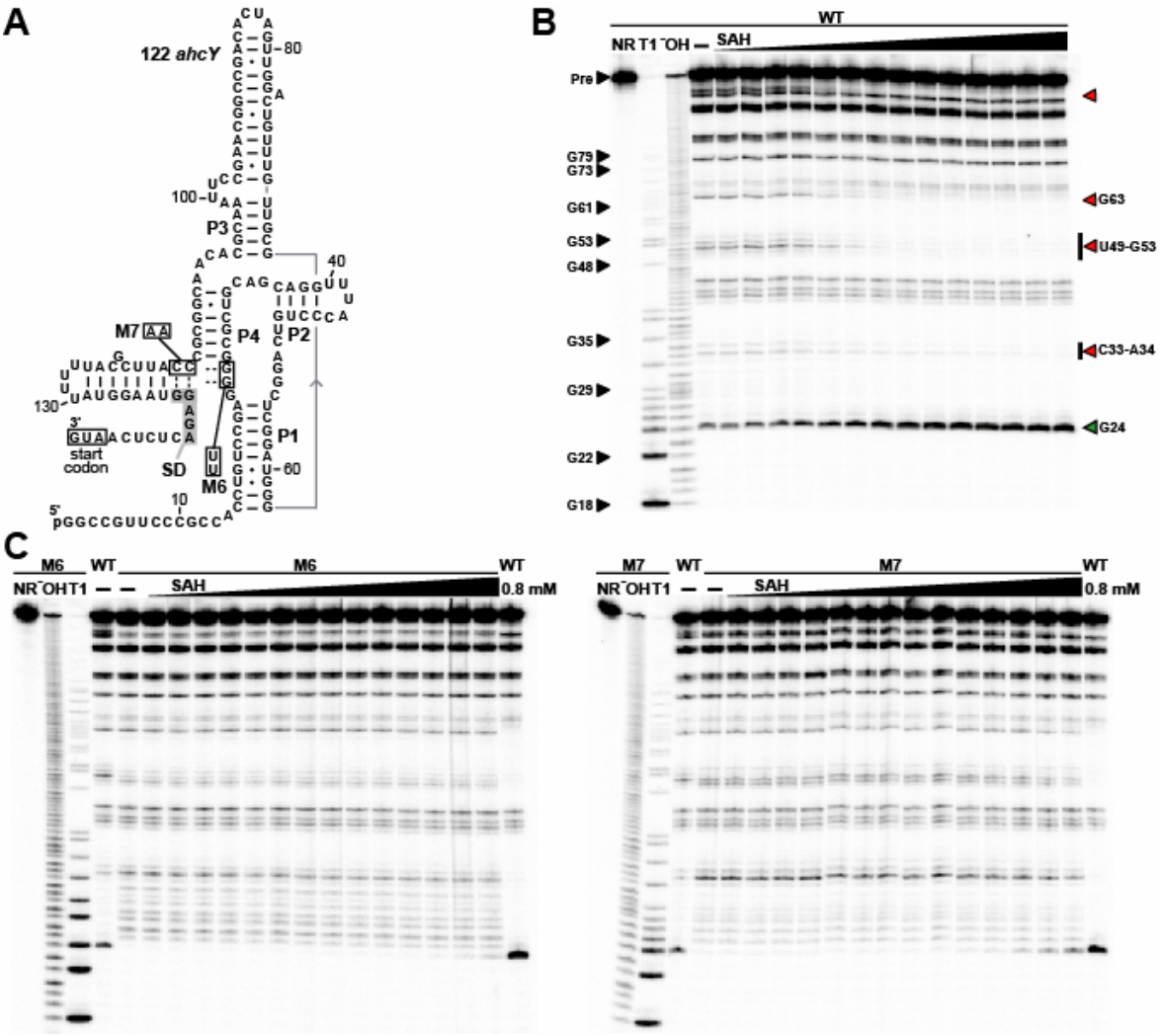

Figure S3. Mutations M6 and M7 Have Similar Effects on Structure Formation of and SAH Binding by the 122 *ahcY* RNA

(A) Sequence and secondary structure model of the 122 *ahcY* RNA, which encompasses an SAH aptamer. Two guanosyl residues were added to the 5′ end to facilitate T7 *in vitro* transcription. Nucleotide changes to generate mutants M6 and M7 are indicated. The last 29 nucleotides of the construct depicted are not present in the 122 *ahcY* construct probed.

(B) Spontaneous cleavage patterns of the wild-type (WT) 122 *ahcY* RNA in the absence (–) or presence of SAH at concentrations ranging from 600 pM to 0.82 mM. Red and green arrowheads identify sites of ligand-mediated decreased or increased spontaneous cleavage, respectively. Other notations are as described in the legend to **Figure 3**.

(C) Spontaneous cleavage patterns of the mutant 122 *ahcY* RNAs M6 and M7 in the absence (–) or presence of SAH. Additional details are as described in **B**.

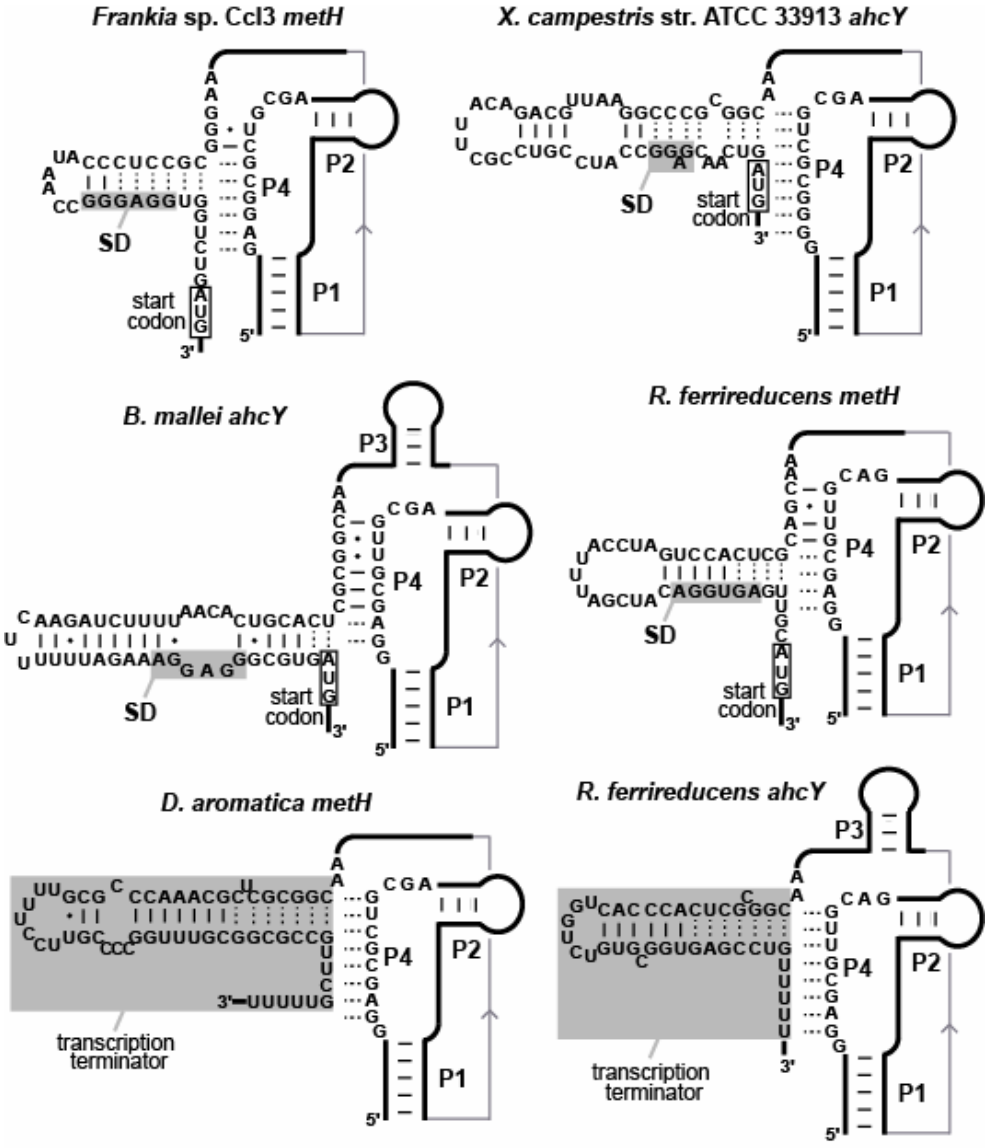

Figure S4. Predicted Expression Platform of Several SAH Riboswitches
Secondary structures of sequences downstream of SAH aptamers were predicted by the mfold (version 3.2) web server using default parameters (Zuker, 2003; Mathews et al., 1999). The transcription terminators (Gusarov and Nudler, 1999; Yarnell and Roberts, 1999) upstream of the *D. aromatica metH* and *R. ferrireducens ahcY* genes were predicted by TransTermHP (version 2.06) with confidence values of 100% and 75%, respectively (Kingsford et al., 2007). Dotted lines identify base pairing potential that could form stem P4 in the presence of SAH. The mfold algorithm can be found at http://frontend.bioinfo.rpi.edu/applications/mfold/cgi-bin/rna-form1.cgi.

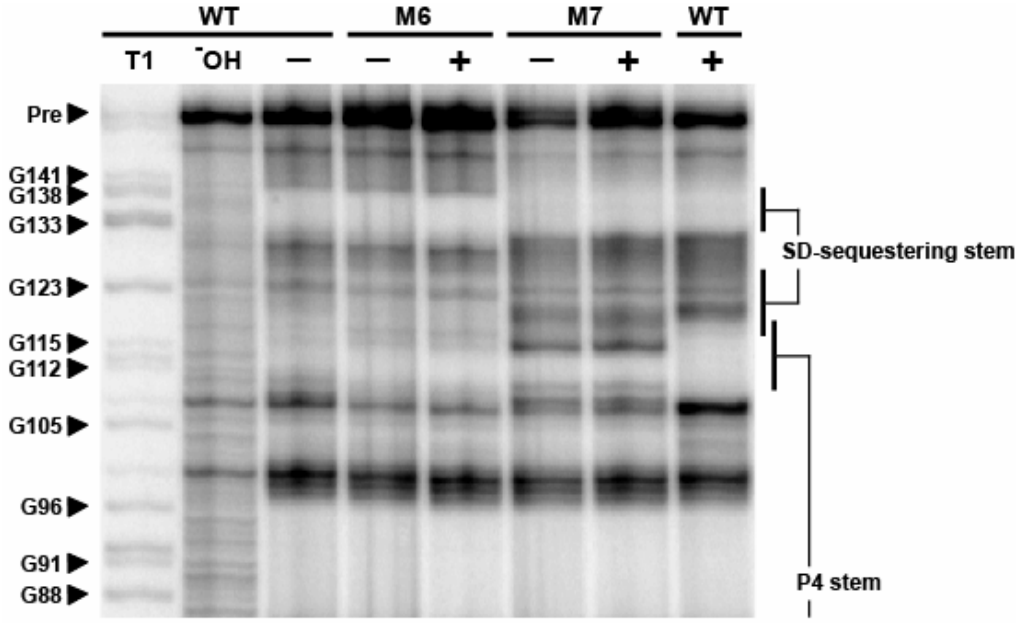

Figure S5. Mutations M6 and M7 Disrupt SAH-Dependent Modulation of the Putative SD-Sequestering Stem in the 151 *ahcY* RNA

In-line probing of the wild-type full-length 151 *ahcY* construct (WT) (see the construct depicted in Fig. S3A) reveals a pattern of spontaneous cleavage products that is consistent with the formation of the SD-sequestering stem in the absence (–) of SAH. The spontaneous cleavage pattern is consistent with SD-sequestering stem disruption in the presence (+) of 800 μM SAH. Regardless of the presence or absence of SAH, the M6 version of this construct retains the spontaneous cleavage pattern observed for the WT RNA in the absence of ligand, indicating that the M6 RNA constantly forms the SD-sequestering stem. In contrast, the M7 RNA exhibits a pattern of spontaneous cleavage that indicates the RNA does not form the SD-sequestering stem either in the presence or absence of SAH. Furthermore, an increase in spontaneous cleavage near nucleotides 115 through 118 of the M7 construct suggests that this RNA also fails to fully form P4, as expected due to the mutation of the C residues at positions 117 and 118.

This interplay of alternative RNA folds is consistent with our hypothesis for the mechanism of gene control by the *ahcY* RNA. However, the in-line probing data indicate that the WT construct has a reduced level of spontaneous cleavage near the SD sequence despite disruption of the SD-sequestering stem in the presence of SAH. Likewise, the M7 construct exhibits a similar in-line probing pattern in this region. Since both constructs express high levels of a reporter gene (Fig. 6), either these constructs are misfolding near the 3′ terminus due to the truncated RNAs used in these assays, or the mechanism of gene control is more complex than the mechanism proposed.

Supplemental References

Eddy, S.R. (2003). INFERNAL. Version 0.55. Distributed by the author. Dept. of Genetics, Washington University School of Medicine. St. Louis, Missouri.

Gerstein, M., Sonnhammer, E.L.L., and Chothia, C. (1994). Volume changes in protein evolution. J. Mol. Biol. *236*, 1067-1078.

Kingsford, C., Ayanbule, K., and Salzberg, S.L. (2007) Rapid, accurate, computational discovery of Rho-independent transcription terminators illuminates their relationship to DNA uptake. Genome Biol. *8*, R22.

Mathews, D.H., Sabina, J., Zuker, M., and Turner, D.H. (1999). Expanded sequence dependence of thermodynamic parameters improves prediction of RNA secondary structure. J. Mol. Biol. *288*, 911-940.

Gusarov, I., and Nudler, E. (1999). The mechanism of intrinsic transcription termination. *Mol*. Cell *3*, 495-504.

Pruitt, K.D., Tatusova, T., and Maglott, D.R. (2007). NCBI reference sequences (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. Nucleic Acids Research *35*, D61-D65.

Tatusov, R.L., Fedorova, N.D., Jackson, J.D., Jacobs, A.R., Kiryutin, B., Koonin, E.V., Krylov, D.M., Mazumder, R., Mekhedov, S.L., Nikolskaya, A.N., *et al.* (2003). The COG database: an updated version includes eukaryotes. BMC Bioinformatics *4*.

Weinberg, Z., Barrick, J. E., Yao, Z., Roth, A., Kim, J.N., Gore, J., Wang, J.X., Lee, E.R., Block, K.F., Sudarsan, N., *et al.* (2007). Identification of 22 candidate structured RNAs in bacteria using the CMfinder comparative genomics pipeline. Nucleic Acids Res. (doi:10.1093).

Weinberg, Z., and Ruzzo, W.L. (2006). Sequence-based heuristics for faster annotation of non-coding RNA families. Bioinformatics *22*, 35-39.

Yao, Z., Barrick, J.E., Weinberg, Z., Neph, S., Breaker, R.R., Tompa, M., and Ruzzo, W.L. (2007). A computational pipeline for high throughput discovery of *cis*-regulatory noncoding RNA in prokaryotes. PLoS Comput. Biol. *3*, e126.

Yarnell, W.S., and Roberts, J.W. (1999). Mechanism of intrinsic transcription termination and antitermination. Science *284*, 611-615.

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