

Evidence for a second class of *S*-adenosylmethionine riboswitches and other regulatory RNA motifs in alpha-proteobacteria

Keith A. Corbino^{*}, Jeffrey E. Barrick[†], Jinsoo Lim^{*}, Rüdiger Welz^{*‡}, Brian J. Tucker[†], Izabela Puskarz^{*},
Maumita Mandal^{*§}, Noam D. Rudnick^{*}, and Ronald R. Breaker^{*}

Addresses: ^{*}Department of Molecular, Cellular and Developmental Biology, [†]Department of Molecular Biophysics and Biochemistry, and [‡]Department of Chemistry, Yale University, P. O. Box 208103, New Haven, Connecticut 06520-8103, USA.

Current address: [§]Department of Physics, University of California, Berkeley, CA 94720-7200

Correspondence: Ronald R. Breaker. E-mail: ronald.breaker@yale.edu. Phone: 203 432-9389. Fax: 203 432-6604.

Additional Data File 1: Figures S1-S5

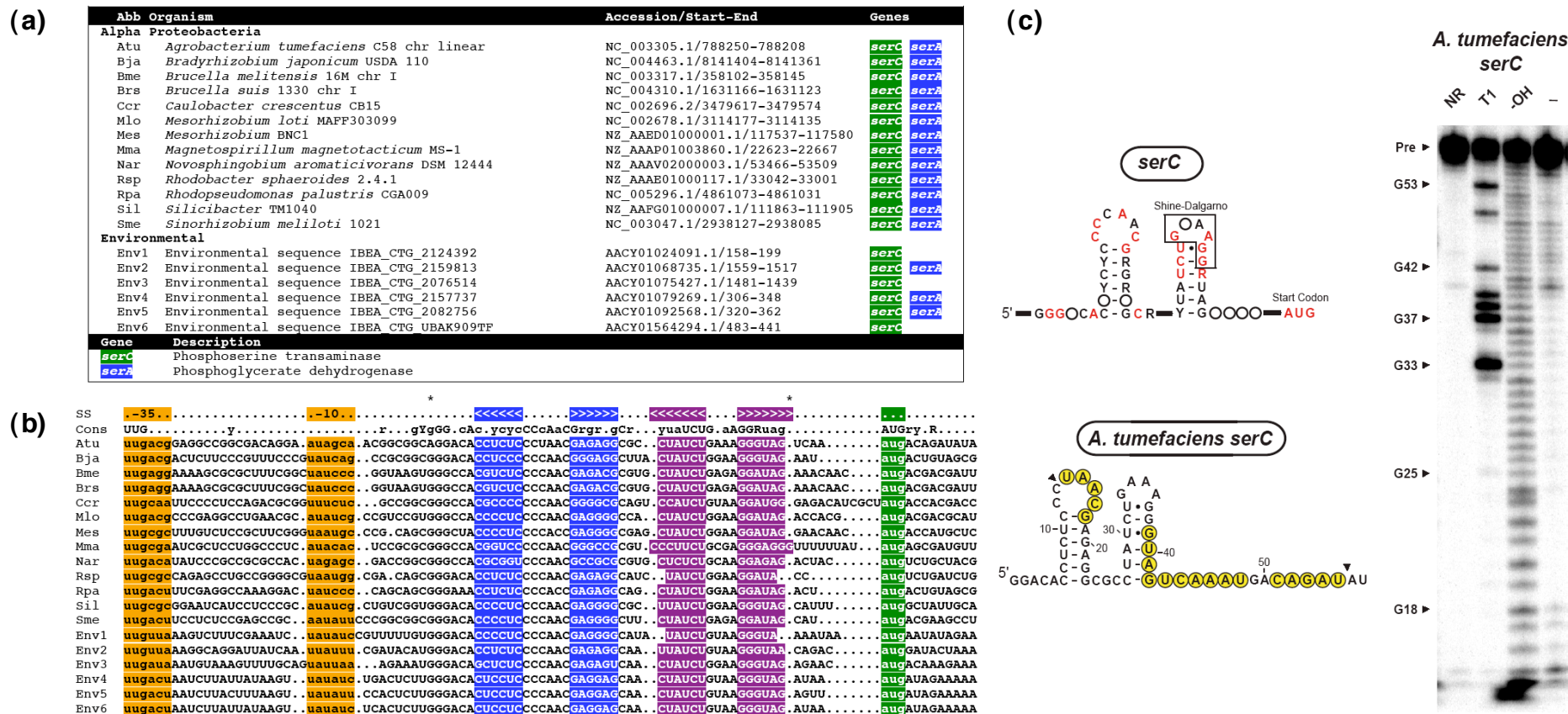


Figure S1

The *serC* Element. **(a)** Species Distribution. Accession numbers refer to GenBank nucleotide records. Functional assignments are provided for the downstream genes that are likely to be regulated if the element is located in the 5' UTR of an mRNA. **(b)** Sequence Alignment. Labels correspond to abbreviations in (a). Periods represent gaps and dashes indicate the end of a sequence record. The secondary structure line (SS) shows conserved base pairing with angle brackets and shading. The consensus line (Cons) highlights positions with >95% (uppercase) and >80% (lowercase) sequence conservation (R = A, G; Y = C, U). Within each sequence, shaded backgrounds represent base pairing predicted for that individual sequence. Putative elements related to transcription and translation initiation are shown in lowercase letters with shaded backgrounds: orange, promoter -35 and -10 boxes; green, start codons. **(c)** Consensus Structure and In-line Probing. Details for the consensus structure are given in Figure 1. Spontaneous cleavage product in the in-line probing gel are mapped onto nucleotides in the experimental RNA construct probed from *A. tumefaciens* with shaded circles. The boundaries of the region where this mapping was possible are demarked by filled triangles. For the in-line probing gels, the lanes are no reaction (NR), partial RNase T1 digestion (T1), partial alkali digestion (-OH), and spontaneous cleavage during a 40 hr incubation (-). The band labeled Pre is the full-length precursor RNA. Some G-specific RNase T1 cleavage products (G18, G20, G22, G23) expected in the T1 lane are missing for *serC*, presumably due to RNA structures that preclude enzyme action.

(a)

Abb	Organism	Accession/Start-End	Genes
Alpha Proteobacteria			
Atu	<i>Agrobacterium tumefaciens</i> str. C58	NC_003305.1/205774-205416	speF
Sme	<i>Sinorhizobium meliloti</i> 1021	NC_003047.1/3105445-3105086	speF
Bme	<i>Brucella melitensis</i> 16M chr II	NC_003318.1/1172829-1172472	speF
Brs	<i>Brucella suis</i> 1330 chr II	NC_004311.1/97128-97485	speF
Mlo	<i>Mesorhizobium loti</i> MAFF303099	NC_002678.1/2390868-2390511	speF
Rpa	<i>Rhodopseudomonas palustris</i> CGA009	NC_005296.1/966567-966102	speF
Bja	<i>Bradyrhizobium japonicum</i> USDA 110	NC_004463.1/8508267-8508777	speF
Bhe	<i>Bartonella henselae</i> str. Houston-1	NC_005956.1/1422014-1421658	speF
Bqu	<i>Bartonella quintana</i>	NC_005955.1/1185319-1184964	speF
Mes	<i>Mesorhizobium</i> sp. BNC1	NZ_AAED01000003.1/4802-5155	speF
Mma	<i>Magnetospirillum magnetotacticum</i> MS-1	NZ_AAAP01003568.1/1-285	speF
Gene Description			
speF	Ornithine decarboxylase		

(b)

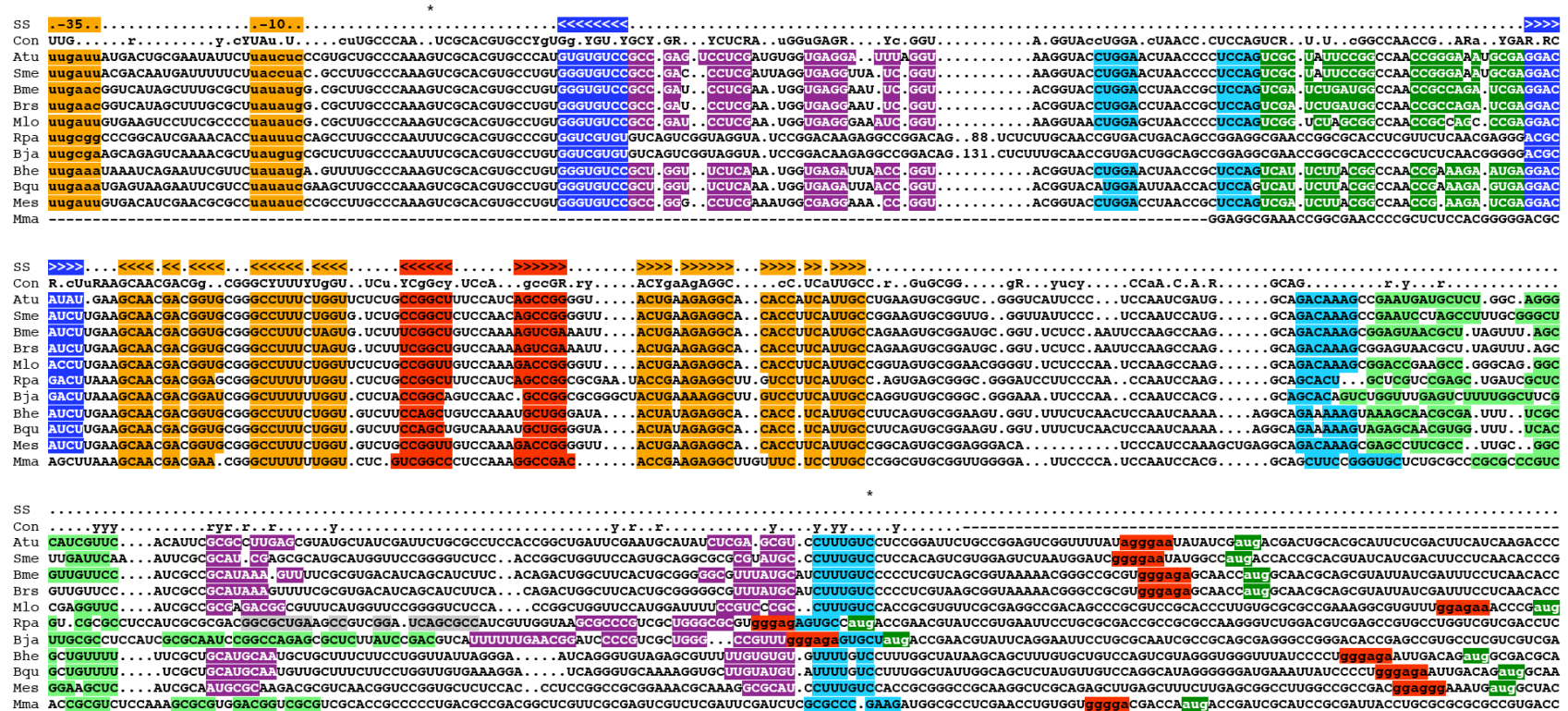
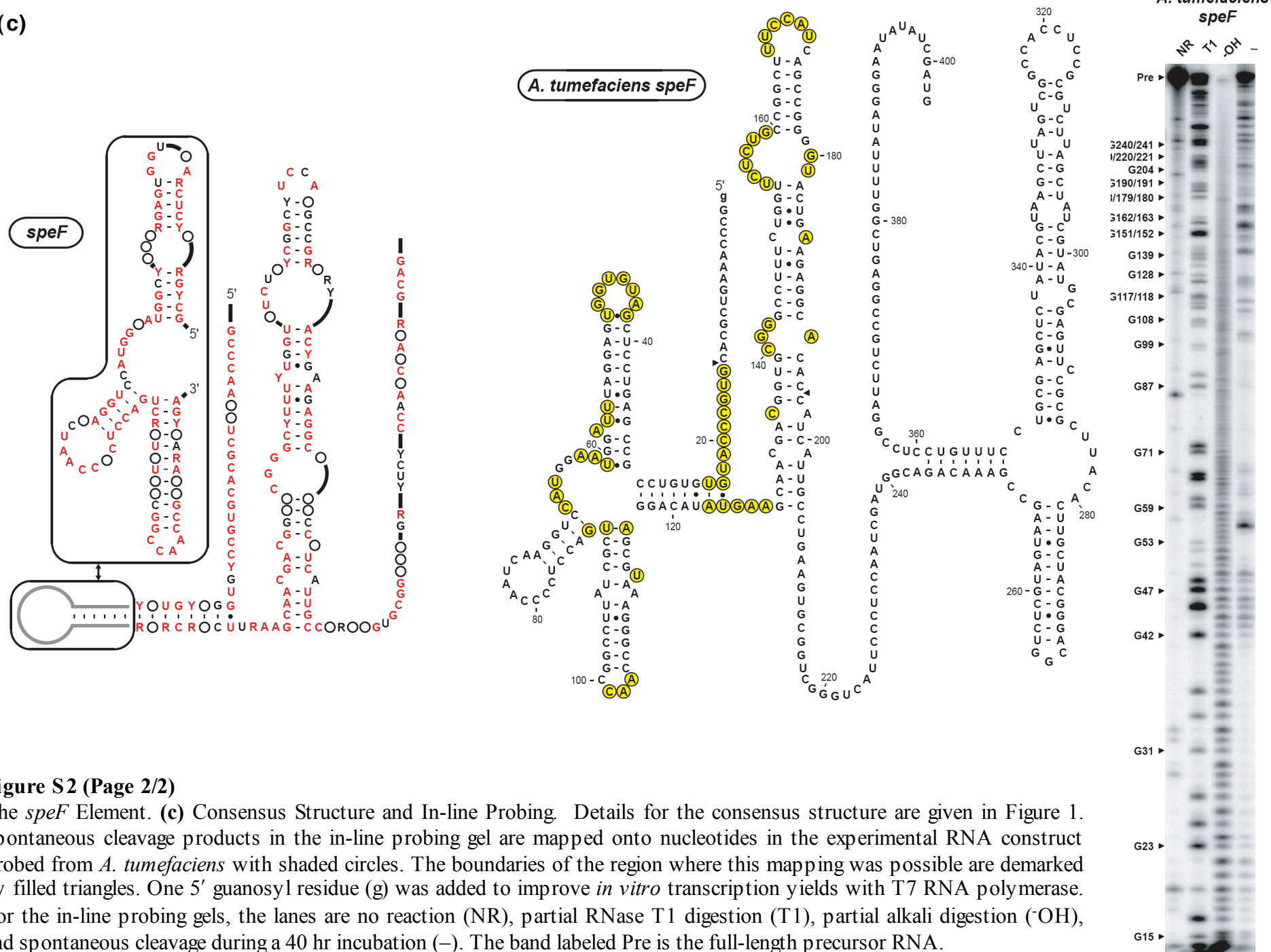


Figure S2 (Page 1/2)

The *speF* Element. (a) Species Distribution. Accession numbers refer to GenBank nucleotide records. Functional assignments are provided for the downstream genes that are likely to be regulated if the element is located in the 5' UTR of an mRNA. (b) Sequence Alignment. Labels correspond to abbreviations in (a). Periods represent gaps and dashes indicate the end of a sequence record. The secondary structure line (SS) shows conserved base pairing with angle brackets and shading. The consensus line (Cons) highlights positions with >95% (uppercase) and >80% (lowercase) sequence conservation (R = A, G; Y = C, U). Within each sequence, shaded backgrounds represent base pairing predicted for that individual sequence. Putative elements related to transcription and translation initiation are shown in lowercase letters with shaded backgrounds: orange, promoter -35 and -10 boxes; red, ribosome binding sites; green, start codons.

(c)



(b)

[illegible]

The *subB* Element. **(a)** Species Distribution. Accession numbers refer to GenBank nucleotide records. Relative directions and general functional assignments according to the COG database are provided for the adjacent upstream and downstream genes. **(b)** Sequence Alignment. Labels correspond to abbreviations in (a). Periods represent gaps and dashes indicate the end of a sequence record. The secondary structure line (SS) shows conserved base pairing with angle brackets and shading. The consensus line (Cons) highlights positions with >95% (uppercase) and >80% (lowercase) sequence conservation (R = A, G; Y = C, U). Within each sequence, shaded backgrounds represent base pairing predicted for that individual sequence. Putative promoter -35 and -10 boxes are shown as lowercase letters with orange backgrounds.

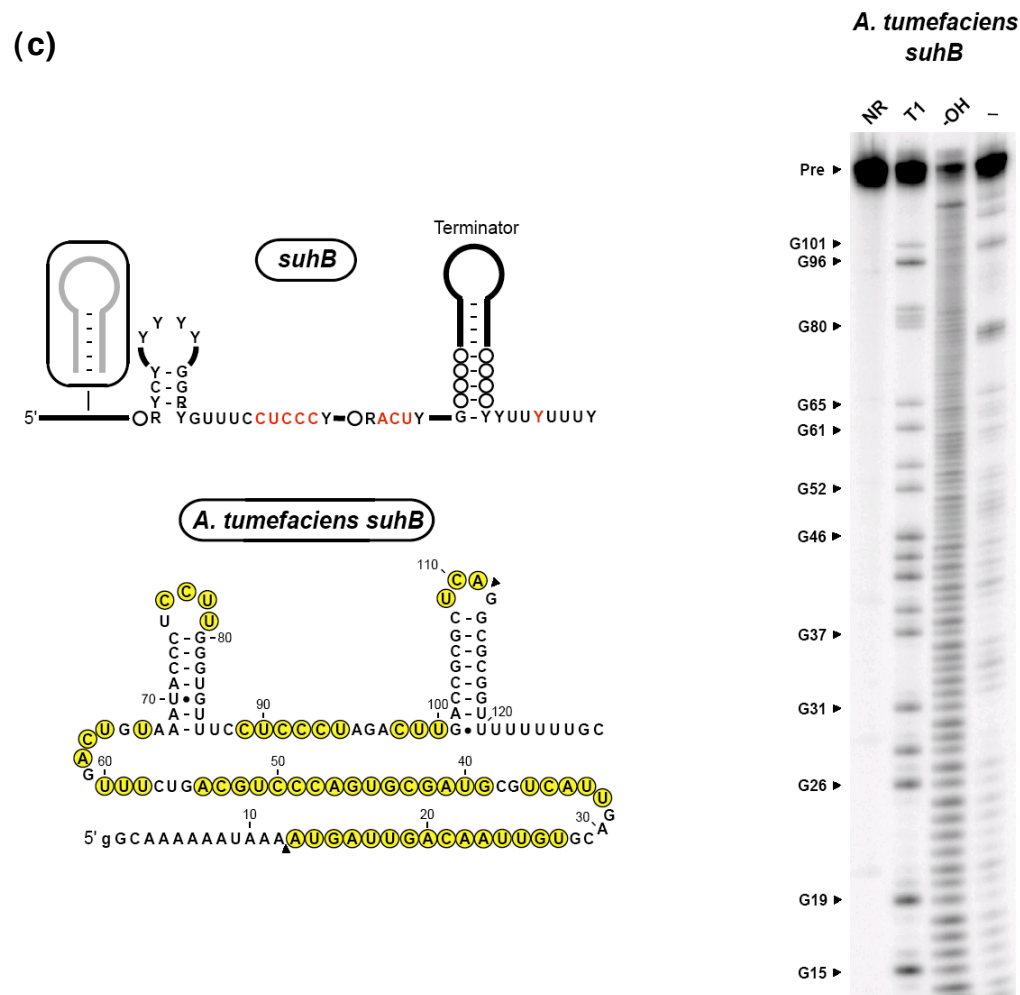


Figure S3 (Page 2/2)

The *suhB* Element. (c) Consensus Structure and In-line Probing. Details for the consensus structure are given in Figure 1. Spontaneous cleavage products in the in-line probing gel are mapped onto nucleotides in the experimental RNA construct probed from *A. tumefaciens* (Atu2) with shaded circles. The boundaries of the region where this mapping was possible are demarked by filled triangles. One 5' guanosyl residue (g) was added to improve *in vitro* transcription yields with T7 RNA polymerase. Note that the initial stem appears to be extended by three closing base pairs compared to the consensus structure in this particular sequence and, as is often seen in intrinsic transcription terminators, two uridines of the final polyU run seem to pair with nucleotides directly 5' of the G-C rich stem. For the in-line probing gels, the lanes are no reaction (NR), partial RNase T1 digestion (T1), partial alkali digestion ($\cdot$ OH), and spontaneous cleavage during a 40 hr incubation (–). The band labeled Pre is the full-length precursor RNA.

(a)

Abb	Organism	Accession/Position	Genes
Alpha Proteobacteria			
Atu1	<i>Agrobacterium tumefaciens</i> C58 chr circular	NC_003304.1/429424-429378	COG0626
Atu2	<i>Agrobacterium tumefaciens</i> C58 chr circular	NC_003304.1/2703363-2703434	COG1897
Bja1	<i>Bradyrhizobium japonicum</i> USDA 110	NC_004463.1/1219858-1219690	COG0626
Bja2	<i>Bradyrhizobium japonicum</i> USDA 110	NC_004463.1/1515006-1515379	COG2021 COG0500
Bja3	<i>Bradyrhizobium japonicum</i> USDA 110	NC_004463.1/7516089-7516237	COG0639
Bme	<i>Brucella melitensis</i> 16M chr I	NC_003317.1/1664142-1664216	COG0626
Brs	<i>Brucella suis</i> 1330 chr I	NC_004310.1/318415-317990	COG0626
Mma1	<i>Magnetospirillum magnetotacticum</i> MS-1	NZ_AAAP01003855.1/12513-12446	COG0192 COG0220
Mma2	<i>Magnetospirillum magnetotacticum</i> MS-1	NZ_AAAP01003864.1/43184-43111	COG0626
Mlo1	<i>Mesorhizobium loti</i> MAFF303099	NC_002678.1/2853515-2853846	COG2021 COG0500
Mlo2	<i>Mesorhizobium loti</i> MAFF303099	NC_002678.1/4543368-4543296	COG0626
Mes1	<i>Mesorhizobium</i> BNC1	NZ_AAED01000001.1/195260-195332	COG1897
Mes2	<i>Mesorhizobium</i> BNC1	NZ_AAED01000008.1/69545-69618	COG0626
Rsp	<i>Rhodobacter sphaeroides</i> 2.4.1	NZ_AAAE01000150.1/27825-27754	COG0626
Rpa1	<i>Rhodopseudomonas palustris</i> CGA009	NC_005296.1/5009663-5009589	COG2021 COG0500
Rpa2	<i>Rhodopseudomonas palustris</i> CGA009	NC_005296.1/5380828-5380777	COG0626
Rru	<i>Rhodospirillum rubrum</i>	NZ_AAAG02000001.1/970168-970115	COG0626
Sme1	<i>Sinorhizobium meliloti</i> 1021	NC_003047.1/580264-580068	COG0626
Sme2	<i>Sinorhizobium meliloti</i> 1021	NC_003047.1/3461576-34618381	COG1897
Beta Proteobacteria			
Bbr	<i>Bordetella bronchiseptica</i> RB50	NC_002927.3/4852879-4852804	COG2021 COG0500
Bpa	<i>Bordetella parapertussis</i> 12822	NC_002928.3/4417692-4417617	COG2021 COG0500
Gamma Proteobacteria			
Cbu	<i>Coxiella burnetii</i> RSA 493	NC_002971.2/1955247-1955190	COG0620 COG0685
Bacteroides			
Bth	<i>Bacteroides thetaiotaomicron</i> VPI-5482	NC_004663.1/4108979-4109039	COG0192
Pgi	<i>Porphyromonas gingivalis</i> W83	NC_002950.2/1994581-1994639	COG0192
COG	Gene	Description	
COG2021	metX	Homoserine acetyltransferase	
COG0646	meth_1	Methionine synthase I (cobalamin-dependent), methyltransferase domain	
COG1410	meth_2	Methionine synthase I, cobalamin-binding domain	
COG0626	metC	Cystathionine beta-lyases/cystathionine gamma-synthases	
COG1897	metA	Homoserine trans-succinylase	
COG0500	bioC	SAM-dependent methyltransferases	
COG2873	-	O-acetylhomoserine sulfhydrylase	
COG0192	metK	S-adenosylmethionine synthetase [metK]	
COG0155	cysI	Sulfite reductase, beta subunit (hemoprotein)	
COG0220	-	Predicted S-adenosylmethionine-dependent methyltransferase	
COG0639	apaH	Diadenosine tetraphosphatase and related serine/threonine protein phosphatases	
COG0620	metE	Methionine synthase II (cobalamin-independent)	
COG0685	metF	5,10-methylenetetrahydrofolate reductase	

Abb	Organism	Accession/Position	Genes
Environmental Sequences			
Env01	IBEA_CTG_2058111	AACY01000458.1/1106-1037	COG2021
Env02	IBEA_CTG_2083395	AACY01001152.1/1304-1233	COG0646 COG1410
Env03	IBEA_CTG_2069629	AACY01013609.1/1550-1623	COG2873
Env04	IBEA_CTG_2079783	AACY01015393.1/1522-1452	COG0646 COG1410
Env05	IBEA_CTG_2092782	AACY01015394.1/1298-1227	COG0646 COG1410
Env06	IBEA_CTG_1978138	AACY01019662.1/986-1056	
Env07	IBEA_CTG_2068602	AACY01020820.1/545-476	COG0155
Env08	IBEA_CTG_2043976	AACY01037785.1/1397-1325	COG2873
Env09	IBEA_CTG_2153715	AACY01038093.1/573-642	COG2021
Env10	IBEA_CTG_2147227	AACY01048873.1/713-779	COG2021
Env11	IBEA_CTG_2155094	AACY01075589.1/2322-2395	COG2873
Env12	IBEA_CTG_1961583	AACY01076673.1/123-192	COG2021
Env13	IBEA_CTG_2151957	AACY01076967.1/2586-2515	COG2021
Env14	IBEA_CTG_2012917	AACY01078968.1/460-390	COG2021
Env15	IBEA_CTG_2157446	AACY01080761.1/1069-999	COG2873
Env16	IBEA_CTG_2159957	AACY01095680.1/1999-1930	COG2021
Env17	IBEA_CTG_2149224	AACY01101458.1/5920-5851	COG2021
Env18	IBEA_CTG_2096709	AACY01103076.1/1423-1354	COG2021
Env19	IBEA_CTG_2150356	AACY01108759.1/4012-3939	COG2873
Env20	IBEA_CTG_2118022	AACY01120549.1/929-862	COG2021
Env21	IBEA_CTG_1956900	AACY01121489.1/451-524	COG2873
Env22	IBEA_CTG_2027402	AACY01135886.1/285-216	COG2021
Env23	IBEA_CTG_2037958	AACY01137881.1/882-811	COG2873
Env24	IBEA_CTG_2064653	AACY01143447.1/813-880	COG0646
Env25	IBEA_CTG_SKAQM84TR	AACY01181281.1/445-378	COG2021
Env26	IBEA_CTG_SKATY55TR	AACY01185668.1/363-291	COG0626
Env27	IBEA_CTG_SKBKZ58TF	AACY01206430.1/333-401	COG2873
Env28	IBEA_CTG_SKBTK336TR	AACY01216447.1/333-263	COG2021
Env29	IBEA_CTG_SLAOG27TF	AACY01230350.1/529-460	COG2021
Env30	IBEA_CTG_SLBMT82TF	AACY01253705.1/36-105	COG2021
Env31	IBEA_CTG_SSNF95TR	AACY01273405.1/372-439	COG2021
Env32	IBEA_CTG_SSBBN24TR	AACY01293033.1/913-844	COG2021
Env33	IBEA_CTG_SSBFE56TR	AACY01303516.1/371-301	COG2021
Env34	IBEA_CTG_SSBFL24TR	AACY01304150.1/460-391	COG0646
Env35	IBEA_CTG_SSBH615TF	AACY01309385.1/703-634	COG2021
Env36	IBEA_CTG_SSBTD84TR	AACY01348042.1/440-510	COG2021
Env37	IBEA_CTG_SSBUT42TR	AACY01352090.1/431-373	COG2021
Env38	IBEA_CTG_SSBW488TR	AACY01356025.1/783-713	COG0646
Env39	IBEA_CTG_SSBWL94TF	AACY01357550.1/484-553	COG2021
Env40	IBEA_CTG_SXAA622TF	AACY01372079.1/300-369	COG2021
Env41	IBEA_CTG_SZAFX30TR	AACY01401139.1/470-401	COG2021
Env42	IBEA_CTG_SZAIM36TR	AACY01409674.1/455-524	COG2021
Env43	IBEA_CTG_SZAT770TR	AACY01441579.1/352-283	COG2021
Env44	IBEA_CTG_UAAS64TR	AACY01453671.1/550-479	COG2021
Env45	IBEA_CTG_UAALM53TF	AACY01485766.1/38-109	COG0646
Env46	IBEA_CTG_UAASF79TR	AACY01494096.1/183-249	COG0626
Env47	IBEA_CTG_UAARI50TF	AACY01503107.1/98-169	COG0646
Env48	IBEA_CTG_UBAAV22TR	AACY01536804.1/768-697	COG0646
Env49	IBEA_CTG_UBAYI05TR	AACY01607341.1/466-537	COG0646
Env50	IBEA_CTG_UDASE14TF	AACY01690877.1/292-222	COG2021
Env51	IBEA_CTG_UDAT846TF	AACY01693958.1/438-508	COG0646
Env52	AMC_Cont262	AADL01000262.1/158-228	

Figure S5 (Page 1/3)

The α -Proteobacterial SAM Riboswitch. **(a)** Species Distribution. Accession numbers refer to GenBank nucleotide records. Functional assignments are provided for the downstream genes that are likely to be regulated if the element is located in the 5' UTR of an mRNA. **(b)** Sequence Alignment. Labels correspond to abbreviations in (a). Periods represent gaps and dashes indicate the end of a sequence record. The secondary structure line (SS) shows conserved base pairing with angle brackets and shading. The consensus line (Cons) highlights positions with >95% (uppercase) and >80% (lowercase) sequence conservation (R = A, G; Y = C, U). Within each sequence, shaded backgrounds represent base pairing predicted for that individual sequence. Putative elements related to transcription and translation initiation are shown in lowercase letters with shaded backgrounds: orange, promoter -35 and -10 boxes; green, start codons.

