www.sciencemag.org/cgi/content/full/science.1215063/DC1

Supporting Online Material for

Widespread Genetic Switches and Toxicity Resistance Proteins for Fluoride

Jenny L. Baker, Narasimhan Sudarsan, Zasha Weinberg, Adam Roth, Randy B. Stockbridge, Ronald R. Breaker*

*To whom correspondence should be addressed. E-mail: ronald.breaker@yale.edu

Published 22 December 2011 on *Science* Express
DOI: 10.1126/science.1215063

This PDF file includes:

Supporting Text

Rationale for testing RNA dinucleotides as possible riboswitch ligands.

The *crcB* motif RNAs are associated with an eclectic list of genes (fig. S1) that did not readily reveal the ligand for this riboswitch candidate. Since the only other riboswitch classes associated with many diverse gene types respond to the bacterial second messenger c-di-GMP (*23, 24*), we used synthetic linear RNA dinucleotides to assess binding of c-di-GMP (regulates a diversity of genes) and another bacterial second messenger called c-di-AMP (implicated in DNA repair control) (*25, 26*).

Gene control by fluoride riboswitches.

Miller assays conducted with ONPG reveal increasing levels of β-galactosidase expression with increasing amounts of fluoride added to liquid medium (Fig. 2B, fig. S8). A series of variant reporter constructs with mutations that alter highly-conserved nucleotides or that disrupt P3 stem formation fail to be activated by fluoride, whereas an altered P3 sequence that retains base pairing exhibits fluoride induction at a level equivalent to WT (fig. S11). These *in vivo* results parallel the fluoride-binding activity of mutants when tested *in vitro*. Moreover, a series of expression platform mutants were generated that yield reporter gene expression results that are consistent with a *B. cereus crcB* riboswitch mechanism involving transcription termination (fig. S9). In total, our findings confirm that fluoride triggers gene expression via distantly related riboswitch representatives from both Gram-positive and Gram-negative bacteria.

Intriguingly, *crcB* KO cells exhibit a detectable level of reporter expression even when the medium is not supplemented with fluoride. This could be due to the presence of low micromolar amounts of fluoride contaminating the reagents used to make the medium. In liquid medium, the WT strain requires approximately two orders of magnitude higher fluoride concentration to trigger reporter gene expression compared to the KO strain (Fig. 2B). These results suggest that *crcB*, the gene most commonly associated with fluoride riboswitches, codes for a protein that reduces the concentration of fluoride in cells, thus avoiding build-up to toxic levels.

Establishing the fluoride contaminant in synthetic RNA dinucleotides samples as the ligand.

Substantial conformational changes (fig. S2) were observed when various synthetic dinucleotide samples were incubated with an RNA construct derived from the *crcB* gene of *P. syringae* (78 Psy; Fig. 1B; Table S1). However, we eventually determined that the conformational changes were caused by fluoride ions. Tetra-*n*-butylammonium fluoride (TBAF) was used to remove the silyl-based protecting groups (*27*) from the ribose 2′ oxygens of the synthetic dinucleotides, and substantial amounts of fluoride remained in the commercial dinucleotide preparations. TBAF and other sources of free fluoride ions such as NaF, triethylamine trihydrofluoride, and potassium bifluoride all trigger identical conformational changes of 78 Psy *crcB* motif RNA (data not shown).

The in-line probing pattern (Fig. 1C) is consistent with the formation of an RNA/fluoride complex with an apparent dissociation constant (K_D) of 60 μM (Fig. 1D). Similar K_D values (~50 μM) for fluoride binding were obtained on analysis of three other *crcB* motif representatives from the bacteria *Thermotoga petrophila* and *Bacillus cereus*, and the archaeon *Thermoplasma volcanium* (fig. S3), which demonstrates that distal homologs of this RNA class also respond to fluoride.

Selective recognition of fluoride by natural RNAs.

Specificity for fluoride binding was established by testing other halogen anions (chloride, bromide and iodide) as well as various other inorganic and organic anions (fig. S5) via in-line probing. Strikingly, the 78 Psy RNA rejects these other anions even when they are present at extremely high concentrations. For example, the 78 Psy structure remains unchanged when exposed to chloride ion concentrations as high as 2.5 M, indicating that the RNA binds fluoride at least 40,000 fold more tightly. Similarly, hydroxide ions do not trigger RNA conformational changes up to a pH of 9.7 (50 μM), at which point the RNA secondary structure becomes denatured. We also determined that carbon monoxide and nitric oxide (as dissolved gases) are rejected by the RNA, despite their small sizes and electronic character. This remarkable selectivity is consistent with a biological role for the RNA as a fluoride sensor.

The effects of Ca^{2+} on fluoride binding to *crcB* RNAs.

Addition of Ca^{2+} in excess over fluoride at the start of an in-line probing assay diminishes RNA structural modulation, while a two-day pre-incubation of the ion mixture results in the

complete loss of modulation (fig. S4). This delay in Ca^{2+}-mediated inhibition of fluoride-dependent structure modulation is consistent with the slow rate of complex formation between calcium and fluoride ions (*28*).

Examining the importance of conserved RNA sequence and structural features for fluoride binding.

Numerous previous studies of riboswitches have proven that the distinct evolutionarily conserved structures and sequences are essential for forming ligand-binding pockets (*29-31*). Therefore, we reasoned that any mutations to the RNA that disrupt conserved structures or that alter any of its strictly-conserved nucleotides should perturb anion binding if fluoride indeed is the natural ligand. Alternatively, if the RNA has an as yet undiscovered natural ligand and only fortuitously binds fluoride, then some mutations that preclude binding of the natural ligand would still permit complex formation with fluoride.

We find that all mutations that alter the strictly-conserved sequence or secondary structure of the RNA likewise adversely affect fluoride binding. For example, mutations that disrupt the P1 stem cause complete loss of fluoride binding, while compensatory mutations that alter the wild-type (WT) sequence but that restore base-pairing also restore binding (fig. S6). Moreover, any mutation to one of the 11 most highly conserved nucleotides diminishes or eliminates fluoride binding, whereas mutations at non-conserved positions have no effect (fig. S7). These findings suggest that fluoride is not a ligand mimic, but rather is the biologically relevant target for *crcB* motif RNAs.

Materials and Methods

Reagents and dinucleotides.

Carbon monoxide was obtained from Tech Air. The nucleotide derivatives 2,8-dihydroxyadenine, 8-hydroxyguanine, 8-hydroxyguanosine and 8-hydroxydeoxyguanosine were purchased from Cayman Chemical, as were spermine NONOate and DAF-2. Formaldehyde (37%) was purchased from J.T. Baker. c-di-GMP, c-di-AMP and pApA were obtained from the BioLog Life Science Institute. ApG, GpG, GpA, pGpA and pApA containing the fluoride

impurity were purchased from Oligos Etc. All other reagents were obtained from Sigma-Aldrich unless otherwise stated.

DNA and RNA oligonucleotides and constructs.

Double-stranded DNAs were made by PCR using synthetic primers (Table S1) and template DNA from bacterial genomic samples or from synthetic DNAs. Synthetic oligonucleotides were obtained from Sigma-Aldrich or the W. M. Keck Foundation at Yale University. RNAs were prepared as described in more detail below by *in vitro* transcription of PCR products using T7 RNA polymerase, and purified by using denaturing (8 M urea) 6% PAGE. Plasmid constructs were prepared by ligating PCR products to specific plasmids that had been pre-digested with appropriate restriction enzymes or by using a TOPO TA Cloning Kit (Invitrogen). Plasmids carrying mutant riboswitch sequences were prepared from the WT construct by using a QuickChange XL Site-directed Mutagenesis kit (Stratagene). Additional details on gene constructs are provided below.

Preparation of RNA oligonucleotides.

To generate *in vitro* transcribed RNA, DNA templates were made by PCR amplification of wild-type plasmid or genomic DNA with the appropriate primers. All primers used to generate wild-type and mutant DNAs are found in Table S1. Transcription was carried out using bacteriophage T7 RNA polymerase (T7 RNAP) in 80 mM N-(2-hydroxyethyl)piperazine-N′-(2-ethanesulfonic acid) (HEPES) (pH 7.5 at 25˚C), 40 mM dithiothreitol (DTT), 24 mM $MgCl_2$, 2 mM spermidine, and 2 mM of each nucleoside 5′-triphosphate (NTP). RNA was purified using denaturing (8 M urea) 6% PAGE. The appropriately sized band was excised and the RNA was eluted from the crushed gel slice using 10 mM Tris-HCl (pH 7.5 at 23˚C), 200 mM NaCl, and 1 mM EDTA (pH 8.0). The RNA was subsequently precipitated with ethanol and pelleted by centrifugation.

To generate 5′ ^{32}P-labeled RNAs, the 5′-terminal phosphate was removed using alkaline phosphatase (Roche Diagnostics) following the manufacturer's protocol. The RNAs were then radiolabeled with [γ-^{32}P] ATP (Perkin Elmer) using T4 polynucleotide kinase (New England Biolabs). The 5′ ^{32}P-labeled RNAs were isolated and purified using the same protocol as described above.

In-line probing.

Conformational changes in the ligand binding domain of a riboswitch can be assessed by using a method called in-line probing (*10, 11*). This method exploits the fact that unstructured RNA regions typically undergo spontaneous cleavage of phosphoester linkages faster than structured regions.

Except where noted, in-line probing experiments were conducted as previously described (*11, 24*). Briefly, *in vitro* transcribed RNA were 5′ ^{32}P-radiolabeled and trace amounts (~1 nM) were incubated at room temperature for 48 hours (unless otherwise indicated) in the presence of 100 mM KCl, 50 mM Tris-HCl (pH 8.3 at 23˚C), and 20 mM $MgCl_2$. Cleavage products were separated by using denaturing 10% PAGE and imaged with a PhosphorImager (Molecular Dynamics). ImageQuant 5.1 was used to establish band intensities. Exceptions to the typical in-line probing methods are detailed in the following sections. Apparent K_D values were determined by varying the amount of ligand in a series of in-line probing reactions and determining the concentration required to cause 50% modulation of spontaneous cleavage. Values for normalized fraction of RNA cleaved at a given site were generated by setting the maximum amount of RNA cleaved among all the concentrations of fluoride tested to a value of 1 and setting the minimum amount of RNA cleaved among all the concentrations of fluoride tested to a value of 0.

In-line probing with calcium chloride.

To confirm that fluoride was the ligand causing RNA structure modulation, in-line probing reactions were conducted in the presence of calcium, which is known to form strong complexes with fluoride (*12*). Two stock solutions were made, one containing only 100 μM NaF, and the other containing 100 μM NaF and 1 mM $CaCl_2$. In-line probing reactions were conducted as usual with some reactions lacking additives (NR, no ligand) and others carrying sodium fluoride (F) or sodium fluoride plus calcium chloride (Ca/F). After two days, the same F or Ca/F stock solutions were used to set up a new set of in-line probing reactions. These additional assays allowed for pre-equilibration of the fluoride and calcium before adding it to the RNA/in-line probing buffer mix. The reaction products were analyzed as usual.

In-line probing at elevated pH.

To test for possible binding of hydroxide ion, the buffer used for in-line probing included 100 mM KCl, 20 mM $MgCl_2$, and 50 mM CHES (pH 9.0, 9.3, 9.7, or 10.0 at 23˚C). Reactions were incubated in this buffer at room temperature for 5 hours and cleavage products were analyzed as described above.

In-line probing with formaldehyde.

To test for possible binding of formaldehyde, HEPES was used in the in-line buffer instead of Tris (50 mM HEPES, pH 8.3 at 23˚C) to prevent reactions between Tris and formaldehyde. A 37% formaldehyde solution (J.T.Baker) was used to make 10X formaldehyde stocks for addition to each in-line probing reaction. Reactions and analyses were otherwise conducted as usual.

An indicator, Schiff reagent for aldehyde detection, was used to establish that a substantial amount of formaldehyde was left in solution after two days (minimal evaporation occurred). Mock in-line probing reactions containing varying concentrations of formaldehyde were prepared in duplicate. To one set, 1 μL of the Schiff reagent for aldehydes was added at time zero. The same was done to the other set after two days. The pink color was similar for the two sets, indicating that a similar amount of formaldehyde was present.

In-line probing with the gases CO and NO.

To test binding of CO gas with the 78 Psy RNA, a variation of the typical in-line probing reaction conditions at pH 9.7 was used. 5′ ^{32}P-labeled RNA was incubated at room temperature for three hours (the reduced time is possible due to the elevated pH) and analyzed as usual. A side-arm flask was capped with a rubber septum and the side-arm was covered with a KimWipe to allow pressure to be released from the system. Carbon monoxide was bubbled through deionized water for 30 minutes before adding this water to the in-line probing reaction. Uncapped 1.5 mL microfuge tubes containing the in-line probing reactions were placed in the side-arm flask. Carbon monoxide was streamed into the flask at 3 psi for 3 hours. In-line probing reaction products were analyzed as usual.

To test NO, spermine NONOate, a compound known to break down into nitric oxide at low pH, was used. Spermine NONOate (Cayman Chemical) was dissolved in 100 mM NaOH immediately before addition to an otherwise typical in-line probing reaction. Dilutions of the stock solution were made with 100 mM NaOH. In-line probing reactions were carried out using

the typical protocol, and the amount of NaOH delivered when adding spermine NONOate did not alter the pH substantially. The colorimetric indicator DAF-2 (Cayman Chemical) was used to monitor NO concentrations. Mock in-line probing reactions containing varying amounts of spermine NONOate were prepared in duplicate, and 1 μL of a DAF-2 solution (1.4 mM in DMSO) was added to each of the tubes in the dilution series either at time zero or after two days. For given concentrations of spermine NONOate, the extents of the color change in duplicate samples were similar, indicating that the levels of dissolved NO did not change significantly over the two day period.

Bacterial strains and culture conditions.

Bacillus subtilis 168 (*trpC*2), *Bacillus cereus* ATCC10987 and the *B. subtilis* integration vector pDG1661 were obtained from the Bacillus Genetic Stock Center (The Ohio State University). *E. coli* strain BW25113 and its various isogenic derivatives including BW25113 Δ*crcB* (kan^r) were obtained from the Coli Genetic Stock Center (Yale University). Reporter vector pRS414 was a gift from R.W. Simons (UCLA). When required, growth medium was supplemented with antibiotics at the following concentrations: carbenicillin, 100 μg mL^{-1}; kanamycin, 30 μg mL^{-1}; chloramphenicol, 5 μg mL^{-1}. To study growth at low pH, LBK medium with appropriate buffer was used as described previously (*32*).

Design of reporter gene constructs.

The sequence containing nucleotides -110 to +1 with respect to the translation start site of the *crcB* gene, encompassing the fluoride-sensing domain and the downstream expression platform, was amplified by PCR as an EcoRI-BamHI fragment from *B. cereus* (ATCC10987) genomic DNA and cloned into the vector pDG1661 to yield a transcriptional fusion with a promoterless *lacZ* gene. The amplified fragment also integrated (via primer) a previously characterized lysine promoter from *B. subtilis* (*33*) to drive expression (primers are listed in Table S1). The resulting construct was integrated into the *amyE* locus of *B. subtilis* as described previously (*33*). Mutations to examine riboswitch control of termination were made by using a three-step PCR (*34*).

The 5′ UTR of COG0038 (*eriC*^{F} homolog) was amplified from *P. syringae* DC3000 genomic DNA as an EcoRI-BamHI fragment and cloned into the translational reporter vector pRS414. Its

gene control function in *E. coli* was established using methods similar to those previously described (*34*). Specifically, the region encompassing nucleotides -248 to +24 with respect to the beginning of the *eriC*^{F} ORF was cloned as a translational fusion in pRS414, wherein the 8th codon of the ORF was fused in-frame to the 9th codon of the *lacZ* reporter gene. Mutations in the regulatory region were generated by a three-step PCR approach using methods described previously (*34*).

Expression analysis of fluoride riboswitch-reporter fusion constructs.

A single colony of *Bacillus subtilis* harboring the *crcB* riboswitch reporter construct was grown overnight in LB with chloramphenicol. The next day, it was subcultured by diluting to 1/200th using a defined glucose glutamate medium with chloramphenicol. The concentrations of added NaF in the medium varied from zero to 30 mM. β-galactosidase reporter gene expression was evaluated by adding X-gal (400 μg mL^{-1}).

E. coli strains BW25113 and its *crcB* KO derivative strain harboring the *eriC*^{F} riboswitch reporter construct were streaked on modified LB plates with carbenicillin (100 μg mL^{-1}) and X-gal (800 μg ml^{-1}). Plates included 0, 0.2 or 50 mM added NaF. Modified LB included 10 mM NaCl. Miller assays for reporter gene expression were conducted in 96-well format as described previously (*39*).

Complementation with *eriC*^{F} gene from *P. syringae*.

The *eriC*^{F} gene from *P. syringae* DC3000 was amplified from genomic DNA by PCR and cloned into the pCR2.1 vector (Invitrogen). The cloned fragment contained the full length *eriC*^{F} ORF, the 5′ UTR features described above, and the putative native promoter to drive expression.

Cell growth curves.

Amounts of cell growth of WT and *crcB* KO *E. coli* strains were measured overnight using a Bioscreen C instrument (Oy Growth Curves Ab Ltd, Finland). The amount of fluoride added to the media was varied to measure its effect on growth at two pH values.

For each bacterial strain, an overnight culture was set up of the desired bacteria the previous day. The majority of the growth curves were conducted in 350 μL LB medium that initially was not supplemented with salt. 2X LB without salt addition was prepared by combining 10 g Bacto

tryptone (pancreatic digest of casein), 5 g Bacto yeast extract, deionized water to 500 mL, and 2 mL 1 N sodium hydroxide (mixture subsequently autoclaved). The desired concentration of NaF or NaCl was subsequently added as designated for each culture after a 1:100 dilution of an overnight culture. Wells with NaF concentrations below 1 mM were supplemented with NaCl to a final concentration of 1 mM Na^+. Each condition was set up in duplicate or triplicate, but data from only one representative well is depicted. Every fifteen minutes the 100-well plate was shaken at medium intensity for 10 seconds and the $O.D._{600}$ was automatically measured for a period of 16 hours. The data was analyzed using Microsoft Excel.

Growth curves to compare pH conditions used 350 µL of LB (neutral pH) or LBK medium (pH 5.5), 3.5 µL of an overnight culture, and 3.5 µL of a 100X sodium fluoride or sodium chloride solution. Growth curves with empty or *eriC*^{F}–containing plasmids included 100 µg mL^{-1} carbenicillin added to the growth medium (2X LB), which was subsequently diluted two-fold to a final concentration of 50 µg mL^{-1}.

Anion flux experiments.

Protein expression, purification, and lipid vesicle reconstitution experiments used reagent grade chemicals obtained from Sigma-Aldrich or Fisher. Lipids were obtained from Avanti and detergents were obtained from Anatrace. The *E. coli* EriC protein was prepared as previously described (*35*). Similarly, a sequence encoding the *P. syringae eriC*^{F} gene was cloned and inserted into the pASK vector with an N-terminal hexahistidine tag (*36*). Protein was expressed in *E. coli* under the same growth conditions used for *E. coli* EriC, except that the induction time was shortened to 1.5 hours. Cells were lysed by sonication and extracted with 40 mM decyl maltoside (DM) for 2 hours at room temperature. After the extract was clarified by centrifugation (12,000 g, 45 minutes), supernatant was passed over cobalt affinity beads (1 mL resin per 1 L bacterial culture), washed with 100 mM NaCl, 20 mM imidazole, and eluted with 400 mM imidazole. After elution, sodium citrate (pH 5) was added to a final concentration of 80 mM. Protein was then purified on a Superdex 200 gel filtration column in 100 mM NaCl, 20 mM sodium citrate (pH 6), and 5 mM DM. $EriC^F$ elutes as a single sharp, symmetrical peak at a position expected for the homodimer. This peak was collected and added to reconstitution buffer containing 20 mg mL^{-1} *E. coli* polar lipid and 20 mM DM, to a final concentration of 10 µg protein mg^{-1} lipid. Detergent was removed by dialysis against the desired intraliposomal solution

at 4°C over 72 hours. Liposomes were stored in aliquots at -80°C until the day of use. Prior to functional assays, liposomes were extruded 21 times through a 400 nm membrane filter.

Anion efflux was monitored as a change in light scattering by liposomes monitored at 600 nm using a fluorimeter. In a typical experiment, liposome samples containing 10 mM sodium citrate (pH 6.5) and 300 mM of either KF or KCl were diluted 1:200 into 2 mL of a degassed isotonic solution containing 10 mM sodium citrate (pH 6.5) and 200 mM K_2SO_4, in a stirred quartz cuvette. The baseline was allowed to stabilize for 20 s before flux was initiated with the addition of 1 μM of the potassium ionophore valinomycin. In these experiments, K^+ and Cl^- or F^- flow down their gradients at a rate limited by ClC-mediated anion permeability. Water follows the ions across the lipid membrane, and the liposomes shrink and flatten, accompanied by a 5-10% increase in 90° light scattering (*37*). Separate control experiments showed that liposomes reconstituted without protein, and liposomes loaded with an impermeant ion, sulfate, do not produce an increase in light scattering upon valinomycin addition. Additional Cl^- efflux experiments monitored with a Cl^--specific electrode (*38*) showed that liposomes release all of the Cl^- along a time course consistent with that measured by the change in light scattering.

E. coli knock out strains and fluoride sensitivity.

Strains that lack the genes for chloride channels (Δ*clcA*, Δ*clcB*) and several other housekeeping genes (Δ*pgi*, Δ*pgm*, Δ*ppx*, Δ*ppk*) do not exhibit increased sensitivity to fluoride (data not shown). These observations suggest that deletion of the *crcB* gene does not cause general distress and indirect sensitivity to toxic agents, but rather its protein product has a direct role in alleviating fluoride toxicity.

Knock-out strains were obtained from the Coli Genetic Stock Center (Yale University). Δ*clcA*, Δ*clcB*, Δ*pgi*, Δ*pgm*, Δ*ppx* and Δ*ppk* were evaluated for fluoride sensitivity in LB with 50 mM NaF. The chloride channel knock-outs (Δ*clcA* and Δ*clcB*) are *eriC* homologs and were tested to determine whether they affect fluoride toxicity compared to the $eriC^F$ from *P. syringae* and the endogenous *crcB*. Neither of the *eriC* homolog KO strains exhibited reporter activity at low fluoride concentrations, unlike Δ*crcB*. Since fluoride is known to inhibit enolase, phosphogluco isomerase (Δ*pgi*) and phosphogluco mutase (Δ*pgm*) knockouts were tested to see if glycolytic intermediates in their respective metabolic pathways might have an effect on the reporter. Again, these KO strains were not unusually sensitive to high fluoride concentrations

and reporter gene expression was not triggered at low fluoride concentrations. Finally, since overproduction of CrcB has been associated with chromosome condensation (*22*), knock outs of two genes known to control chromosome condensation , *ppx* (exopolyphosphatase) and *ppk* (polyphosphate kinase), were tested. Neither KO strain exhibits reporter gene activation at low fluoride concentrations.

Nucleic acid sequence databases.

Analysis was performed on sequences from the bacterial and archaeal subsets of version 44 of the RefSeq database (*40*), as well as environmental metagenome sequences from acid mine drainage (*41*), air (*42*), Global Ocean Survey scaffolds (*43, 44*), ground water (*45*), gutless sea worms (*46*), hot springs (*47*), human gut (*48-50*), hydrothermal vent (*51*), kimchi (*52*), lake sediment (*53*), marine sequences (*54, 55*), mouse gut (*56*), sludge (*57*), soil (*58*), termite hindgut (*59*), wallaby gut (*60*), whalefall (*58*) and others (*61*). These and other environmental sequences were identified and downloaded from the GenBank (*62*), IMG/M (*63*), CAMERA (*63*) or MG-RAST (*64*) web sites, or from sources specific to the publication (*50*). Where available, protein-coding genes were extracted from the annotation downloaded with the sequences, or predicted using MetaGene (*65*), as described previously (*8*). Conserved protein domains were predicted using the Conserved Domain Database version 2.25 (*66*). When two predicted domains overlapped, the prediction with the better E-value was used, except that "specific hits" (*66*) were always selected over other hits regardless of relative E-values.

Homology search for fluoride riboswitches.

Nearly 2200 representatives of fluoride riboswitches have been identified wherein each adopts either a two- (69%) or three-stem (31%) junction carrying more than a dozen highly-conserved nucleotides (Fig. 1A). Homology searches were conducted using the cmsearch program from the Infernal software package (*67*). A multiple-sequence alignment of the RNAs was edited using the RALEE (*68*) extension to the Emacs text editor. Searches were performed iteratively, initially using the previously established alignment of the *crcB* RNA motif (*8*). A previously described search strategy was applied (*69*) wherein some searches were performed on sequences within 2 kb upstream of genes commonly associated with fluoride riboswitches.

Because the high diversity of sequences within the relatively small RNA might blur sequence conservation in related sequences, we performed some additional search strategies. First, fluoride riboswitches present in the RefSeq database were partitioned by their phyla according to RefSeq, and these alignment partitions were used as separate queries with the cmsearch program. Second, fluoride riboswitches were partitioned into separate alignments using the CD-HIT program (*70*) with threshold 0.6, and searched separately. Third, all known fluoride riboswitches were used as queries in a nucleotide BLAST (*71*) search. These three search strategies increased the number of known fluoride riboswitches by less than 10%, but did help to fully uncover previously undetected fluoride riboswitches in the genus *Streptomyces*. The BLAST searches helped to find truncated fluoride riboswitches, or sequences similar to borderline fluoride riboswitch matches. In some cases these latter sequences matched the consensus very poorly and revealed that the borderline matches were probably false positives. The RNA consensus diagram was initially drawn using R2R (*72*) based on the alignment resulting from all searches. The alignment of fluoride riboswitches is available upon request.

Genes regulated by fluoride riboswitches.

One or more consecutive genes were predicted to reside in an operon regulated by a fluoride riboswitch when the first gene is found within 700 base pairs of the consensus fluoride binding domain, all genes are transcribed in the same direction as this domain and the maximum distance between the genes is 100 base pairs. Protein domains predicted to be encoded by these genes were identified as above. Genes encoded on the opposite strand to the riboswitch were never counted, even though there is precedent for such genes to be regulated by riboswitches (*73*).

In fig. S1, we used the numbers of operons containing one or more of genes encoding domains in the relevant category; this definition avoids overcounting categories present in multiple genes within the same operon (*e.g.*, formate hydrogen lyase). When many closely related organisms are sequenced, genes regulated in these organisms can be overrepresented. For example, 31 inositol monophosphatase genes are regulated by fluoride riboswitches, but these genes are present only in strains of *Enterococcus faecalis* as the third gene in predicted operons, and thus are not likely to be significant. To reduce the impact of these effects, operons were weighted based on the uniqueness of the fluoride riboswitch predicted to regulate them. These weights were derived by applying the GSC algorithm (*74*) as implemented in the Infernal

software package (*67*) to the alignment of fluoride riboswitches. Gene categories were ignored when they occurred in fewer than 1% of operons, using weighted frequencies. For the small gray pie chart in the upper, left corner of fig. S1, weighted frequencies were also used, but genes were counted individually (*i.e.*, not considered in the context of operons).

Among genes predicted to be regulated by fluoride riboswitches, we noticed some correlations. "Ni/Fe-hydrogenase III" and "radical SAM" domains are always encoded in operons that also encode formate hydrogen lyase genes, so we grouped these proteins together. We also grouped the "class II terpene cyclase-like" domain (Conserved Domain Database accession cd00688) under the "COG1689" category, since all terpene cyclase-like proteins regulated by the fluoride riboswitch also matched the COG1689 domain, but the reverse was not always true. We noted that the presence of a gene encoding COG1993 is highly correlated with the occurrence of another gene in the operon encoding CrcB or $EriC^F$. When weighted by the GSC algorithm as above, 80% of COG1993 genes are predicted as cotranscribed with a *crcB* gene, and 8% of operons with COG1993 genes also have $eriC^F$ genes. However, we continued to treat COG1993 separately from other domains, because it is correlated with multiple rather than a single other domain. Similarly, the "haloacid dehalogenase" and "universal stress protein" domains were also highly correlated with other domains, with 10% or fewer of the corresponding genes appearing in an operon in the absence of genes from one of the other 11 most common categories.

Each category in the pie chart was defined as the union of certain accessions in the Conserved Domain Database, as below:

Category name	Conserved Domain Database accessions
COG1689	cd00688, COG1689
COG1993	COG1993, pfam02641
CrcB	COG0239, PRK14195, PRK14196, PRK14202, PRK14215, PRK14228, TIGR00494
Enolase	cd03313, pfam03952, PTZ00081
EriC	cd00400, cd03682
Fe-S cluster (E.g. NifU)	COG0822, COG1142, TIGR03652
Formate hydrogen lyase (Includes Ni/Fe-hydrogenase III and Radical SAM)	cd01335, COG0650, COG0651, COG1143, COG3260, COG3261, COG3262, COG4237, pfam00329, PRK06459
Haloacid dehalogenase	cd01427, pfam08282

(Full CDD name: "haloacid dehalogenase-like hydrolase")	
Major Facilitator Superfamily	cd06174, pfam07690
MutS (Note: might only correspond to the ATPase domain of MutS; this is the only domain matched when proteins are searched against Pfam)	cd03243, cd03283, COG0249
Na^+/H^+ antiporter (Note: includes K^+/H^+ antiporters)	COG0025, COG0475, pfam00999, TIGR00932
Universal Stress Protein	cd00293, pfam00582

A detailed listing of fluoride riboswitches and their associated genes within RefSeq version 32 is available from a previous report (8).

Rho-independent transcription terminators were predicted using RNIE (*75*).

Analysis of EriC proteins.

Initial analysis was conducted using the previously published alignment of *crcB* RNAs (*8*) restricted to sequences in RefSeq version 32. Genes predicted to be regulated by these fluoride riboswitches were extracted as above. Those genes predicted to belong to the EriC model in version 2.08 of the Conserved Domain Database were identified, and their protein sequences were retrieved. This was called the "positive set". Other *eriC* genes found within the same organisms as those in the positive set were extracted with their protein sequences; these formed the "negative set". All protein sequences were aligned to the Pfam 23.0 model PF00654 (*16*) using the HMMER version 3.0 (http://hmmer.janelia.org) program hmmsearch. We manually compared the positive and negative set proteins, and observed numerous differences, the most noticeable of which corresponded to certain residues in the previously established selectivity filter (*18*). One fluoride riboswitch-associated protein (out of 103 proteins) did not have the typical sequence features and was removed from the positive set. Also, one protein in the negative set out of 263 was similar in sequence to the positive set and was removed.

We then used the positive set proteins to search the proteins in RefSeq version 44 and metagenome sequences. The HMM from the Pfam version 24.0 model PF00654 was also used to search this protein database using the hmmsearch program with the score cutoff encoded by Pfam, and proteins scoring below this cutoff were discarded from further analysis. To classify proteins that are similar to the positive set, whether or not they are associated with an RNA, we

subtracted the hmmsearch score obtained in the PF00654 search from the hmmsearch score obtained in the positive set search. This defines the "heuristic subfamily score". Proteins with scores greater than zero favor classification as putative fluoride channels. (However, the heuristic subfamily score is not a true LOD score because the null model probabilities in each hmmsearch score are not necessarily equal.) We observed that virtually all EriC proteins associated with fluoride riboswitches have the sequences GREGT or GREGV and GXVXP in their selectivity filter, where X represents any amino acid. (Among proteins whose genes are in completely sequenced genomes, 68 out of 70 have these features.) We removed putative fluoride channels that lacked these sequence features or that appeared to us to be truncated on their N- or C-termini. These defined a higher confidence set of putative fluoride channels. Sequences were automatically aligned to the earlier model using hmmsearch.

To generate a phylogenetic tree, we began with the alignment of proteins matching the PF00654 model of EriC generated by hmmsearch (see above). Only proteins in completed genome sequences were considered further, since prediction of riboswitch regulation might be difficult in incomplete or environmental sequence. However, we did the phylogenetic analysis described below with all proteins clustered to 60% identity, and obtained similar results.

Protein sequences were eliminated if they had more than 15 consecutive gaps on their N-terminus or 6 consecutive gaps on their C-terminus. These numbers were chosen empirically based on which sequences we judged to be truncated, and the numbers might need to change if a different set of proteins were searched. Proteins can be truncated because they come from environmental sequence fragments, because of incorrect start codon predictions or because they do not fully match the computer's model. We clustered these proteins into sets that were 80% identical in sequence using CD-HIT (*70*), and took one representative from each cluster. In the resulting alignment, columns that were more than 50% gaps were eliminated. Then we inferred a phylogenetic tree and branch significance using version "phyml-20100914" of PhyML (*76*) with the following command line:

phyml -i alignment.phylip --rand_start --n_rand_starts 10 -d aa -f e -t e -v e -a e -s SPR -o tlr --no_memory_check -b -4

The predicted tree was drawn using the Interactive Tree of Life web site (*76*). The branching confidence values are based on the likelihood ratio tests (*77*) specified by the "-b -4" command-line option in PhyML.

In fig. S14, we predicted selectivity filter residues in EriC^F proteins based on the alignments produced by the hmmsearch program against the Pfam EriC model. However, we shifted the alignment of the five residues aligned to the sequence GSGIP in chloride channels by 2 positions. This shift resulted in the aligning of five residues that were more highly conserved among EriC^F proteins than those in the original alignment, and therefore more consistent with the likely constraints of a selectivity filter. Also, the Y445 position (*18*) is not part of the Pfam model, but a longer alignment of EriC proteins led to the alignment of a tyrosine in EriC^F proteins to this position. Biochemical or structural studies will be necessary to determine the correct alignment of selectivity filters in the EriC^F proteins. However, any of the possible alignments lead to the conclusion that EriC^F proteins have several distinct amino acids in the selectivity filter positions. An alignment of EriC^F proteins is available upon request.

Distribution of CrcB and EriC^F proteins and fluoride riboswitches

A previously established phylogenetic tree of the three domains (*78*) was downloaded from the iTOL web site (*79*). To analyze CrcB proteins, species names on the tree were compared to those in the Pfam database (*16*). Strain designations were ignored because they were inconsistently named between the two datasets. Tree species that did not have a match in the Pfam entry for ribosomal protein S12 (Pfam accession PF00164) were assumed to be absent from or un-mappable to the underlying Pfam database, and were removed from the final tree image. Species matching a CrcB protein (Pfam accession PF02537) were marked appropriately.

EriC^F proteins and fluoride riboswitches were predicted as described above within the RefSeq database. Species names were extracted from the RefSeq database, and matched with species on the phylogenetic tree, again ignoring strain designations. Tree species that did not match any species in the RefSeq database were removed from the tree.

These data were used to annotate the tree of life using the iTOL web site. The resulting tree image was downloaded from the web site, and edited manually. Branches were removed when their lengths were too short to be distinguished in the final figure.

Hydrogen bonding of fluoride ions and its implication in RNA-fluoride binding

Due to their strong electronegativity, fluoride ions have been known to be potent hydrogen bond accepters (*80, 81*). This is supported by the fact that the F-H-F^- bond is the strongest hydrogen bond in existence (*81*). Additionally, it has been noted that a key difference between fluoride and chloride ions is the stronger hydrogen bonding capabilities of F^- (*82*). These observations suggest that hydrogen bonding may play an important role in the ability of the RNA aptamer to selectively sense fluoride over chloride and other negatively charged species.

Table S1. Sequences of DNA primers used in this study. The T7 RNA polymerase promoter is italicized, the lysine promoter is bolded, and any restriction sites are underlined.

	DNA sequence (5′ to 3′)	Use
1	*taatacgactcactata*ggatcggcgcattggagatggcattcc	Amplification of 78 Psy from *P. syringae* DC3000 genomic DNA
2	caggtttctgtaggcatcatc	Amplification of 78 Psy from *P. syringae* DC3000 genomic DNA
3	*taatacgactcactata*ggatcggcgcattggag	In-line probing, WT 78 Psy
4	caggtttctgtagggttcatcagctgctacggg	In-line probing, M1 78 Psy
5	cggcgcattggagatgggtttcctccattaacaaaccg	In-line probing, M2 78 Psy; Two-step PCR, Psy-pCR2.1, M2
6	cggtttgttaatggaggaaacccatctccaatgcgccg	In-line probing, M2 78 Psy; Two-step PCR, PsyM1-pCR2.1, M2
7	*taatacgactcactata*ggactaacacccgtggaatggcgtttc	In-line probing, WT 74 Tvo
8	cacccgtggaatggcgtttccactcgtacaaaccgccagctccag	In-line probing, WT 74 Tvo
9	tttcaatcaggcgtcatcagtcagctggagctggcggtttgtacg	In-line probing, WT 74 Tvo
10	*taatacgactcactata*ggataacccaacgggcgatgaggcccg cccaa	In-line probing, WT 62 Tpe
11	cagtagaggccatcagccctcttcagggcagtttgggcgggcctc atcg	In-line probing, WT 62 Tpe
12	*taatacgactcactata*ggatcggcgcattggagccggcattcct ccattaaca	In-line probing, M3 78 Psy; Two-step PCR, Psy-pCR2.1, M3
13	caggtttctgtgggcatcatcagctgctacgg	In-line probing, M4 78 Psy
14	caggtttctgtacgcatcatcagctgctacggg	In-line probing, M5 78 Psy
15	caggtttctgtaggcatcatctgctgctacgggcgcagcggttt	In-line probing, M6 78 Psy; Two-step PCR, Psy-pCR2.1, M6
16	*taatacgactcactata*ggatcggcgcattgaagatggcattcctt cattaacaaaccgctgcgcc	In-line probing, M7 78 Psy
17	gatggcattcctccattaacataccgctgcgcccgtagcagc	In-line probing, M8 78 Psy
18	gctgctacgggcgcagcggtatgttaatggaggaatgccatc	In-line probing, M8 78 Psy

19	ggcattcctccattaacaatccgctgcgcccgtagcagctg	In-line probing, M9 78 Psy
20	cagctgctacgggcgcagcggattgttaatggaggaatgcc	In-line probing, M9 78 Psy
21	caggtttctgtaggcatcagcagctgctacgggcgcagcg	In-line probing, M10 78 Psy
22	caggtttctgtaggcatcttcagctgctacgggcgcagc	In-line probing, M11 78 Psy
23	cattggagatggcattcctccactaacaaaccgctgcgcccg	In-line probing, M12 78 Psy
24	cgggcgcagcggtttgttagtggaggaatgccatctccaatg	In-line probing, M12 78 Psy
25	ggagatggcattcctccattcacaaaccgctgcgcccgtag	In-line probing, M13 78 Psy
26	ctacgggcgcagcggtttgtgaatggaggaatgccatctcc	In-line probing, M13 78 Psy
27	gatggcattcctccattaactaaccgctgcgcccgtagcag	In-line probing, M14 78 Psy
28	ctgctacgggcgcagcggttagttaatggaggaatgccatc	In-line probing, M14 78 Psy
29	caggtttctgtaggcatcatcagctgctactggcgcagcggtttgtt aatgg	In-line probing, M15 78 Psy
30	gcgcccgtagcagctgatgaaccctacagaaacctgatcaaacc	Two-step PCR, Psy-pCR2.1, M1
31	ggtttgatcaggtttctgtagggttcatcagctgctacgggcgc	Two-step PCR, Psy-pCR2.1, M1
32	gctaagatcggcgcattggagccggcattcctccattaacaaacc	Two-step PCR, Psy-pCR2.1, M3
33	ggtttgttaatggaggaatgccggctccaatgcgccgatcttagc	Two-step PCR, Psy-pCR2.1, M3
34	ccgtagcagctgatgatgcccacagaaacctgatcaaaccagg	Two-step PCR, Psy-pCR2.1, M4
35	cctggtttgatcaggtttctgtgggcatcatcagctgctacgg	Two-step PCR, Psy-pCR2.1, M4
36	aaaccgctgcgcccgtagcagcagatgatgcctacagaaacctg	Two-step PCR, Psy-pCR2.1, M6
37	cattcctccattaacaaaccggagcgcccgtagcagctgatgatg	Two-step PCR, Psy-pCR2.1, M16
38	catcatcagctgctacgggcgctccggtttgttaatggaggaatg	Two-step PCR, Psy-pCR2.1, M16
39	caaaccggagcgcccgtagctcctgatgatgcctacagaaac	Two-step PCR, PsyM16-pCR2.1, M17
40	gtttctgtaggcatcatcaggagctacgggcgctccggtttg	Two-step PCR, PsyM16-pCR2.1, M17
41	aaggaattc**aaaaataatgttgtccttttaaataagatctgataa aatgtgaactaa**atgtaataattataggcgatggagttcg	Amplification of *B. cereus* ATCC10987 *crcB* motif and expression platform
42	aaggatccctctttaaatagcttgctcaaaaaaatagactcct	Amplification of *B. cereus* ATCC10987 *crcB* motif and expression platform
43	atgcgaattctttggccctctttcgtaagcggtga	Cloning of the *eriC*^{F} gene from *P. syringae* DC3000 (full length into pCR2.1; fusion to *lacZ* gene into pRS414)
44	gacgggatccccaggtcgtcgaaatttagacatttggcc	Cloning of the *eriC*^{F} gene from *P. syringae* DC3000 into pRS414 to form an in-frame fusion to the *lacZ* gene around the 8th codon
45	gcgaaagctttcatttctcttcccccgcaagagtc	Cloning of the full length *eriC*^{F} gene from *P. syringae* DC3000 into pCR2.1
46	ccgcttaatcagcgggttttttttggtctttagtgtgcggttgaggccg a	Verifying the deletion of the *crcB* gene in *E. coli*
47	cggcgttacacttatcactcatacaaatcaaatagcaggattttgca	Verifying the deletion of the *crcB*

	gtg	gene in *E. coli*
48	*taatacgactcactata*ggtaggcgatggagttcgccataaac	In-line probing, Bce RNA, full and short terminators
49	tagactcctaccagtagtgatac	In-line probing, Bce RNA, full terminator
50	accagtagtgatactggtagg	In-line probing, Bce RNA, short terminator
51	aaggaattc**aaaaataatgttgtccttttaaataagatctgataa aatgtgaactaa**atgtaataattataggcgatgatgttcgccataa acg	Two-step PCR, Bce-pDG1661, M18, M21, M23
52	ctgcttagctaatgacatctaccagtatcactactggt	Two-step PCR, Bce-pDG1661, M19, M21, M22, M23
53	accagtagtgatactggtagatgtcattagctaagcag	Two-step PCR, Bce-pDG1661, M19, M21, M22, M23
54	aaggatccctctttaaatagcttgctcaaaaaaatagacauctacc agtagtgatac	Two-step PCR, Bce-pDG1661, M20, M22, M23

fig. S1. Genes most commonly associated with fluoride riboswitches in bacteria and archaea as classified by the Conserved Domain Database (*66*). Top: 62% of all riboswitch-associated genes fall into twelve common categories as noted with sections proportional to the number of operons encoding the specified protein. These genes are predominantly bacteria, since bacterial genome sequences significantly outnumber archaeal genome sequences. MFS: major facilitator superfamily; USP: universal stress protein. Bottom: fluoride riboswitch-associated genes from archaea. A detailed listing of fluoride riboswitches and their associated genes within RefSeq version 32 is available from a previous report (8).

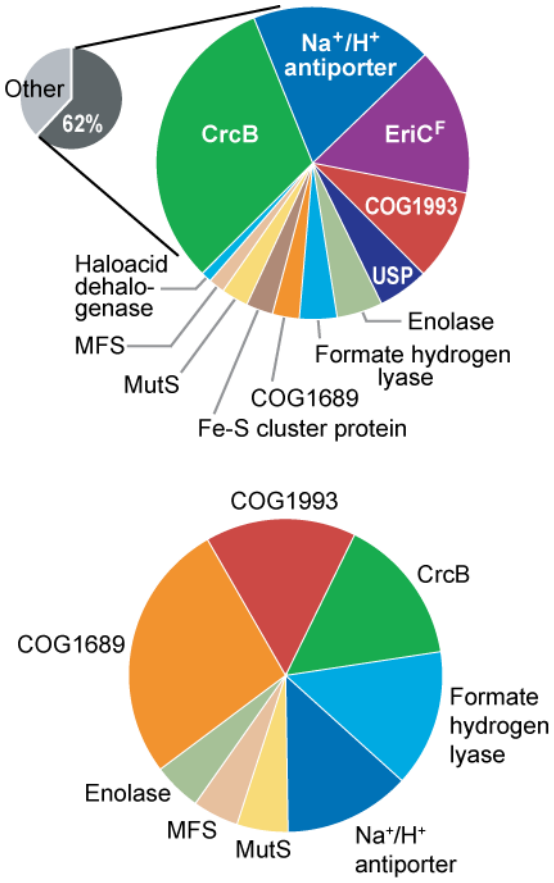


fig. S2. Initial in-line probing assay of a *crcB* motif RNA incubated with either a synthetic dinucleotide or sodium fluoride. Depicted is a PAGE image of in-line probing reactions with WT 78 Psy RNA incubated in the presence of 14 μM pApA (AA) or 1 mM NaF (F). Other annotations are as described in the legend to Fig. 1C. The identical pattern of spontaneous cleavage products in the AA and F lanes is consistent with fluoride as the ligand.

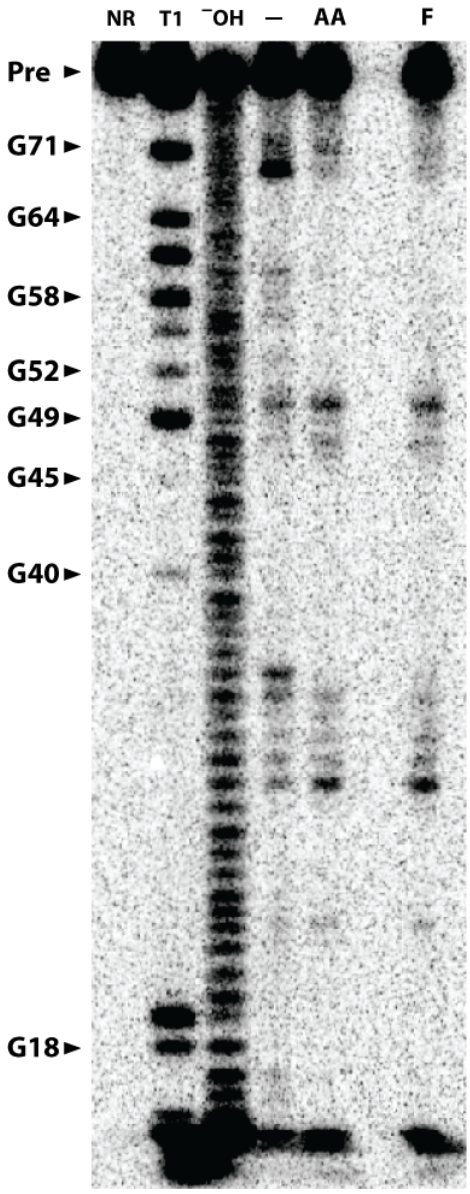


fig. S3. Sequences, structures and in-line probing assays of *crcB* motif RNAs from the bacterium *Thermotoga petrophila* and the archaeon *Thermoplasma volcanium* incubated with sodium fluoride. (**A**) Sequence and secondary structure model for the *Thermotoga petrophila crcB* motif RNA. Annotated sites of spontaneous cleavage are derived from B. (**B**) PAGE image of in-line probing reactions with WT 62 Tpe RNA incubated with various concentrations of fluoride ranging from 1 μM to 8 mM (K_D ~50 μM). (**C**) Sequence and secondary structure model for the *Thermoplasma volcanium crcB* motif RNA. Annotated sites of spontaneous cleavage are derived from D. (**D**) PAGE image of in-line probing reactions with WT 74 Tvo RNA incubated with various concentrations of fluoride ranging from 1 μM to 8 mM (K_D ~50 μM). Other annotations are as described in the legend to Fig. 1B and Fig. 1C.

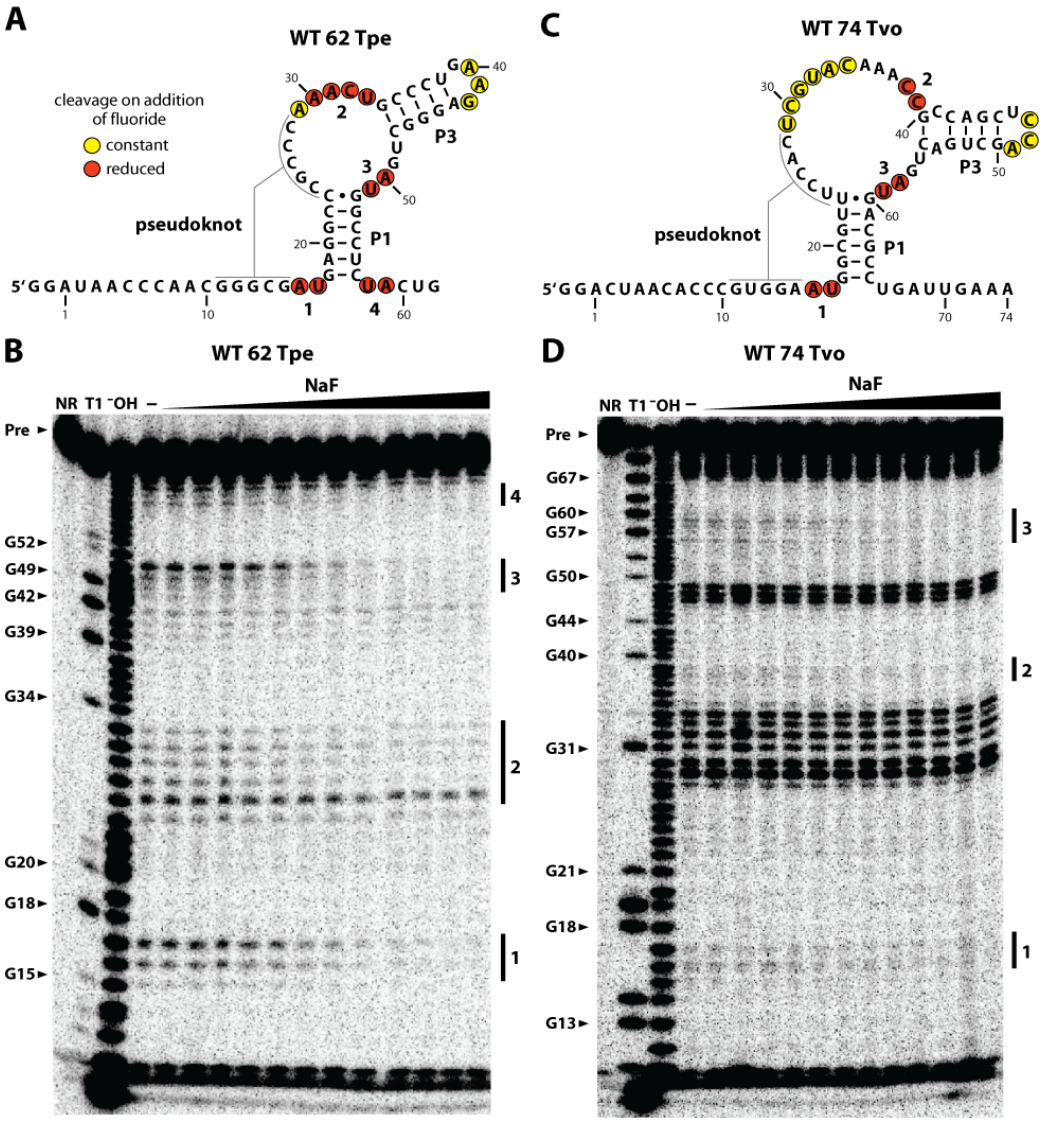


fig. S4. In-line probing assays with 78 Psy RNA and fluoride in the presence or absence of divalent calcium ions. (**A**) PAGE image of in-line probing reactions conducted without added ions (□), with 100 μM fluoride (F), or with 1 mM $CaCl_2$ and 100 μM fluoride (Ca/F). Zero and two days refer to preincubation times (see Methods). Other annotations are as described in the legend to Fig. 1C. (**B**) Plot of the band intensities for regions 1 and 2 denoted in A. Intensities in the (□) and (F) lanes were normalized to 1 and 0, respectively.

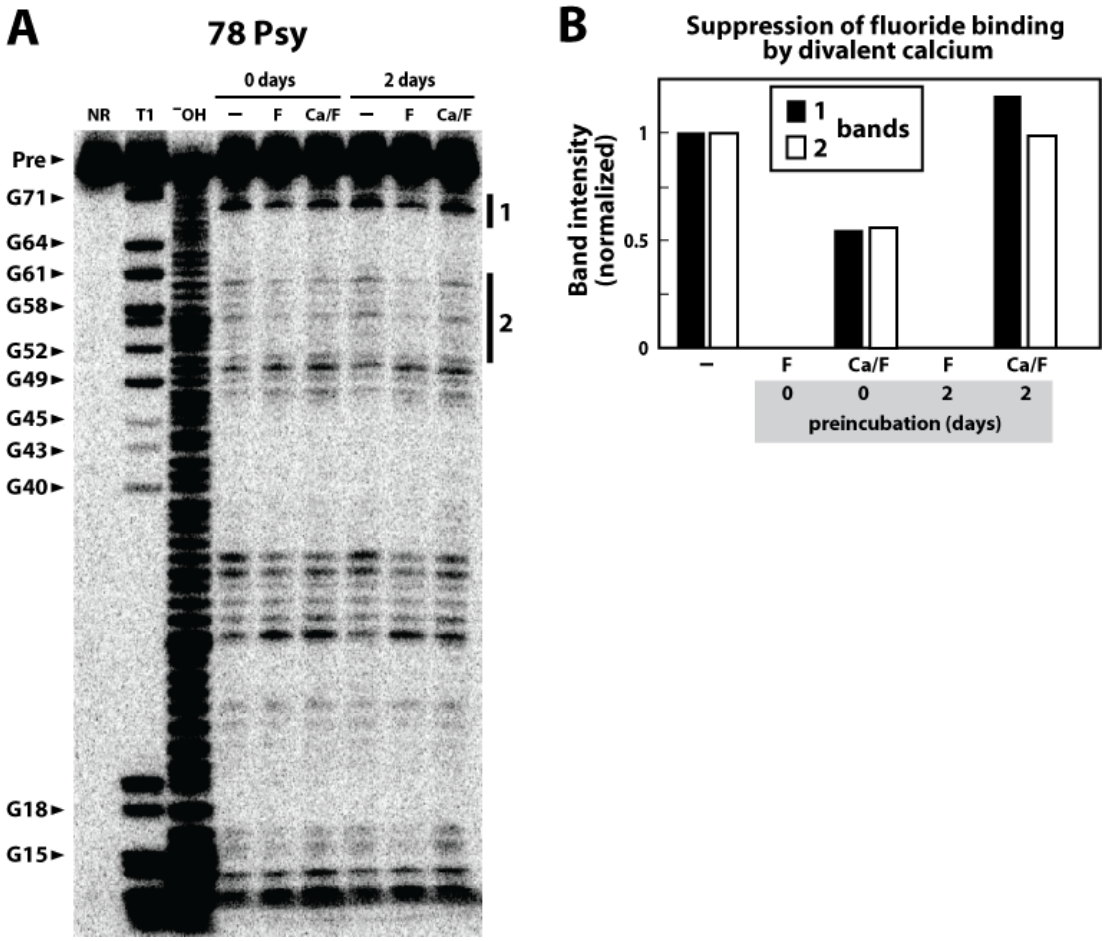


fig. S5. Compounds tested for binding by fluoride riboswitches. No evidence of binding was observed for any chemical species tested, except fluoride ions.

Halides	Small anions		Gases
F^-	$^*OH^-$	$^*HCOO^-$	$^*CO$
$^*Cl^-$	SH^-	HCO_3^-	$^*NO$
Br^-	CN^-	SO_4^{2-}	
I^-	SCN^-	$^*H_2PO_4^-$	
	NO_3^-	$HAsO_4^{2-}$	
	NO_2^-		

*All ligands tested at 10 mM except these six: Cl^-, up to 2.5 M; OH^-, up to 100 μM (pH 10); $HCOO^-$, NO and $H_2PO_4^-$, 1 mM; CO, ~0.9 mM.

Other Compounds	
pApA	nicotinic acid
pGpA	niacinamide
pAp	methylthioadenosine
c-di-GMP	cystathionine
c-di-AMP	methionine
GTP	histidine
TPP	spermidine
SAM	3-phosphoglycerate
SAH	methylglyoxal
5′-deoxyadenosine	carbamoyl phosphate
*2,8-dihydroxyadenine	ribose-5′-phosphate
**8-hydroxyguanine	glucose-6-phosphate
**8-hydroxyguanosine	fructose-6-phosphate
**8-hydroxydeoxyguanosine	fructose-1,6-biphosphate
***formaldehyde	phosphoribosyl pyrophosphate
****trifluoroacetic acid	phosphoenolpyruvate
pyridoxamine	FAD
pyridoxic acid	
pyridoxal phosphate	

All ligands tested at 1 mM except the following: *50 μM; **100 μM; ***10 mM; ****1.3 mM.

fig. S6. Analysis of the role of the P1 stem for fluoride binding by mutation/compensation mutation and in-line probing. (**A**) Sequence and secondary structure of the 78 Psy RNA construct and mutations used to evaluate the importance of P1. For each mutant, the boxed nucleotides on the RNA structure are mutated as depicted in the connected boxes. M1 is expected to disrupt formation of the P1 stem, and thus reduce affinity for fluoride, while M2 is expected to restore P1 formation and fluoride binding. (**B**) PAGE images of in-line probing reactions with WT 78 Psy RNA and mutants M1 and M2. RNAs were incubated in the presence or absence of fluoride delivered by the use of a pApA solution as described in the legend to fig. S1. Intervening lanes (carrying other assay conditions) between the three marker lanes and the two in-line probing assay lanes for each RNA were deleted for clarity. Other annotations are as described in the legend to Fig. 1.

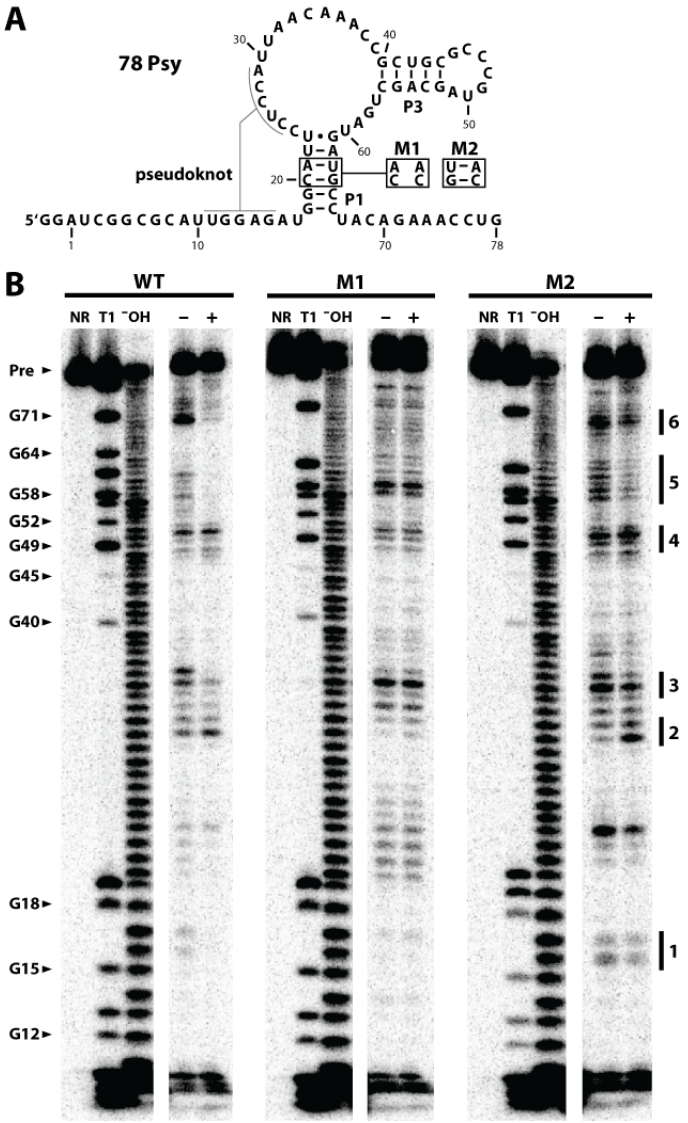


fig. S7. Analysis of various fluoride RNA mutants by in-line probing. (**A**) Sequence and secondary structure of the 78 Psy RNA construct and various mutations used to evaluate the importance of certain nucleotides. Eleven of the most conserved nucleotides are depicted in red, while other nucleotides are black. Since conserved nucleotides are generally required to form binding pockets (see Supplementary text), alteration of these conserved nucleotides should disrupt binding, if fluoride is the natural ligand. However, if fluoride is merely a mimic of the natural ligand, then mutations to some parts of the binding pocket would not disrupt fluoride binding. For each mutant, the boxed nucleotides on the RNA structure are mutated as depicted in the connected boxes. (**B**) PAGE images of in-line probing reactions with two representative mutants that either do (M11) or do not (M14) disrupt function when incubated with 1 mM fluoride. In nearly all instances, those mutations (M3-M11) that alter the base identities of highly-conserved nucleotides result in a complete loss of fluoride binding. Exceptions are M7, which exhibits normal affinity, and M8, which exhibits only a 10-fold loss of affinity (data not shown). Mutation M7 preserves Watson-Crick base pairing potential, which might explain why this change is tolerated under our experimental conditions. Mutations to nonconserved nucleotides (M12-M15) do not disrupt fluoride binding. These results are consistent with the hypothesis that fluoride is the natural ligand of these riboswitches. Annotations are as described in the legend to Fig. 1. The extra band in the M11 RNase T1 lane (position 34) does not reflect the presence of a G, but rather appears to be a hot-spot for spontaneous cleavage in most constructs.

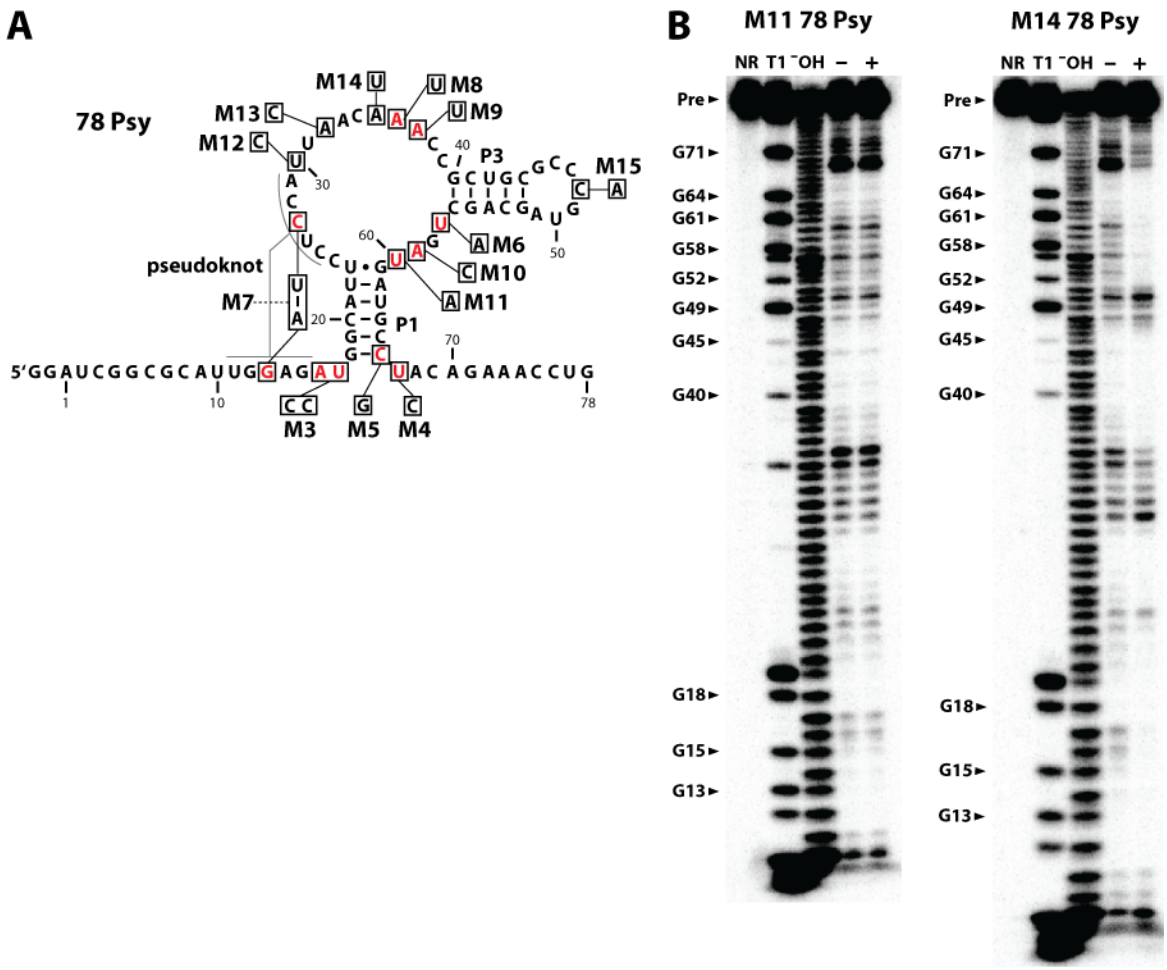

A
78 Psy
M14
M8
M13
M9
M12
P3
M15
pseudoknot
M7
M6
M10
M11
P1
5'GGAUCGGCGCAUUGGAGAAU
UACAGAAACCUG
M3
M5
M4
B
M11 78 Psy
M14 78 Psy
NR T1 ⁻OH – +
Pre
G71
G64
G61
G58
G52
G49
G45
G40
G18
G15
G13

fig. S8. Fluoride-induced reporter gene activity controlled by the riboswitch associated with *B. cereus crcB*. (**A**) Liquid cultures of *B. subtilis* cells transformed with a plasmid carrying the fluoride riboswitch from the *B. cereus crcB* gene fused to *lacZ*. Expression levels of *lacZ* are indicated by the blue color due to β-galactosidase activity on X-gal (5-bromo-4-chloro-3-indolyl-beta-D-galactopyranoside). (**B**) Plot of the β-galactosidase reporter activity versus fluoride concentration in liquid medium supporting growth of transformed *B. subtilis* cells as quantified using Miller assays.

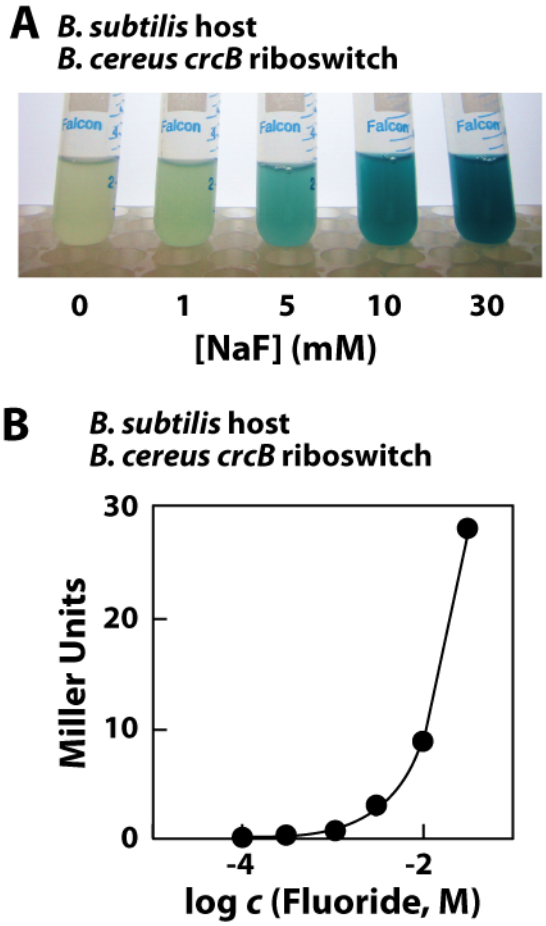


fig. S9. Mutants of the Bce riboswitch (also see fig. S10) used to assess the expression platform mechanism. (**A**) Mutations M18 through M23 sequentially disrupt and compensate base pairing in the aptamer P1 stem and/or terminator stem. Mutated positions are depicted in the context of the secondary structure models for the fluoride-bound (left) and the fluoride-free (right) states. (**B**) Plot of the levels of β-galactosidase activity in cells containing the DNA of WT or M18 through M23 Bce constructs fused to a *lacZ* reporter in the presence of 40 mM NaCl or NaF in minimal medium. Note that constructs with mutations favoring constitutive terminator formation (M18 and M22) lack fluoride responsiveness and fail to express the reporter gene. Constructs with mtations that disrupt terminator formation (M19, M20, M21) substantially reduce fluoride responsiveness and yield constitutively high reporter gene expression. Finally, construct M23 combines three mutations that individually disrupt normal fluoride responsiveness, yet yields near wild-type performance by restoring aptamer, antiterminator, and terminator base pairing despite using a nucleotide sequence different than that of the WT RNA. Miller unit values are the mean of three replicates +/- the standard deviation.

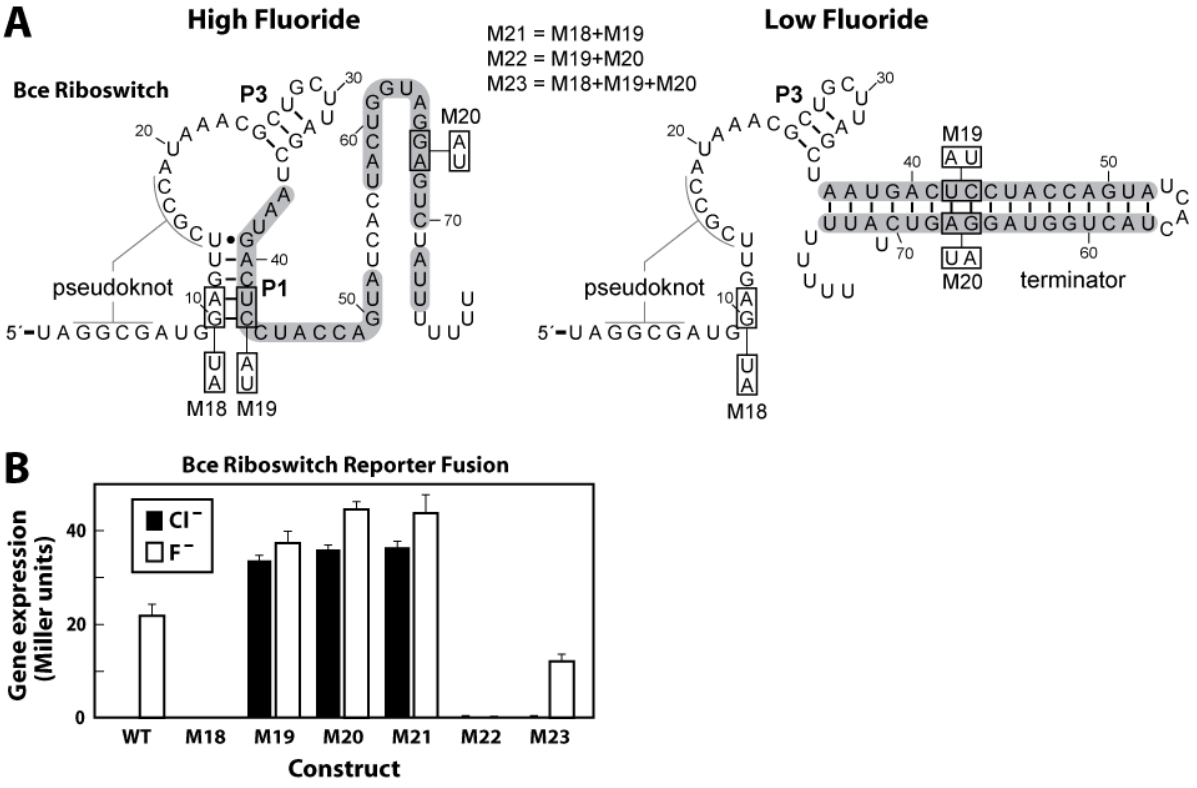


fig. S10. Evidence for a transcription termination mechanism to activate gene expression upon fluoride binding. (**A**) In-line probing analysis of the full-length fluoride riboswitch (ligand-binding aptamer and expression platform) from the *B. cereus crcB* gene (see C). Details are as described for Fig. 1C. The pattern of spontaneous cleavage products reveals no changes in structure upon fluoride addition, and is consistent with an RNA that cannot bind fluoride because it is thermodynamically trapped in a structure that includes formation of a strong terminator stem (see C). (**B**) In-line probing analysis of a truncated fluoride riboswitch (aptamer and partial expression platform) from the *B. cereus crcB* gene (see D). Details are as described for Fig. 1C. Deletion of a portion of the terminator stem creates a weaker stem that allows interchange between aptamer-ligand formation and terminator stem formation (see D). The pattern of spontaneous cleavage products reveals changes in both the aptamer structure and a portion of the terminator stem upon fluoride addition. The inset depicts a region of the gel at greater contrast, and shows modulations around nucleotides 47 and 48, within the terminator stem. The nucleotides involved in the 3′ side of the stem are too close to the 3′ end of the molecule to resolve clearly. (**C**) Sequence and secondary structure model for the full riboswitch construct examined in A. The run of U residues typically located after a rho-independent terminator stem was not included in the construct. (**D**) Sequence and secondary structure models for the truncated riboswitch construct examined in B.

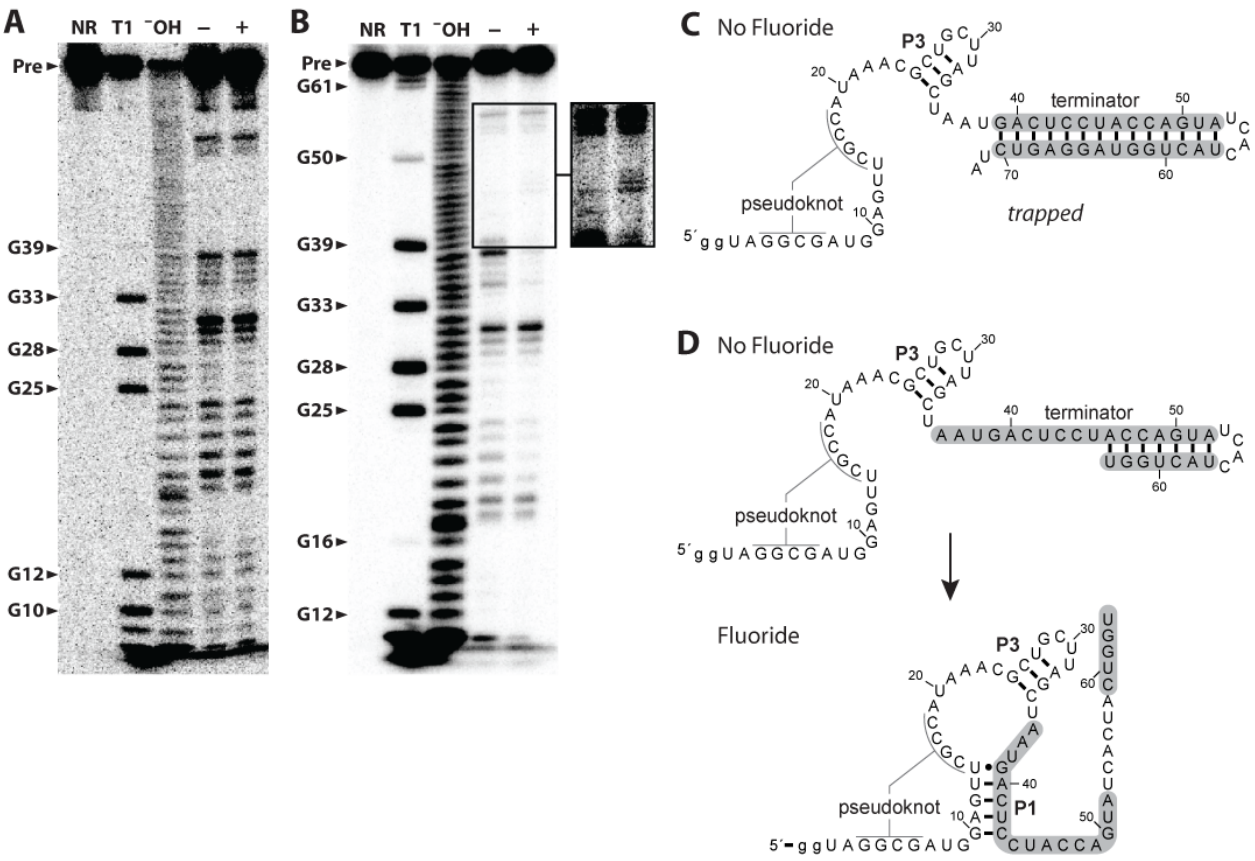

A
NR T1 ⁻OH – +
Pre
G39
G33
G28
G25
G12
G10
B
NR T1 ⁻OH – +
Pre
G61
G50
G39
G33
G28
G25
G16
G12
C No Fluoride
P3
terminator
pseudoknot
5′ g g U A G G C G A U G
trapped
D No Fluoride
P3
terminator
pseudoknot
5′ g g U A G G C G A U G
Fluoride
P3
P1
pseudoknot
5′ g g U A G G C G A U G

fig. S11. Assessment of gene control by WT and mutated variants of a *Pseudomonas syringae* fluoride-responsive riboswitch in *E. coli*. (**A**) Sequences and secondary structures of the wild-type fluoride riboswitch aptamer from the *eriC* gene of *P. syringae* and various mutants fused to the β-galactosidase reporter gene (*lacZ*). Eleven of the most conserved nucleotides are colored red. Diagonal line reflects the fusion of the first eight codons of *eriC* to the ninth codon of the *lacZ* gene. For each mutant, the boxed nucleotides on the RNA structure are mutated as depicted in the connected boxes. (**B**) Plot of the level of β-galactosidase activity in cells containing the reporter constructs depicted in A in the presence of 40 mM fluoride. Miller unit values are the mean of three replicates +/- the standard deviation. The Miller unit values determined for all six constructs in the presence of chloride were approximately zero (data not shown). As expected, only the WT riboswitch and the M17 mutant with restored ability to form the conserved secondary structure show an ability to increase gene expression in response to fluoride.

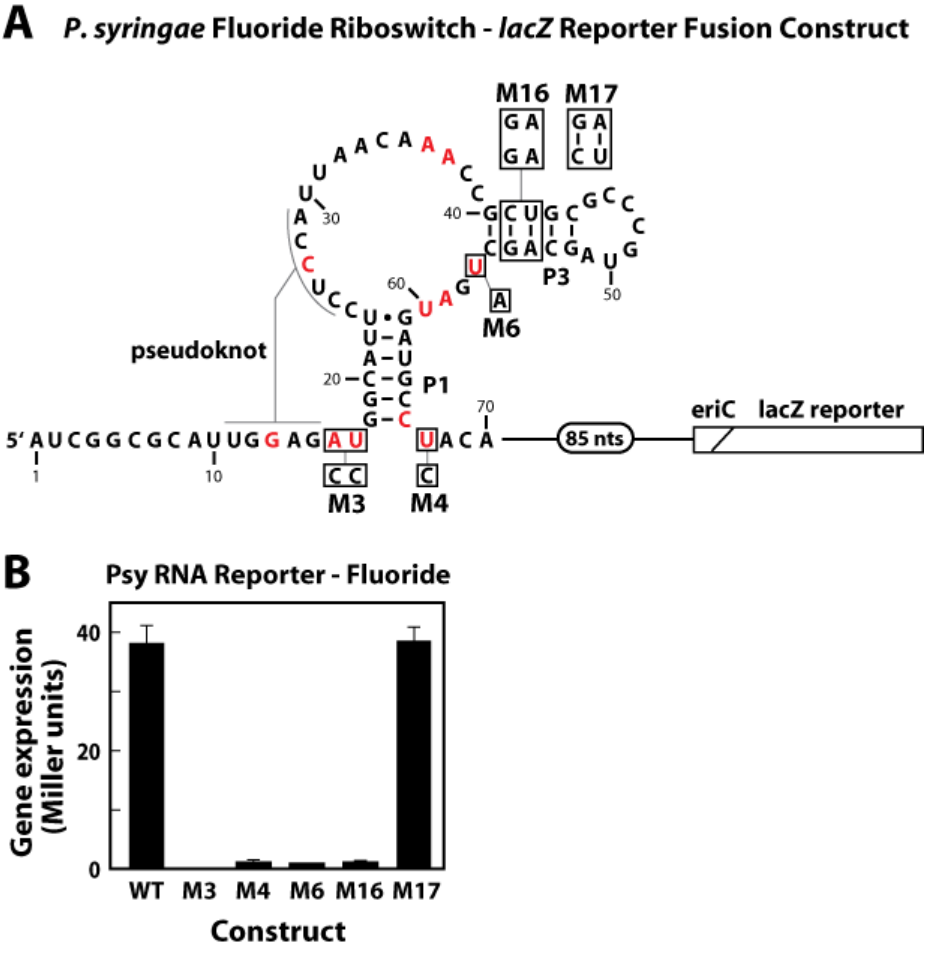


fig. S12. Comparison of the growth of *E. coli* WT and *crcB* KO cells in the presence of various concentrations of fluoride. (**A**) Comparison of $O.D._{600}$ for WT and *crcB* KO *E. coli* cells after 16 hr incubation. (**B**) Comparison of $O.D._{600}$ for WT and *crcB* KO *E. coli* cells after 5 hr incubation in media at pH 7.5 or 5.5. Experimental details are as described for the data in Fig. 3.

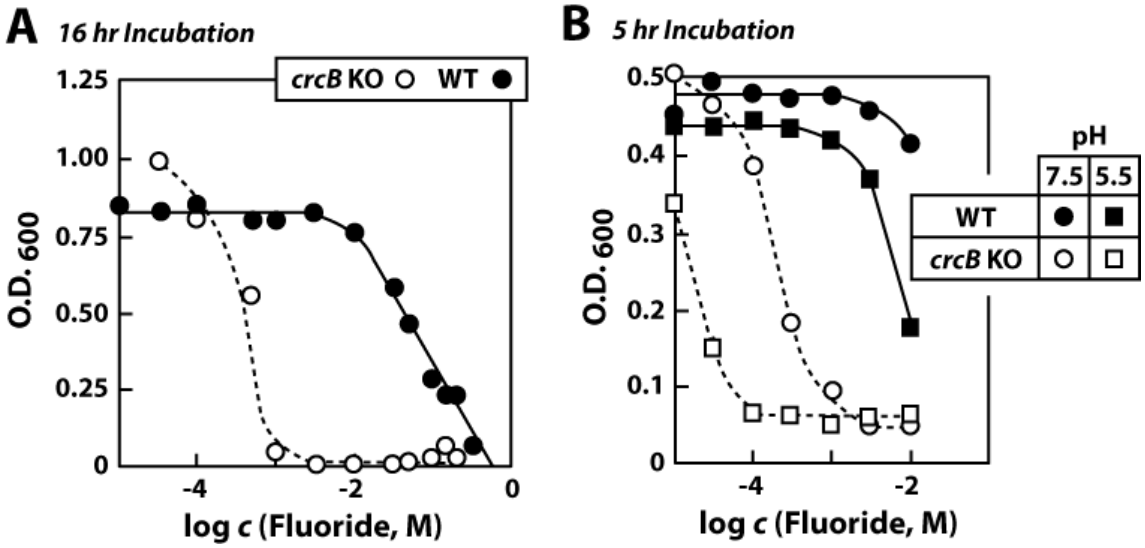


fig. S13. Clustering of EriC proteins, including variant proteins (termed EriC^F) associated with fluoride riboswitches. Plot of the sequence similarity of EriC proteins in bacteria and archaea reveals that many examples of an evolutionarily distinct collection of representatives (dark and light blue) are associated with fluoride riboswitches. A single riboswitch-associated protein in *Hydrogenobaculum* sp. Y04AAS1 (marked with an asterisk) does not fit the normal characteristics. In *Streptococcus* species whose genomes have been sequenced, *crcB* or *eriC^F* genes are always located adjacent to *pheA* (chorismate mutase) (fig. S18). The reason for this correlation remains unclear. This phylogenetic tree suggests that other *eriC* genes with sequences similar to those associated with fluoride riboswitches may also transport fluoride. The branch marked "100%" has an estimated confidence of 100% in correctly splitting EriC and EriC^F proteins according to their evolutionary history.

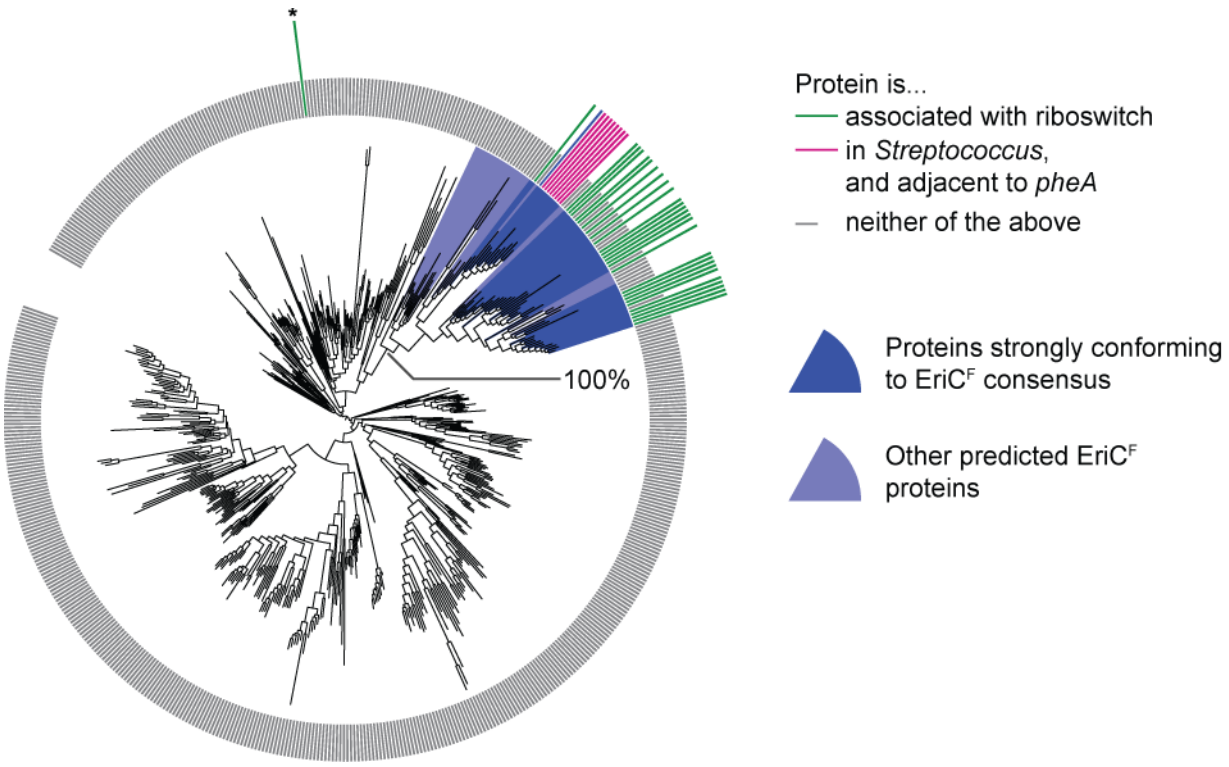


fig. S14. Amino acid alignments of the conserved channel-forming residues of several EriC proteins reveal unique sequences among representatives associated with fluoride riboswitches (boxed) compared to those known to transport chloride. Experimentally confirmed (Cl^-, NO_3^-) or hypothesized (F^-) substrates are indicated. Note that other alignments of residues are possible, but they imply at least as many changes have occurred in the selectivity filter (see explanation in Materials and Methods).

Substrate	Organism	"Selectivity Filter" Residues			
Cl^-	human (ClC-1)	GSGIP	GKEGP	GGFMP	Y
Cl^-	human (ClC-5)	GSGIP	GKEGP	GLFIP	Y
Cl^-	*E. coli*	GSGIP	GREGP	GIFAP	Y
NO_3^-	*Arabidopsis*	GPGIP	GKEGP	GIFAP	Y
F^-	*Pseudomonas syringae*	GNNLI	GREGT	GEVTP	Y
F^-	*Clostridium difficile*	GMNLI	GREGV	GEVTP	Y
F^-	*Streptococcus mutans*	GMGLI	GREGV	GEVTP	Y

Red conserved in Cl^- channels

Blue differ from Cl^- channels

fig. S15. Rescue of *E. coli crcB* KO cells from extreme fluoride toxicity in liquid (**A**) or on solid (**B**) media by transformation with the *eriC*^{F} gene from *P. syringae*.

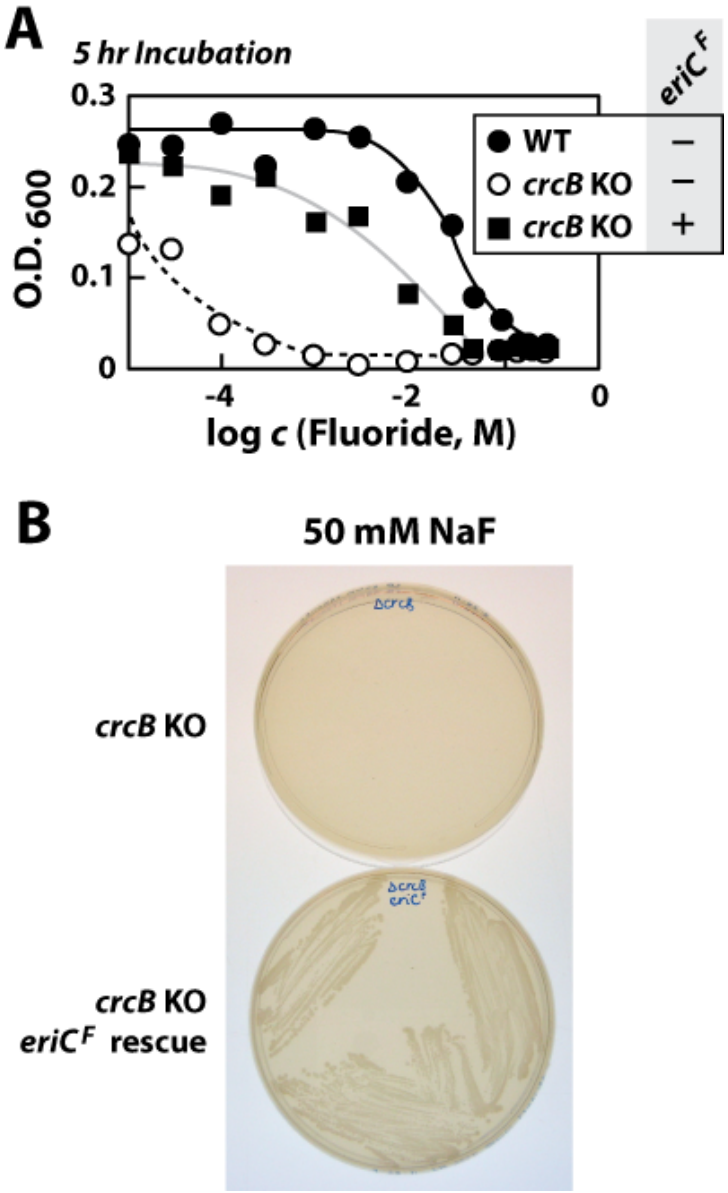


fig. S16. Distribution of fluoride riboswitches and putative fluoride transporters among organisms from the three domains of life. Selected taxa are depicted based on a previous tree of life (*78*).

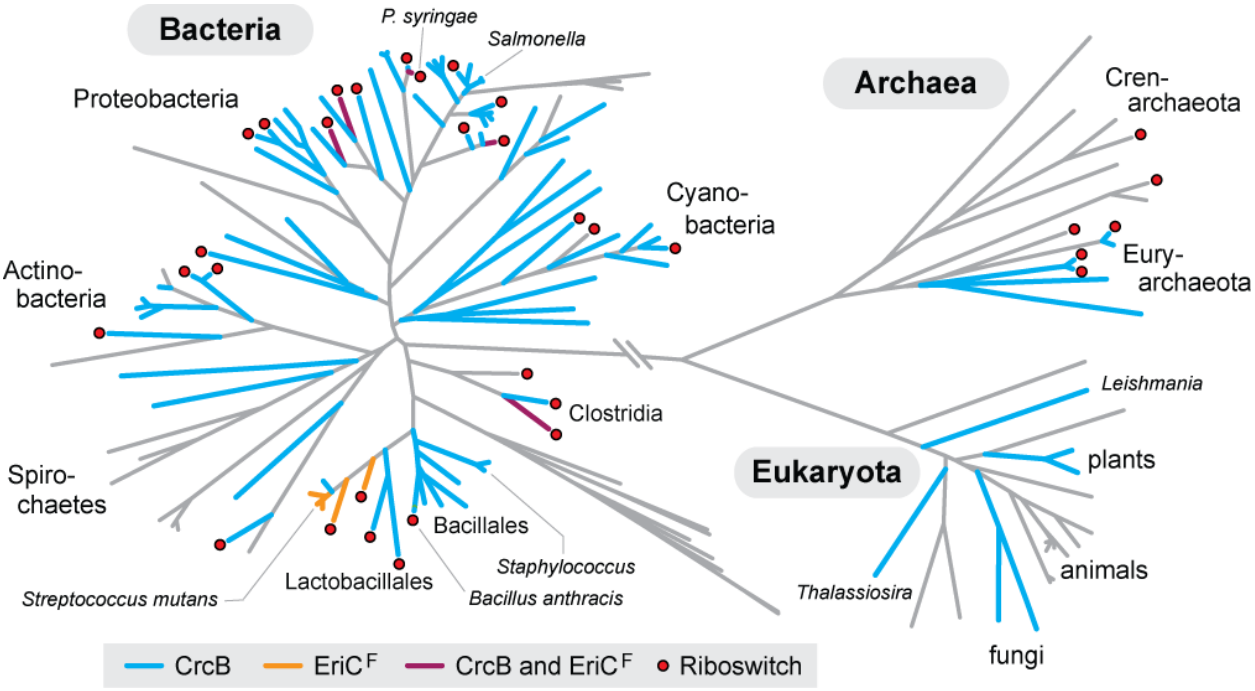


fig. S17. CrcB proteins are broadly associated with fluoride riboswitches. Plot of the sequence similarity of CrcB proteins in bacteria and archaea reveals that many distantly-related examples are associated with fluoride riboswitches.

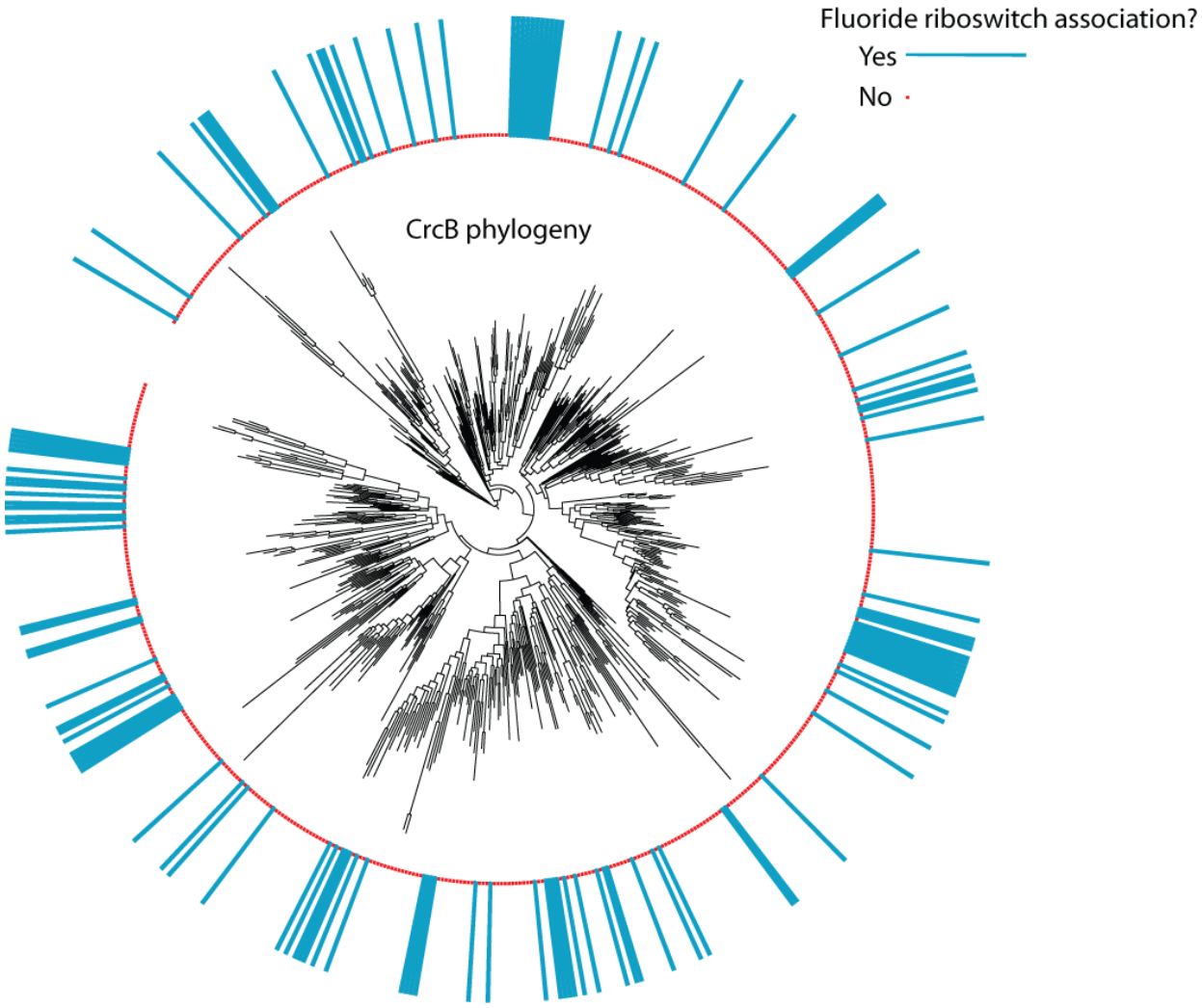
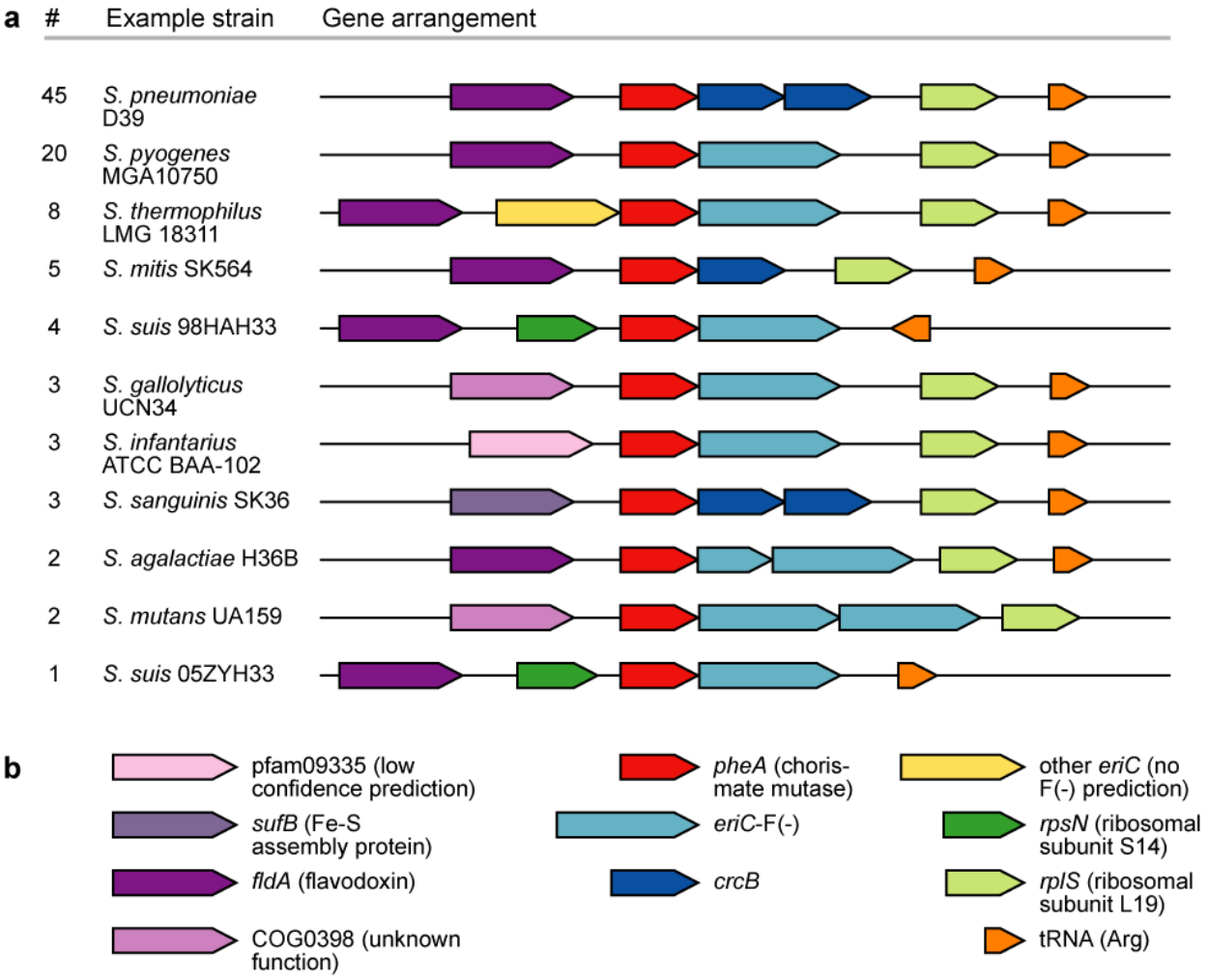


fig. S18. Locations of *crcB* and *eriC*^{F} genes in *Streptococcus*. (**A**) Distinct gene arrangements surrounding all *crcB* and *eriC*^{F} genes in sequenced *Streptococcus* species. "#": number of organisms with this arrangement, "Example strain": one of the organisms. Genes are colored by class (see B). Corresponding regions harboring transposases and one partial sequence fragment are ignored. Lengths of genes and gaps are not to scale, as exact distances vary per organism. However, the first *eriC*^{F} gene in *S. agalactiae* is unusually small, as shown. A tRNA(Arg) gene is located roughly 3,000 base pairs downstream of the depicted region in *S. mutans*. As the diagram shows, all sequenced *Streptococcus* species have either a *crcB* or an *eriC*^{F} gene (but never both) flanked by highly similar gene arrangements. (**B**) Color codes of gene families used in A.



References and Notes

1. R. J. Lesher, G. R. Bender, R. E. Marquis, Bacteriolytic action of fluoride ions. *Antimicrob. Agents Chemother.* **12**, 339 (1977). Medline

2. M. Maltz, C. G. Emilson, Susceptibility of oral bacteria to various fluoride salts. *J. Dent. Res.* **61**, 786 (1982). doi:10.1177/00220345820610062701 Medline

3. R. E. Marquis, S. A. Clock, M. Mota-Meira, Fluoride and organic weak acids as modulators of microbial physiology. *FEMS Microbiol. Rev.* **26**, 493 (2003). doi:10.1111/j.1574-6976.2003.tb00627.x Medline

4. R. S. Levine, The action of fluoride in caries prevention. A review of current concepts. *Br. Dent. J.* **140**, 9 (1976). doi:10.1038/sj.bdj.4803690 Medline

5. H. Koo, Strategies to enhance the biological effects of fluoride on dental biofilms. *Adv. Dent. Res.* **20**, 17 (2008). doi:10.1177/154407370802000105 Medline

6. I. R. Hamilton, Biochemical effects of fluoride on oral bacteria. *J. Dent. Res.* **69**, (Spec No), 660, discussion 682 (1990). Medline

7. C. Van Loveren, Antimicrobial activity of fluoride and its in vivo importance: Identification of research questions. *Caries Res.* **35**, 65 (2001). doi:10.1159/000049114 Medline

8. Z. Weinberg *et al*., Comparative genomics reveals 104 candidate structured RNAs from bacteria, archaea, and their metagenomes. *Genome Biol.* **11**, R31 (2010). doi:10.1186/gb-2010-11-3-r31 Medline

9. Supporting text and materials and methods are available as supporting material on *Science* Online.

10. G. A. Soukup, R. R. Breaker, Relationship between internucleotide linkage geometry and the stability of RNA. *RNA* **5**, 1308 (1999). doi:10.1017/S1355838299990891 Medline

11. E. E. Regulski, R. R. Breaker, In-line probing analysis of riboswitches. *Methods Mol. Biol.* **419**, 53 (2008). doi:10.1007/978-1-59745-033-1_4 Medline

12. J. T. Dobbins, H. A. Ljung, A system of qualitative analysis for the anions. *J. Chem. Educ.* **12**, 586 (1935). doi:10.1021/ed012p586

13. P. Auffinger, L. Bielecki, E. Westhof, Anion binding to nucleic acids. *Structure* **12**, 379 (2004). doi:10.1016/j.str.2004.02.015 Medline

14. K. H. Hu *et al*., Overproduction of three genes leads to camphor resistance and chromosome condensation in *Escherichia coli*. *Genetics* **143**, 1521 (1996). Medline

15. M. Rapp, E. Granseth, S. Seppälä, G. von Heijne, Identification and evolution of dual-topology membrane proteins. *Nat. Struct. Mol. Biol.* **13**, 112 (2006). doi:10.1038/nsmb1057 Medline

16. R. D. Finn *et al*., The Pfam protein families database. *Nucleic Acids Res.* **38**, D211 (2010). doi:10.1093/nar/gkp985 Medline

17. K. Matulef, M. Maduke, The CLC ‘chloride channel’ family: Revelations from prokaryotes. *Mol. Membr. Biol.* **24**, 342 (2007). doi:10.1080/09687680701413874 Medline

18. R. Dutzler, E. B. Campbell, M. Cadene, B. T. Chait, R. MacKinnon, X-ray structure of a ClC chloride channel at 3.0 Å reveals the molecular basis of anion selectivity. *Nature* **415**, 287 (2002). doi:10.1038/415287a Medline

19. A. Accardi, C. Miller, Secondary active transport mediated by a prokaryotic homologue of ClC Cl^- channels. *Nature* **427**, 803 (2004). doi:10.1038/nature02314 Medline

20. R. Gälli, T. Leisinger, Specialized bacterial strains for the removal of dichloromethane from industrial waste. *Conserv. Recy.* **8**, 91 (1985). doi:10.1016/0361-3658(85)90028-1

21. S. Vuilleumier, H. Sorribas, T. Leisinger, Identification of a novel determinant of glutathione affinity in dichloromethane dehalogenases/glutathione *S*-transferases. *Biochem. Biophys. Res. Commun.* **238**, 452 (1997). doi:10.1006/bbrc.1997.7309 Medline

22. S. Silver, Bacterial resistances to toxic metal ions - a review. *Gene* **179**, 9 (1996). doi:10.1016/S0378-1119(96)00323-X Medline

23. N. Sudarsan *et al*., Riboswitches in eubacteria sense the second messenger cyclic di-GMP. *Science* **321**, 411 (2008). doi:10.1126/science.1159519 Medline

24. E. R. Lee, J. L. Baker, Z. Weinberg, N. Sudarsan, R. R. Breaker, An allosteric self-splicing ribozyme triggered by a bacterial second messenger. *Science* **329**, 845 (2010). doi:10.1126/science.1190713 Medline

25. G. Witte, S. Hartung, K. Büttner, K. P. Hopfner, Structural biochemistry of a bacterial checkpoint protein reveals diadenylate cyclase activity regulated by DNA recombination intermediates. *Mol. Cell* **30**, 167 (2008). doi:10.1016/j.molcel.2008.02.020 Medline

26. U. Römling, Great times for small molecules: c-di-AMP, a second messenger candidate in bacteria and archaea. *Sci. Signal.* **1**, pe39 (2008). doi:10.1126/scisignal.133pe39 Medline

27. F. Wincott *et al*., Synthesis, deprotection, analysis and purification of RNA and ribozymes. *Nucleic Acids Res.* **23**, 2677 (1995). doi:10.1093/nar/23.14.2677 Medline

28. F. F. Feagin, M. Gonzalez, B. G. Jeansonne, Kinetic reactions of calcium, phosphate, and fluoride ions at the enamel surface-solution interface. *Calcif. Tissue Res.* **10**, 113 (1972). doi:10.1007/BF02012541 Medline

29. R. K. Montange, R. T. Batey, Riboswitches: Emerging themes in RNA structure and function. *Annu. Rev. Biophys.* **37**, 117 (2008). doi:10.1146/annurev.biophys.37.032807.130000 Medline

30. A. Serganov, The long and the short of riboswitches. *Curr. Opin. Struct. Biol.* **19**, 251 (2009). doi:10.1016/j.sbi.2009.02.002 Medline

31. J. Zhang, M. W. Lau, A. R. Ferré-D'Amaré, Ribozymes and riboswitches: Modulation of RNA function by small molecules. *Biochemistry* **49**, 9123 (2010). doi:10.1021/bi1012645 Medline

32. L. M. Stancik *et al*., pH-dependent expression of periplasmic proteins and amino acid catabolism in *Escherichia coli*. *J. Bacteriol.* **184**, 4246 (2002). doi:10.1128/JB.184.15.4246-4258.2002 Medline

33. N. Sudarsan, J. K. Wickiser, S. Nakamura, M. S. Ebert, R. R. Breaker, An mRNA structure in bacteria that controls gene expression by binding lysine. *Genes Dev.* **17**, 2688 (2003). doi:10.1101/gad.1140003 Medline

34. A. Nahvi *et al.*, Genetic control by a metabolite binding mRNA. *Chem. Biol.* **9**, 1043 (2002). doi:10.1016/S1074-5521(02)00224-7 Medline

35. A. Accardi, L. Kolmakova-Partensky, C. Williams, C. Miller, Ionic currents mediated by a prokaryotic homologue of CLC Cl^- channels. *J. Gen. Physiol.* **123**, 109 (2004). doi:10.1085/jgp.200308935 Medline

36. M. Maduke, D. J. Pheasant, C. Miller, High-level expression, functional reconstitution, and quaternary structure of a prokaryotic ClC-type chloride channel. *J. Gen. Physiol.* **114**, 713 (1999). doi:10.1085/jgp.114.5.713 Medline

37. A. J. Jin, D. Huster, K. Gawrisch, R. Nossal, Light scattering characterization of extruded lipid vesicles. *Eur. Biophys. J.* **28**, 187 (1999). doi:10.1007/s002490050199 Medline

38. M. Walden *et al.*, Uncoupling and turnover in a Cl^-/H^+ exchange transporter. *J. Gen. Physiol.* **129**, 317 (2007). doi:10.1085/jgp.200709756 Medline

39. N. Sudarsan *et al.*, Tandem riboswitch architectures exhibit complex gene control functions. *Science* **314**, 300 (2006). doi:10.1126/science.1130716 Medline

40. K. D. Pruitt, T. Tatusova, D. R. Maglott, NCBI reference sequences (RefSeq): A curated non-redundant sequence database of genomes, transcripts and proteins. *Nucleic Acids Res.* **35**, (Database issue), D61 (2007). doi:10.1093/nar/gkl842 Medline

41. G. W. Tyson *et al.*, Community structure and metabolism through reconstruction of microbial genomes from the environment. *Nature* **428**, 37 (2004). doi:10.1038/nature02340 Medline

42. S. G. Tringe *et al.*, The airborne metagenome in an indoor urban environment. *PLoS ONE* **3**, e1862 (2008). doi:10.1371/journal.pone.0001862 Medline

43. D. B. Rusch *et al.*, The Sorcerer II Global Ocean Sampling expedition: Northwest Atlantic through eastern tropical Pacific. *PLoS Biol.* **5**, e77 (2007). doi:10.1371/journal.pbio.0050077 Medline

44. J. C. Venter *et al.*, Environmental genome shotgun sequencing of the Sargasso Sea. *Science* **304**, 66 (2004). doi:10.1126/science.1093857 Medline

45. C. L. Hemme *et al.*, Metagenomic insights into evolution of a heavy metal-contaminated groundwater microbial community. *ISME J.* **4**, 660 (2010). doi:10.1038/ismej.2009.154 Medline

46. T. Woyke *et al.*, Symbiosis insights through metagenomic analysis of a microbial consortium. *Nature* **443**, 950 (2006). doi:10.1038/nature05192 Medline

47. W. P. Inskeep *et al.*, Metagenomes from high-temperature chemotrophic systems reveal geochemical controls on microbial community structure and function. *PLoS ONE* **5**, e9773 (2010). doi:10.1371/journal.pone.0009773 Medline

48. S. R. Gill *et al.*, Metagenomic analysis of the human distal gut microbiome. *Science* **312**, 1355 (2006). doi:10.1126/science.1124234 Medline

49. K. Kurokawa *et al*., Comparative metagenomics revealed commonly enriched gene sets in human gut microbiomes. *DNA Res.* **14**, 169 (2007). doi:10.1093/dnares/dsm018 Medline

50. J. Qin *et al*.; MetaHIT Consortium, A human gut microbial gene catalogue established by metagenomic sequencing. *Nature* **464**, 59 (2010). doi:10.1038/nature08821 Medline

51. W. J. Brazelton, J. A. Baross, Abundant transposases encoded by the metagenome of a hydrothermal chimney biofilm. *ISME J.* **3**, 1420 (2009). doi:10.1038/ismej.2009.79 Medline

52. J. Y. Jung *et al*., Metagenomic analysis of kimchi, a traditional Korean fermented food. *Appl. Environ. Microbiol.* **77**, 2264 (2011). doi:10.1128/AEM.02157-10 Medline

53. M. G. Kalyuzhnaya *et al*., High-resolution metagenomics targets specific functional types in complex microbial communities. *Nat. Biotechnol.* **26**, 1029 (2008). doi:10.1038/nbt.1488 Medline

54. K. T. Konstantinidis, J. Braff, D. M. Karl, E. F. DeLong, Comparative metagenomic analysis of a microbial community residing at a depth of 4000 meters at station ALOHA in the North Pacific subtropical gyre. *Appl. Environ. Microbiol.* **75**, 5345 (2009). doi:10.1128/AEM.00473-09 Medline

55. E. A. Dinsdale *et al*., Microbial ecology of four coral atolls in the Northern Line Islands. *PLoS ONE* **3**, e1584 (2008). doi:10.1371/journal.pone.0001584 Medline

56. P. J. Turnbaugh *et al*., An obesity-associated gut microbiome with increased capacity for energy harvest. *Nature* **444**, 1027 (2006). doi:10.1038/nature05414 Medline

57. H. García Martín *et al*., Metagenomic analysis of two enhanced biological phosphorus removal (EBPR) sludge communities. *Nat. Biotechnol.* **24**, 1263 (2006). doi:10.1038/nbt1247 Medline

58. S. G. Tringe *et al*., Comparative metagenomics of microbial communities. *Science* **308**, 554 (2005). doi:10.1126/science.1107851 Medline

59. F. Warnecke *et al*., Metagenomic and functional analysis of hindgut microbiota of a wood-feeding higher termite. *Nature* **450**, 560 (2007). doi:10.1038/nature06269 Medline

60. P. B. Pope *et al*., Adaptation to herbivory by the Tammar wallaby includes bacterial and glycoside hydrolase profiles different from other herbivores. *Proc. Natl. Acad. Sci. U.S.A.* **107**, 14793 (2010). doi:10.1073/pnas.1005297107 Medline

61. E. A. Dinsdale *et al*., Functional metagenomic profiling of nine biomes. *Nature* **452**, 629 (2008). doi:10.1038/nature06810 Medline

62. D. A. Benson, I. Karsch-Mizrachi, D. J. Lipman, J. Ostell, D. L. Wheeler, GenBank. *Nucleic Acids Res.* **36**, D25 (2008). doi:10.1093/nar/gkm929 Medline

63. S. Sun *et al*., Community Cyberinfrastructure for Advanced Microbial Ecology Research and Analysis: The CAMERA resource. *Nucleic Acids Res.* **39**, D546 (2011). doi:10.1093/nar/gkq1102 Medline

64. F. Meyer *et al*., The metagenomics RAST server - a public resource for the automatic phylogenetic and functional analysis of metagenomes. *BMC Bioinformatics* **9**, 386 (2008). doi:10.1186/1471-2105-9-386 Medline

65. H. Noguchi, J. Park, T. Takagi, MetaGene: Prokaryotic gene finding from environmental genome shotgun sequences. *Nucleic Acids Res.* **34**, 5623 (2006). doi:10.1093/nar/gkl723 Medline

66. A. Marchler-Bauer *et al*., CDD: A Conserved Domain Database for the functional annotation of proteins. *Nucleic Acids Res.* **39**, D225 (2011). doi:10.1093/nar/gkq1189 Medline

67. E. P. Nawrocki, D. L. Kolbe, S. R. Eddy, Infernal 1.0: Inference of RNA alignments. *Bioinformatics* **25**, 1335 (2009). doi:10.1093/bioinformatics/btp157 Medline

68. S. Griffiths-Jones, RALEE—RNA Alignment editor in Emacs. *Bioinformatics* **21**, 257 (2005). doi:10.1093/bioinformatics/bth489 Medline

69. Z. Weinberg *et al*., Identification of 22 candidate structured RNAs in bacteria using the CMfinder comparative genomics pipeline. *Nucleic Acids Res.* **35**, 4809 (2007). doi:10.1093/nar/gkm487 Medline

70. W. Li, A. Godzik, Cd-hit: A fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* **22**, 1658 (2006). doi:10.1093/bioinformatics/btl158 Medline

71. S. F. Altschul *et al*., Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Res.* **25**, 3389 (1997). doi:10.1093/nar/25.17.3389 Medline

72. Z. Weinberg, R. R. Breaker, R2R—software to speed the depiction of aesthetic consensus RNA secondary structures. *BMC Bioinformatics* **12**, 3 (2011). doi:10.1186/1471-2105-12-3 Medline

73. G. André *et al*., S-box and T-box riboswitches and antisense RNA control a sulfur metabolic operon of *Clostridium acetobutylicum*. *Nucleic Acids Res.* **36**, 5955 (2008). doi:10.1093/nar/gkn601 Medline

74. M. Gerstein, E. L. Sonnhammer, C. Chothia, Volume changes in protein evolution. *J. Mol. Biol.* **236**, 1067 (1994). doi:10.1016/0022-2836(94)90012-4 Medline

75. P. P. Gardner, L. Barquist, A. Bateman, E. P. Nawrocki, Z. Weinberg, RNIE: Genome-wide prediction of bacterial intrinsic terminators. *Nucleic Acids Res.* **39**, 5845 (2011). doi:10.1093/nar/gkr168 Medline

76. S. Guindon, O. Gascuel, A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst. Biol.* **52**, 696 (2003). doi:10.1080/10635150390235520 Medline

77. M. Anisimova, O. Gascuel, Approximate likelihood-ratio test for branches: A fast, accurate, and powerful alternative. *Syst. Biol.* **55**, 539 (2006). doi:10.1080/10635150600755453 Medline

78. F. D. Ciccarelli *et al*., Toward automatic reconstruction of a highly resolved tree of life. *Science* **311**, 1283 (2006). doi:10.1126/science.1123061 Medline

79. I. Letunic, P. Bork, Interactive Tree Of Life (iTOL): An online tool for phylogenetic tree display and annotation. *Bioinformatics* **23**, 127 (2007). doi:10.1093/bioinformatics/btl529 Medline

80. J. Emsley, Very strong hydrogen bonding. *Chem. Soc. Rev.* **9**, 91 (1980). doi:10.1039/cs9800900091

81. J. W. Larson, T. B. McMahon, *J. Am. Chem. Soc.* **105**, 2294 (1983).

82. J. Emsley, M. Arif, P. A. Bates, M. B. Hursthouse, Hydrogen bonding between free fluoride ions and water molecules: two x-ray structures. *J. Mol. Struct.* **220**, 1 (1990). doi:10.1016/0022-2860(90)80094-Z