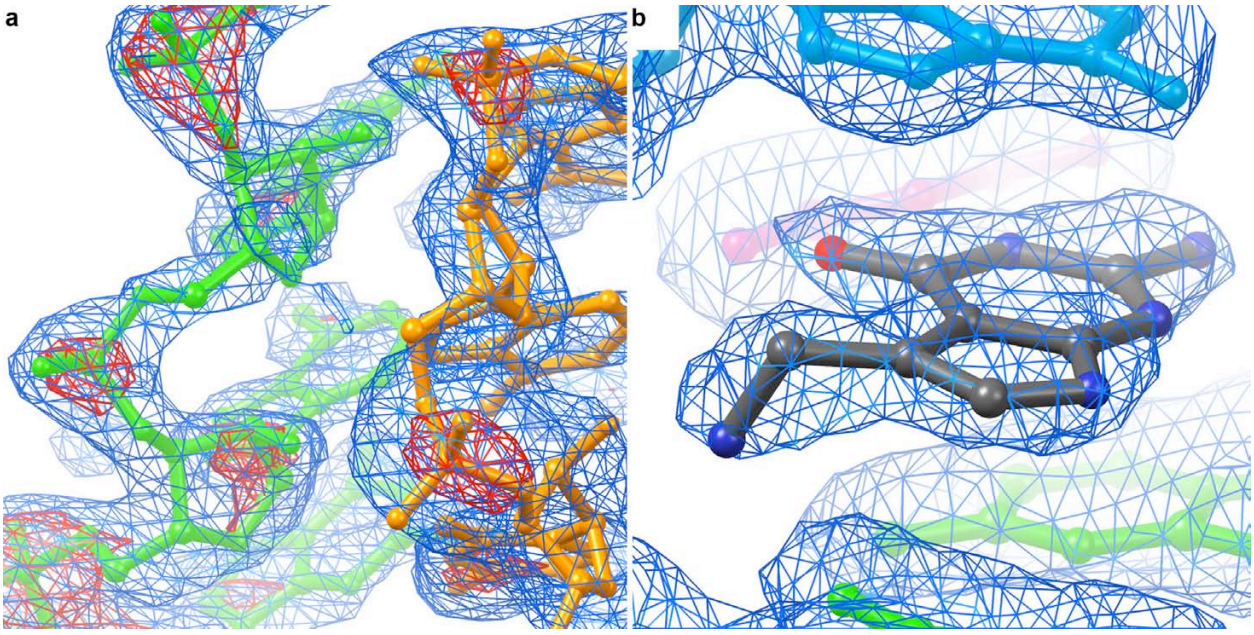
Supplementary Material

Cocrystal structure of a class-I preQ$_1$ riboswitch reveals a pseudoknot recognizing an essential hypermodified nucleobase

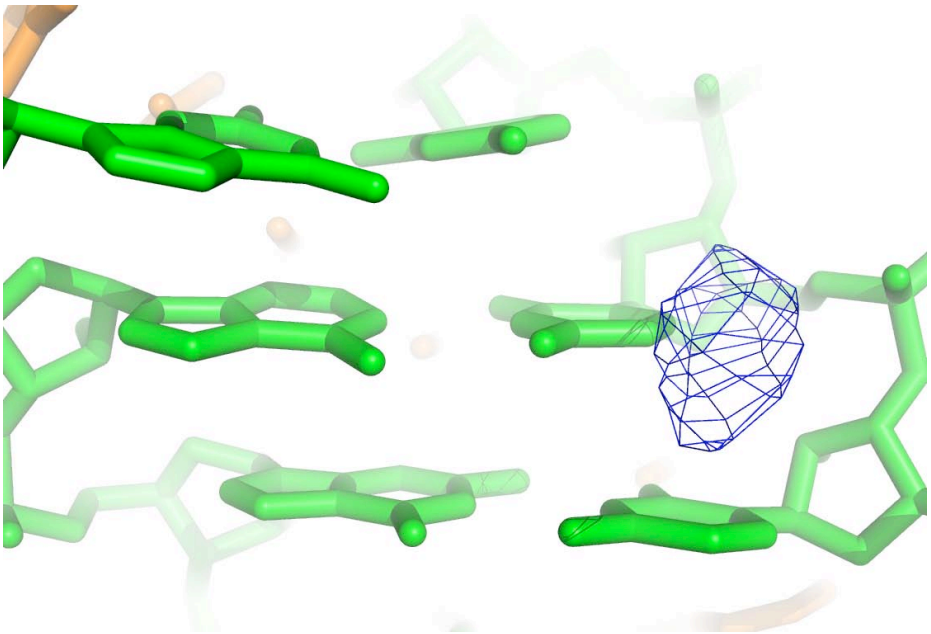
Daniel J. Klein[1,3], Thomas E. Edwards[1,3] & Adrian R. Ferré-D'Amaré[1,2]*

[1]Division of Basic Sciences and [2]Howard Hughes Medical Institute, Fred Hutchinson Cancer Research Center, 1100 Fairview Avenue North, Seattle, WA 98109-1024, USA. [3]Present addresses: Merck Global Structural Biology, Merck & Co., West Point, PA 19486, USA (D.J.K.); deCODE Biostructures, 7869 Northeast Day Road West, Bainbridge Island, WA 98110, USA (T.E.E.)

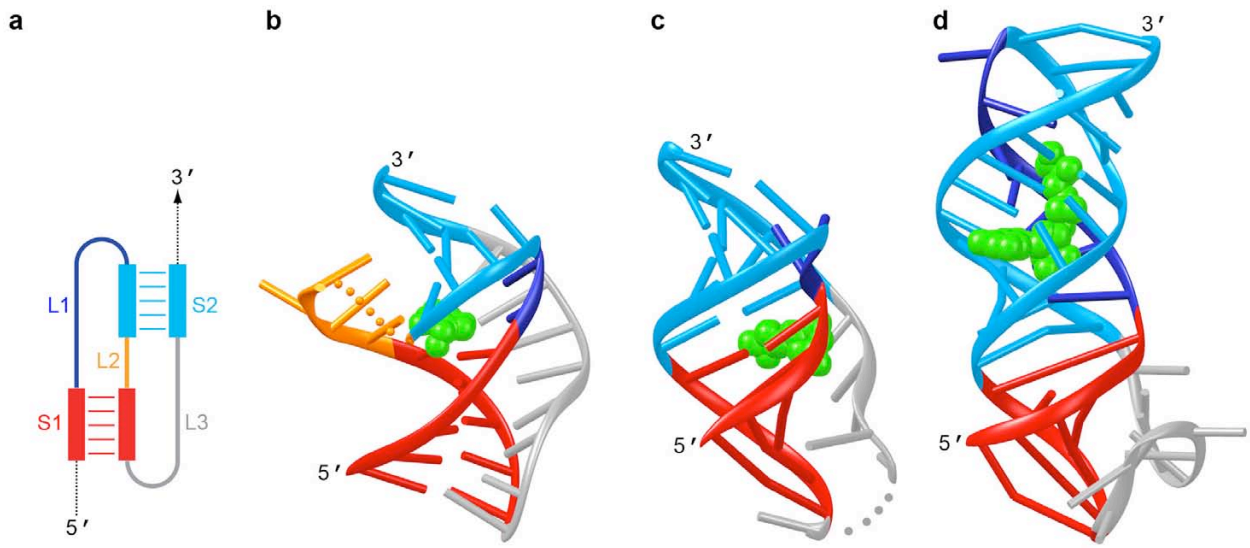
*Address correspondence to Adrian R. Ferré-D'Amaré, e-mail: aferre@fhcrc.org, Telephone: (206) 667-3622, Fax: (206) 667-3331



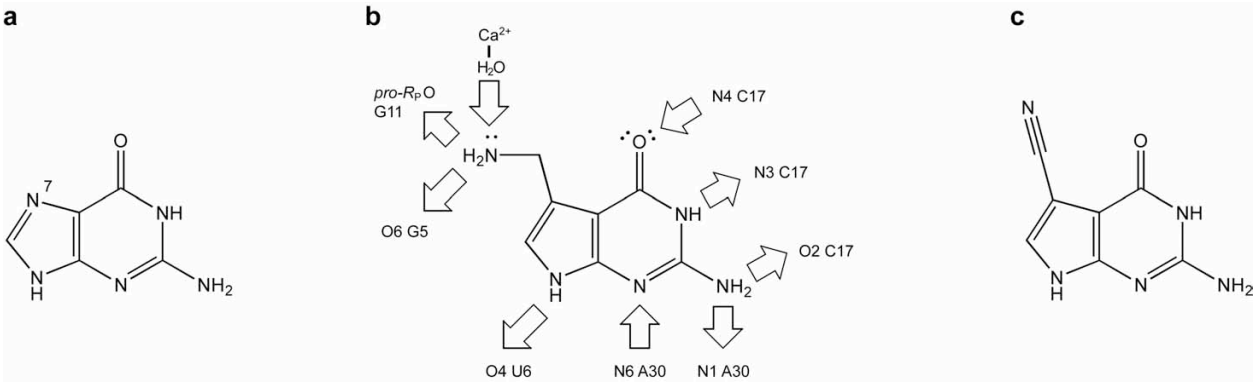
Supplementary Figure 1. Simulated-annealing omit electron density for Crystal I. (**a**) Portion of the composite simulated annealing-omit $2|F_o|-|F_c|$ electron density map calculated with the refined Crystal I model and all reflections extending to 2.2 Å resolution, contoured at 1.0 and 3.0 σ (blue and red mesh, respectively). (**b**) Portion of an omit-anneal $2|F_o|-|F_c|$ electron density map calculated with the refined Crystal I model and all reflections extending to 2.2 Å resolution, by omitting the $preQ_1$ molecule and a spherical cushion of 2 Å radius around it prior to simulated annealing, contoured at 1.0 σ.



Supplementary Figure 2. Evidence for correct sequence register of the crystallographic model. Portion of the anomalous difference Fourier synthesis calculated with amplitudes from Crystal III, and phases from the molecular replacement solution using the refined structure from Crystal I as the search model (see **Supplementary Methods**, online), contoured at 7 σ (blue mesh). The 10.5 σ peak is consistent with the presence of a 5-iodouracil at this position of the RNA, confirming the sequence register of the crystallographic model.



Supplementary Figure 3. Comparison of the structures of three independently evolved small molecule binding H-type (or ABAB) pseudoknots. (**a**) Schematic depiction of the structural elements of an H-type pseudoknot (after Supplementary ref. 1) (**b**) Cartoon representation of the structure of the wild-type *B. subtilis* preQ$_1$ riboswitch. The RNA is colored as in (a); the bound ligand is in green. Uniquely among the three RNA structures in this figure, this riboswitch RNA has an L2 loop (yellow) of non-zero length. (**c**) Cartoon representation of the structure of the biotin aptamer[2] (PDB code 1F27). The bound biotin is colored green. In this crystal structure, there is a covalent break in the RNA backbone (gray spheres). Note binding of biotin between helices S1 and S2, despite the RNA having an L2 of zero length. (**d**) Cartoon representation of the structure of the class II S-adenosylmethionine (SAM-II) riboswitch[3] (PDB code 2QWY), colored as in the preceding panels. This pseudoknotted RNA also has L2 of length zero. The adenine moiety of S-adenosylmethionine binds into a pocket formed by a gap in the sequence of helix S2.



Supplementary Figure 4. Discrimination between G, $preQ_1$ and $preQ_0$ by the *B. subtilis* class-I riboswitch. (**a**) Chemical structure of G. The nitrogen at position 7, a potential hydrogen bond acceptor, is indicated. (**b**) Structure of $preQ_1$. Note the aminomethyl group that is attached to a *carbon* at position 7. The $preQ_1$ riboswitch needs to discriminate against guanine, and does so employing the full hydrogen bonding potential of the amine of the aminomethyl group. Thick arrows denote hydrogen bonds made by the $preQ_1$ riboswitch to $preQ_1$. Except for one lone pair of the carbonyl oxygen at position 6, every possible hydrogen bond donor and acceptor of the small molecule is recognized by the riboswitch. (**c**) Structure of $preQ_0$, the immediate biosynthetic precursor to $preQ_1$. This compound has a nitrile at position 7. Presumably, the riboswitch discriminates against $preQ_0$ because the nitrile group cannot donate hydrogen bonds, and because its linear structure clashes against the binding pocket of the riboswitch which is shape-complementary to the bent (due to the tetrahedral methylene group) structure of the aminomethyl group.

Start Codon PreQ1 switch Possible antiterminator pairs Terminator

>Lysinibacillus sphaericus C3-41 (NC_010382.1)
Residues shown: 3036970…3037103
Downstream gene: NADPH-dependent 7-cyano-7-deazaguanine reductase
ON:

```
agaggttcat agctgatacc ctctataaaa aactatggaa acatgacaat ttgccatgtc
catatcggac gtggcttttt tatgtttaga cgtatgtttt aatagatgaa ttggtaaagc
ctctcttgaa agtg
```

OFF:

```
agaggttcat agctgatacc ctctataaaa aactatggaa acatgacaat ttgccatgtc
catatcggac gtggcttttt tatgtttaga cgtatgtttt aatagatgaa ttggtaaagc
ctctcttgaa agtg
```

>Staphylococcus saprophyticus subsp. Saprophyticus ATCC 15305 (NC_007350.1)
Residues shown: 2076129…2076272
Downstream gene: Hypothetical protein YP_302098.1
ON:

```
agaggttcct agctgatacc ctctataaaa aactagacac atgtacaacg tctgtctttt
ttatagagat aggcgttttt ttatgcgctt atctaaaccc tgtaccagtt agttcgacta
tttttaagga gtgcaagaca aatg
```

OFF:

```
agaggttcct agctgatacc ctctataaaa aactagacac atgtacaacg tctgtctttt
ttatagagat aggcgttttt ttatgcgctt atctaaaccc tgtaccagtt agttcgacta
tttttaagga gtgcaagaca aatg
```

>Oceanobacillus iheyensis HTE831 (NC_004193.1)
Residues show: 2245681…2245793
Downstream gene: 7-cyano-7deazaguanine reductase
ON:

```
agaggttctt agcttcaacc ctctataaaa aactaaggac aaacaatgat ttattcaagt
catgttcctt agggacatg actttttta tggaaaagga ggtaataagt atg
```

OFF:

```
agaggttctt agcttcaacc ctctataaaa aactaaggac aaacaatgat ttattcaagt
catgttcctt agggacatg actttttta tggaaaagga ggtaataagt atg
```

>Geobacillus thermodenitrificans NG80-2 (NC_009328.1)
Residues shown: 940287…940404
Downstream gene: aluminum resistance protein
ON:

```
     agagg tttctagcaa aaccctctat aaaaaactag ggacggctgt atcctagggg
tatggccttt tttgttttgg aaacctcttc ttctgtcgat tcagaaggag gttagtcaag
atg
```

OFF:

```
     agagg tttctagcaa aaccctctat aaaaaactag ggacggctgt atcctagggg
tatggccttt tttgttttgg aaacctcttc ttctgtcgat tcagaaggag gttagtcaag
atg
```

>Bacillus pumilus SAFR-032 (NC_009848.1)
Residues shown: 1324519…1324720
Downstream gene: transcriptional regulator
ON

```
     agagg ttctagcaaa accctctata aaaaactaag gacttgccgt atccttggat
gtggcatttt ttatttttta tggctgggac acctcataca catgtaaagg tgtcttttct
tttgagaaaa ataggggttt tcatgagcct ttcagatgca gctttgtacg atggtataga
tctgcaaaag gaggagaaca gaagatg
```

OFF

```
     agagg ttctagcaaa accctctata aaaaactaag gacttgccgt atccttggat
gtggcatttt ttatttttta tggctgggac acctcataca catgtaaagg tgtcttttct
tttgagaaaa ataggggttt tcatgagcct ttcagatgca gctttgtacg atggtataga
tctgcaaaag gaggagaaca gaagatg
```

>Bacillus subtilis subsp. subtilis str. 168 (NC_000964.2)
Residues shown: 1438587…1438758
Downstream gene: ykvJ
ON:

```
     agagg ttctagctac accctctata aaaaactaag gacgagctgt atccttggat
acggcctttt ttatgttttt ctagagcacc ttccgaaaaa aggtgttttt ttgcgtgaat
tagctgtagc gatgtctctc gccggcgttt ttattgcgga gaaggagagg aatcatg
```

OFF:

```
     agagg ttctagctac accctctata aaaaactaag gacgagctgt atccttggat
acggcctttt ttatgttttt ctagagcacc ttccgaaaaa aggtgttttt ttgcgtgaat
tagctgtagc gatgtctctc gccggcgttt ttattgcgga gaaggagagg aatcatg
```

Supplementary Figure 5. Sequence analysis of selected 5'-UTRs that contain class I, sub-type II preQ$_1$ riboswitches. Each of these UTRs appear to control gene expression by transcriptional attenuation in a preQ$_1$-dependent manner. Highlighted residues in each sequence include the preQ$_1$ binding domain (yellow), anti-terminator helix (red), terminator helix (cyan) and start codon (green).

Although this figure is limited to class-I preQ$_1$ riboswitches, the sequestration of a 3' single-stranded extension into the riboswitch as a function of metabolite binding could be a strategy more widely employed by other riboswitches. Indeed class-II preQ1 riboswitches are thought to fold into H-type pseudoknots [4].

At present there is no evidence whether the two putative secondary structures shown in **Fig. 1c** interconvert, or whether formation of either (in the absence or presence of sufficient concentration of preQ1, respectively) is under kinetic control, as in the case of the FMN riboswitch studied by Wickiser et al. [5]. Therefore, the secondary structures of the UTR shown in that figure remain purely hypothetical.

Supplementary Table 1. Data collection and refinement statistics.

	Crystal I	Crystal II	Crystal III
Data collection			
Space group	*R*32	$P3_221$	*R*32
Cell dimensions			
a, *b*, *c* (Å)	76.81, 76.81, 85.90	78.21, 78.21, 85.22	78.40, 78.40, 459.94
α, β, γ (°)	90, 90, 120	90, 90, 120	90, 90, 120
Resolution (Å)	30.0-2.2 (2.3-2.2)*	30-2.85 (2.95-2.85)	30-3.0 (3.1-3.0)
R_{merge}	8.5 (41.4)	12.6 (46.0)	12.5 (34.8)
$\langle I \rangle / \langle \sigma(I) \rangle$	23.1 (2.2)	23.7 (3.4)	28.1 (8.1)
Completeness (%)	95.0 (69.1)	96.8 (68.3)	97.5 (86.3)
Redundancy	8.8 (2.6)	11.3 (9.2)	12.3 (9.6)
Refinement			
Resolution (Å)	30.0-2.2	30-2.85	
No. reflections	4592 (483)†	7122 (390)†	
R_{work} / R_{free} (%)	23.0/26.8	21.4/26.7	
No. atoms			
RNA	614	1991	
$PreQ_1$	13	39	
Ion	4	9	
Water	62	21	
B-factors			
RNA	34.2	52.5	
$PreQ_1$	22.7	43.2	
Ion	25.7	60.2	
Water	34.8	30.7	
R.m.s. deviations			
Bond lengths (Å)	0.00363	0.00527	
Bond angles (°)	1.3252	1.3270	

*A single crystal was used for each dataset. *Values in parentheses are for highest-resolution shell.
†Values in parentheses are for the cross-validation set.

Supplementary Table 2. Tertiary interactions formed by L1, L2, and L3 residues of the preQ$_1$ riboswitch (nucleotides from other segments of the RNA identified in black).

Pair partners	Distance (Å)	Type of interaction
U6 N3 – A29 N7	2.7	*Cis* Watson-Crick • Hoogsteen pair
U6 O4 – A29 N6	2.8	
U7 N3 – A30 N7	3.1	*Cis* Watson-Crick • Hoogsteen pair
U7 O4 – A30 N6	3.1	
G11 N2 – A16 N1	3.1 [a]	*Trans* sugar edge • Watson-Crick pair
G11 N3 – A16 N6	3.0 [a]	
A14 N1 – U32 O2'	2.8 [a]	"Reverse Type II" A-minor motif
A14 N6 – U32 O2	3.3 [a]	
A23 N1 – G2 O2'	2.9	"Reverse Type II" A-minor motif
A23 N6 – G2 N3	3.1	
U24 O4 – G2 N2	3.0	Single H-bond
A25 N1 – C21 O2'	2.6	Single H-bond; part of triple with below.
A25 N6 – A3 N3	2.9	*Cis* Hoogsteen•sugar edge; with preceding,
A25 N7 – A3 O2'	3.2	part of inclined Hoogsteen • minor groove
A26 N1 – U20 O2'	3.1	*Cis* Watson-Crick•sugar edge; with next,
A26 N6 – U20 O2	2.7	part of inclined Hoogsteen • minor groove
A26 N6 – G4 N3	2.9	*Cis* Hoogsteen•sugar edge; with preceding,
A26 N7 – G4 O2'	2.6	part of inclined Hoogsteen • minor groove
A27 N1 – C19 O2'	2.7	*Cis* Hoogsteen•sugar edge; together with below,
A27 N6 – C19 O2	2.9	part of a triple.
A27 N6 – G4 N2	3.2	Single H-bond; part of a triple with above
A28 N1 – G5 N2	3.0	*Trans* Watson-Crick • sugar edge
A28 N6 – G5 N3	3.1	
A28 N6 – G5 O2'	2.9	

[a]These distances are from the Crystal II structure (chain B). All others are from the Crystal I structure.

Supplementary Methods

Crystallization and data collection

Three 34-nt oligonucleotides of sequence 5'-AGA GGU UCU AGC UAC ACC CUC UAU AAA AAA CUA A-3' (Crystal I), 5'-AGA GGU UCU AGC ACA UCC CUC UAU AAA AAA CUA A-3´ (Crystal II), and 5'-AGA GGU UCU AGC ACA UCC CXC UAU AAA AAA CUA A-3' (where X denotes 5-iodouridine, Crystal III) were purchased from Dharmacon, Inc., deprotected following manufacturer's instructions, and used without further purification. The underlined 4-nt segment differs between the RNA that yielded Crystal I and the two other RNAs. The RNAs used for Crystal II and Crystal III correspond to the wild-type *B. subtilis queC* 5'-UTR sequence[6] (excepting the 5-iodouridine in the Crystal III RNA). The sequence of the RNA used for Crystal I is a *B. subtilis queC* 5'-UTR sequence variant reported by [7]. Although the underlined nucleotides are disordered in our Crystal I structure (see below), this sequence variant produced better diffracting crystals with a smaller unit cell than the wild-type sequence (**Supplementary Table 1**, online). Purified $preQ_1$ was a gift from Professor Raven Huang (University of Illinois at Urbana-Champaign).

All crystals were grown by vapor diffusion at 296 K. For Crystal I, an RNA solution containing 0.37 mM RNA, 0.5 mM $preQ_1$, 10 mM $MgCl_2$, 10 mM Hepes-KOH pH 7.5, and 1 mM spermine was incubated for 15 minutes at 315 K prior to mixing with an equal volume of reservoir solution comprised of 30% (v/v) polyethylene glycol (PEG) 200, 50 mM $CaCl_2$, 100 mM Hepes-KOH pH 7.0. For crystal II, an RNA solution containing 0.37 mM RNA, 0.5 mM $preQ_1$, 10 mM $MgCl_2$, 10 mM Hepes-KOH, pH 7.5, and 1 mM spermine was incubated for 15 minutes at 315 K prior to mixing with an equal volume of reservoir solution comprised of 25% (v/v) PEG monomethyl ether (MME) 550, 50 mM $CaCl_2$, 100 mM Hepes-KOH pH 7.0. For crystal III, an RNA solution containing 0.37 mM RNA, 0.5 mM $preQ_1$, 10 mM $MgCl_2$, 10 mM Hepes-KOH pH 7.5, 50 mM $CaCl_2$ was incubated for 15 minutes at 315 K, and briefly centrifuged to remove precipitated RNA prior to mixing with an equal volume of reservoir solution comprised of 35% (v/v) pentaerythritol propoxylate (17/8 PO/OH) (Hampton Research) and 0.2 M KCl. Crystals (unit cell parameters in **Supplementary Table 1**, online) appeared within one week and grew to maximum dimensions of 150x150x30 μm^3.

Crystal I was transferred to 50% (v/v) PEG 200, 50 mM $CaCl_2$, 100 mM Hepes-KOH pH 7.0 prior to flash-freezing by plunging into liquid nitrogen while supported in a nylon loop. Crystal II was transferred to 45% (v/v) PEG 550 MME, 50 mM $CaCl_2$, 100 mM Hepes-KOH pH 7.0 prior to flash-freezing. Crystal III flash frozen directly from mother liquor by plunging into liquid nitrogen while supported in a nylon loop. Diffraction data were collected in oscillation mode at 100 K at beamline 5.0.1 of the Advanced Light Source using 1.0 Å X-radiation, Lawrence Berkeley National Laboratory for Crystal I, and with a copper rotating anode source (1.5418 Å X-radiation) for Crystals II and III, and reduced with the HKL package[8].

Structure determination and refinement

The riboswitch structure in Crystal I was determined by molecular replacement with PHASER[9,10] using two A-form RNA duplexes of sequence 5'-GAGGG-3' and 5'-UAG-3' (and their complements) as independent starting search models. Simulated annealing, energy minimization, and individual isotropic *B*-factor refinement[11] against all data (with a random 10%

excluded for cross-validation) to 2.2 Å of the top solution (*Z*-scores for rotation and translation searches of 3.2 and 4.9, respectively; overall log-likelihood gain of 140) resulted in a model with free-*R* factor of 50.1%, and an electron density map with additional interpretable features. Eighteen iterative rounds of manual model building[12], simulated annealing, energy minimization, and individual isotropic *B*-factor refinement[11] produced the current model of the single riboswitch RNA in the asymmetric unit of this crystal form (statistics in **Supplementary Table 1** online). A composite simulated annealing omit $2|F_o|-|F_c|$ map (ref [11]) calculated with the final model shows clear electron density for all the RNA, except residues 12-16, inclusive (**Supplementary Fig. 1a** online). The bound $preQ_1$ molecule has electron density that clearly indicates the conformation of its aminomethyl group in an anneal-omit $2|F_o|-|F_c|$ map (**Supplementary Fig. 1b** online). Analysis of dissolved crystals on 8M Urea-PAGE demonstrated that the RNA in the crystalline state is intact (not shown). Therefore, the lack of electron density for residues 12-16 indicates crystallographic disorder, not RNA scission. The σA coordinate error for this structure is 0.32 Å (0.38 Å for the cross-validation set).

The riboswitch structure in Crystal II was solved by molecular replacement with PHASER[9] using the refined Crystal I model to search for the three crystallographically independent riboswitch RNAs in the asymmetric unit of Crystal II. The initial solution had *Z*-scores for rotation and translation searches of 4.6 and 31.4, respectively, and an overall log-likelihood gain of 1428. Iterative rounds of manual model building[12], simulated annealing, energy minimization, and individual isotropic *B*-factor refinement[11] produced the current model, which comprises the complete RNA chain, except residues 12 and 13 and the nucleobase of residue 15, which are presumed disordered (statistics in **Supplementary Table 1** online). The σA coordinate error for this structure is 0.65 Å (0.76 Å for the cross-validation set).

The sequence register of the crystallographic model was confirmed by carrying out molecular replacement against the Crystal III dataset with the refined model from Crystal I as the search model. Anomalous difference Fourier syntheses with the model phases from the top PHASER solution (*Z*-scores for rotation and translation searches of 4.6 and 18.6, respectively, and an overall log-likelihood gain of 288) show an electron density feature with a peak height of 10.5 standard deviations above mean peak height (σ) adjacent to the 5 position of residue U20 (**Supplementary Fig. 2** online).

Figures 1a, 1b, and 2b depict chain B from the Crystal II structure of the wild-type *B. subtilis queC* 5'-UTR sequence RNA[6]. Figures 1c, 1d, and 2a depict the model from the Crystal I structure of the *B. subtilis queC* 5'-UTR sequence variant reported by Barrick et al.[7] Figures were prepared with RIBBONS[13] and PyMOL[14]. Solvent accessible surface areas of $preQ_1$ were calculated using CNS[11] with a probe radius of 1.4 Å (and the Crystal I structure, for the bound state).

Supplementary references

1. Aalberts, D.P. & Hodas, N.O. *Nucleic Acids Res* **33**, 2210-2214 (2005).
2. Nix, J., Sussman, D. & Wilson, C. *J Mol Biol* **296**, 1235-1244 (2000).
3. Gilbert, S.D., Rambo, R.P., Van Tyne, D. & Batey, R.T. *Nat Struct Mol Biol* **15**, 177-182 (2008).
4. Meyer, M.M., Roth, A., Chervin, S.M., Garcia, G.A. & Breaker, R.R. *RNA* **14**, 685-695 (2008).
5. Wickiser, J.K., Winkler, W.C., Breaker, R.R. & Crothers, D.M. *Mol Cell* **18**, 49-60 (2005).
6. Roth et al. *Nat Struct Mol Biol* **14**, 308-317 (2007).
7. Barrick, J.E. et al. *Proc Natl Acad Sci USA* **101**, 6421-6426 (2004).
8. Otwinowski, Z. & Minor, W. *Meth Enzymol.* **276**, 307-326 (1997).
9. McCoy, A.J., Grosse-Kunstleve, R.W., Storoni, L.C. & Read, R.J. *Acta Crystallogr D* **61**, 458-464 (2005).
10. Robertson, M.P. & Scott, W.G. *Acta Crystallogr D* **64**, 738-744 (2008).
11. Brünger, A.T. et al. *Acta Crystallogr D* **54**, 905-921 (1998).
12. Jones, T.A., Zou, J.Y., Cowan, S.W. & Kjeldgaard, M. *Acta Crystallogr A* **47**, 110-119 (1991).
13. Carson, M. *Meth Enzymol.* **277**, 493-505 (1997).
14. DeLano, W.L. *The PyMOL Molecular Graphics System*, (DeLano Scientific, San Carlos, 2002).