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

Riboswitches are mRNA domains that bind metabolites and modulate gene expression in *cis*. We report cocrystal structures of a remarkably compact riboswitch (34 nucleotides suffice for ligand recognition) from *Bacillus subtilis* selective for the essential nucleobase $preQ_1$ (7-aminomethyl-7-deazaguanine). These reveal a previously unrecognized pseudoknot fold, and suggest a conserved gene-regulatory mechanism whereby ligand binding promotes sequestration of an RNA segment that otherwise assembles into a transcriptional anti-terminator.

Queuosine (Q) is a post-transcriptional modification of the wobble position of GUN anticodons of certain bacterial and eukaryal tRNAs. It is important for translational fidelity (reviewed in ref. 1). During tRNA maturation, a transglycosylase replaces the guanine at this pre-tRNA position with free $preQ_1$, which is subsequently elaborated into Q. The 5'-untranslated region (UTR) of an operon encoding Q biosynthetic enzymes in many bacteria harbors a $preQ_1$-specific riboswitch, consistent with regulation of the pathway by the last small-molecule intermediate in the biogenesis of Q (ref. 2). Sequence differences define two sub-types of class-I $preQ_1$ riboswitches, which probably employ the same metabolite recognition motif[2]. Structurally distinct (class-II) $preQ_1$-responsive riboswitches were recently described[3].

Sequence analyses led Roth *et al*., to propose that class-I $preQ_1$ riboswitches consist simply of a stem-loop followed by a short single-stranded segment[2]. Characterization of the *B. subtilis queC* 5'-UTR demonstrated that an RNA as short as 34 nucleotides (nt) binds $preQ_1$ (K_d ~20 nM), discriminating between $preQ_1$ and G through the 7-aminomethyl group unique to the former[2]. To elucidate how such a small RNA achieves high affinity binding of an essential metabolite, and to provide a framework to understand how this riboswitch modulates gene expression, we determined $preQ_1$ complex structures of the 34-nt core of the class I, sub-type II *B. subtilis* $preQ_1$ riboswitch[2] and of a sequence variant[4], at 2.85 and 2.2 Å resolution, respectively **(Supplementary Figs. 1 and 2**, **Supplementary Table 1**, and **Supplementary Methods** online, PDB 3FU2 and 3FU4). Structure determination was by molecular replacement using arbitrary RNA duplexes as initial search models, similarly to what was recently reported for two ribozymes[5,6].

The *B. subtilis queC* $preQ_1$ riboswitch folds into an H-type pseudoknot (**Figs. 1a,b**). Two stems separated by three loops define this most abundant type of pseudoknot. In H-type pseudoknots, loops L1 and L3 lie in the major and minor grooves of stems S2 and S1, respectively. L1 and L3 are typically uracil and adenine rich, respectively (reviewed in ref. 7). The $preQ_1$ riboswitch conforms to these trends with an L1 of two uracils, and an L3 of seven adenines and one uracil. U6 and U7 of L1 make Watson-Crick•Hoogsteen pairs with A29 and A30, the last two nucleotides of L3. L3 residues often form triplex interactions[7]. L3 of the $preQ_1$ riboswitch is noteworthy for the sheer density of these, including inclined adenosine Hoogsteen face interactions (different from the *glmS* and SAM-II riboswitch "inclined A-minor" motif[8,9]), interactions with L1, and with $preQ_1$ (**Fig. 1c**, and **Supplementary Table 2** online). As in other H-type pseudoknots[10], divalent cations stabilize the sharp turn after L1 (**Fig. 1d**).

Over 75% of H-type pseudoknots[7] (including the biotin aptamer[11] and the SAM-II riboswitch[9]; **Supplementary Fig. 3** online) have 0-nt L2, and an additional 10% have 1-nt L2. A minimal L2 allows efficient coaxial stacking of S1 and S2 (ref. 7). In contrast, the $preQ_1$ riboswitch has a 6-nt L2 (**Fig. 1a**). Nonetheless, intercalation of $preQ_1$ at the interhelical interface (between G11 and the G5•C18 pair) maintains continuous stacking of the two stems (**Figs. 1b,2a**). C17, the last L2 residue, Watson-Crick pairs with $preQ_1$, as predicted[2]. A16, stacks on C17 and also pairs with G11 (at the bottom of S2). C15 is extruded into solvent, and A14 stacks on A16. C12 and U13 are disordered in our wild-type structure. In addition to pairing with C17, $preQ_1$ makes a *trans* sugar edge•Watson-Crick pair with A30, and hydrogen bonds with its N9 to U6. The aminomethyl group of $preQ_1$ hydrogen bonds to G5, the *pro-*R_P phosphate oxygen of G11, and a hydration water of a Ca^{2+} ion. Thus, nine of the ten potential hydrogen-bonding groups of $preQ_1$

are recognized by the riboswitch, which buries 92% of the solvent accessible surface area of its ligand.

Despite the similarity between $preQ_1$ and G (**Supplementary Fig. 4**, online), the riboswitches that recognize these two metabolites appear to have evolved independently. The G riboswitch binds its ligand in a pocket formed by a 3-helix junction[12]. Structural and biochemical data suggest that formation of helix P1 of this junction depends strongly on G binding (reviewed in ref. 13). P1 stabilization promotes transcriptional termination by precluding formation of an anti-terminator stem-loop. Analogous transcriptional attenuation mechanisms have been postulated for riboswitches of most classes, excepting the *glmS* riboswitch-ribozyme[13].

Genetic control by the $preQ_1$ riboswitch was assumed to be at the transcriptional level, owing to the presence of a nearby terminator stem[2]. However, unlike the case with most known riboswitches that utilize transcriptional attenuation, it was not obvious from the proposed secondary structure of the $preQ_1$ riboswitch how it might adopt alternative folds corresponding to genetic on/off states[2]. Our discovery of a pseudoknot in the $preQ_1$ riboswitch suggests a model in which transcriptional terminator formation depends on stabilization of helix S2 (**Fig. 2b**). Without formation of S2, nucleotides at the 3'-end of the $preQ_1$ riboswitch could form part of an anti-terminator stem (**Fig. 2c** and **Supplementary Fig. 5**, online), thereby promoting synthesis of the full-length mRNA by RNA polymerase. On the basis of our structure and the previously reported in-line probing data[2], we speculate that $preQ_1$ binding promotes formation of helix S2 in a manner analogous to how G stabilizes helix P1 of the G riboswitch. Therefore, S2 in the $preQ_1$

riboswitch is functionally equivalent to P1 in the G riboswitch, and S1 of the $preQ_1$ riboswitch plays no direct role in gene regulation.

Figure 1. $PreQ_1$ riboswitch structure. (**a**) Secondary structure. Thin lines denote connectivity; outline letters, disordered nucleotides (base-pairing symbols, ref. 14). Not shown are tertiary interactions between S1 and L3. (**b**) Structure cartoon. Gray, yellow, and red spheres depict disordered portion of L2, Ca^{2+}, and water, respectively. (**c**) Select L3 tertiary interactions. (**d**) Partially hydrated Ca^{2+} ions stabilize the L1 turn. (a) and (b) depict the wild-type[2] structure; (c) and (d) that of the sequence variant[4].

Figure 2. $PreQ_1$ recognition and control of gene expression. (**a**) Phylogenetically conserved[2] binding pocket. (**b**) In-line probing data[2] mapped onto the structure. Blue: nucleotides with reduced scission in the presence of $preQ_1$. Crystallographically disordered C12 and U13 (red spheres) exhibit increased scission in the presence of $preQ_1$ (ref. 2). (**c**) Gene regulation. Absent $preQ_1$, one S2 strand (pink) instead forms part of an anti-terminator. $PreQ_1$ stabilizes S2, and allows formation of the terminator. (a) and (b) depict the sequence variant[4] and wild-type[2] structures, respectively.

Accession codes. Protein Data Bank: Coordinates and structure factors for the wild-type and sequence variant *B. subtilis* $preQ_1$ riboswitch-$preQ_1$ complexes have been deposited with accession codes 3FU2 and 3FU4, respectively.

Note: Supplementary information is available on the *Nature Structural & Molecular Biology* website.

Acknowledgements

We thank the staff of ALS beamline 5.0.1 and J. Bolduc for assistance with synchrotron and home laboratory data collection, respectively; R. Huang (University of Illinois at Urbana-Champaign) for his generous gift of $preQ_1$; and N. Baird, T. Hamma, C. Hoang, N. Kulshina, J. Pitt, J. Posakony, A. Roll-Mecak, and H. Xiao for discussions. D.J.K. and T.E.E. were Damon Runyon Fellows. This work was supported by grants from the Damon Runyon Cancer Research Foundation (DRG-1863-05 to D.J.K. and DRG-1844-04 to T.E.E.), the Howard Hughes Medical Institute (A.R.F. is an Investigator of the HHMI), the NIH (K99 GM084076-01 to D.J.K.), and the W.M. Keck Foundation (A.R.F. was a Distinguished Young Scholar in Medical Research).

1. Iwata-Reuyl, D. *Bioorg Chem* **31**, 24-43 (2003).

2. Roth et al. *Nat Struct Mol Biol* **14**, 308-317 (2007).

3. Meyer, M.M., Roth, A., Chervin, S.M., Garcia, G.A. & Breaker, R.R. *RNA* **14**, 685-695 (2008).

4. Barrick, J.E. et al. *Proc Natl Acad Sci USA* **101**, 6421-6426 (2004).

5. Robertson, M.P. & Scott, W.G. *Science* **315**, 1549-1553 (2007).

6. Xiao, H., Murakami, H., Suga, H. & Ferré-D'Amaré, A.R. *Nature* **454**, 358-361 (2008).

7. Aalberts, D.P. & Hodas, N.O. *Nucleic Acids Res* **33**, 2210-2214 (2005).

8. Klein, D.J. & Ferré-D'Amaré, A.R. *Science* **313**, 1752-1756 (2006).

9. Gilbert, S.D., Rambo, R.P., Van Tyne, D. & Batey, R.T. *Nat Struct Mol Biol* **15**, 177-182 (2008).

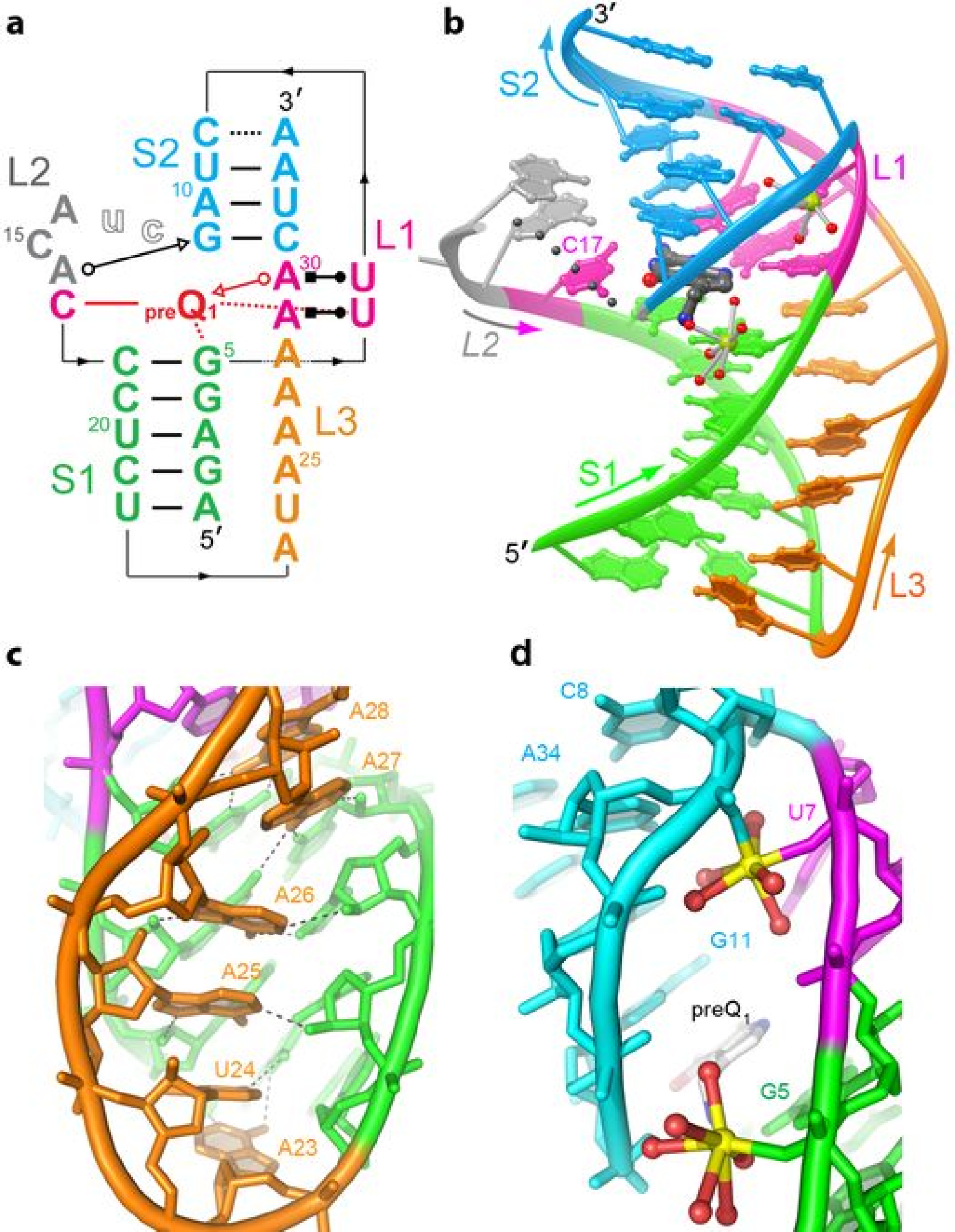
10. Egli, M., Minasov, G., Su, L. & Rich, A. *Proc Natl Acad Sci USA* **99**, 4302-4307 (2002).

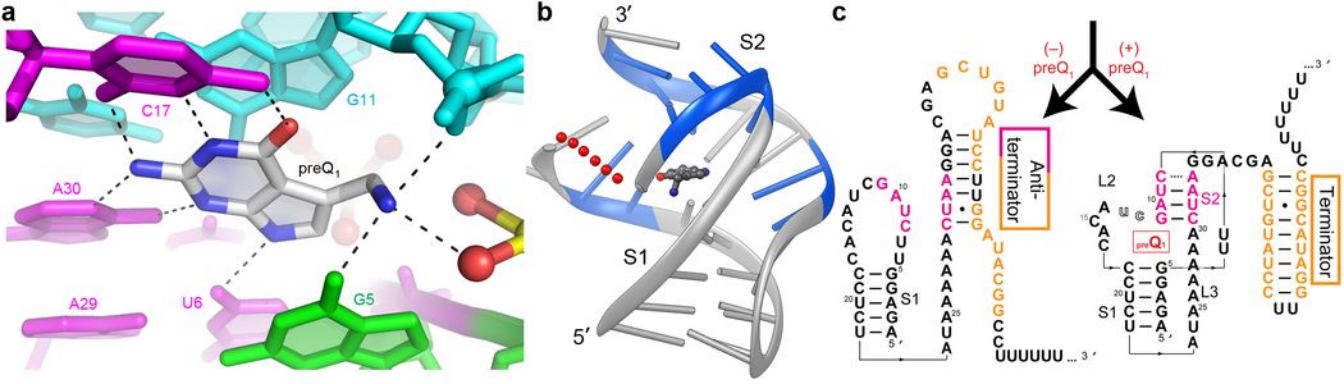
11. Nix, J., Sussman, D. & Wilson, C. *J. Mol. Biol.* **296**, 1235-1244 (2000).

12. Batey, R.T., Gilbert, S.D. & Montange, R.K. *Nature* **432**, 411-415 (2004).

13. Edwards, T.E., Klein, D.J. & Ferré-D'Amaré, A.R. *Curr Op Struct Biol.* **17**, 273-279 (2007).

14. Leontis, N.B. & Westhof, E. *RNA* **7**, 499-512 (2001).


a
3′
S2
L2
L1
L3
S1
5′
preQ1
b
3′
S2
L1
C17
L2
S1
5′
L3
c
A28
A27
A26
A25
U24
A23
d
C8
A34
U7
G11
preQ1
G5


a
C17
G11
preQ1
A30
A29
U6
G5
b
3′
S2
S1
5′
c
(–)
preQ1
(+)
preQ1
Anti-
terminator
L2
L3
S1
S2
Terminator