Cocrystal structure of a class-I preQ₁ riboswitch reveals a pseudoknot recognizing an essential hypermodified nucleobase

Daniel J. Klein¹,³, Thomas E. Edwards¹,³ & Adrian R. Ferré-D'Amaré¹,²*

¹Division of Basic Sciences and ²Howard Hughes Medical Institute, Fred Hutchinson Cancer Research Center, 1100 Fairview Avenue North, Seattle, WA 98109-1024, USA. ³Present 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 \textit{cis}. We report cocrystal structures of a remarkably compact riboswitch (34 nucleotides suffice for ligand recognition) from \textit{Bacillus subtilis} selective for the essential nucleobase \textit{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 preQ₁, which is subsequently elaborated into Q. The 5'‐untranslated region (UTR) of an operon encoding Q biosynthetic enzymes in many bacteria harbors a preQ₁-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 preQ₁ riboswitches, which probably employ the same metabolite recognition motif². Structurally distinct (class-II) preQ₁-responsive riboswitches were recently described³.

Sequence analyses led Roth et al., to propose that class-I preQ₁ riboswitches consist simply of a stem-loop followed by a short single-stranded segment². Characterization of the B. subtilis queC 5'-UTR demonstrated that an RNA as short as 34 nucleotides (nt) binds preQ₁ (Kₐ ~20 nM), discriminating between preQ₁ and G through the 7-aminomethyl group unique to the former². 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 preQ₁ complex structures of the 34-nt core of the class I, sub-type II B. subtilis preQ₁ riboswitch² and of a sequence variant⁴, 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 ribozymes⁵,⁶.
The *B. subtilis* queC preQ₁ 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 preQ₁ 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. L₃ of the preQ₁ 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), interactions with L₁, and with preQ₁ (Fig. 1c, and Supplementary Table 2 online). As in other H-type pseudoknots, divalent cations stabilize the sharp turn after L₁ (Fig. 1d).

Over 75% of H-type pseudoknots (including the biotin aptamer and the SAM-II riboswitch; 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 preQ₁ riboswitch has a 6-nt L₂ (Fig. 1a). Nonetheless, intercalation of preQ₁ 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 preQ₁, as predicted. 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, preQ₁ makes a trans sugar edge×Watson-Crick pair with A₃₀, and hydrogen bonds with its N₉ to U₆. The aminomethyl group of preQ₁ hydrogen bonds to G₅, the *pro-R*ₚ phosphate oxygen of G₁₁, and a hydration water of a Ca²⁺ ion. Thus, nine of the ten potential hydrogen-bonding groups of preQ₁
are recognized by the riboswitch, which buries 92% of the solvent accessible surface area of its ligand.

Despite the similarity between preQ\textsubscript{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\textsuperscript{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 \textit{glmS} riboswitch-ribozyme\textsuperscript{13}.

Genetic control by the preQ\textsubscript{1} riboswitch was assumed to be at the transcriptional level, owing to the presence of a nearby terminator stem\textsuperscript{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\textsubscript{1} riboswitch how it might adopt alternative folds corresponding to genetic on/off states\textsuperscript{2}. Our discovery of a pseudoknot in the preQ\textsubscript{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\textsubscript{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\textsuperscript{2}, we speculate that preQ\textsubscript{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\textsubscript{1}
riboswitch is functionally equivalent to P1 in the G riboswitch, and S1 of the preQ₁ riboswitch plays no direct role in gene regulation.
**Figure 1.** PreQ₁ 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, Ca²⁺, and water, respectively. (c) Select L3 tertiary interactions. (d) Partially hydrated Ca²⁺ ions stabilize the L1 turn. (a) and (b) depict the wild-type² structure; (c) and (d) that of the sequence variant⁴.

**Figure 2.** PreQ₁ recognition and control of gene expression. (a) Phylogenetically conserved² binding pocket. (b) In-line probing data² mapped onto the structure. Blue: nucleotides with reduced scission in the presence of preQ₁. Crystallographically disordered C12 and U13 (red spheres) exhibit increased scission in the presence of preQ₁ (ref. 2). (c) Gene regulation. Absent preQ₁, one S2 strand (pink) instead forms part of an anti-terminator. PreQ₁ stabilizes S2, and allows formation of the terminator. (a) and (b) depict the sequence variant⁴ and wild-type² structures, respectively.
**Accession codes.** Protein Data Bank: Coordinates and structure factors for the wild-type and sequence variant *B. subtilis* preQ₁ riboswitch-preQ₁ 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 preQ₁; 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).