Supplementary Information for:

Molecular basis of cobalamin-dependent RNA modification

Daniel P. Dowling[1,2], Zachary D. Miles[3], Caroline Köhrer[4], Stephanie J. Maiocco[5], Sean J. Elliott[5], Vahe Bandarian[3], Catherine L. Drennan[1,2,4,*]
[1] Howard Hughes Medical Institute, Massachusetts Institute of Technology, Cambridge MA, 02139, USA
[2] Department of Chemistry, Massachusetts Institute of Technology, Cambridge, MA 02139, USA
[3] Department of Chemistry, University of Utah, Salt Lake City, UT 84112, USA
[4] Department of Biology, Massachusetts Institute of Technology, Cambridge, MA 02139, USA
[5] Department of Chemistry, Boston University, Boston, Massachusetts 02215, USA

* To whom correspondence should be addressed. Tel: 617-253-5622; Fax: 617-258-7847; Email: cdrennan@mit.edu

Present Addresses:
Daniel P. Dowling, Department of Chemistry, University of Massachusetts, Boston, MA 02125, USA
Zachary D. Miles, Scripps Institution of Oceanography, University of California-San Diego, La Jolla, CA 92037, USA

Table S1: Data collection, MAD (SeMET) phasing and refinement statistics for the structure of *B. subtilis* QueG without RNA.

		SeMET-labeled QueG. Structure is substrate-free with glycerol in active site (PDB ID 5D08)	
Data collection			
Space group		$P\ 2_1\ 2_1\ 2_1$	
Cell dimensions			
a, *b*, *c* (Å)		84.5, 95.8, 111.2	
α, β, γ (°)		90, 90, 90	
	Peak[$]	*Inflection*	*Remote*
Wavelength	0.9793	0.9795	0.9421
Resolution (Å)	50 – 2.05 (2.12 – 2.05)*	50 – 1.75 (1.81 – 1.75)*	50 – 1.97 (2.04 – 1.97)*
R_{sym}	0.114 (0.566)*	0.084 (0.578)*	0.083 (0.558)*
$I / \sigma I$	17.2 (3.8)*	14.0 (2.6)*	14.9 (2.5)*
Completeness (%)	95.5 (97.2)*	94.8 (93.0)*	94.0 (93.1)*
Redundancy	6.4 (6.2)*	4.7 (4.3)*	4.6 (4.4)*
Refinement			
Resolution (Å)	46.4 - 1.75		
No. reflections	86,580		
R_{work} / R_{free}[&]	0.167/ 0.196		
No. atoms			
Protein	5,923		
Cbl	182		
[4Fe-4S]	32		
glycerol	12		
Phosphate ions	10		
Water	639		
B-factors (Å^2)			
Protein	27.6		
Cbl	17.4		
[4Fe-4S]	23.0		
glycerol	21.5		
Phosphate ions	19.4		
Water	36.3		
R.m.s deviations			
Bond lengths (Å)	0.005		
Bond angles (°)	1.0		

[$]Friedel pairs are not merged.

*Values in parentheses are for highest-resolution shell.

[&]R_{free} determined using a 5% test set of all reflections.

Table S2: Data collection and refinement statistics for *B. subtilis* QueG – RNA complexes

	2.10-Å resolution of QueG with a cleaved ACSL (PDB ID 5D0A)	2.65-Å resolution of QueG with a cleaved ACSL (PDB ID 5T8Y)	QueG bound to an intact ACSL that was oQ-modified [#] (PDB ID 5D0B)
Data collection			
Space group	$P\,2_1$	$P\,2_1$	$P\,2_1$
Cell dimensions			
a, *b*, *c* (Å)	90.6, 111.1, 96.0	44.7, 111.0, 96.0	44.6, 110.8, 99.0
α, β, γ (°)	90.0, 104.9, 90.0	90.0, 99.1, 90.0	90.0, 95.2, 90.0
Resolution (Å)	50 – 2.10 (2.18 – 2.10)*	50 – 2.65 (2.74 – 2.65)*	50 – 2.65 (2.74 – 2.65)*
R_{sym}	0.081 (0.394)*	0.113 (0.451)*	0.154 (0.337)*
I / σI	15.3 (2.6)*	12.0 (1.95)*	10 (3.25)*
CC1/2	0.97 (0.89)*	0.96 (0.896)*	0.92 (0.87)*
CC*	0.99 (0.97)*	0.99 (0.972)*	0.97 (0.97)*
Completeness (%)	97.4 (89.3)*	95.1 (79.0)*	92.8 (68.3)*
Redundancy	3.4 (2.9)*	4.1 (3.7)*	5.3 (3.3)*
Refinement			
Resolution (Å)	46.4 – 2.10	47.9 – 2.65	32.1 – 2.65
No. reflections	103,952	26,643	26,329
R_{work} / R_{free} [&]	0.184 / 0.225	0.237 / 0.267	0.192 / 0.222
No. atoms			
Protein	11,915	5,869	5,918
RNA	507	107	368
Cbl	364	182	182
[4Fe-4S]	64	32	32
glycerol	24	–	6
Phosphate ions	20	35	10
Water	794	14	70
B-factors (Å^2)			
Protein	35.1	71.1	31.3
RNA	75.6	93.2	39.6
Cbl	28.7	66.3	24.3
[4Fe-4S]	34.9	74.6	25.1
glycerol	32.2	–	36.3
Phosphate ions	30.1	80.8	31.3
Water	40.0	66.9	27.3
R.m.s. deviations			
Bond lengths (Å)	0.009	0.003	0.005
Bond angles (°)	1.1	0.7	0.9

[#]Two crystals were used to generate this dataset; R_{merge} reported.
*Highest-resolution shell is shown in parentheses.
[&]R_{free} determined using a 5% test set of all reflections.

Table S3: *B. subtilis* QueG – RNA interface interactions, determined by PISA server for the QueG-intact RNA anticondon stem loop structure.

Hydrogen-bonding distances

RNA residue [atom]	Dist. [Å]	QueG residue [atom]
Q 34[OH]	3.20	GLN 220[N]
U 35[O4]	3.05	LYS 222[NZ]
A 29[OP2]	3.11	ASN 280[N]
A 29[OP1]	2.96	ASN 280[ND2]
G 30[OP2]	3.07	ASN 280[ND2]
G 27[O3']	3.06	ARG 281[NH1]
C 28[OP2]	2.47	ARG 281[NH1]
C 28[OP1]	2.82	ARG 281[NH2]
U 35[O4]	3.05	ARG 295[NH1]
G 30[OP1]	3.14	LYS 297[NZ]
G 30[OP2]	3.07	LYS 297[NZ]
A 31[OP1]	3.01	LYS 298[NZ]

All Interacting residues

QueG Residue	RNA Residue	Interaction Type
PHE 49	Q 34	VDW
HIS 106	Q 34	VDW
ASP 134	Q 34	Close Contact/ Proposed General Acid
THR 135	Q 34	VDW
GLN 220	Q 34	H/ Proposed Substrate Binding
LYS 222	U 35	H
TYR 238	Q 34	VDW
ILE 278	A 29	VDW
SER 279	A 29	VDW
ASN 280	A 29 G 30	H
ARG 281	G 27 C 28	H
PHE 283	A 36	VDW
LYS 284	A 36	VDW
GLY 288	A 36	VDW
HIS 289	A 36	VDW
SER 293	A 36	VDW
TRP 294	Q 34 U 35	VDW
ARG 295	U 35	H
GLY 296	A 36	VDW
LYS 297	G 30	H
LYS 298	A 31	H
ASP 327	G 30	VDW
PRO 328	G 30	VDW
ARG 329	G 30 A 31	VDW/ Possible Salt Bridges

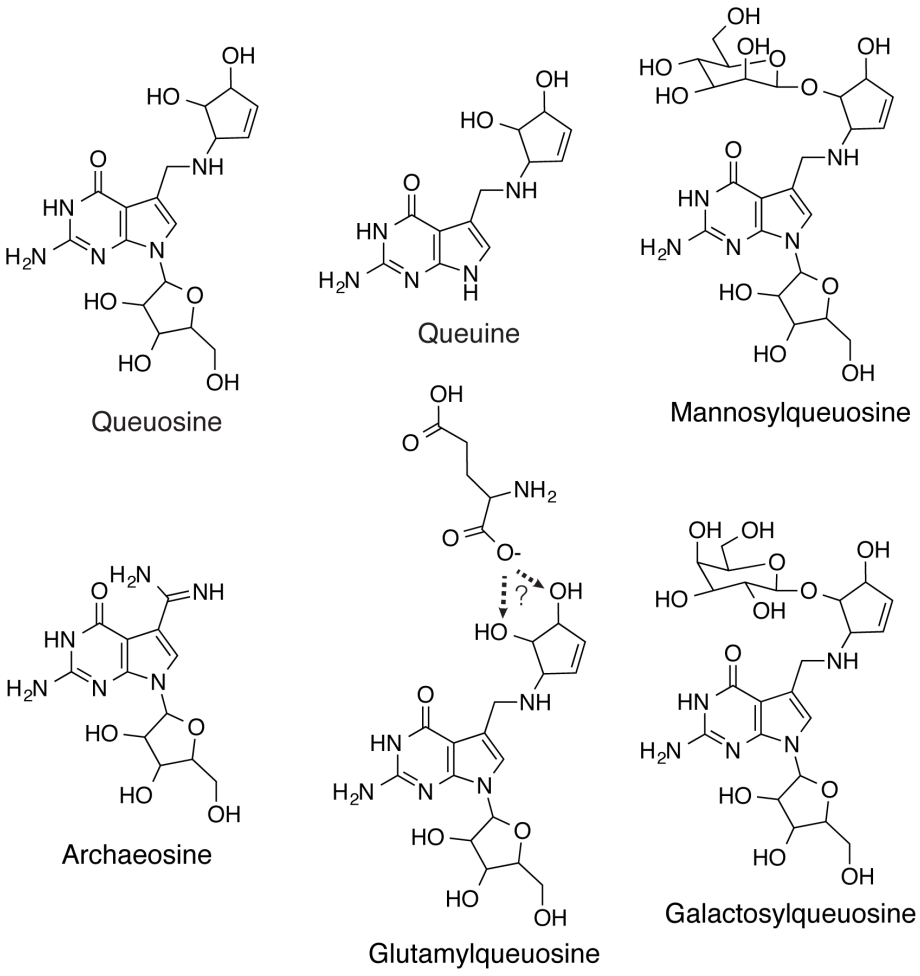

Figure S1: Chemical diagram of known forms of the Q nucleoside, displayed with additionally modified forms existing in eukaryotes (e.g., mannosyl- and galactosylqueuosine) and bacteria (e.g., glutamylqueuosine), and the related archaeal 7-deazaguanine modification (archaeosine).

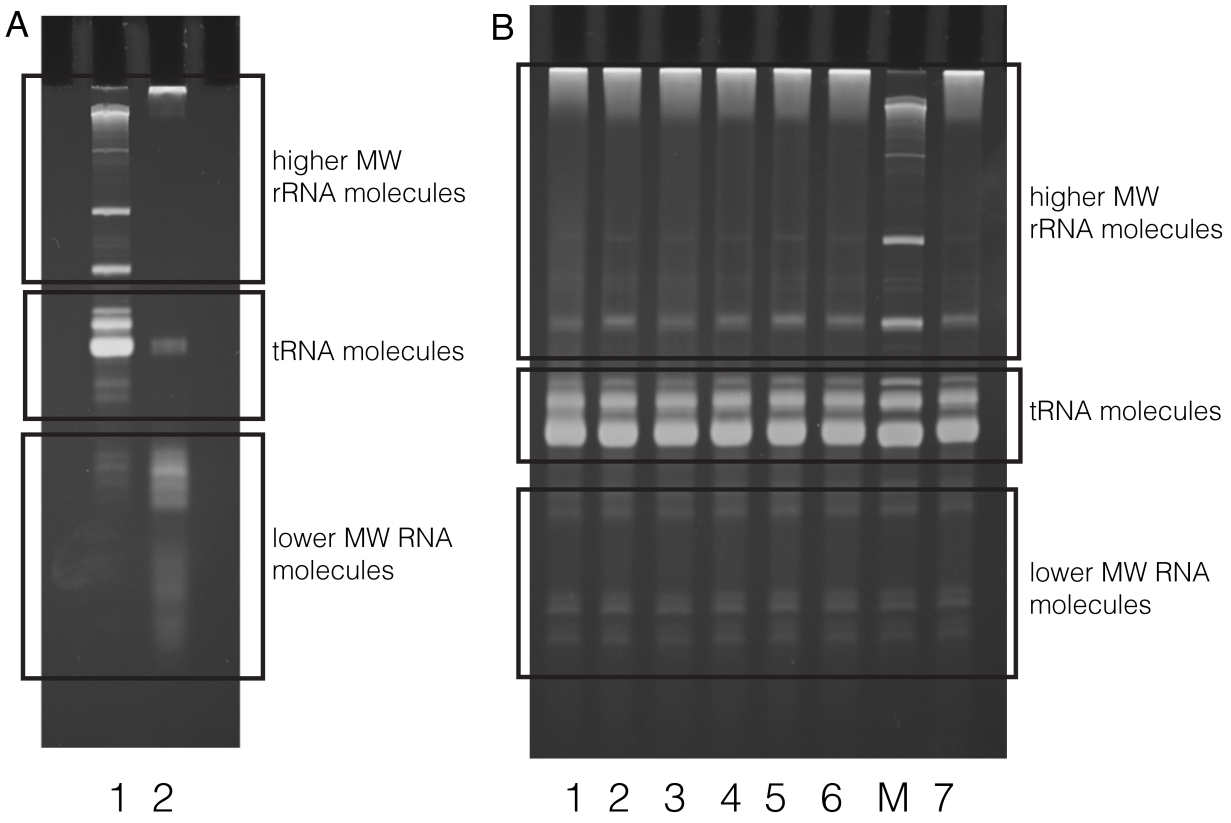

Figure S2: Determination and minimization of RNase contaminating activity within purified QueG protein samples. (**A**) Denaturing 8% PAGE/7M urea analysis of tRNA-enriched total RNA from *E. coli* MG1655, both without (lane 1) and with (lane 2) incubation in the presence of QueG, 37°C, at concentrations similar to those used for crystallization experiments. Lane 2 shows addition of protein sample causes RNA degradation. Gels were stained with ethidium bromide for detection of RNA degradation. (**B**) Incubation at 37°C of RNA samples with QueG in the presence of different amounts of murine RNase inhibitor to protect the RNA from degradation. Lane 1, 40 RNase units; lane 2, 4 units; lane 3, 0.4 units; lane 4, 0.04 units; lane 5, 0.004 units; lane 6, 0.0004 units; lane M, MG1665 total RNA; lane 7, buffer control in the absence of QueG. Even the smallest amount of RNase inhibitor tested, 0.0004 units, was enough to considerably protect RNA samples from degradation compared to panel (A).

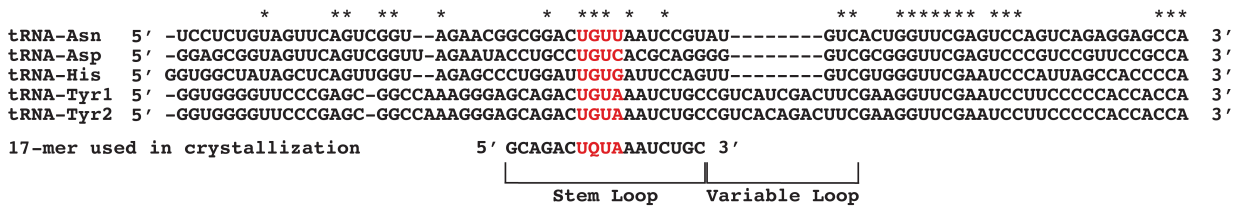

Figure S3: tRNA sequence analysis for QueG substrate tRNAs (G/Q-U-N anticodons) from *E. coli*. Alignments generated using clustalX. Nucleotides U-33 to N-36 of the anticodon stem loop (ACSL) are colored red. Asterisks (*) signify strictly conserved nucleotides in the alignment. The modified ACSL derived from $tRNA^{Tyr}$ used in crystallization is also shown.

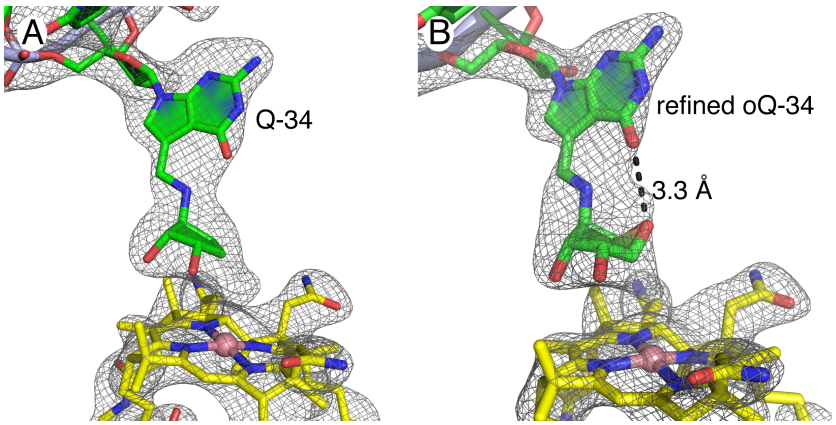

Figure S4: The product queuosine in the intact ACSL structure is the best fit to the electron density. (**A**) Close-up view of the Q-34 base, positioned above the Cbl molecule with SigmaA-weighted simulated annealing omit (mFo-DFc) electron density displayed at 3σ level as gray mesh. Q is a good fit to the density. (**B**) Same view as in (A), displaying SigmaA-weighted simulated annealing omit (mFo-DFc) electron density with oQ refined in the structure. There is no electron density for the epoxide oxygen of oQ. Electron density is colored and displayed as in (A).

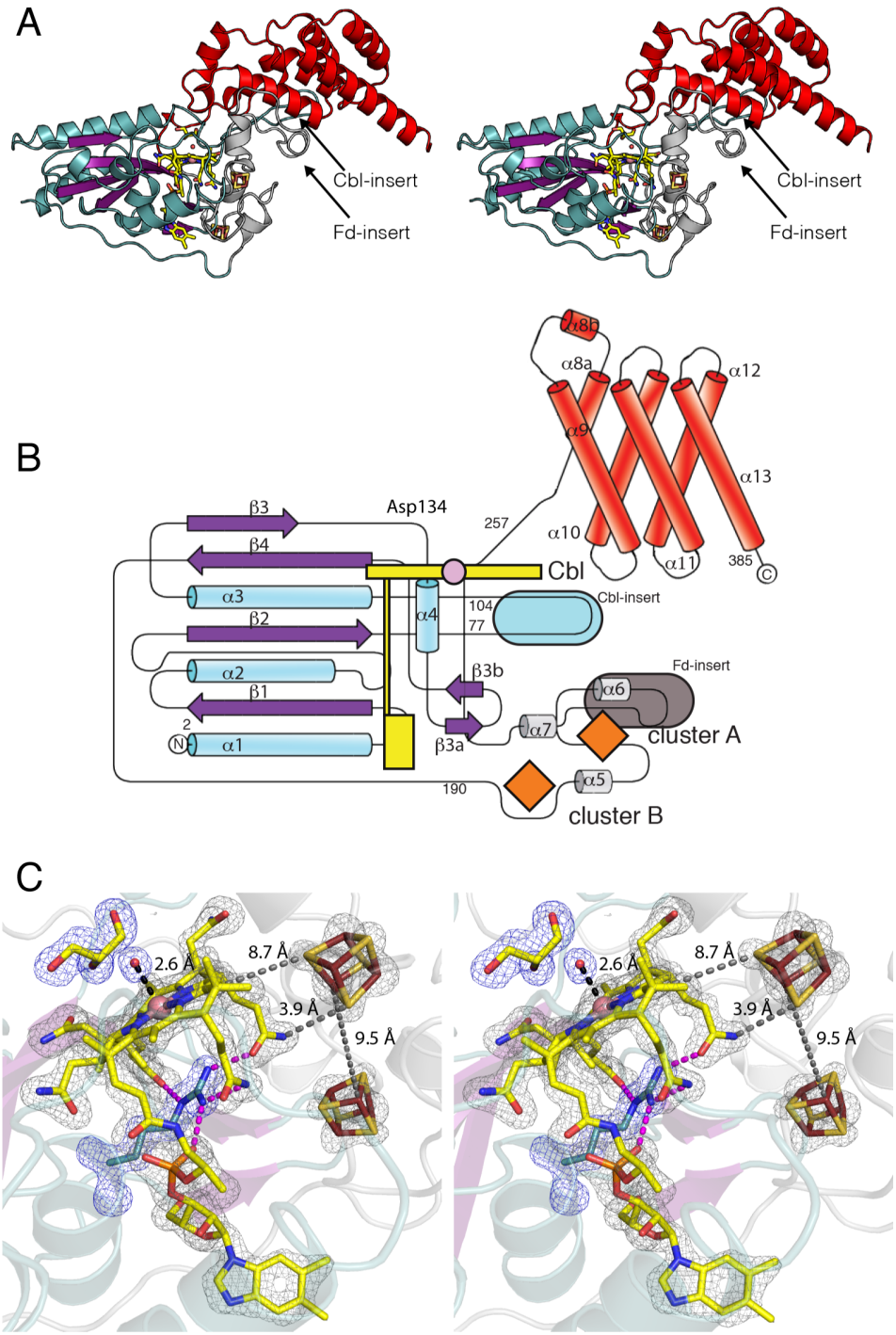

Figure S5: Structure of QueG. (**A**) Stereoview of QueG with the N-terminal Cbl-binding domain colored with blue helices and purple β-strands, the Fd-domain in gray, and the RNA-binding domain in red. (**B**) QueG topology diagram. (**C**) Stereoview showing sigmaA-weighted simulated annealing omit (mFo-DFc) difference electron density for cofactors, glycerol, and Arg141 in substrate-free QueG structure, displayed with the following color and contour levels: Cbl: gray, 5σ; water ligand, glycerol and Arg141: blue, 5σ; [4Fe-4S] clusters: gray, 10σ.

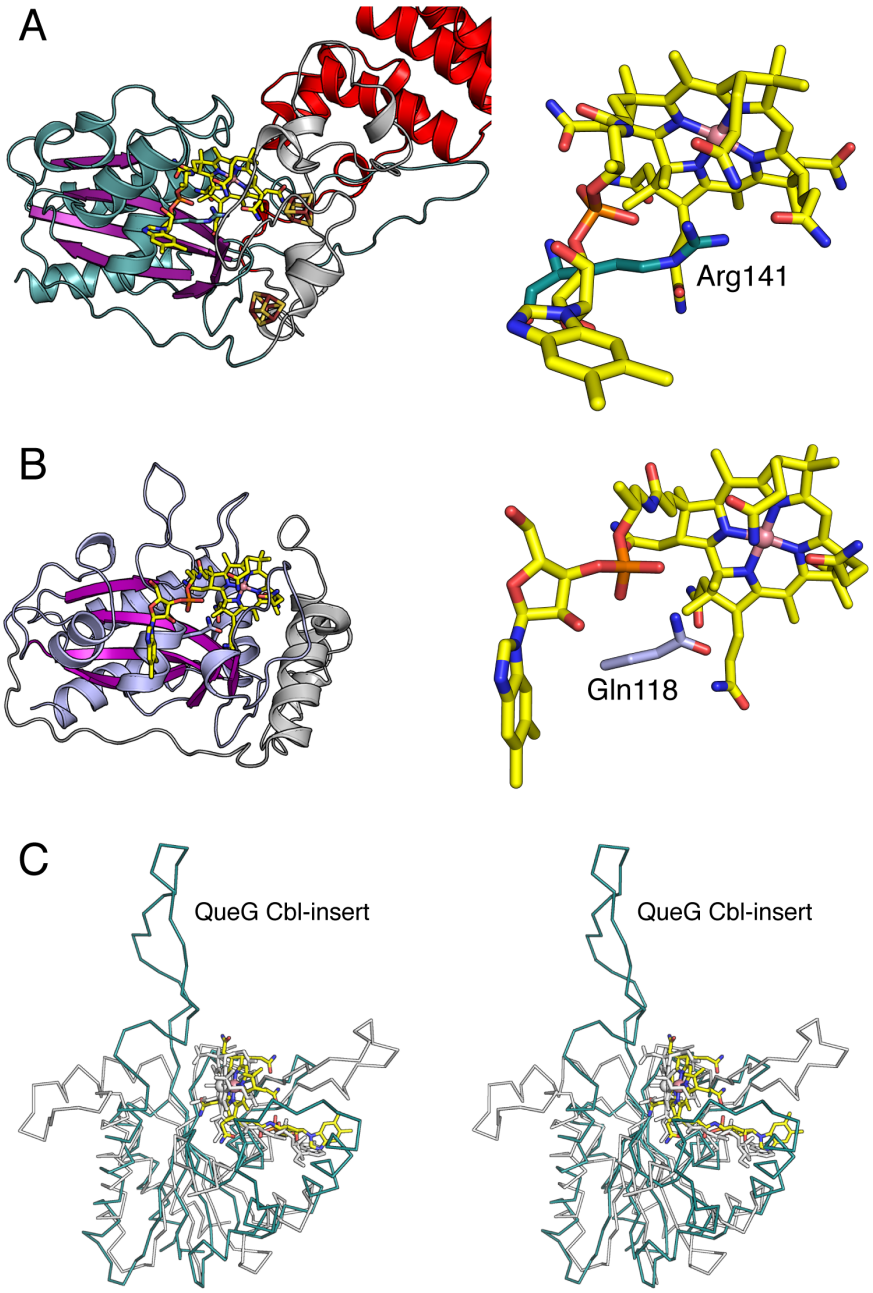

Figure S6: Structural comparison of QueG and CblC. **(A)** QueG colored as in Figure S5A. **(B)** CblC (PDB accession code 3SOM) colored with Cbl-binding domain in purple/magenta and C-terminal cap in gray. **(C)** Stereoview of an overlay of QueG (teal) with CblC (gray), displayed as ribbons and rotated ~90° counterclockwise from panels (A) and (B). The Cbl-binding domains display a R.M.S. deviation of 2.9 Å for 162 Cα atoms as determined by the DALI server.

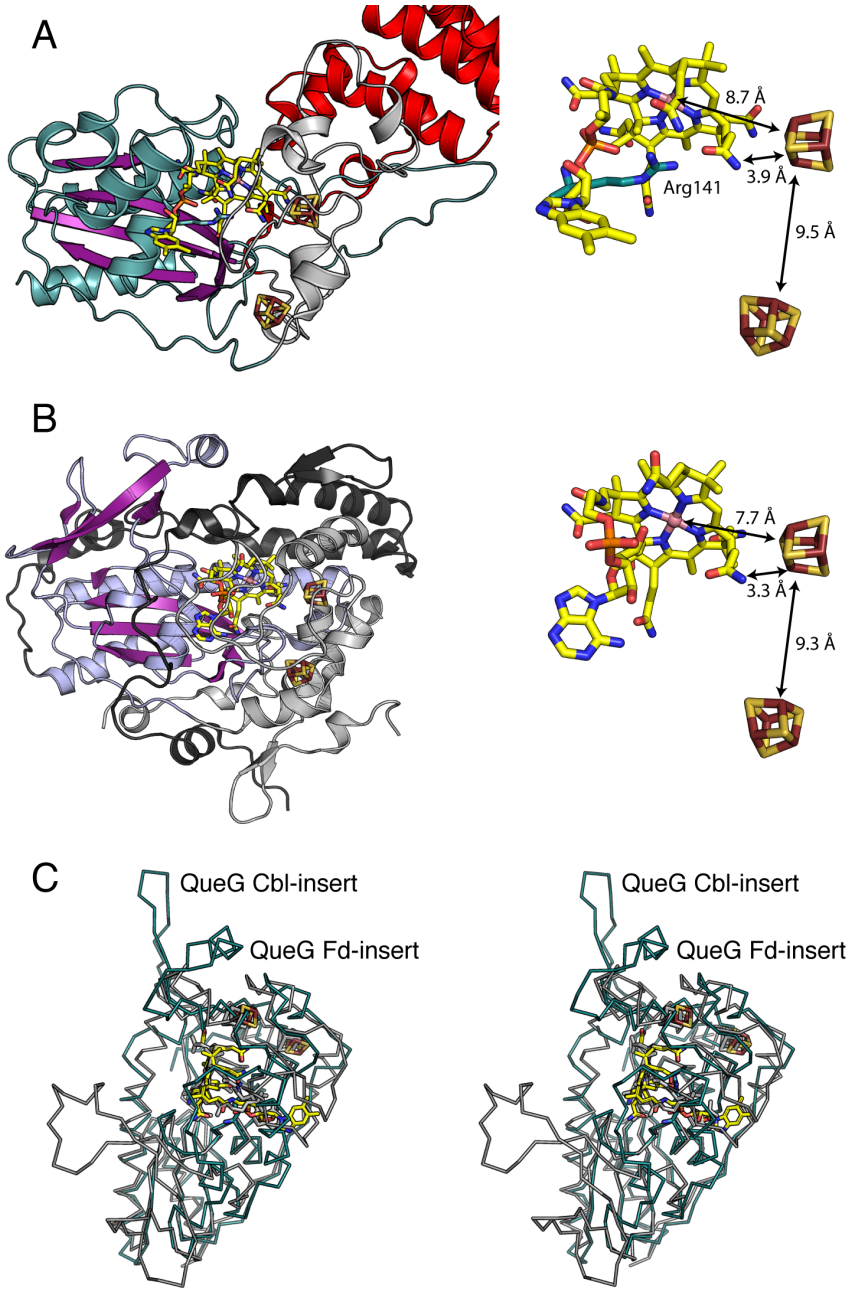

Figure S7: Structural comparison of QueG and PceA. (**A**) QueG colored as in Figure S5A. (**B**) PceA (PDB accession code 4UR0) colored with Cbl-binding domain in purple/magenta; Fd-domains in light gray; and N-terminus in dark gray. (**C**) Stereoview of an overlay of QueG (colored teal) with PceA (colored gray), displaying only the Cbl-binding and Fd-domains as ribbons (R.M.S. deviation of 2.6 Å for 197 Cα atoms) for simplicity and rotated ~90° counterclockwise from panels (A) and (B).

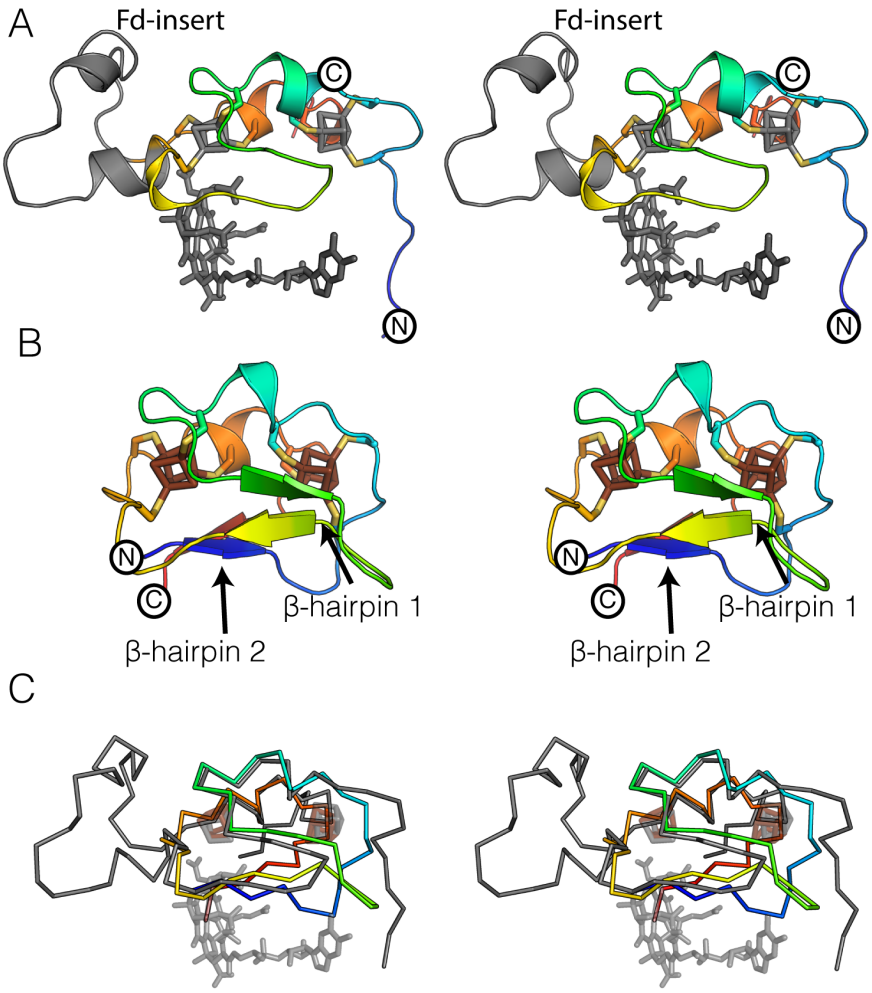

Figure S8: Structural comparison of the QueG [4Fe-4S] cluster-binding domain and ferredoxin from *Clostridium acidurici*, displaying three noticeable differences: the first β-hairpin in QueG is shortened to a small loop; a platform, on which the RNA-binding domain of QueG sits, is created through a 24-residue insertion between the fifth and sixth cysteines of the cluster-binding motif; and the Cbl positioning in QueG replaces the second ferredoxin β-hairpin. (**A**) Stereoview ribbon depiction of the QueG Fd-domain colored from N- to C-termini in a gradient of blue to red. The 24-residue insertion, Cbl and [4Fe-4S] clusters in QueG are colored gray. (**B**) Stereoview ribbon depiction of *C. acidurici* ferredoxin (PDB accession code 1FDN) colored as above except that clusters are colored brown. (**C**) Stereoview depiction of the cluster-binding domain of QueG (colored gray) superimposed with ferredoxin from *C. acidurici* (colored as gradient of blue to red) and displaying a R.M.S.D. of 3.8 Å for 41 Cα atoms. The first hairpin (hairpin 1) of ferredoxin is replaced with a shorter loop, and the second hairpin (hairpin 2) of ferredoxin is missing in QueG, and instead replaced by the Cbl-binding site.

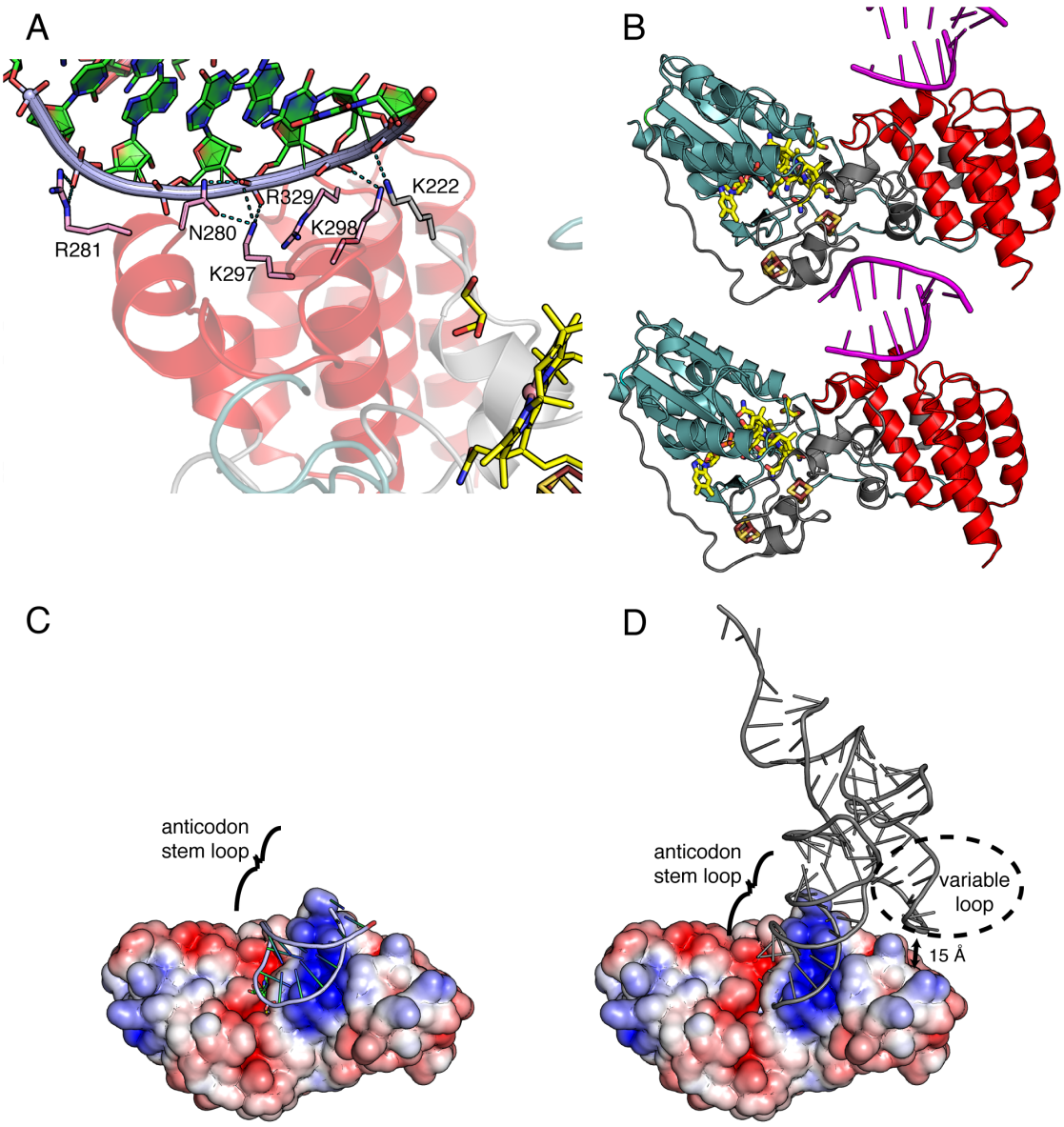

Figure S9: QueG:tRNA interactions. (**A**) 2.10 Å resolution structure of QueG bound to an ACSL (green) that is cleaved. Anticodon bases (G-27 to U-33) pack against basic residues of the QueG RNA-binding domain (red). (**B**) The 3´-end bases (A-38 to C-43 in magenta) are involved in lattice contacts in this 2.10 Å resolution QueG structure. QueG is colored as in Figure S5A. (**C**) Electrostatic surface calculation for QueG with an intact ACSL. Surface is displayed as a gradient from red to blue (-5 to +5 K_bT e_c^{-1}). (**D**) Modeled $tRNA^{Tyr}$ binding to QueG based on structure of QueG with an intact ACSL. $tRNA^{Tyr}$ was obtained from PDB accession code 3UZ6, in which it was bound to the A-site of the 30S ribosomal subunit. In this model, the $tRNA^{Tyr}$ variable loop is ~15 Å away from QueG.

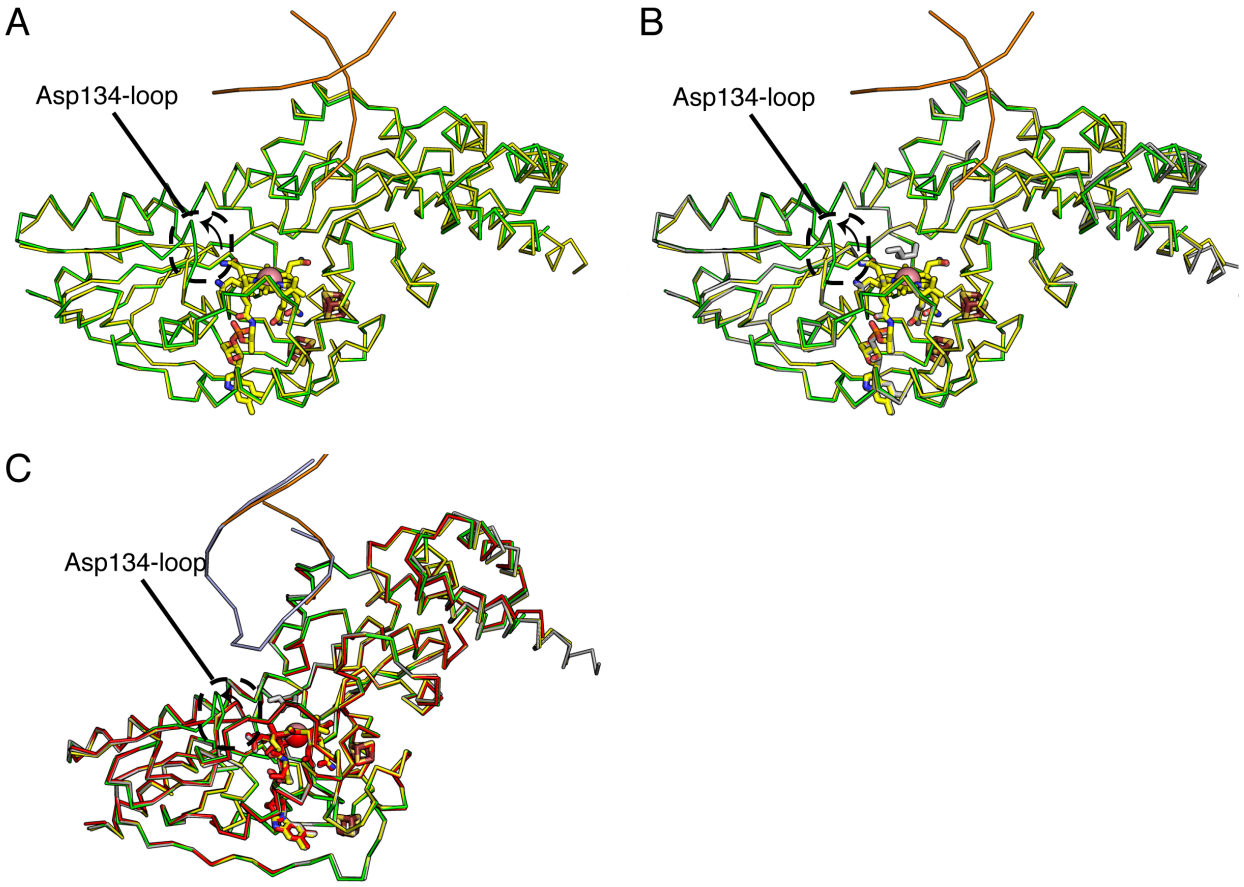

Figure S10: Comparison of all four QueG structures. (**A**) A superposition of two QueG structures: the 2.10 Å resolution structure of QueG with a cleaved ACSL (yellow and orange for protein and RNA, respectively), and the monomer of the 2.65 Å resolution QueG structure that has the D134-loop flipped out of the active site (green); R.M.S.D of ~0.5 Å for ~370 Cα atoms. (**B**) A superposition of three QueG structures: the two from panel A with the 1.75 Å resolution substrate-free QueG structure with glycerol bound (gray); R.M.S.D of ~0.2 and ~0.5 Å for ~370 Cα atoms, respectively. (**C**) A superposition of the four structures: the three from panel B with the 2.65 Å resolution structure of QueG bound to an intact ACSL (red for protein and light blue for ACSL); R.M.S.D of ~0.45 Å, ~0.5 Å, and ~0.5 Å for ~370 Cα atoms, respectively.

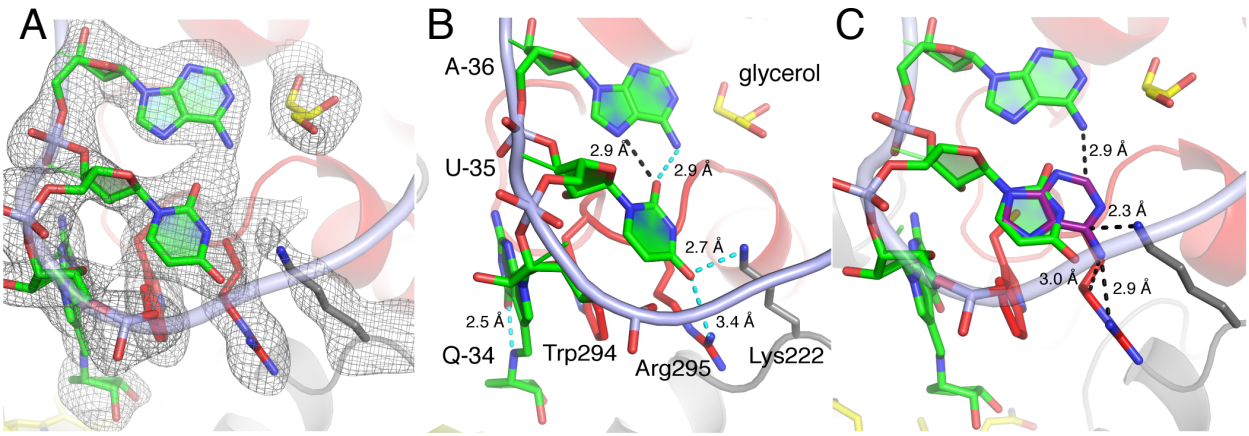

Figure S11: QueG – ACSL interactions. (**A**) SigmaA-weighted simulated annealing omit (mFo-DFc) electron density at 3σ level and displayed as gray mesh for the anticodon loop bound to QueG and Lys222, Trp294 and Arg295, with residual omit electron density near A-36 refined as a glycerol molecule. (**B**) Interactions between U-35 and protein. Same view as in (A). Dashed lines in cyan indicate potential hydrogen bonds and the dashed line in black is a close contact. (**C**) Protein binding pocket is too small to accommodate a purine base (magenta carbons). Adenine was manually overlaid with U-35, and resulting close contacts to A-36, Lys222 and Arg295 are displayed as black dashed lines with distances indicated.

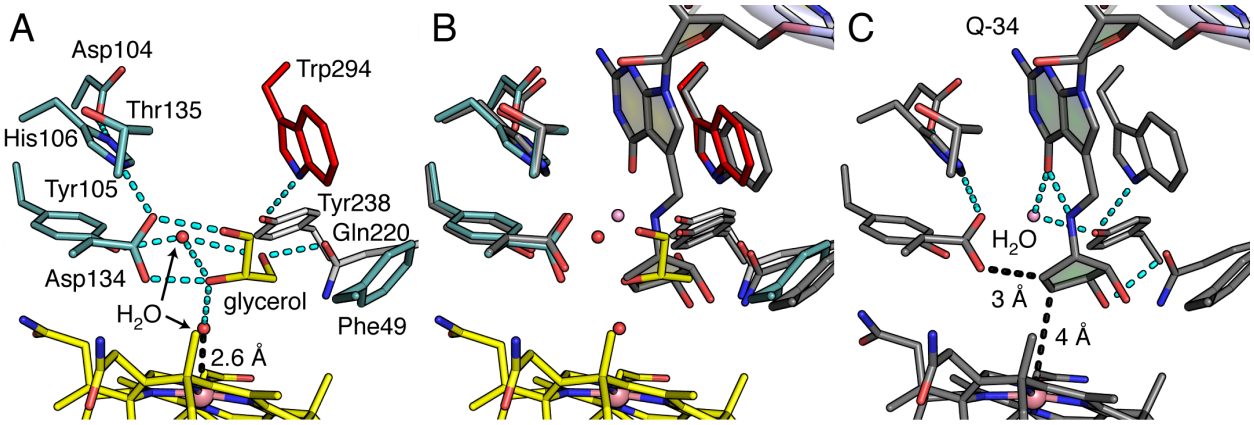

Figure S12: QueG active site comparisons. (**A**) Glycerol binding to QueG in the absence of substrate. (**B**) Overlay of QueG structure with glycerol (yellow carbons) and water (red sphere) and QueG structure bound to an intact ACSL (gray carbons) and water (pink). (**C**) Q-binding in structure of QueG with an intact ACSL. Hydrogen-bonding and close van der Waals interactions are displayed as dashed cyan and black lines, respectively.

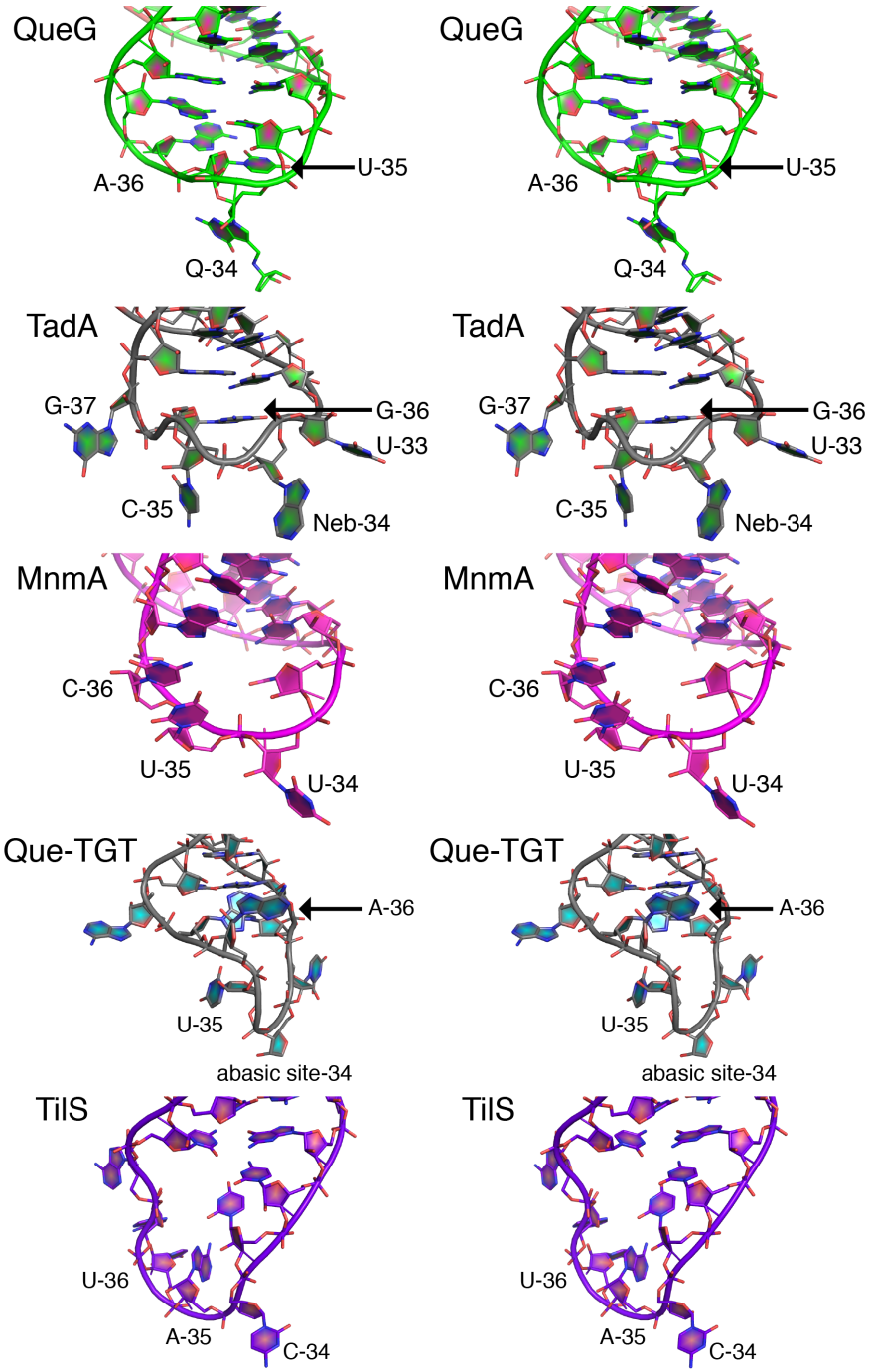

Figure S13: Stereoview comparisons of QueG with other position 34 RNA-modifying enzyme co-crystal structures, exemplifying how each respective modifying enzyme induces a different orientation of the ACSL for defining recognition and specificity (PDB accession codes 2B3J, 2DER, 1Q2R, and 3A2K for tRNA adenosine deaminase (TadA), thiouridylase (MnmA), queuosine tRNA-guanine transglycosylase (Que-TGT), and lysidine synthetase (TilS), respectively). Only the ACSL portion of the RNA is shown for clarity.

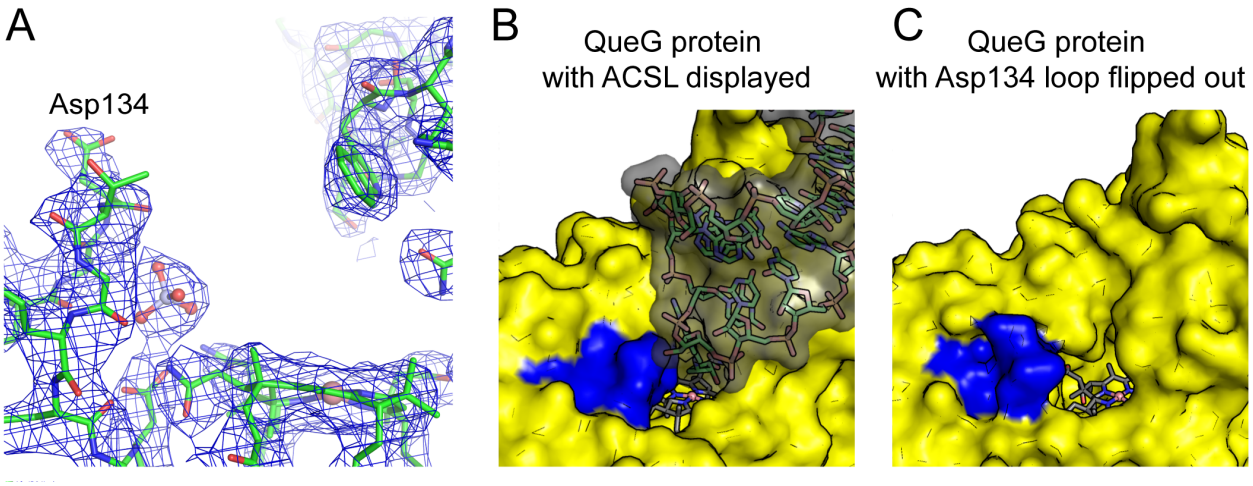

Figure S14: The flexible Asp134-containing loop could play a role in closing/opening the active site. (**A**) The 2.65 Å resolution structure of QueG with a cleaved ACSL shows the Asp134-loop in a different conformation with Asp134 pointing out of the active site and a phosphate molecule from the crystallization buffer binding where Asp134 is located in the three other QueG structures. 2Fo-Fc electron density is displayed as blue mesh at 1 σ, and protein is colored with green carbons. (**B**) Space filling depiction of QueG (yellow) bound to an intact ACSL showing that the Asp134-loop (colored in dark blue) is partially responsible for sealing the active site. RNA is covered in gray transparent surface and carbons of Cbl and RNA are colored gray and green, respectively. (**C**) Space filling depiction of the 2.65 Å-resolution structure of QueG (yellow) with a cleaved ACSL that shows that when the Asp134-loop is flipped out (dark blue), the Cbl (carbons in gray) is more accessible.

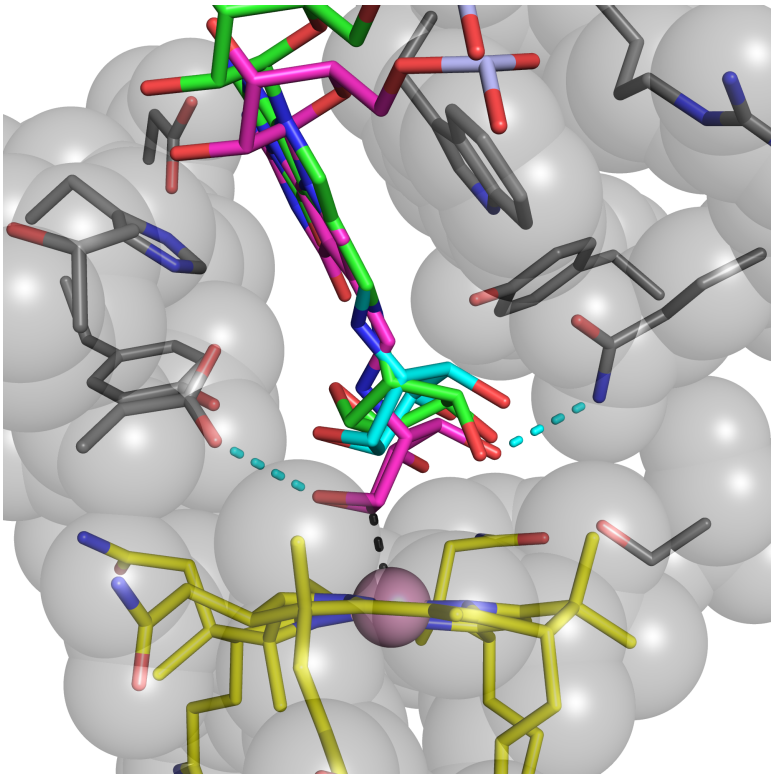

Figure S15. Modeling of Cbl-substrate intermediates. Formation of a Cbl-substrate adduct would require both rotation of substrate torsion angles and translation into the active site (protein residues in gray are surrounded by a gray van der Waals surface). An attempt to refine the oQ substrate against the data (oQ carbons colored green) positions the epoxide ~4.0 Å from the Cbl cobalt (pink sphere). See Figure S4 for electron density. Further rotation of substrate torsion angles of the aminomethyl linker leads to a distance from the substrate epoxide to Cbl cobalt of ~3.7 Å (oQ carbons colored cyan). In order to model a covalent adduct, further rotations of the substrate aminomethyl torsion angles, accompanied with an ~1.5 Å rigid shift of the RNA base (oQ carbons colored magenta) towards the Cbl (yellow), would position the epoxide ~2.3 Å from Cbl cobalt (black dashed line). Hydrogen bonding interactions for the modeled magenta covalent adduct (~3 Å) are displayed as cyan dashed lines.