

**Structure of the bifunctional methyltransferase YcbY (RlmKL) that adds the
m⁷G2069 and m²G2445 modifications in *Escherichia coli* 23S rRNA.**

Supplementary Information

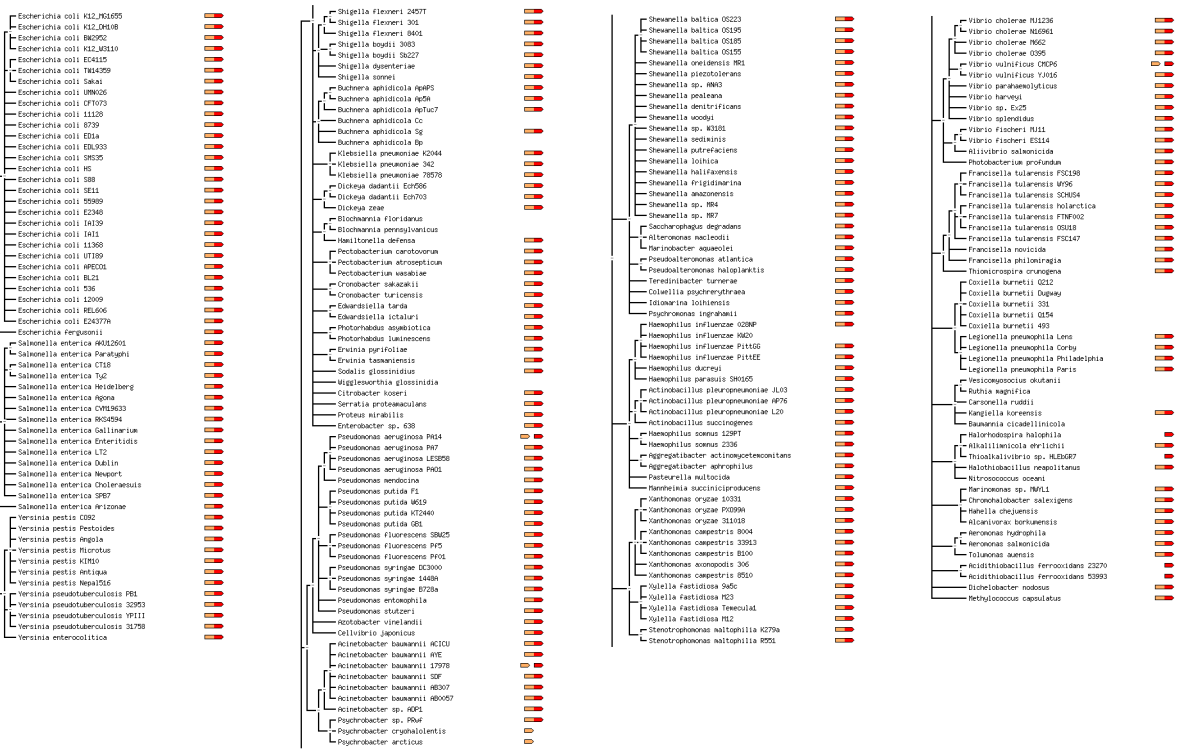
Kai-Tuo Wang, Benoit Desmolaize, Jie Nan, Xiao-Wei Zhang, Lan-Fen Li
Stephen Douthwaite, & Xiao-Dong Su

Figure S1: Bioinformatics analysis of YcbY fusion orthologs in Gammaproteobacteria and the two separate orthologs in other bacteria. (A) Gram-negative Gammaproteobacteria, where 213 out of 238 sequenced species have orthologs of the YcbY fusion. (B) in the Beta- and Deltaproteobacteria, and the Bacilli branch of Firmicutes, most of the sequenced species have two separate genes. These are exemplified by the *B. subtilis* genes *ypsC* and *ywbD* and the *S. mutans* genes *smu472* and *smu776*. Gene fusion data were analyzed in StringDB (<http://string.embl.de/>).

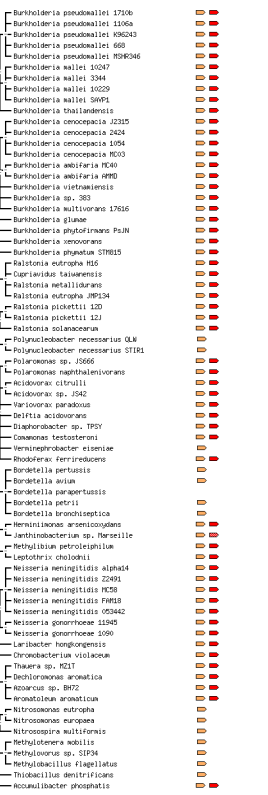
Only two of the genomes of Gammaproteobacteria, *Psychrobacter cryohalolentis* and *Psychrobacter arcticum*, have a separate ortholog comprising the N-terminus without a corresponding separate ortholog of the C-terminus of YcbY (and this is possibly a misannotation, as these species also have a complete YcbY fusion). Three Gram-negative genomes have separate YcbY-N and YcbY-C orthologs, including *Vibrio vulnificus* CMCP6, although here again the genes are possibly incorrectly annotated. *Pseudomonas aeruginosa* PA14 and *Acinetobacter baumannii* ATCC 17978 have separate genes for YcbY-N and YcbY-C. Sixteen genomes in the Gammaproteobacteria do not have any ortholog of YcbY.

In other groups, only six bacteria have YcbY as a fusion protein. Three of these are also Gram-negative bacteria (*Magnetococcus* sp MC-1, *Desulfotalea psychrophila*, and *Pelobacter curbinolicus*), while the other three are Gram-positives (*Eggerthella lenta*, *Slackia heliotrinireducens* and *Atopobium parvulum*). Alphaproteobacteria do not possess orthologs of YcbY, while Beta- and Deltaproteobacteria have the same gene arrangement as in the Gram-positive Firmicutes branch of Bacilli. The NMB0445 and NMB1367 orthologs of *Neisseria meningitidis* were analyzed in a contemporary study (1) and shown to have function methyltransferase functions identical to YcbY-N and YcbY-C, respectively.

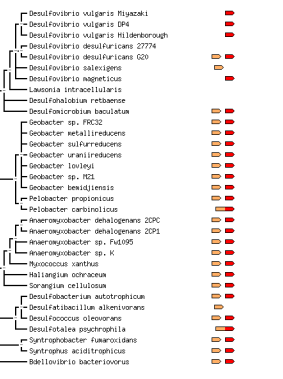
A Gamma-proteobacteria



B Beta-proteobacteria



Delta-proteobacteria



Bacilli branch of Firmicutes



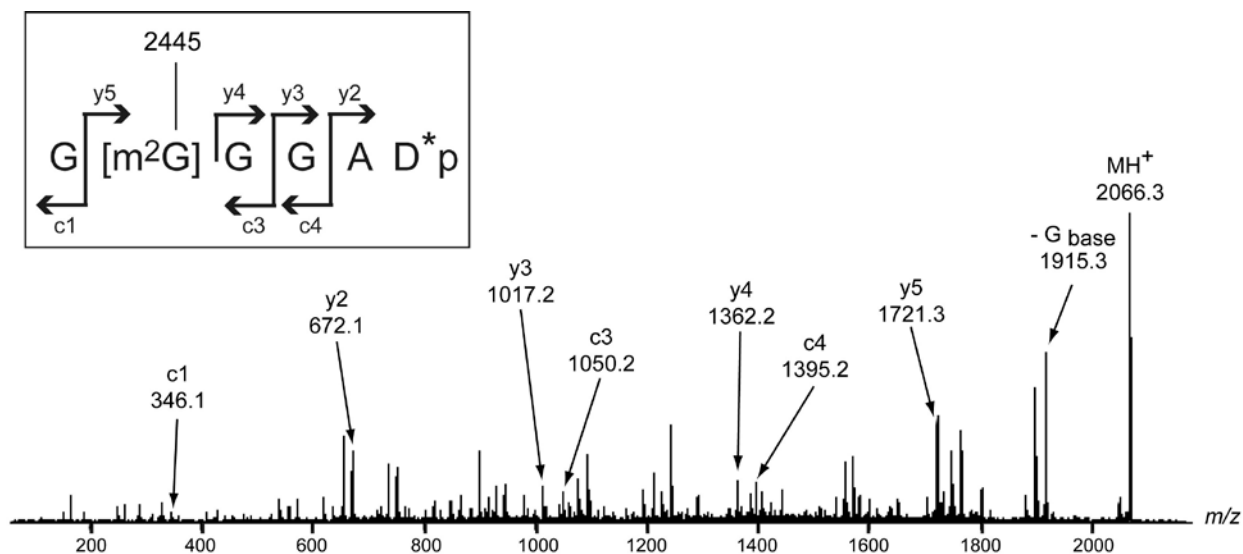


Figure S2: MALDI tandem mass spectrum of the 23S rRNA region containing m²G2445

Prior to analysis, the 23S rRNA sequence 2422-2467 containing m²G2445 was dried under vacuum and RNase A was dried separately. Each was resuspended in 2 μ l of heavy water (>95% ¹⁸O) and mixed. Digestion produced the fragment GGGADp (G2444 to D2449, boxed) containing the m²G2445 and a second modification with an initial mass of 2066.3 m/z (MH⁺). This was an additional 2 m/z greater than the fragment produced without heavy water, and helped to discriminate the y ions in the fragmentation pattern. D* indicates that the y2 ion carried 14 Da of extra mass that was possibly another methyl group attached to D2449; the location of this extra mass was confirmed by tandem MS analysis of an overlapping fragment (Figure S3). The ions are annotated according to McLuckey *et al.*(2).

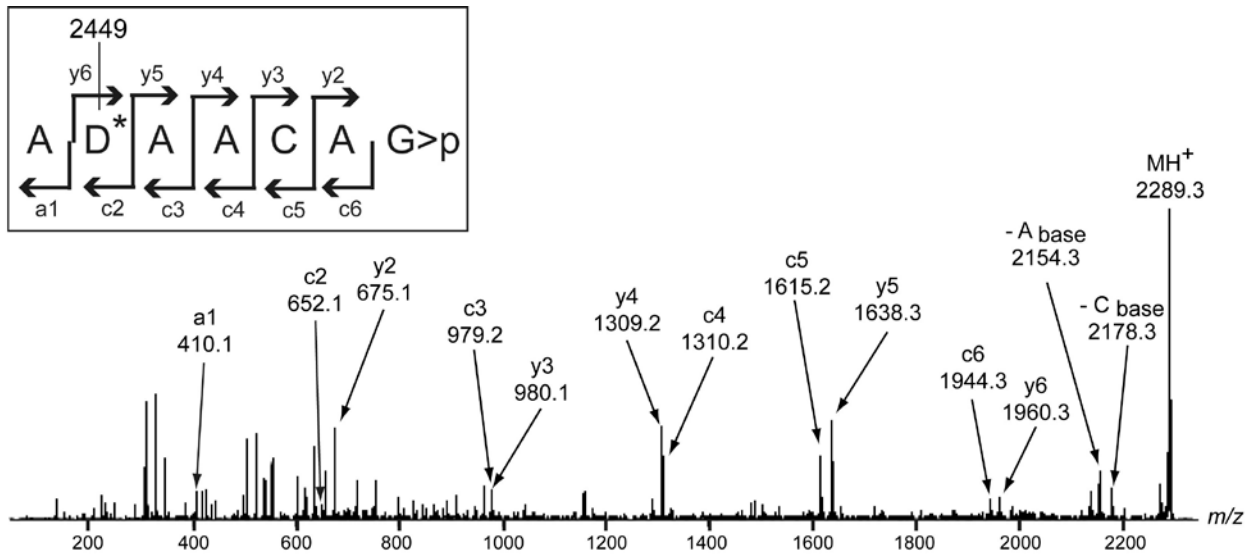


Figure S3: MALDI tandem mass spectrum of the 23S rRNA fragment containing D2449

The nucleotide was analyzed after generation of RNase T1 digestion products from the C2422 to C2467 region of rRNA from the $\Delta ycbY$ strain. The RNase T1 oligo ADAACAG>p corresponds to nucleotides 2448 to 2454 (box). D2449 was shown to carry the 14 Da of extra mass, which is presumed to be a methylation, and this modification is unconnected with the function of YcbY. The ion fragments are annotated as above (2).

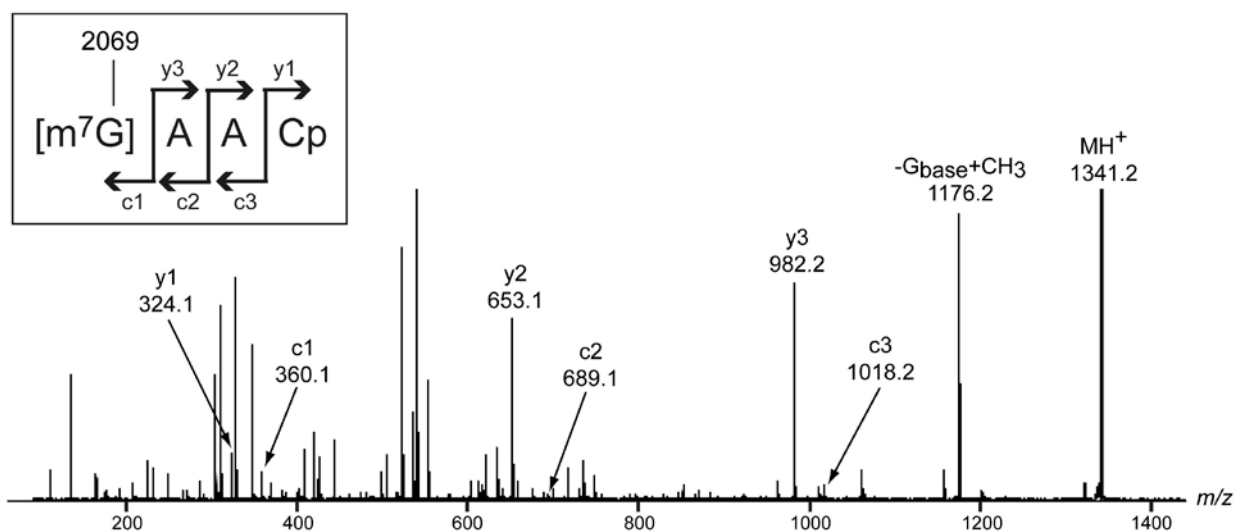


Figure S4: MALDI tandem mass spectrum of 23S rRNA containing m⁷G2069

Nucleotide m⁷G2069 was analyzed after generation of RNase A digestion products from the C2044 to C2093 region from the wild-type rRNA strain; m⁷G2069 is contained in the GAACp fragment (nucleotides 2069 to 2072, box). The initial mass of the fragment was 1341.2 *m/z* (MH⁺). The ion fragmentation pattern is annotated as above (2).

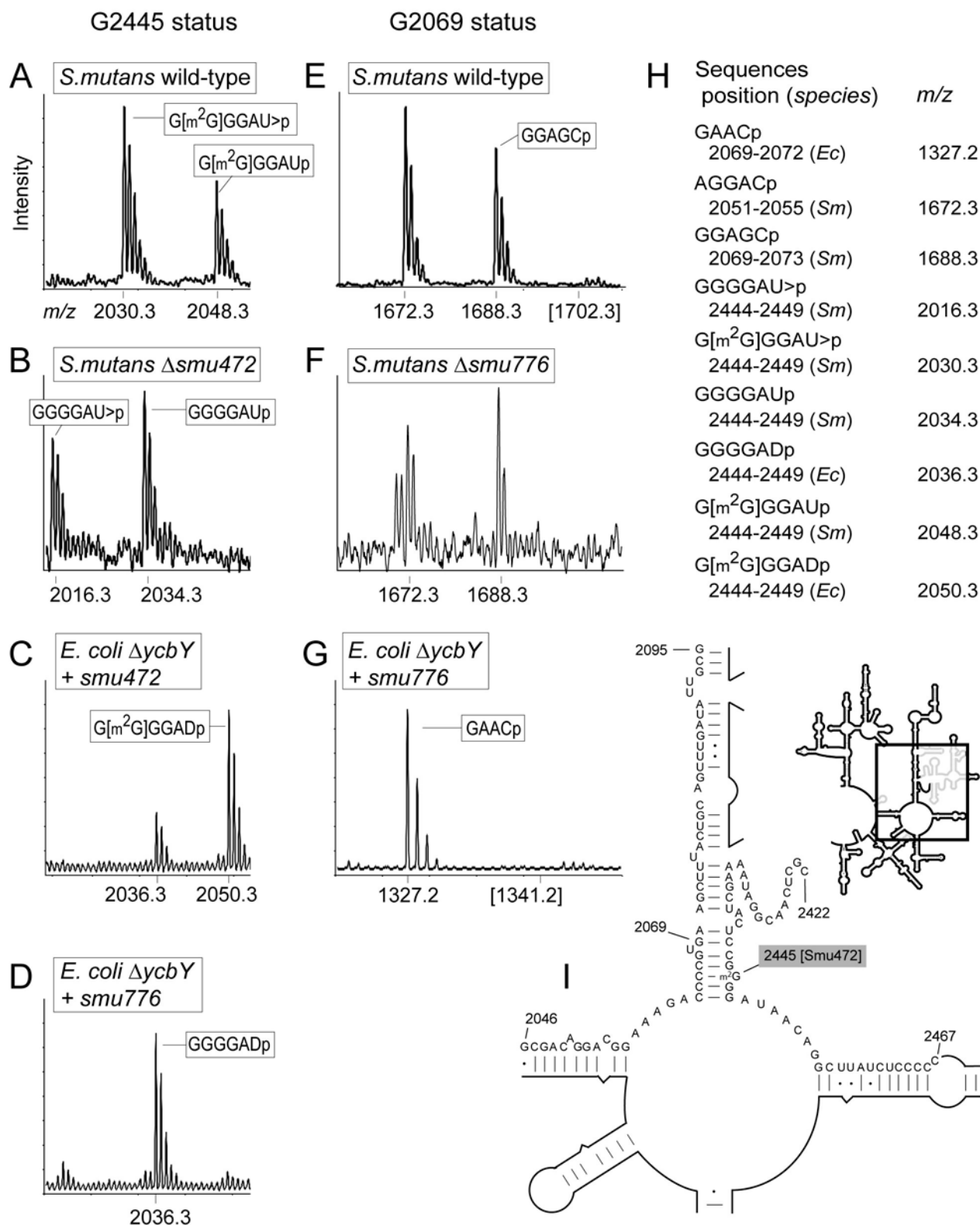


Figure S5: MALDI-MS analysis of rRNA modification by Smu472 and Smu776.

(A-B) Isolation of the C2422 to C2467 region from *S. mutans* 23S rRNA followed by RNase A digestion generates GGGGAU fragments containing nucleotide G2445. In these spectra, digestion

gave rise to a mixture of linear phosphates (p) and 2'-3'-cyclic phosphates (>p); their sequences are identical, although the latter is missing the equivalent of a water molecule and is 18 Da lighter. The m²G2445 modification is present and stoichiometric in the *S. mutans* wild-type strain (A). The modification is lost in the *smu472*-null strain (B). (E-F) Isolation of the G2046 to G2095 region of *S. mutans* followed by RNase A digestion generate a fragment GGAGC with the nucleotide G2069. There is no m⁷G2069 modification in the *S. mutans* wild-type strain (E) or the *smu776*-null strain (F). Complementation of the *E. coli ycbY*-null strain with active copies of *smu472* partially restored methylation at m²G2445 (C); whereas *smu776* had no effect on G2445 (D) or G2069 (G). Sequences and theoretical mass/charges (*m/z*) of all the RNase fragments are shown in (H). The enlarged secondary structure (boxed in the outline) of the *S. mutans* 23S rRNA is showed (I). Nucleotides are numbered according to the *E. coli* rRNA structure (3-4).

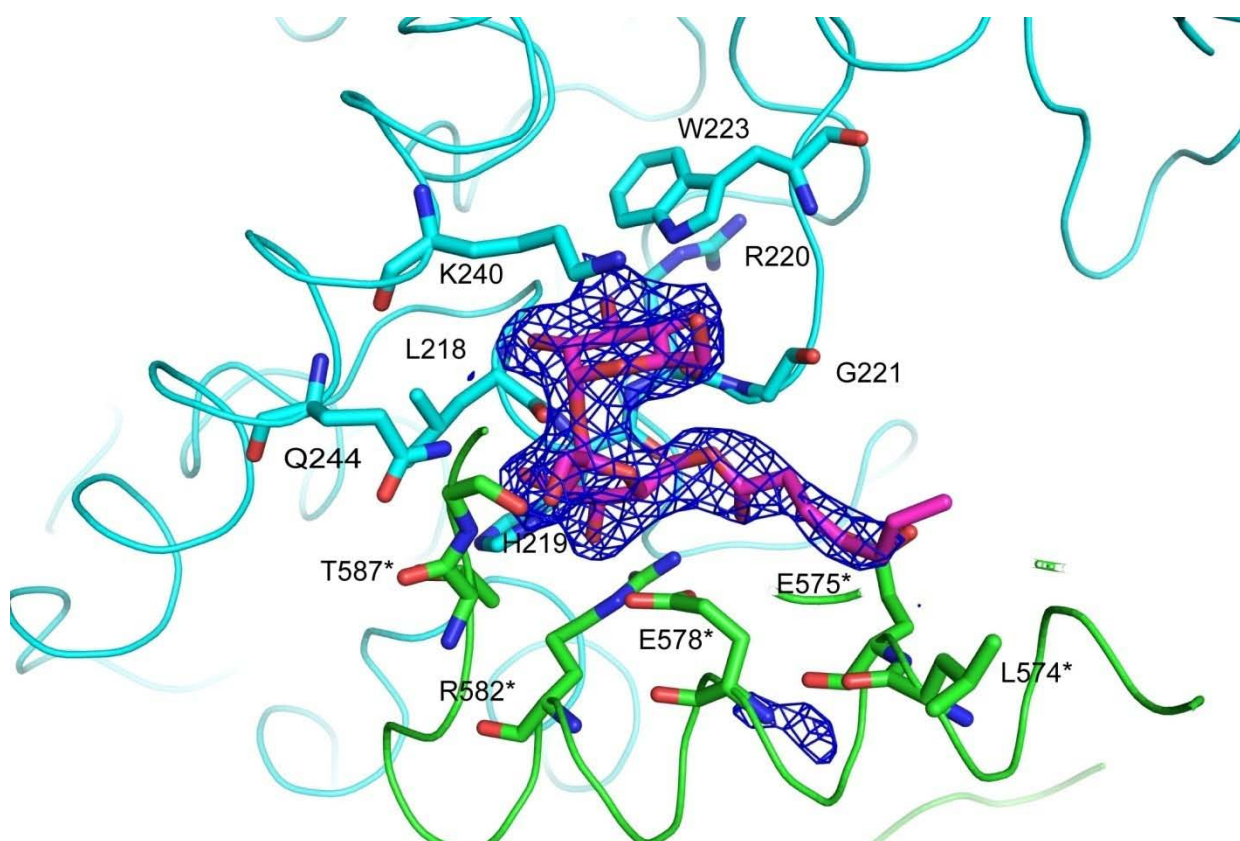


Figure S6: NCS related YcbY monomers

Monomers in the crystal are shown in ribbon and colored cyan and green. The interaction residues

are shown in stick mode and labeled. The asterisk (*) indicates the NCS related monomer. $|Fo-Fc|$ electron density map was calculated with the protein model excluding n-octanoylsucrose and contoured at 4.0σ .

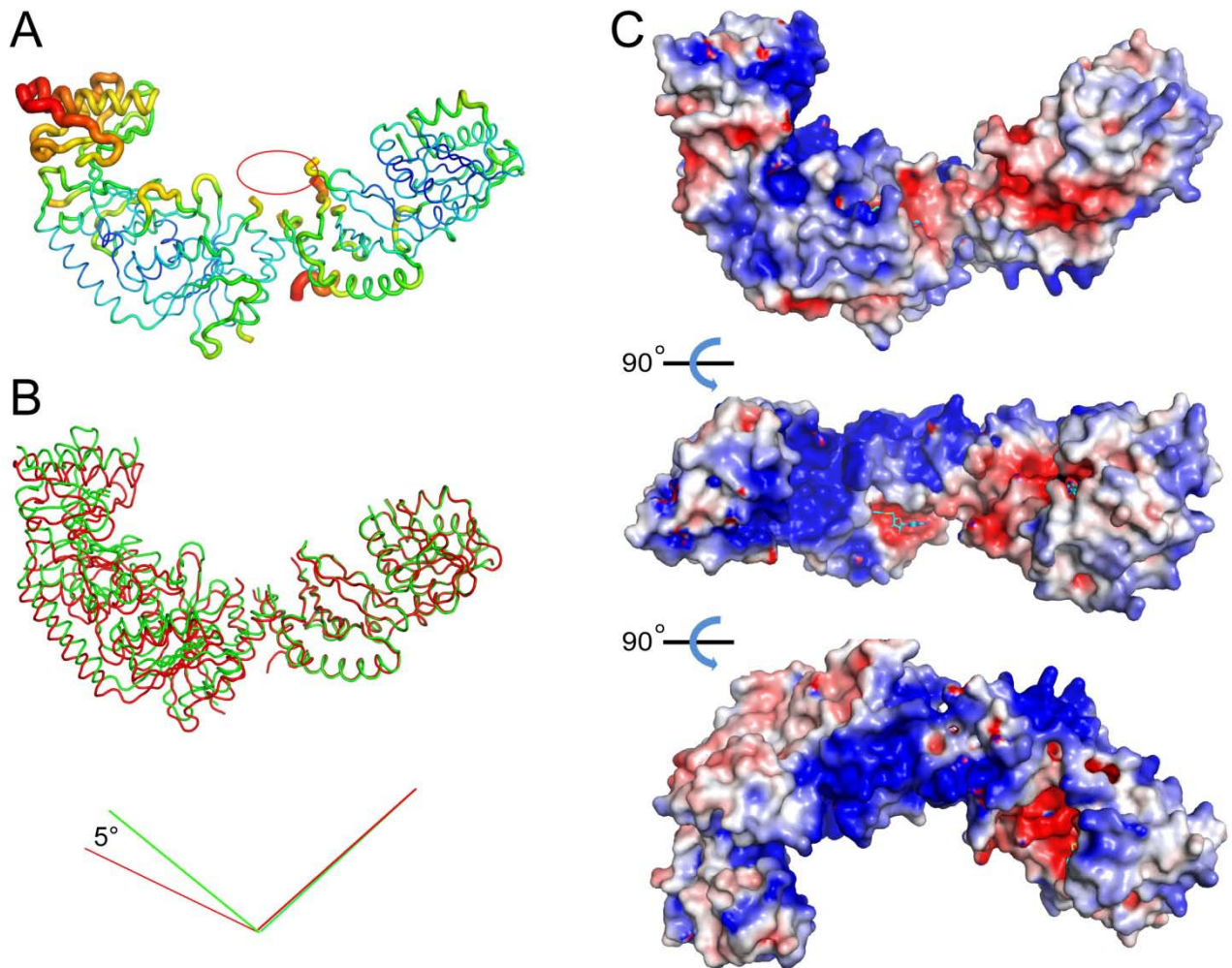


Figure S7: Surface potential on the YcbY structure

(A) B-factor presentation of the YcbY monomer showing unstable regions in the THUMP and EEHEE domains. The loops encircle in red could not be modeled. (B) Open angle difference between NCS related monomers aligned over the C-terminal halves. (C) Surface potential representation of YcbY in three 90 degree steps with SAH molecules shown as sticks. The YcbY structure has an overall concaved shape with a central cleft that is large enough to accommodate double-stranded RNA. It is feasible that the enzyme could bind in minor groove of helix 74 and flip

the two target bases into their respective active sites during one binding event (see main text). The protein structure is highly flexible and the angle between YcbY-N and YcbY-C differs by 5 degrees in the NCS related monomers. Several other features complicate building of an unambiguous model for RNA-protein interaction. For instance, the THUMP domain has high B-factors in the one monomer, and is barely visible in the other where no supporting interactions occur during crystal packing. Furthermore, the EEHEE domain and two loop regions (residues 446~452 and residues 484~495) could not be modeled in either monomer. These domains would feasibly become more stably structured upon RNA binding, and better resolution must therefore await YcbY-RNA co-crystal studies.

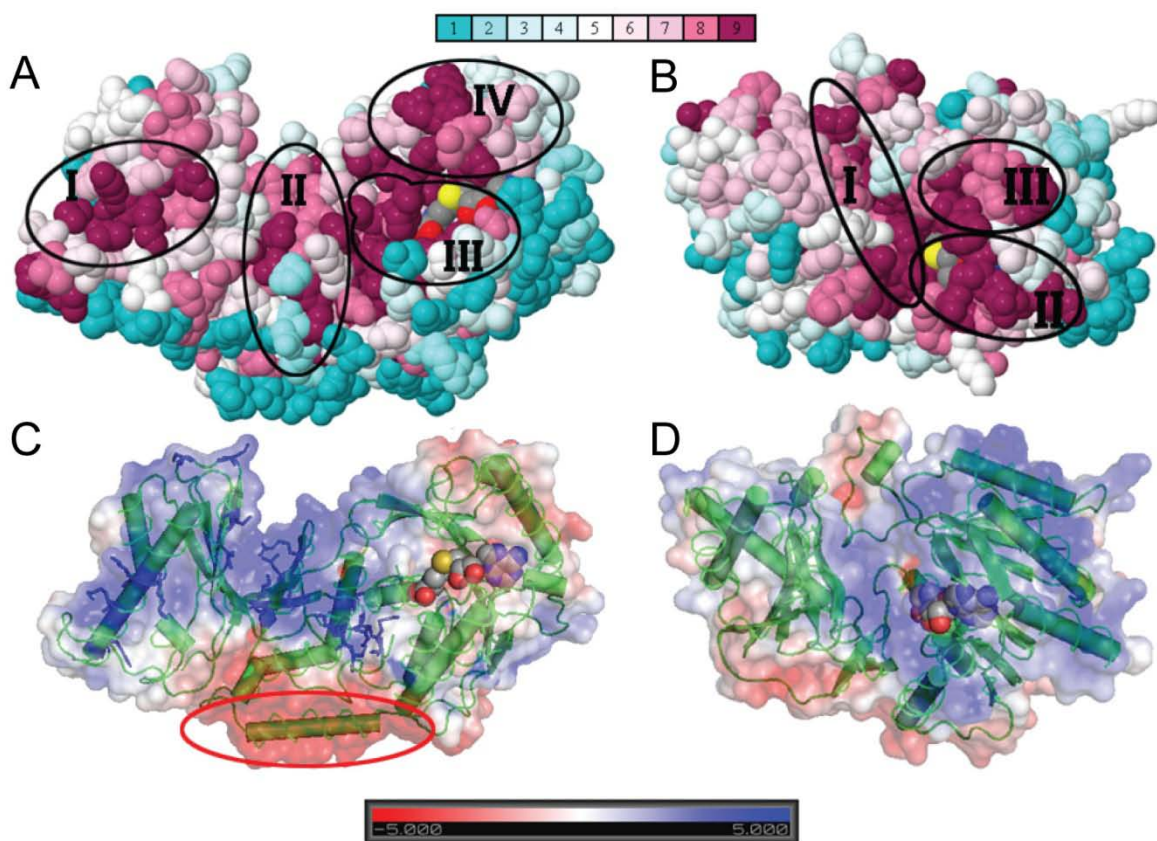


Figure S8: Residue conservation and surface potentials of Smu472 and Smu776.

The relative conservation of residues in (A) Smu472 and (B) Smu776 without the PUA domain is

shown with surface residues coloured according to their conservation levels calculated in ConSurf Server (5). (A) The encircled regions in the THUMP domain (I) and in the central groove (II) have concentrated positive charged residues that could interact with phosphates in the RNA backbone. Region III represents the conserved residues that directly bind SAH. Residues 307NPPYGERL314 in region IV constitute the most conserved motif of Smu472, and are situated in the loop between β 4-N and α D-N of the N-terminal methyltransferase fold. (B) Conserved residues in the cleft between EEHEE and the methyltransferase domain (I) on the surface of Smu776. Region II binds the cofactor. Region III represents a conserved insertion motif between β 4-C and α D-C in the C-terminal methyltransferase fold where a β -hairpin interacts directly with the base moiety of SAH forming a lid-like structure over the cofactor; this region appears to be important for catalysis. Surface potential representation of (C) Smu472 and (D) Smu776 without the PUA domain showing positively charged regions (blue) that could serve to bind the RNA backbone. The red circle indicates insertion sequences that are highly acidic in Smu472 but not in YcbY-N.

Table S1: Crystal growth, data collection, structure determination and refinement statistics

	Smu472	Smu776	YcbY-SAH	YcbY-SAM
Reservoir conditions	0.1 M Bicine pH9.0, 2% v/v Dioxane, 10% w/v PEG20000	0.1 M Na cacodylate pH 6.5, 0.2M Mg(OAc) ₂ 20% PEG8000	0.1 M HEPES pH 7.0, 12% (w/v) PEG 8000	0.1 M HEPES pH 7.0, 12% (w/v) PEG 8000
Additive	none	none	6% (w/v) n-Octanoylsuc	6% (w/v) n-Octanoylsuc
Data collection	Photon Factory Japan	Home X-ray source Cu anode	SSRF, Shanghai China	SSRF, Shanghai China
Structure determination methods	Se-SAD	SIRAS	MR	MR
Space group	P2 ₁ 2 ₁ 2 ₁	C2	P2 ₁	P2 ₁
Unit cell dimensions				
<i>a, b, c</i> , Å	65.25 67.35 91.85	165.68 50.823 59.57	73.646 140.834 102.866	73.724 140.586 102.097
α, β, γ , °	90.0 90.0 90.0	90.00 107.14 90.00	90.00 102.30 90.00	90.00 101.88 90.00
Resolution, Å	50-1.97	50-2.20 (2.30-2.20)	50-2.20 (2.24-2.20)	50-2.60 (2.66-2.60)
R_{merge}^a , %	11.0 (27.0)	3.99(14.93)	12.1 (68.0)	8.8 (29.3)
I/σ	26.3 (4.9)	10.73(2.68)	16.4 (1.6)	23.7 (4.84)
completeness, %	99.6 (96.6)	90.8 (71.9)	96.8 (88.2)	94.7 (76.5)
Refinement resolution, Å	50-1.97 (2.02-1.97)	50-2.20 (2.33-2.20)	20-2.20 (2.26-2.20)	20-2.60 (2.67-2.60)
R_{cryst}^b , %	18.1 (28.0)	17.3 (18.7)	21.1 (35.0)	18.2 (21.9)
R_{free}^c , %	23.8 (40.7)	23.2 (24.9)	25.8 (39.0)	24.7 (31.0)

Table S2: Deoxyoligonucleotide primers used for rRNA isolation and for PCR analyses.

Primer name	Sequence [5' - 3']	Application
p87	GCCCTCAACTTTATGACATG	Checking the integrity of <i>ycbY</i> gene
p88	GCGCTACAATCAAATCCTTGC	Checking the integrity of <i>ycbY</i> gene
p90	GATATGAATTCTCTGTTTGCCAGTAC- GGCCCGTGGGCTGGAAGAGCTGTTA- AAACTGAAGTGTAGGCTGGAGC- TGCTTCG	Knock-out of <i>ycbY</i> gene; complementary sequence for hybridization to kanamycin resistance cassette (bold)
p91	CTTTCAGGCTGCGGTAATCAGCCA- GCAGTTGTGGATCTGGCGGTTACG- GGCGAAATCCTGCATATGAATATC- CTCCTTAGTTC	Knock-out of <i>ycbY</i> gene; complementary sequence to kanamycin cassette in bold
Ec2069rt	GCTCCCACCTATCCTACACATC	Reverse transcriptase extension to investigate methylation at G2069
Ec2445rt	CTTGGGCGGTATCAGCCTG	Reverse transcriptase extension to investigate methylation at G2445
Ec2069	CAGTGTC AAGCTATAGTAAAGGTTC- ACGGGGTCTTTCCGTCTTGCCGCGG	Isolation of <i>E. coli</i> 23S rRNA around position G2069 for MS
Ec2445	GGGCGGTATCAGCCTGTTATCCCC- GGAGTACCTTTTATCCGTTGAGCG	Isolation of <i>E. coli</i> 23S rRNA around position G2445 for MS
Sm2069	GCAATATCAAACCTGCAGTAAAGCT- CCATGGGGTCTTTCCGTCTTGTCGC	Isolation of <i>S. mutans</i> 23S rRNA around position G2069 for MS
Sm2445	GGGGGAGATAAGCCTGTTATCCCC- AGGGTAGCTTTTATCCGTTGAGCG	Isolation of <i>S. mutans</i> 23S rRNA around position G2445 for MS

References

1. Kimura, S., Ikeuchi, Y., Kitahara, K., Sakaguchi, Y., Suzuki, T., and Suzuki, T. (2012) Base methylations in the double-stranded RNA by a fused methyltransferase bearing unwinding activity, *Nucleic Acids Res* **in press**.
2. McLuckey, S. A., Van Berkel, G. J., and Glish, G. L. (1992) Tandem mass spectrometry of small multiply charged oligonucleotides, *J. Am. Soc. Mass Spectrom* **3**, 60-70.
3. Cannone, J. J., Subramanian, S., Schnare, M. N., Collett, J. R., D'Souza, L. M., Du, Y., Feng, B., Lin, N., Madabusi, L. V., Müller, K. M., Pande, N., Shang, Z., Yu, N., and Gutell, R. R. (2002) The Comparative RNA Web (CRW) Site: an online database of comparative sequence and structure information for ribosomal, intron, and other RNAs, *BMC Bioinformatics* **3**, 2.
4. Noller, H. F. (1984) Structure of ribosomal RNA., *Ann. Rev. Biochem.* **53**, 119-162.
5. Ashkenazy, H., Erez, E., Martz, E., Pupko, T., and Ben-Tal, N. (2010) ConSurf 2010: calculating evolutionary conservation in sequence and structure of proteins and nucleic acids, *Nucleic Acids Res* **38**, W529.