Published in final edited form as:
J Biol Chem. 2007 February 23; 282(8): 5880–5887. doi:10.1074/jbc.M608214200.

Methyltransferase That Modifies Guanine 966 of the 16 S rRNA: *FUNCTIONAL IDENTIFICATION AND TERTIARY STRUCTURE*[*]

Dmitry V. Lesnyak[‡,1], **Jerzy Osipiuk**[§,1], **Tatiana Skarina**[¶], **Petr V. Sergiev**[‖], **Alexey A. Bogdanov**[‖], **Aled Edwards**[¶], **Alexei Savchenko**[¶], **Andrzej Joachimiak**[§,2], and **Olga A. Dontsova**[‖,3]

[‡] Department of Bioinformatics and Bioengineering, Moscow State University, Moscow 119992, Russia

[§] Midwest Center for Structural Genomics and Structural Biology Center, Biosciences Division, Argonne National Laboratory, Argonne, Illinois 60439

[¶] Banting and Best Department of Medical Research, University of Toronto, Toronto, Ontario M5G IL6, Canada

[‖] Department of Chemistry and A.N. Belozersky Institute of Physico-Chemical Biology, Moscow State University, Moscow 119992, Russia

Abstract

N^2-Methylguanine 966 is located in the loop of *Escherichia coli* 16 S rRNA helix 31, forming a part of the P-site tRNA-binding pocket. We found *yhhF* to be a gene encoding for m^2G966 specific 16 S rRNA methyltransferase. Disruption of the *yhhF* gene by kanamycin resistance marker leads to a loss of modification at G966. The modification could be rescued by expression of recombinant protein from the plasmid carrying the *yhhF* gene. Moreover, purified m^2G966 methyltransferase, in the presence of *S*-adenosylomethionine (AdoMet), is able to methylate 30 S ribosomal subunits that were purified from *yhhF* knock-out strain *in vitro*. The methylation is specific for G966 base of the 16 S rRNA. The m^2G966 methyltransferase was crystallized, and its structure has been determined and refined to 2.05 Å. The structure closely resembles RsmC rRNA methyltransferase, specific for m^2G1207 of the 16 S rRNA. Structural comparisons and analysis of the enzyme active site suggest modes for binding AdoMet and rRNA to m^2G966 methyltransferase. Based on the experimental data and current nomenclature the protein expressed from the *yhhF* gene was renamed to RsmD. A model for interaction of RsmD with ribosome has been proposed.

Ribosomal RNAs form the main functional core of the ribosome, a universal molecular machine for protein synthesis. It is believed that primordial ribosome was initially composed entirely of RNA (1) and proteins were added later in evolution. A set of four ribonucleotides used in rRNA is rather limited for the full range and fine tuning of ribosomal functions especially when compared with 20 amino acids the proteins are composed of. The diversity in

[*]This work was supported by grants from Howard Hughes Medical Institute, Russian Foundation for Basic Research, Leading Scientific Schools, National Institutes of Health Grants GM074942 and GM62414, and by the United States Department of Energy, Office of Biological and Environmental Research, under Contract W-31-109-Eng-38.

[2]To whom correspondence may be addressed: Midwest Center for Structural Genomics and Structural Biology Center, Biosciences Division, Argonne National Laboratory, 9700 South Cass Ave., Bldg. 202, Argonne, IL 60439. andrzejj@anl.gov.
[3]To whom correspondence may be addressed: A.N. Belozersky Institute of Physico-Chemical Biology (Laboratory Bldg. A), Moscow State University, Moscow 119992, Russia. Tel.: 7-495-9395418; Fax: 7-495-9393181; dontsova@genebee.msu.su.
[1]These two authors contributed equally to this work.

The atomic coordinates and structure factors (code 2FPO) have been deposited in the Protein Data Bank, Research Collaboratory for Structural Bioinformatics, Rutgers University, New Brunswick, NJ (http://www.rcsb.org/).

RNA could be increased through the usage of a set of modified nucleotides (2). A number of such nucleotides have been found in functional ribosomal RNA molecules (as well as other RNA). It was also shown that nucleotide modification is an important component of the ribosome maturation process (3). The nucleotide modifications can involve bases and ribose. The most common in bacteria is base methylation, but other substitutions have also been reported (4). The specific functional role of nucleotide modifications is not always obvious, but their clustering in the ligand-binding and catalytic centers (5) suggests their involvement in the vital ribosome functions. Even more striking is the fact that many of the nucleotide methylations are performed by site-specific RNA methyltransferases, at a rather significant metabolic cost to the cell.

The surrounding of the P-site of the ribosome is particularly rich in modified bases. Several RNA small patches consisting entirely of modified bases have been found. For example, helix 31 of 16 S rRNA contains two modified bases in a row: m^2G966 and m^5C967 (Fig. 1*A*). m^2G966 is in direct contact with P-site-bound tRNA as revealed by a footprinting assay (6) and later structural studies (7,8). It is located in the head of the small ribosomal subunit, above the P-site-bound tRNA and forms a direct contact with the tip of the anticodon loop (7,8) (Fig. 1*B*). Interestingly, despite the lack of conservation of the nucleotide 966 among the kingdoms of life (9), and phenotypically silent character of its mutants (10), the modified nature of this base is conserved (11,12). A protein with RNA methyltransferase activity, specific for m^2G966, was partially purified from *Escherichia coli* extracts (13). It was shown to act on assembled 30 S subunits but not protein-free 16 S rRNA as substrate. Together ribosomal proteins S7 and S19 were shown to be sufficient to render 16 S rRNA a substrate for the (guanine-N^2)-methyltransferase (13). Both proteins affect helix 31 folding, as was shown by footprinting (14). On the contrary, RNA methyltransferase, specific for C967 methylation, was shown to modify naked 16 S rRNA but not 30 S subunits (13). The addition of S7 and S19 together is sufficient to impede C967 methylation.

While a gene coding for m^5C967 methyltransferase, *rsmB*, was identified twice independently (15,16), an enzyme specific for m^2G966 modification, *rsmD*, was suggested to be encoded by the *ygjO* gene on the basis of sequence similarity (17) with m^2G1207 methyltransferase (18). Here we present evidence that in *E. coli* the *yhhF* gene, not *ygjO*, is encoding for m^2G966 specific methyltransferase RsmD.

We have also determined the crystal structure of RsmD at 2.05 Å resolution. The structure of RsmD is highly similar to that of RsmC protein (19), whose modification target m^2G1207 is located very close in the spatial structure of the small ribosomal subunit. Comparison of the RsmD structure to that of the RsmB methyltransferase (20), specific for the methylation of adjacent residue with different base specificity (m^5C967), made it clear why RsmB modify protein-free 16 S rRNA, while RsmD acts on the assembled 30 S subunit.

EXPERIMENTAL PROCEDURES

rRNA Preparation and Reverse Transcription

E. coli cells of the strain harboring replacements of the *rsmD*(*yhhF*) gene by the kanamycin resistance gene, the same strain carrying the plasmid expressing *rsmD* (*yhhF*), and the parental BW25141 strain (21–23) were harvested at the mid-log growth phase. After ultrasonic disruption and sedimentation of debris, ribosomes were pelleted by ultracentrifugation in the Ti70 rotor at 30,000 rpm, for 3 h. The ribosomal pellet was resuspended in the high salt buffer (HepesK, pH 7.8, 10 mM $MgCl_2$, 500 mM NH_4Cl) and centrifuged through the 20% sucrose cushion at 40,000 rpm, for 18 h. 30 S ribosomal subunits were prepared by ultracentrifugation through the 10–30% sucrose density gradient in SW28 rotor at 20,000 rpm for 18 h. 30 S peak was collected, and the subunits were sedimented by ultracentrifugation in Ti70 rotor at 30,000

rpm, for 18 h. Ribosomal RNA was prepared from the ribosomes by SDS/phenol extraction and ethanol precipitation. Reverse transcription was performed as described (24). The same procedure was used for preparation of rRNA from ygjO and ycbY knock-out strains. The primer used for reverse transcription was complementary to the 16 S rRNA nucleotides 995–1011. Products of reverse transcription were separated by electrophoresis in the 10% denaturing polyacrylamide gel and visualized with a phosphorimager.

Protein Expression and Purification, in Vivo Methylation Assay

Recombinant RsmD protein was prepared from M15 *E. coli* cells, harboring pCA24N plasmid with the *rsmD*(*yhhF*) gene cloned under control of T5lac promoter (23). Cells were grown in LB medium at 37 °C until A_{600} 0.5 and induced by isopropyl isopropyl β-D-thiogalactopyranoside, 0.5 mM f.c. After induction, cells were grown for an additional 3 h and lysed by sonication in a buffer of 50 mM TrisCl pH 7.5, 10 mM $MgCl_2$, 100 mM NH_4Cl. Protein purification on nickel-nitrilotriacetic acid-agarose (Qiagen) was done in the same buffer, containing 10 mM imidazole. After washing three times with the same buffer containing 30 mM imidazole, the protein was eluted by increasing imidazol concentration up to 200 mM. To eliminate endogenous AdoMet[4] molecule the protein was incubated with 2 mM SAH in the 50 mM Tris-Cl, pH 7.5, 10 mM $MgCl_2$, 100 mM NH_4Cl for 1 h at 37 °C. Then RsmD protein was dyalized, and its purity was proven by SDS-PAGE. *In vitro* methylation assay was performed in a buffer containing 50 mM HepesK, pH 7.6, 2.5 mM $Mg(OAc)_2$, 20 mM NH_4Cl, 100 mM KCl, 8 mM β-mercaptoethanol, 0.2 mM AdoMet. We used 0.1 pmol of RsmD protein and 15 pmol of either rRNA or 30 S subunits for 50-μl reaction mixture. The reaction was stopped after 1 h at 37 °C.

Growth Rate and Growth Competition

Growth rate measurement was performed in LB medium at 37 °C. Three colonies of either the BW25141 parental strain or the *rsmD*(*yhhF*) deletion strain were used for inoculation of three independent 5-ml overnight cultures. The next day, each of these cultures was used for inoculation of separate 500-ml containers of fresh LB medium. Initial density of the cultures after inoculation was A_{600} 0.01. Cell growth was monitored by spectrophotometer until stationary phase.

A growth competition experiments were performed using LB medium at 30, 37, and 42 °C or in M9 medium at 37 °C. Equal amounts of kanamycin-sensitive BW25141 parental strain and kanamycin-resistant *rsmD*(*yhhF*) deletion strain cells were used for simultaneous inoculation of the same 1 ml of LB or M9 medium. At each 24-h growth cycle the medium was diluted 1000-fold, so that 1 ml of the medium with bacteria was transferred into a tube with 1 ml of fresh medium. At the same time, serial dilutions of the bacterial culture were made until a dilution of $1:10^{-7}$ was reached. Equal 10-μl aliquots from the diluted cultures were plated to LB agar and LB agar supplemented with kanamycin (50 μg/ml). Plates were incubated at 37 °C overnight, and the colonies were counted and compared.

Protein Cloning Expression and Purification for Crystallization

The open reading frame of the *rsmD*(*yhhF*) gene was amplified by polymerase chain reaction from *E. coli* genomic DNA (ATCC). The gene was cloned into the NdeI and BamHI sites of a modified pET15b cloning vector (Novagen) in which the tobacco etch virus protease cleavage site replaced the thrombin cleavage site and a double stop codon was introduced downstream from the BamHI site. This construct provides for an N-terminal His_6-tag separated from the gene by a tobacco etch virus protease recognition site (ENLYFQG). The fusion protein was

[4]The abbreviations used are: AdoMet, *S*-adenosylomethionine; SAD, single-wavelength anomalous diffraction; PDB, Protein Data Bank.

overexpressed in *E. coli* BL21-Gold (DE3) (Stratagene) harboring an extra plasmid encoding three rare tRNAs (AGG and AGA for Arg, ATA for Ile). The cells were grown in LB medium at 37 °C to an A_{600} of 0.6 and protein expression induced with 0.4 mM isopropyl β-D-thiogalactopyranoside. After induction, the cells were incubated with shaking overnight at 15 °C. The harvested cells were resuspended in binding buffer (500 mM NaCl, 5% glycerol, 50 mM Hepes, pH 7.5, 5 mM imidazole), flash-frozen in liquid N_2, and stored at −70 °C. The thawed cells were lysed by sonication after the addition of 0.5% Nonidet P-40 and 1 mM each of PMSF and benzamidine. The lysate was clarified by centrifugation (30 min at 27,000 × *g*) and passed through a DE52 column pre-equilibrated in binding buffer. The flow-through fraction was then applied to a metal chelate affinity column charged with Ni^{2+}. The His_6-tagged protein was eluted from the column in elution buffer (500 mM NaCl, 5% glycerol, 50 mM Hepes, pH 7.5, 250 mM imidazole) and the tag then cleaved from the protein by treatment with recombinant His-tagged tobacco etch virus protease. The cleaved protein was then resolved from the cleaved His_6-tag and the His_6-tagged protease by flowing the mixture through a second Ni^{2+} column.

The RsmD protein was dialyzed in 10 mM Hepes, pH 7.5, 500 mM NaCl, and concentrated using a BioMax concentrator (Millipore). Before crystallization, any particulate matter was removed from the sample by passing through a 0.2-μm Ultra-free-MC centrifugal filter (Millipore). Selenomethionine-enriched protein was produced according to previously described procedure (25–28) and purified under the same conditions as the native protein. The reducing reagent tris(2-carboxyethyl)phosphine (0.5 mM) was added to all purification buffers.

Protein Crystallization

The protein was crystallized by the vapor diffusion method in hanging drops by mixing 2 μl of the protein solution with 2 μl of 0.2 M sodium tartrate, 20% polyethylene glycol 3350, and equilibrated at 20 °C over 500 μl of this solution. Before data collection, crystals were soaked for 20 s in a cryoprotectant consisting of 10% glycerol, 10% sucrose, and 10% ethylene glycol in the crystal mother liquor and then flash-frozen in liquid nitrogen.

Data Collection

Diffraction data were collected at 100 K at the 19ID beamline of the Structural Biology Center at the Advanced Photon Source, Argonne National Laboratory. Single-wavelength anomalous diffraction (SAD) data to 2.05 Å were collected at selenium-peak wavelength from a selenomethionine substituted protein crystal as described earlier by Walsh *et al.* (27). One crystal was used for data collection at 100 K, and a total oscillation range of 360° was obtained using inverse beam geometry. Data were integrated, reduced, and scaled with the HKL2000 suite (28). Data statistics are summarized in Table 1.

Structure Determination and Refinement

The structure was determined by SAD phasing using HKL2000_ph suite (29) incorporating SHELXD (30), MLPHARE (31), DM (31), and SOLVE/RESOLVE (32) programs. The initial model was further improved by ARP/wARP autobuilding (33). The final model was refined using the Refmac 5 program (34) of the CCP4 program suite (31). The crystallographic structure was built, analyzed, and validated using programs Coot (35) and PROCHECK (36). Final analysis of the structure was performed by SSM (37), DALI (38), and ProFunc (39) servers. Final refinement statistics are presented in Table 2.

Protein-rRNA Docking

GRAMM program (40) was used for geometric docking of 16 S RNA fragment (PDB code 2AW7, residues 940–985) to the RmsD protein structure. The parameters of the grid step,

repulsion, and attraction double range were equal to 3.4, 10, and 0.5, respectively. The projection was black and white. The best docking model served as a guideline for manual docking of RmsD protein to the entire 16 S RNA molecule using program Coot (35).

RESULTS

Identification of the Methyltransferase, Specific for G966 of the 16 S rRNA

N^2-Methylguanosine specifically interrupts the primer-extension by avian myeloblastosis virus reverse transcriptase by impairing the m^2G-C Watson-Crick base pair formation in the reverse transcriptase active site. We used the primer-extension method for the detection of the methylation at position G966 of 16 S RNA. Reverse transcription of the 16 S rRNA extracted from the wild type strain stops after the template residue 967 (Fig. 2, *lane 1*), unable to incorporate dC opposite to the G966, when G966 is methylated at the 2-amino group. Ribosomes were prepared from the strains, carrying disruption of the candidate genes for the G966 methylation: *yhhF*, *ygjO*, and *ycbY*. Ribosomal RNA was isolated and served as a template for reverse transcription. A complete loss of the primer-extension stop correspondent to the m^2G966 was observed only for the rRNA, extracted from the strain, harboring the *yhhF* gene disruption (Fig. 2, *lane 2*). In contrast to the prediction (17), the *ygjO* gene of *E. coli* encodes methtyltransferase that is not responsible for m^2G966 formation (Fig. 2, *lane 3*). Recently, the true *ygjO* target was identified to be m^2G1835 of the 23 S rRNA (41). The G966 modification also remained intact in the rRNA extracted from the *ycbY* knock-out strain (Fig. 2, *lane 4*). *ycbY* encodes another ribosomal m^2G methyltransferase, responsible for the 23 S rRNA m^2G2445 formation (42). We propose that, according to the accepted nomenclature described in a recent publication on identification of m^5C1407 specific methyltransferase (43) and previous characterization of partially purified enzymatic activity (13), the *yhhF* gene is now to be renamed to *rsmD*.

In Vivo Complementation of the rsmD Gene Inactivation by Supplying the rsmD Gene on the Plasmid

It might be imagined that multiple genes can be inactivated during the PCR-based gene replacement by the kanamycin resistance cassette, surrounded by the homology regions, similar to the ones flanking *rsmD* gene, or change in the genomic context might influence the surrounding genes. To eliminate a possibility that the change in the DNA region other than *rsmD* could lead to the loss of G966 methylation, we carried out *in vivo* complementation experiments. A plasmid, carrying *rsmD* gene under control of the isopropyl *β*-D-thiogalactopyranoside-inducible T5*lac* promoter (23), was introduced to the strain with the chromosome copy of the *rsmD* gene replaced by the kanamycin resistance gene. Expression of the *rsmD* gene from the plasmid efficiently complemented chromosomal *rsmD* gene disruption (Fig. 2, *lane 5*).

Recombinant RsmD Protein Can Methylate G966 in Vitro

Although it was shown that the *rsmD* gene product is necessary for m^2G966 formation, there is still a chance that other genes are also required. To establish that the *rsmD* gene product is solely responsible for the methylation of the 16 S rRNA G966 residue, we performed *in vitro* methylation of either 16 S rRNA or 30 S ribosomal subunits by purified recombinant RsmD protein. RsmD protein carrying C-terminal His_6 tag was expressed in *E. coli* and purified by affinity chromatography on nickel-nitrilotriacetic acid-agarose. The protein was >95% pure according to the SDS-polyacrylamide gel electrophoresis. Its molecular mass corresponded to the predicted 22 kDa.

Both protein-free ribosomal RNA and ribosomal subunits purified from the strain, defective in the *rsmD* gene, were tested as *in vitro* substrates for G966 methylation. The products of

methylation were analyzed by reverse transcription. Recombinant RsmD protein efficiently methylated G966 of the assembled 30 S subunits *in vitro* in the presence of AdoMet (Fig. 2, *lane 6*). No reverse transcription stop was observed in the absence of AdoMet (Fig. 2, *lane 7*), giving additional proof that the stop was caused by the methylation but not nucleolytic cleavage. As was shown earlier for partially purified enzymatic activity (13), protein-free 16 S rRNA was not a substrate for RsmD (Fig. 2, *lane 8*).

Influence of the 16 S rRNA G966 Methylation on the Growth of E. coli Cells

Doubling time of the strain JW3430 (21), carrying the inactive *rsmD* gene (42 ± 1 min), was essentially the same as that of the parental (22) *E. coli* strain BW25141 (41 ± 1 min). To distinguish slight variations in the growth characteristics, we performed a growth competition assay. Equal numbers of the kanamycin-sensitive parental strain cells and kana-mycin-resistant *rsmD* knock-out strain cells (A_{600} 0.001) were mixed and incubated for 24 h at 30, 37, and 42 °C in rich LB or minimal M9 medium (37 °C only). At the end of the growth cycle 1/1000th of the medium containing the mixture of the cells was used for inoculation of the fresh portion of LB or M9 at 1000 times dilution. The cells amplify 1000-fold at each growth cycle or undergo ~10 doublings. After each growth cycle, the cells were serially diluted and plated on either LB agar or LB agar plates supplemented with kanamycin. Kanamycin-resistant cell titer related to the total cell titer is presented (Fig. 3). The decrease of *rsmD* knock-out cell proportion in the mixture was very slow, but detectable, and suggests that G966 modification modestly affects cell fitness. In M9 minimal medium no disadvantage of the *yhhF* knock-out strain was detected over 80 cell doublings.

Structure of RsmD rRNA Methyltransferase

The structure of RsmD protein (198 amino acids) (Fig. 4, *A* and *B*) has been determined by the SAD method and refined to 2.05 Å (Table 1). The protein crystallizes with six molecules in the asymmetric unit, providing a total of 18 selenium sites for use in phasing. Twelve of these selenium sites were located using SHELXD program (30) implemented in HK2000_ph suite (29). SAD phases were calculated with the program MLPHARE of the CCP4 program suite (31) with parameters of phasing power and figure of merit equal to 1.53 and 0.368, respectively. Solvent flattening was performed with the program DM of the CCP4 program suite (31) improving FOM to 0.700 value. The initial structure model was built with SOLVE/RESOLVE (32) program and completed with ARP/wARP program (33).

The RsmD protein closely resembles numerous methyltransferases, including rRNA methyltransferases, m^2G1207 RsmC (Z-score and root mean square deviation equal to 17.5 and 2.8 Å, respectively) (Fig. 5*A*) and C-terminal domain of m^5C967 RsmB (Z-score and root mean square deviation equal to 13.3 and 2.6 Å, respectively) (Fig. 5*B*). The overall structure consists of 8-stranded β-sheet flanked on both sides by α-helices (Fig. 4*A*). The superimposition of methyltransferase structures (PDB codes 1I9G, 1T43, 1VE3, 1VID, and 1WY7), containing AdoMet or its analogs, identified a cavity formed by loops β4-α4, β5-α5, and β6-α6 as AdoMet-binding sites (Fig. 6*A*). The main part of AdoMet fits very well in the RsmD cavity, suggesting a universal binding mode of AdoMet in this class of methyltransferases. The upper ridge of the cavity is formed by DPPF motif (residues 127–130) matching perfectly conserved motif D/N/S-P-P-Y/F (44) (Fig. 6*A*). This motif is known to be involved in AdoMet binding, forming methyltransferase active sites (44), and is involved in catalysis. The main part of the methionine side chain of AdoMet overhangs into a large cavity, which is most likely a binding site for substrate, guanine 966 (Fig. 6*A*). The cavity is large enough to accommodate a guanine ring and position its amino group next to the active site residues. The AdoMet sulfur atom and reactive methyl group are located on the edge between AdoMet- and guanine-binding pockets (Fig. 6*A*).

Proposed Model of rRNA/RsmD Interactions in the Active Site

The GRAMM program (37) was used for geometric docking of *E. coli* 16 S RNA (45) (PDB code 2AW7) to the RsmD protein structure. Various lengths of rRNA, ranging from 5 to about 200 residues, were tested. We also examined different protein models including each of the protein chains from PDB deposits as well as models with deleted parts of N terminus (residues 10–35). We assumed that a flexible protein N terminus could obstruct access to the active site in the structure. The best result was obtained for 940–985 fragment of rRNA and the protein with deleted 35 N-terminal residues (model consisting of residues 36–192). Very similar results were also achieved with protein lacking residues 22–35. The best model places the G966 guanine ring partially submerged inside the substrate-binding pocket and positions the target N^2 atom inside the protein active center (Fig. 6, *A* and *B*). The N^2 atom is located within a distance of 1.9–4.7 Å from atoms essential for catalysis: AdoMet C^{ε} atom, Asp-127 side chain oxygen and Pro-128 carbonyl group. These values are in the acceptable range for approximate presentation of molecular interactions. There are no significant collisions in the obtained model. However, after including all N-terminal residues, potential collisions are observed for residues 10–13 and 20–29. We believe that protein N terminus, especially the mentioned residues, must acquire different conformation for RNA binding. Moreover, positive surface electrostatic potential of N terminus and α3-helix (residues 85–97) should be crucial for effective rRNA binding (Fig. 6*A*).

The whole 16 S RNA molecule was superposed over the rRNA fragment used in the docking to show a hypothetical RsmD protein-ribosome complex (Fig. 6*C*). In this model, the protein is positioned between two major domains of 16 S rRNA, head and body, next to 16 S rRNA neck, at a place of the P-site-bound tRNA. In our model, there are some overlaps of surface protein residues with rRNA residues of head and body. Nevertheless, due to known flexibility of the neck (45–47), more room for the enzyme is likely to be provided by opening space between the head and body. At the same time, small shifts of the helix 31 could allow better fit of the RsmD without loss of the G966 positioning in the protein active site.

DISCUSSION

In our study we discovered that the *yhhF* gene (renamed *rsmD*, according to Ref. 43) of *E. coli* is encoding a methyltransferase responsible for the m^2G966 formation in the 16 S rRNA of 30 S. The structure of RsmD protein was determined and the putative complex of RsmD protein and 30 S subunit was modeled. The model shows how AdoMet and guanine could fit into the active site of the RsmD methyltransferase (Fig. 6. *A* and *B*). The location of the methyl group donor appears to be well conserved among the different methyltransferases, while the position of guanine moiety places the N^2 atom in the direct proximity of the methyl group of AdoMet (Fig. 6, *A* and *B*). This may imply that base flipping of the target residue, occurring in the case of DNA methyltransferases (48), is not crucial for activity of RsmD methyltransferase. This assumption is supported by the observed solvent accessibility of G966 guanine ring, resulting from three-dimensional conformation of the region adjacent to the base observed in 30 S subunit structures deposited in the PDB. Another feature of DNA methyltransferases, the flipping of active site loop, may explain the minor collisions of rRNA with the N-terminal part of the RsmD protein. The flipping of the 21-residue loop and locking target base in the active center was observed by NMR studies of HhaI DNA methyltransferase in solution (49). If the N-terminal part of the protein (approximately residues 1–36) is flexible, then this loop may move and convert protein into the open form, which does not block access of rRNA to the active center. After binding substrate, the loop may flip back and turn into the closed form ready for the methyl transfer reaction.

The biological function of many modified bases in rRNA is rather poorly understood. However, it is absolutely clear that modified nucleotides form clusters within the functional centers of

the ribosome (5) and are extremely scarce in the regions outside the enzymatic and ligand-binding sites. Modifications are often highly conserved among the species, and many organisms maintain modification function at significant metabolic expense. The 16 S rRNA nucleotide m^2G966 is located in the very important functional region. Together with its adjacent m^5C967, it is in direct contact with the very tip of the anticodon loop of the P-site-bound tRNA (7,8) (Fig. 1*B*). It is protected from chemical modification by tRNA bound to the P-site (6). Surprisingly, 16 S rRNA residue 966 is not very conserved. It shows a preference to be guanosine in bacteria and chloroplasts. In archaea and eukaryotes it is usually uridine, while it is mainly adenosine in mitochondria (9). According to NCBI (50), *E. coli* RsmD protein is a member of COG0742, nearly universal among bacteria but completely absent in archaea and eukarya. RsmD ortholog distribution agrees with conservation of its target nucleotide. In agreement with the lack of conservation mutations of m^2G966 in *E. coli*, this does not lead to significant growth retardation (10). Mutant 30 S subunits are normally incorporated into the 70 S ribosome pool and show no functional defects. The base 966 is modified in a wide range of organisms and was shown to be acp^3U in archaea (11) and $m^1acp^3\psi$ in eukaryotes (12). In the case of m^2G966, lack of modification leads to a small decrease in the cell fitness, as evidenced from growth competition assays (Fig. 3). It is unlikely that any vital function of the ribosome is affected. Remarkably, all of these modifications influence Watson-Crick positions within the modified bases. One can imagine that such modifications might be useful to prevent the formation of some unwanted alternative secondary or tertiary structures. Alternatively, the hydrophobic modification at nucleotide 966 might be more important than the nature of the nucleotide it is attached to. This small hydrophobic patch could make a hydrophobic interaction with the anticodon of the P-site-bound tRNA.

A set of enzymes, such as RsmB, responsible for m^5C967 modification could utilize protein-free rRNA as a substrate (13). In contrast, the other modification enzymes, such as the RsmD identified here, could use assembled subunits as a substrate. *In vivo*, the real substrates are likely to be represented by a set of early and late assembly intermediates. However, it is notable that rRNA methyltransferases could not usually recognize both protein-free rRNA and assembled subunits. There is also a shortage of information regarding methyltransferases that recognize exclusively assemble intermediates and fail to recognize either rRNA or fully assembled particles. The cause and purpose of such a division is not clear, although comparison of different methyltransferase structures (Fig. 5) with location of their target nucleotides (Fig. 1*B*) may offer a chance to solve this puzzle. At present, all three methyltransferases acting on the nucleotide residues located in the head of the small subunit are solved. RsmC is responsible for m^2G1207 formation in the helix 34 (18). It also utilizes assembled 30 S subunits as a substrate (18) and has a structure highly similar to that of RsmD (Fig. 5*A*). RsmB (Fig. 5*B*) modifies nucleotide C967 (15,16), adjacent to m^2G966 modified by RsmD. However, it uses protein-free 16 S rRNA as a substrate (13). Careful examination of G1207, G966, and C967 location in the three-dimensional structure of the subunit (7,8,45) reveals that both G1207 and G966 are exposed to solvent, although they are located in a deep cavity harboring mRNA and tRNA anticodons in the course of translation (Fig. 1*B*). Usually an enzyme binds its substrate in a cavity to achieve good substrate recognition. Here the enzyme is bound in the cavity of the substrate (Fig. 6*C*). Such arrangement allows good enzyme-substrate recognition while keeping enzymes themselves minimal. In contrast, C967 residue in the 30 S structure is less accessible for the methyltransferase, since it is shielded by its neighboring nucleotide G966 (Fig. 1*B*). Consequently, m^5C967 formation could more easily be accomplished at an earlier stage, before the head of the 30 S subunit is assembled. Recognition of protein-free RNA target imposes a requirement for more sophisticated RsmB enzyme. RsmB is likely to accommodate 16 S rRNA helix 31 substrate to the positively charged cleft in the enzyme (20). RsmB methyltransferase possess an additional NusB-like N-terminal domain, which is most likely to recognize other 16 S rRNA elements to reach necessary substrate specificity. At the same time, existence of this very N-terminal domain in RsmB would cause a clash with the major mass

of the 30 S subunit head upon attempt of RsmB to fit to the assembled subunit. RsmD protein does not have N-terminal extension and could easily fit into the cleft, occupied by a P-site-bound tRNA (Fig. 6*C*). It might be, however, that 16 S rRNA secondary structure changes upon assembly, rather than nucleotide accessibility, could influence the RsmB and RsmD substrate selectivity.

A remarkable structure similarity is evident from the comparison of RsmD and RsmC (19) tertiary structures (Fig. 5*A*). It is easy to understand, since both enzymes utilize the same binding cleft of the assembled 30 S subunit. Their target residues, m^2G1207 and m^2G966, are located close in the tertiary structure (Fig. 1*B*). Surprisingly, there is a relatively low level of sequence similarity between RsmC and RsmD. It prevented the *rsmD*(*yhhF*) gene from being computationally predicted as encoding for G966 methyltransferase RsmD, while YgjO protein, highly similar to RsmC by the sequence comparison, was mistakenly suggested to modify G966 (17).

Acknowledgments

We thank Dr. H. Mori for providing us with the knock-out strains and a plasmid, encoding yhhF; Andrea M. Cipriani for critical reviews; and Youngchang Kim, Ruslan Sanishvili, and Elena Evdokimova for preliminary work on the protein crystallization.

References

1. Noller HF, Woese CR. Science 1981;212:403–411. [PubMed: 6163215]
2. Levy M, Miller SL. J Mol Evol 1999;48:631–637. [PubMed: 10229566]
3. Cunningham PR, Richard RB, Weitzmann CJ, Nurse K, Ofengand JT. Biochimie (Paris) 1991;73:789–796.
4. McCloskey JA, Rozenski J. Nucleic Acids Res 2005;33:D135–D138. [PubMed: 15608163]
5. Mueller F, Brimacombe R. J Mol Biol 1997;271:524–544. [PubMed: 9281424]
6. Moazed D, Noller HF. J Mol Biol 1990;211:135–145. [PubMed: 2405162]
7. Yusupov MM, Yusupova GZ, Baucom A, Lieberman K, Earnest TN, Cate JH, Noller HF. Science 2001;292:883–896. [PubMed: 11283358]
8. Petry S, Brodersen DE, Murphy FV IV, Dunham CM, Selmer M, Tarry MJ, Kelley AC, Ramakrishnan V. Cell 2005;123:1255–1266. [PubMed: 16377566]
9. Cannone JJ, Subramanian S, Schnare MN, Collett JR, D'Souza LM, Du Y, Feng B, Lin N, Madabusi LV, Muller KM, Pande N, Shang Z, Yu N, Gutell RR. BioMed Central Bioinformatics 2002;3:15.
10. Jemiolo DK, Taurence JS, Giese S. Nucleic Acids Res 1991;19:4259–4265. [PubMed: 1714565]
11. Youvan DC, Hearst JE. Nucleic Acids Res 1981;9:1723–1741. [PubMed: 6164994]
12. Kowalak JA, Bruenger E, Crain PF, McCloskey JA. J Biol Chem 2000;275:24484–24489. [PubMed: 10818097]
13. Weitzmann C, Tumminia SJ, Boublik M, Ofengand J. Nucleic Acids Res 1991;19:7089–7095. [PubMed: 1766869]
14. Powers T, Changchien LM, Craven GR, Noller HF. J Mol Biol 1988;200:309–319. [PubMed: 3373531]
15. Gu XR, Gustafsson C, Ku J, Yu M, Santi DV. Biochemistry 1999;38:4053–4057. [PubMed: 10194318]
16. Tscherne JS, Nurse K, Popienick P, Michel P, Sochacki M, Ofengand J. Biochemistry 1999;38:1884–1892. [PubMed: 10026269]
17. Bujnicki JM, Rychlewski L. BMC Bioinformatics 2002;3:10. [PubMed: 11929612]
18. Tscherne JS, Nurse K, Popienick P, Ofengand J. J Biol Chem 1999;274:924–929. [PubMed: 9873033]
19. Huang L, Hung L, Odell M, Yokota H, Kim R, Kim SH. J Struct Funct Genomics 2002;2:121–127. [PubMed: 12836702]

20. Foster PG, Nunes CR, Greene P, Moustakas D, Stroud RM. Structure (Camb) 2003;11:1609–1620. [PubMed: 14656444]
21. Baba T, Ara T, Hasegawa M, Takai Y, Okumura Y, Baba M, Datsenko KA, Tomita M, Wanner BL, Mori H. Mol Systems Biol 2006;2:0008.
22. Datsenko KA, Wanner BL. Proc Natl Acad Sci U S A 2000;97:6640–6645. [PubMed: 10829079]
23. Kitagawa M, Ara T, Arifuzzaman M, Ioka-Nakamichi T, Inamoto E, Toyonaga H, Mori H. DNA Res 2005;12:291–299. [PubMed: 16769691]
24. Sergiev PV, Bogdanov AA, Dahlberg AE, Dontsova O. J Mol Biol 2000;299:379–389. [PubMed: 10860746]
25. O'Gara M, Adams GM, Gong W, Kobayashi R, Blumenthal RM, Cheng X. Eur J Biochem 1997;247:1009–1018. [PubMed: 9288926]
26. Qoronfleh MW, Ho TF, Brake PG, Banks TM, Pulvino TA, Wahl RC, Eshraghi J, Chowdhury SK, Ciccarelli RB, Jones BN. J Biotechnol 1995;39:119–128. [PubMed: 7755966]
27. Walsh MA, Dementieva I, Evans G, Sanishvili R, Joachimiak A. Acta Crystallogr Sect D Biol Crystallogr 1999;55:1168–1173. [PubMed: 10329779]
28. Otwinowski Z, Minor W. Methods Enzymol 1997;276:307–326.
29. Terwilliger TC. Acta Crystallogr Sect D Biol Crystallogr 2003;59:38–44. [PubMed: 12499537]
30. Schneider TR, Sheldrick GM. Acta Crystallogr Sect D Biol Crystallogr 2002;58:1772–1779. [PubMed: 12351820]
31. Collaborative Computational Project Number 4. Acta Crystallogr Sect D Biol Crystallogr 1994;50:760–763. [PubMed: 15299374]
32. Minor W, Cymborowski M, Otwinowski Z, Chruszcz M. Acta Crystallogr Sect D Biol Crystallogr 2006;62:859–866. [PubMed: 16855301]
33. Perrakis A, Morris R, Lamzin VS. Nat Struct Biol 1999;6:458–463. [PubMed: 10331874]
34. Murshudov GN, Vagin AA, Dodson EJ. Acta Crystallogr Sect D Biol Crystallogr 1997;53:240–255. [PubMed: 15299926]
35. Emsley P, Cowtan K. Acta Crystallogr Sect D Biol Crystallogr 2004;60:2126–2132. [PubMed: 15572765]
36. Laskowski RA, MacArthur MW, Moss DS, Thornton JM. J Appl Crystallogr 1993;26:283–291.
37. Krissinel E, Henrick K. Acta Crystallogr Sect D Biol Crystallogr 2004;60:2256–2268. [PubMed: 15572779]
38. Holm L, Sander C. Nucleic Acids Res 1998;26:316–319. [PubMed: 9399863]
39. Laskowski RA, Watson JD, Thornton JM. J Funct Genomics 2003;4:167–177.
40. Katchalski-Katzir E, Shariv I, Eisenstein M, Friesem AA, Aflalo C, Vakser IA. Proc Natl Acad Sci U S A 1992;89:2195–2199. [PubMed: 1549581]
41. Sergiev PV, Lesnyak DV, Bogdanov AA, Dontsova OA. J Mol Biol 2006;364:25–31.
42. Lesnyak DV, Sergiev PV, Bogdanov AA, Dontsova OA. J Mol Biol 2006;364:20–25. [PubMed: 17010378]
43. Andersen NM, Douthwaite S. J Mol Biol 2006;359:777–786. [PubMed: 16678201]
44. Jeltsch A. Chembiochem 2002;3:274–293. [PubMed: 11933228]
45. Schuwirth BS, Borovinskaya MA, Hau CW, Zhang W, Vila-Sanjurjo A, Holton JM, Cate JH. Science 2005;310:827–834. [PubMed: 16272117]
46. Gabashvili IS, Agrawal RK, Grassucci R, Frank J. J Mol Biol 1999;286:1285–1291. [PubMed: 10064696]
47. Ogle JM, Murphy FV IV, Tarry MJ, Ramakrishnan V. Cell 2002;111:721–732. [PubMed: 12464183]
48. Cheng X, Roberts RJ. Nucleic Acids Res 2001;29:3784–3795. [PubMed: 11557810]
49. Klimasauskas S, Szyperski T, Serva S, Wuthrich K. EMBO J 1998;17:317–324. [PubMed: 9427765]
50. Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. Nucleic Acids Res 1997;25:3389–3402. [PubMed: 9254694]

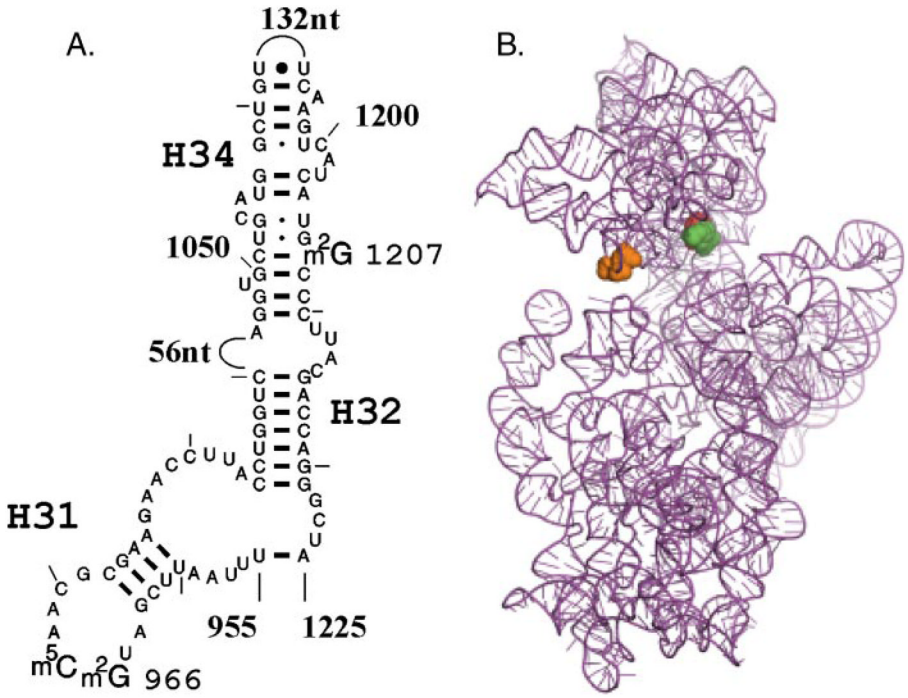

FIGURE 1. Location of the m^2G966 in the 30 S subunit

A, secondary structure of *E. coli* 16 S rRNA part (9) encompassing helices 31, 32, and 34 (*H31*, *H32*, *H34*). Location of the modified bases m^2G966, m^5C967, and m^2G1207 is marked. *B*, structure of the *E. coli* 16 S rRNA (PDB code 2AW7). Methylated residues m^2G966, m^5C967, and m^2G1207 are shown as oversized van der Waals spheres and colored *green*, *red*, and *orange* accordingly (in this view, residue 967 is located behind residue 966).

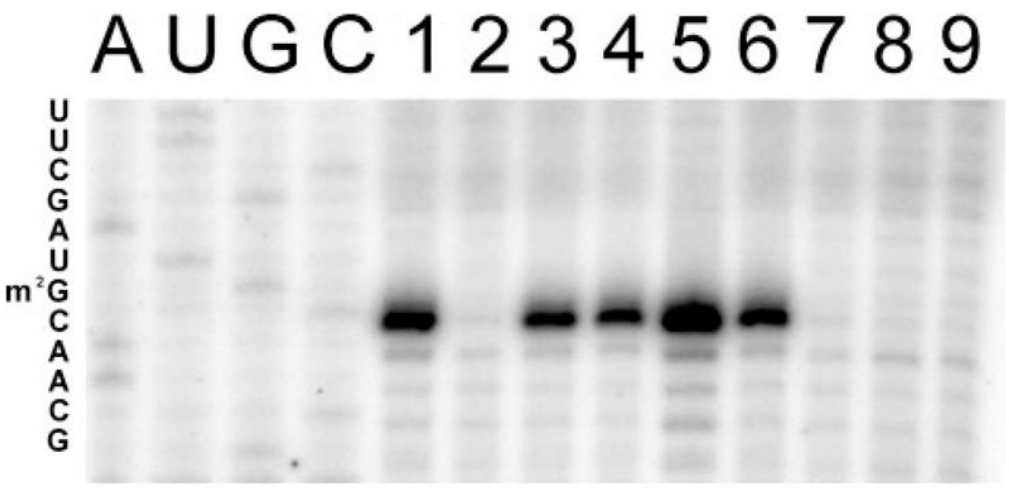

FIGURE 2. Reverse transcription analysis of G966 methylation

"A, U, G, and C" correspond to sequencing lanes. For the sequence clarity rDNA was used as a template for sequencing. Primer extension products in *lanes 1–9* correspond to: the wild-type rRNA (*lane 1*), rRNA extracted from *rsmD*(*yhhF*) knock-out strain (*lane 2*); rRNA extracted from *rlmG*(*ygjO*) knock-out strain (*lane 3*), rRNA extracted from *rlmL*(*ycbY*) knock-out strain (*lane 4*), rRNA extracted from *rsmD* knock-out strain transformed by *rsmD* expression plasmid (*lane 5*), 30 S subunits from *rsmD*(*yhhF*) knock-out strain modified by RsmD protein *in vitro* in the presence of AdoMet (*lane 6*); 30 S subunits from *rsmD*(*yhhF*) knock-out strain modified by RsmD protein *in vitro* in the absence of AdoMet (*lane 7*), 16 S rRNA from *rsmD* (*yhhF*) knock-out strain modified by RsmD protein *in vitro* in the presence of AdoMet (*lane 8*), and 16 S rRNA from *rsmD*(*yhhF*) knock-out strain modified by RsmD protein *in vitro* in the absence of AdoMet (*lane 9*). Primer complementary to positions 995–1011 of the 16 S rRNA was used.

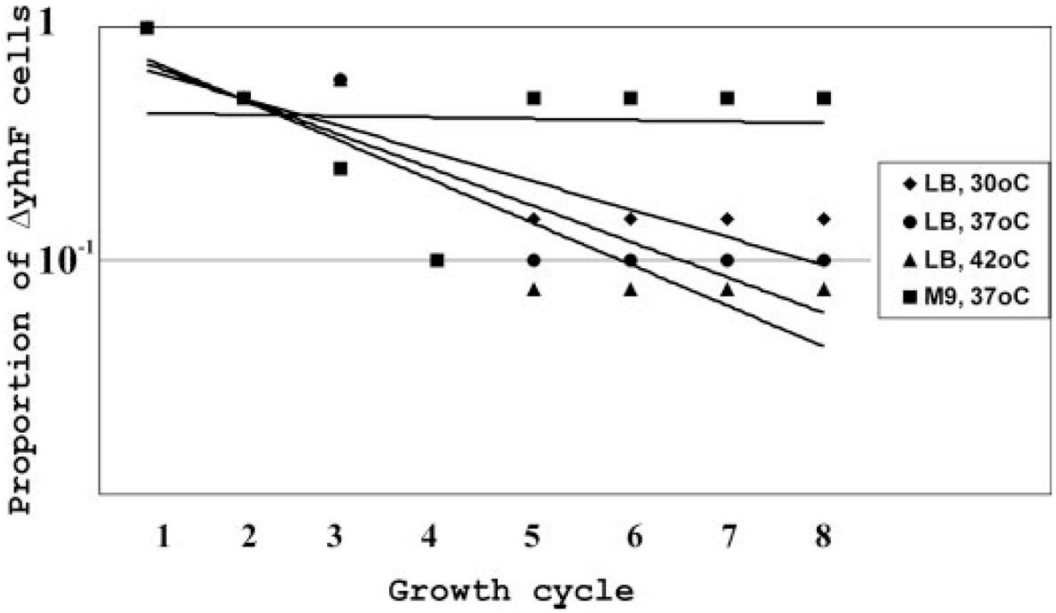

FIGURE 3. Growth competition between the wild type strain and *rsmD* knock-out strain

Shown are a proportion of the *rsmD* knock-out strain cells in the mixture to the wild type calls. Each point corresponds to a 24-h growth cycle. One cycle encompasses ~10 cell doublings.

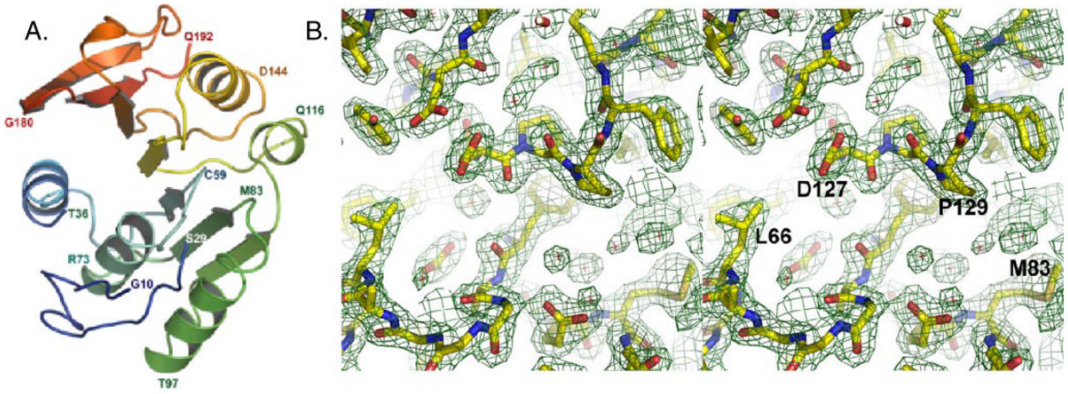

FIGURE 4. Structure of the RsmD protein

A, the RsmD protein monomer structure shown as a schematic in rainbow colors with N terminus and C terminus marked in *blue* and *red*, respectively (only chain A of the PDB code 2FPO deposit is shown). *B*, stereo view of representative electron density map showing the RsmD active-site region.

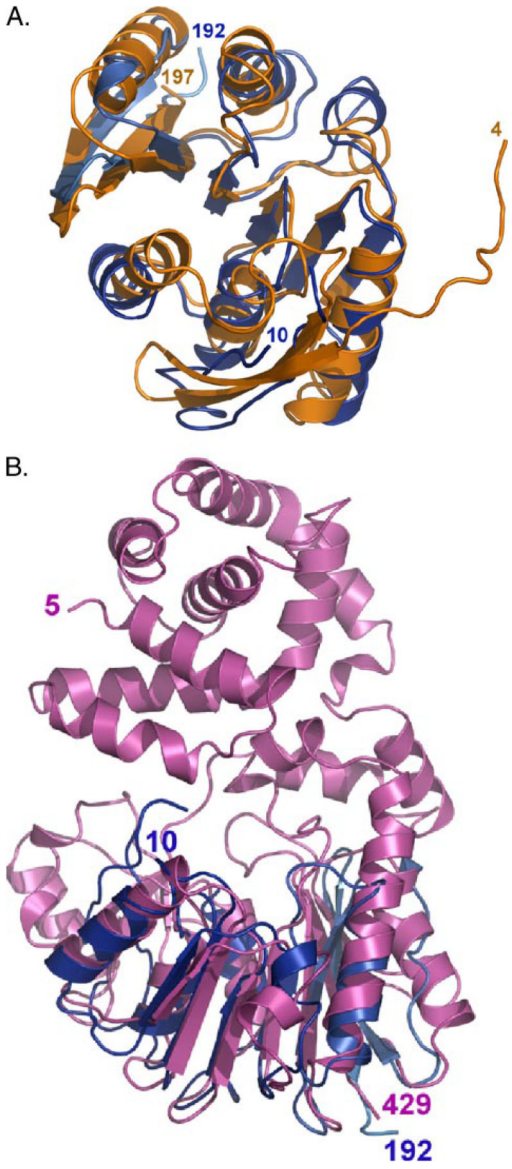

FIGURE 5.
Structure alignment of RsmD protein (shown in "*blue*") to related rRNA methyltransferases RsmC (shown in *orange*) (*A*) and RsmB (shown in *magenta*) (*B*).

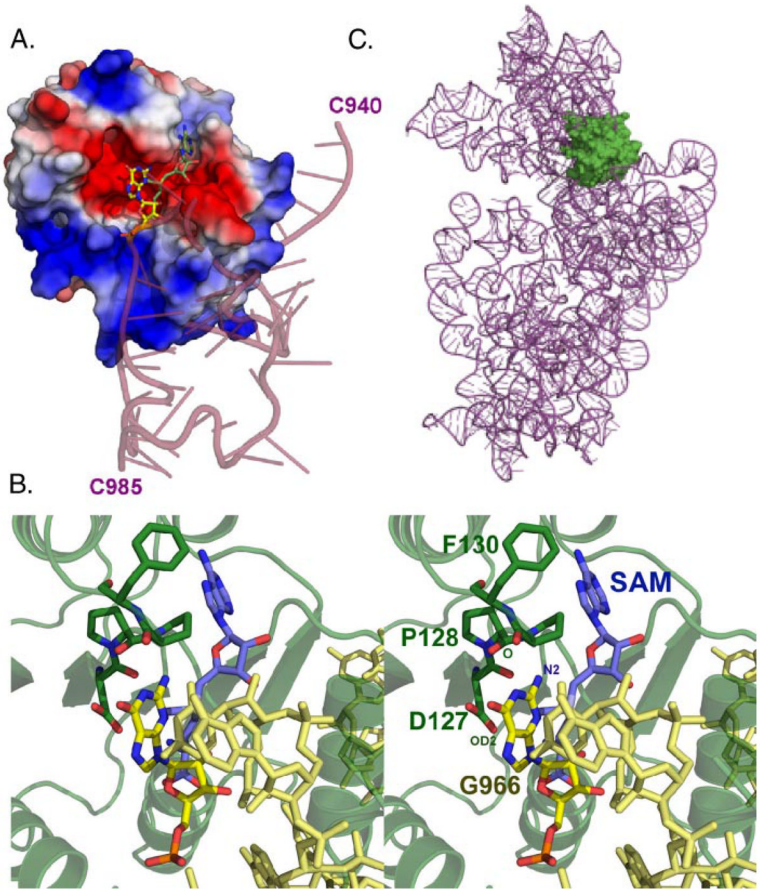

FIGURE 6. Docking of the substrates to RsmD protein

A, overview of docked AdoMet and helix 31 of the 16 S rRNA inside the substrate-binding site of RsmD protein. *B*, stereo view of AdoMet and Gr966 RNA residue positions inside the active site of the methyltransferase. The atoms involved in reaction are marked, N^2 atom of G966 rRNA residue, side chain oxygen atom of Asp-127, and main chain carboxyl oxygen of Pro-128. *C*, overview of RsmD protein docked in the tRNA-binding cleft of the *E. coli* 30 S ribosomal subunit (45) (PDB code 2AW7).

TABLE 1

Data collection statistics

X-ray wavelength	0.97956
Space group	C2
Unit cell dimensions	a = 70.504 Å, b = 96.912 Å, c = 193.961Å; $\alpha = \gamma = 90°$, $\beta = 92.21°$
Resolution	38.76-2.05 Å (2.11-2.05 Å)
No. of unique reflections[a]	80096 (5233)
Completeness[a]	97.5% (76.8%)
R-merge[a]	0.077 (0.623)
Mean I/σ(I)[a]	26.2 (1.89)
B value from Wilson plot	42.0 Å^2
Protein molecules per asymmetric unit	6
No. of protein residues per molecule	198

[a]The numbers in parentheses correspond to the highest resolution shell.

TABLE 2

Refinement statistics

Resolution range (Å)	38.76-2.05 Å
Reflections	79,903
σ cutoff	None
R-value (*R*-work) (%)	18.3
Free *R*-value (%)	21.7
R-value (all) (%)	18.4
Root mean square deviations from ideal geometry	
Bond length (Å)	0.017
Angle (degrees)	1.480
Chiral (Å)	0.103
No. of atoms	
Protein	8445
Chloride ion	6
Ethylene glycol	1
Water	390
Mean *B*-factor (Å^2)	
Protein atoms	33.6
Protein main chain	32.7
Protein side chain	34.6
Solvent	39.0
Ramachandran plot statistics	
Residues in most favored regions (%)	89.8
Residues in additional allowed regions (%)	10.0
Residues in generously allowed regions	0.2
Residues in disallowed region (%)	0.0