



RNA
A PUBLICATION OF THE RNA SOCIETY

Structural insights into the function of 23S rRNA methyltransferase RlmG (m²G1835) from *Escherichia coli*

Heng Zhang, Zeng-Qiang Gao, Yong Wei, et al.

RNA 2012 18: 1500-1509 originally published online July 2, 2012

Access the most recent version at doi:[10.1261/rna.033407.112](https://doi.org/10.1261/rna.033407.112)

**Supplemental
Material**

<http://rnajournal.cshlp.org/content/suppl/2012/06/12/rna.033407.112.DC1.html>

References

This article cites 26 articles, 6 of which can be accessed free at:
<http://rnajournal.cshlp.org/content/18/8/1500.full.html#ref-list-1>

**Email alerting
service**

Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or [click here](#)



Publish your data
using Ambion® products

ambion®
by life technologies™

To subscribe to RNA go to:
<http://rnajournal.cshlp.org/subscriptions>

Structural insights into the function of 23S rRNA methyltransferase RlmG (m²G1835) from *Escherichia coli*

HENG ZHANG,¹ ZENG-QIANG GAO,¹ YONG WEI,^{1,2} WEN-JIA WANG,¹ GUANG-FENG LIU,¹ ELEONORA V. SHTYKOVA,³ JIAN-HUA XU,¹ and YU-HUI DONG^{1,4}

¹Beijing Synchrotron Radiation Facility, Institute of High Energy Physics, Chinese Academy of Sciences, Beijing 100049, People's Republic of China

²School of Life Sciences, University of Science and Technology of China, Hefei 230027, People's Republic of China

³Institute of Crystallography, Russian Academy of Sciences, 117333 Moscow, Russia

ABSTRACT

RlmG is a specific AdoMet-dependent methyltransferase (MTase) responsible for N²-methylation of G1835 in 23S rRNA of *Escherichia coli*. Methylation of m²G1835 specifically enhances association of ribosomal subunits and provides a significant advantage for bacteria in osmotic and oxidative stress. Here, the crystal structure of RlmG in complex with AdoMet and its structure in solution were determined. The structure of RlmG is similar to that of the MTase RsmC, consisting of two homologous domains: the N-terminal domain (NTD) in the recognition and binding of the substrate, and the C-terminal domain (CTD) in AdoMet-binding and the catalytic process. However, there are distinct positively charged protuberances and a distribution of conserved residues contributing to the charged surface patch, especially in the NTD of RlmG for direct binding of protein-free rRNA. The RNA-binding properties of the NTD and CTD characterized by both gel electrophoresis mobility shift assays and isothermal titration calorimetry showed that NTD could bind RNA independently and RNA binding was achieved by the NTD, accomplished by a coordinating role of the CTD. The model of the RlmG-AdoMet-RNA complex suggested that RlmG may unfold its substrate RNA in the positively charged cleft between the NTD and CTD, and then G1835 disengages from its Watson-Crick pairing with C1905 and flips out to insert into the active site. Our structure and biochemical studies provide novel insights into the catalytic mechanism of G1835 methylation.

Keywords: RNA methylation; methyltransferase; AdoMet binding domain; RNA recognition domain; RNA binding

INTRODUCTION

In all living organisms, post-transcriptional modifications of ribosomal RNA (rRNA) nucleotides are a common mechanism fine-tuning the process of protein synthesis (Grosjean 2005). The majority of these modifications are phylogenetically conserved and located in and around the conserved decoding and peptidyltransferase active sites of the ribosome. These modifications play an important role for modifications in ribosome activity, such as the fine-tuning of local rRNA structure, 30S subunit assembly, and antibiotic resistance (Ofengand and Fournier 1998; Decatur and Fournier 2002; Douthwaite et al. 2005; Chow et al. 2007; Connolly et al. 2008; Xu et al. 2008).

In recent years, several complete modification maps of bacteria, such as *Escherichia coli* and *Thermus thermophilus* have been determined, and the modifying enzymes are quite conserved (Ofengand and Del Campo 2004; Guymon et al. 2006; Purta et al. 2009). This suggests that there are common recognition mechanisms and common functional requirements conserved since divergence from their last common ancestor (Demirci et al. 2010). Though many structures of the rRNA methyltransferases (MTases) have been solved, little is known about the process in which they bind the methyl group donor AdoMet and specifically recognize certain nucleotides and then transfer a methyl group to the nucleotide. Only the structures of 5-methyluridine MTase RumA (m⁵U1939) and pseudouridine synthase RluA (ψ746), in complex with their respective rRNA substrates, clearly demonstrate the specific process (Lee et al. 2005; Hoang et al. 2006).

N²-methylation of guanosine residues is the second most abundant rRNA modification pattern after pseudouridylation in *E. coli*, and four of five m²G nucleotides have been

⁴Corresponding author

E-mail dongyh@ihep.ac.cn

Article published online ahead of print. Article and publication date are at <http://www.rnajournal.org/cgi/doi/10.1261/rna.033407.112>.

determined in the 16S rRNA (modified by RsmC and RsmD) and 23S rRNA (modified by RlmG and RlmL) (Sergiev et al. 2007). RlmG is a specific AdoMet-dependent MTase responsible for N^2 -methylation of G1835 in 23S rRNA (Sergiev et al. 2006). The methylated G1835 is located in the 50S side of intersubunit bridge B2, which is a functionally extremely important region of the ribosome. Recombinant RlmG is able to methylate in vitro naked 23S rRNA but not assembled 50S subunits purified from the *rlmG*-knockout ($\Delta rlmG$) strain. Growth competition assays reveal that the lack of the G1835 methylation causes growth retardation, especially at temperatures higher than optimal, and in poor media (Sergiev et al. 2006). Recently, methylation of G1835 was found to specifically enhance association of ribosomal subunits and provide a significant advantage for bacteria in osmotic and oxidative stress (Osterman et al. 2011).

The structures of RsmC in both *E. coli* (RsmC-Ec, the apo-form) and *T. thermophilus* (RsmC-Tt, in complex with cofactor and putative substrate guanosine) have been solved (Sunita et al. 2007; Demirci et al. 2008). Their structures are very similar and reveal two homologous domains with nearly identical folds tandemly duplicated within a single polypeptide, and one of the domains is important for the folding of the other. The structure of RlmG has been predicted to be closely related to RsmC and also composed of two domains (Bujnicki and Rychlewski 2002; Sunita et al. 2007). However, the real substrate of RsmC in vivo is the 30S subunit assembly intermediate, while RlmG can only recognize protein-free 23S rRNA (Tscherne et al. 1999; Sergiev et al. 2006). Therefore, their different substrate specificity combined with the different location of target bases in the ribosome may cause distinct structure and functional characterization especially in their substrate-recognizing domains.

To gain insights into the substrate RNA recognition mechanism, we determined the crystal structures of the RlmG monomer in a binary complex with AdoMet and characterized its RNA-binding properties, thereby providing insights into its catalytic mechanism. Although the overall structure is similar to RsmC, there are distinct positively charged protuberances and a distribution of conserved residues contributing to the charged surface patch especially in the N-terminal domain (NTD) of RlmG for direct binding of protein-free rRNA. The RNA-binding experiments showed the recognition and binding of the target G1835 and its surrounding nucleotides was achieved by the NTD, accomplished by a coordinating role of the C-terminal domain (CTD). The model of the RlmG-AdoMet-RNA complex for substrate recognition has been proposed.

RESULTS

Overall structure of RlmG-AdoMet complex

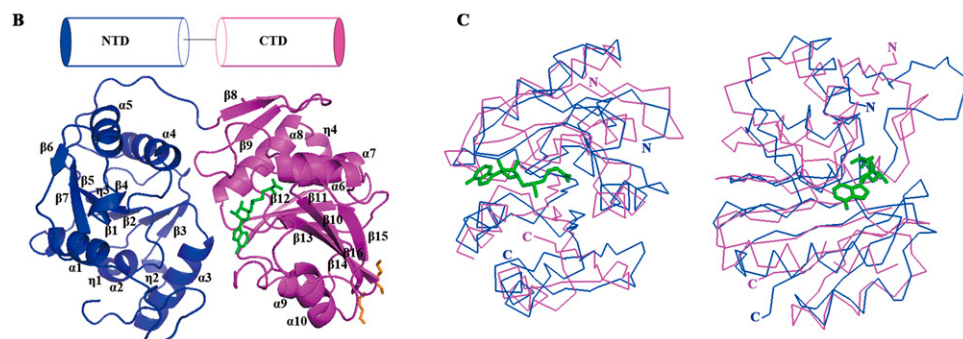
The crystal structure of RlmG in complex with AdoMet was solved by the single-wavelength anomalous dispersion (SAD)

method from synchrotron data using Se-Met-labeled protein and was refined to a final R/R_{free} factor of 0.20/0.25 at 2.3 Å. The asymmetric unit contains one RlmG molecule with overall dimensions of $\sim 67 \times 42 \times 72$ Å, comprising 365 residues from Arg9 to Lys373 and a total of 130 water molecules. The residues from 309 to 315 located in the RlmG $\beta 13$ - $\alpha 9$ loop near the active site (Fig. 1A), which may be flexible in the absence of substrate, were not observed in the electron density map and not included in the current model. RlmG is composed of two structurally related domains of a mixed α/β fold, with amino acids 8–180 and 191–373 forming the N-terminal domain and C-terminal domain, respectively, connected by a short linker region (Fig. 1B), probably the result of a gene duplication event followed by domain subfunctionalization. The NTD beginning with a long loop (residues 8–29) consists of eight helices (five α - and three η -) and seven β -sheets, and the CTD domain consists of six helices (five α - and one η -) and nine β -sheets. Both domains contain an additional β -hairpin forming a highly conserved hydrophobic interface with the core domain similar to other MTases.

The NTD is a putative RNA-binding domain where the protein-free RNA substrate was recognized and bound. A similar domain has been predicted as the substrate-binding region in RsmC-Tt and RsmC-Ec (Sunita et al. 2007; Demirci et al. 2008). The CTD contains all of the common SpoU sequence and structural motifs that define the global structure of these enzymes. There is one AdoMet molecule bound to the CTD domain, as well as two PEG molecules from the precipitant used for crystallization. Although there is only 15% amino acid identity between the NTD and CTD, the secondary structure elements generally align very well with a least squares root mean square deviation (RMSD) of 2.4 Å for 142 C α atoms, and both of them form a canonical Rossmann-like MTase fold. Meanwhile, some distinct variations are observed, especially in the loop regions flanking the cofactor- and substrate-binding sites of the catalytic domain (Fig. 1C). According to DALI structure searches (http://ekhidna.biocenter.helsinki.fi/dali_server), the closest homolog of the NTD is the putative MTase MT0146 from *Methanobacterium thermoautotrophicum* (PDB entry 1l3b) with a DALI Z-score of 17.0 and with a RMSD of 2.6 Å for 157 C α atoms, but the closest homolog of the CTD is the putative MTase MJ0882 from *Methanococcus jannaschii* (PDB entry 1dus) with a DALI Z-score of 23.3 and a RMSD of 1.5 Å for 163 C α atoms.

Monomer status of RlmG in solution

The SAXS experiment was applied to study the solution structure of RlmG, and the results are shown in Supplemental Fig. S1. The fit of the theoretical curve of the crystal structure of RlmG to the experimental data is very good with a discrepancy value of 1.712 (Supplemental Fig. S1, curve2), indicating that the crystal structure is consistent with the SAXS data. A low resolution model is also built by



1502 RNA, Vol. 18, No. 8

the SAXS data (Supplemental Fig. S1, left, below inset a), and the dummy atom model practically coincides with the crystal structure (Supplemental Fig. S1, left, below inset b). These two methods of SAXS were used independently, and both are consistent with RlmG being a monomer in solution. The monomer status determined by SAXS is also consistent with observations from the gel filtration column (data not shown), as well as the analysis of intermolecular contacts in the crystal. We can conclude that the active biological unit of RlmG is very possibly a monomer *in vivo*.

AdoMet binding in the active site of the CTD domain

Similar to class I MTases RsmC-Ec and RsmC-Tt, composed of two homologous domains (Sunita et al. 2007; Demirci et al. 2008), the AdoMet molecule is bound tightly in a canonical conformation in the pocket mainly consisting of the loops $\beta 10$ – $\alpha 7$, $\beta 12$ – $\beta 13$, $\beta 13$ – $\alpha 9$ and $\beta 11$ of the conserved CTD domain, forming a negative-deep groove (Figs. 1A, 2A). The main part of the methionine side chain of AdoMet overhangs into a large cavity, which is most likely a binding site for the substrate, guanine 1835 (Fig. 2A). The cavity is large enough to accommodate a guanine ring and position its amino group next to the active site residues. The directly interacting amino acids surrounding the AdoMet binding region (within 3.8 Å) include the side chains of G236, D259, N288, N289, N305, and P306 (Fig. 2A,B), and most of them are conserved residues in RlmG orthologs (Fig. 1A). The side chains of F208, D234, N239, and I242 contact with AdoMet indirectly via well-ordered water molecules (Fig. 2A).

N305 in the identical conserved proline-rich motif NPPF (N305–F308) located in the loop between $\beta 13$ and $\alpha 9$ among m^2G MTases, interacts strongly with the guanine N2 atom of

the reactive methyl group (3.0 Å), carboxyl group (3.0 Å), and hydroxyl group (3.5 Å) of methionine of AdoMet. This motif is known to be involved in AdoMet binding, forming MTase active sites (Sergiev et al. 2007), and is involved in catalysis. G236, in the conserved glycine-rich DXGXGXG MTase signature motif (D234–G240) located in the loop $\beta 10$ – $\alpha 7$, directly interacts with the amino group of methionine of AdoMet, which also interacts with D234 and I242 via solvent water molecules. F208 in the loop $\beta 9$ – $\alpha 6$ is conserved among RlmG orthologs and approaches the AdoMet sulfur atom, providing a hydrophobic aromatic stacking, which may be important to stabilize the substrate of AdoMet. Both ribose hydroxyl groups form hydrogen bonds (2.8 Å) to D259 in the consensus sequence 259D*S261. The N6 and N1 atoms of the adenine group are coordinated by the hydrogen bonds to N289 (3.2 Å) and N288 (3.6 Å), respectively.

In vitro expression of the NTD and CTD and their interactions with RNA

To characterize and confirm the predicted role in substrate binding based on our RlmG structure, we designed and constructed two deletion mutants corresponding to the isolated NTD and CTD (amino acids 1–180 and 191–373, respectively). Initially, the CTD expressed in the pET21a vector could not be well-purified (some molecular chaperones were unable to be removed), and it degraded easily. Therefore, we attempted to express it in pET28a-SUMO with cleavable ubiquitin-like-specific protease 1 to improve its proper folding of fusion partners and the solubility. The result showed the SUMO-CTD was purified well but unstable after the SUMO tag was cleaved. Our results suggest that the apo-form CTD has lost the ability to fold on its own and requires a prefolded “intramolecular

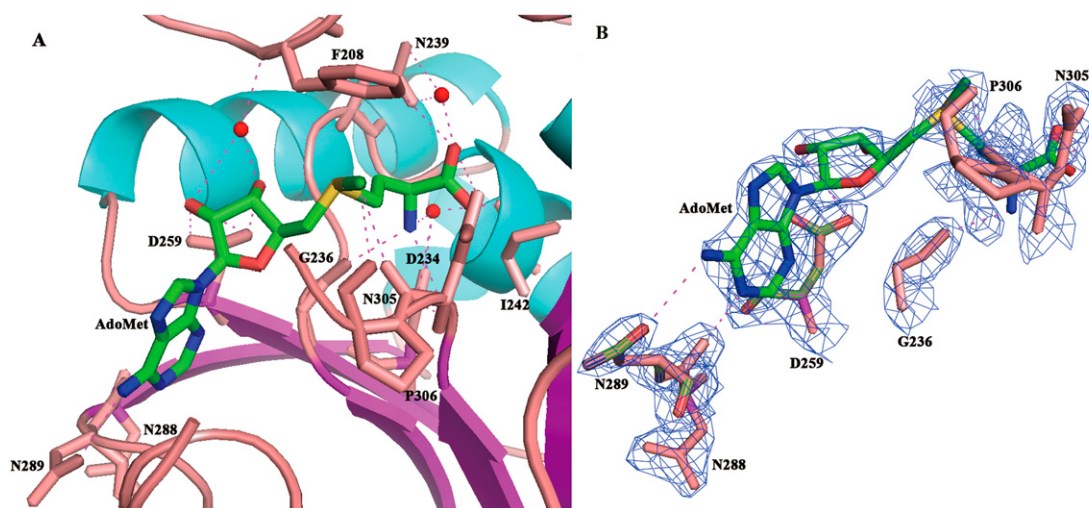


FIGURE 2. AdoMet binding in the active site. AdoMet is shown in green sticks, α -helix and β -sheet are shown in blue and magenta, respectively. Solvent water molecules are shown as red spheres. (A) Interaction between AdoMet and the residues in binding pocket. (B) Stereo view of the electron density of AdoMet and its interacting side chains of the residues from a 2Fo-Fc map contoured at 2.0 σ .

chaperone” localized at its N terminus like the NTD or another well-folded domain such as SUMO, similar to the status of the CTD in RsmC-Ec (Sunita et al. 2007), while the NTD was expressed and purified well in pET21a.

Electrophoresis mobility shift assays (EMSA) were used to examine the RlmG-rRNA interaction using the 21-nt hairpin structure containing the G1835 target site (Fig. 3A). As expected, we observed a major complex with a 1:1 protein:rRNA stoichiometry for the 21-nt rRNA fragment in full-length and the NTD, and the functional complex is formed by RlmG monomer and equal amount of target rRNA (Fig. 3B). The SUMO-CTD could not bind RNA under this condition. Meanwhile, no obvious shift difference of the enzyme-RNA complexes with increasing protein concentration (from 1:1 to 10:1) was observed. These results proved that RlmG could efficiently bind substrate RNA in vitro, as was reported previously (Sergiev et al. 2006), and more importantly, NTD could bind substrate RNA independently.

The binding affinities (K_a) of RlmG-rRNA were further quantitatively analyzed by isothermal titration calorimetry (ITC), and these experiments indicate a single binding site for all the samples. Both native RlmG ($K_a = 1.68 \times 10^5 \text{ M}^{-1}$) and NTD ($K_a = 4.40 \times 10^4 \text{ M}^{-1}$) could efficiently bind substrate RNA, while SUMO-CTD could not (Supplemental Fig. S2), consistent with the EMSA results above. The control experiment for apo-form SUMO binding the RNA fragment also showed there was no interaction between them (Supplemental Fig. S2). Considering the lower K_a of NTD than that of the full-length, we propose that the CTD can also bind the RNA fragment only in the presence of the NTD, although it is still stable after being fused with the other chaperones such as SUMO.

We are unable to get sigmoidal bindings, as the RNA may not be very stable under this condition, and therefore, the

enthalpy (ΔH) and entropy (ΔS) cannot be reliably estimated. However, our results were obtained under the same condition and could reflect their difference on binding affinities of the substrate RNA to a great degree.

Proposed model of RlmG-rRNA interactions

To gain structural insight into how RlmG might recognize its specific target RNA, we constructed a model of the RlmG-AdoMet-rRNA complex using the 10-nt rRNA (G1831–G1840) from the structure of *E. coli* 50S subunit (PDB entry 2AW4) as the substrate (Fig. 4A). Molecular dynamics results indicated that the model was both structurally feasible and energetically stable. The minimized model demonstrated that the RNA stem-loop would easily fit into the cleft with the electropositive surface mainly generated by the NTD (Fig. 4B,C), consistent with the EMSA and ITC experiment results that NTD plays the dominant role in RNA binding. In this model, the distance between the AdoMet methyl group and the substrate N2 atom is 3.5 Å, which is essential for catalysis. The conserved motif NPPF (N305–F308) is involved in forming the MTase active site pocket. Similar to the interactions of those residues with the putative substrate GMP in RsmC-Tt (Demirci et al. 2008), P306 and P307 interact with the substrate guanosine, and F308 may provide the base stacking interaction for the substrate guanine in RlmG (Fig. 4D).

DISCUSSION

Functional characterization of the NTD and CTD of RlmG in substrate binding

Although the methylation of G1835 was observed when protein-free 23S rRNA was used as a substrate by reverse

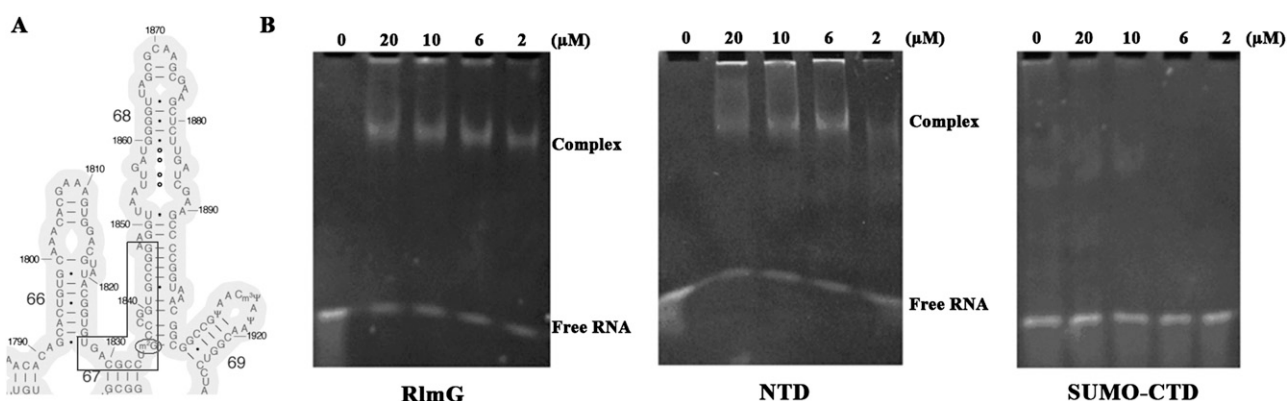


FIGURE 3. rRNA binding experiment. (A) A schematic of the secondary structure of segmental helix 66–68 of *E. coli* 23S rRNA from the Center for Molecular Biology of the RNA website (<http://www.http://rna.ucsc.edu/rnacenter>). m^2 G1835 is indicated by an ellipse and the 21-nt RNA (1827–1847) is labeled with a box as the substrate in this study. (B) Binding of the full-length, NTD, and SUMO-CTD to substrate RNA determined by EMSA. The RNA samples (2 μM final concentration) were annealed at 50°C for 10 min and then cooled at room temperature for another 10 min to form a hairpin structure. Protein(full-length, NTD, and CTD)-RNA complexes were prepared by adding protein at 0, 2, 6, 10, and 20 μM final concentration and incubated for 30 min at room temperature.

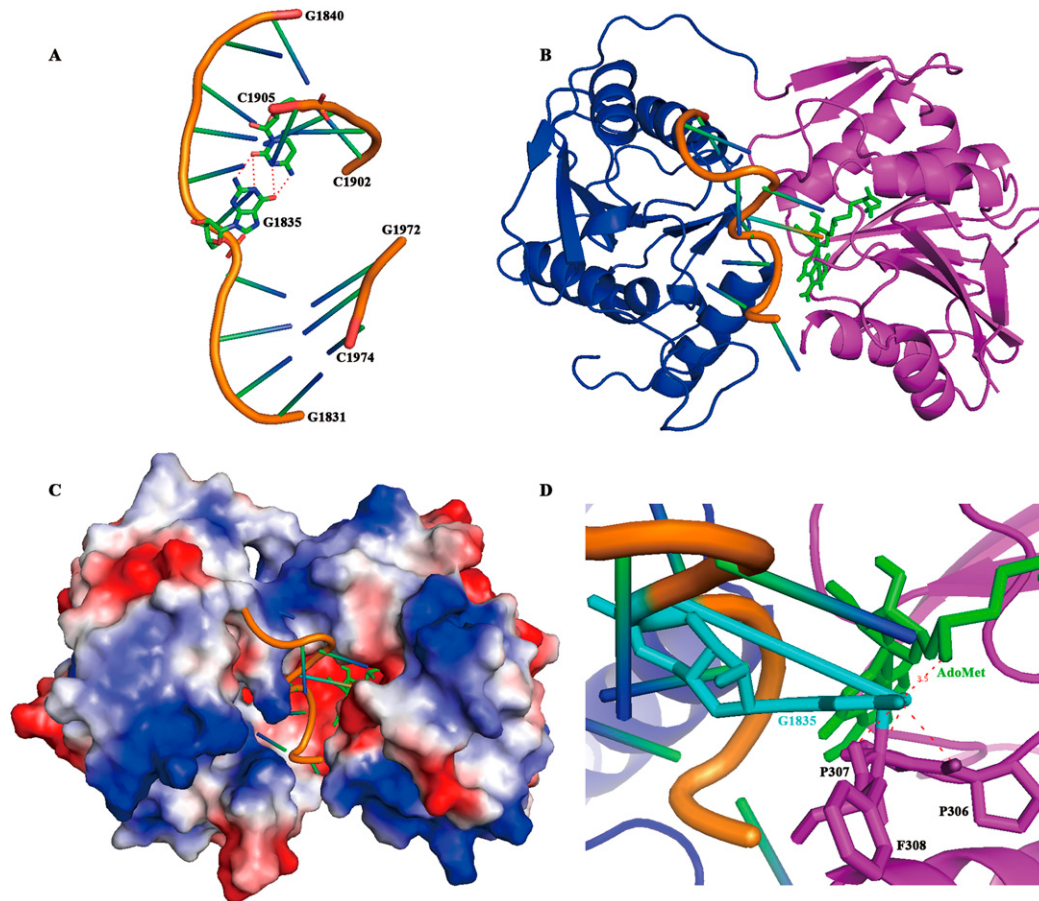


FIGURE 4. Docking model of the RlmG-AdoMet-rRNA complex. The NTD, CTD, and 10-nt RNA backbone containing G1835 are shown in blue, magenta, and yellow, respectively. (A) Tertiary structure of the substrate rRNA fragment (G1831–G1840) and the pairing fragments (C1902–C1905 and G1972–C1974) from the structure of the *E. coli* 50S subunit (PDB entry 2AW4). G1835 and its pairing C1905 are shown as green sticks. (B) The stereo diagram showing a schematic representation of the model for RlmG binding to the substrate RNA. (C) A molecular surface representation of RlmG, colored by its local electrostatic potential (blue, +7KT; red, –7KT) with the substrate RNA. The portion of the active site cleft that directly contacts the modeled RNA is predominantly electropositive. (D) The interacting residues with G1835. AdoMet and G1835 are shown in green and cyan sticks, respectively. The distance between the AdoMet methyl group and the N2 atom of G1835 is 3.5 Å.

transcription experiments (Sergiev et al. 2006), there is still no direct evidence on its binding to the substrate RNA fragment. In this study, from EMSA results, both RlmG and NTD could bind to the RNA fragment containing G1835, providing direct evidence that RlmG can efficiently bind rRNA and that NTD is necessary in recognition and binding of the target and its surrounding nucleotides as an RNA-binding domain during the methylation of G1835. Moreover, the binding affinity of NTD is obviously weaker than that of full-length, indicating that the CTD may play a coordinated role to help the NTD bind substrate RNA in vivo, even it may not be directly involved in RNA recognition. Besides, the instability of the CTD and the lower binding affinity of the NTD also suggested a physical linkage, such as the possibility that the loop connecting them is also required for their cooperation. To our surprise, the DNA fragment (corresponding to the 21-nt RNA, except that U is replaced with T) could also be bound efficiently by RlmG (data not shown), a property that may be utilized under specific con-

ditions such as higher temperatures and poor media during cellular growth. This property has only been observed in the RsmD-like MTase Rv2966c from *M. tuberculosis* recently (Kumar et al. 2011).

Comparisons of the NTD and CTD between RlmG and RsmCs with different substrate specificity

The tertiary structures alignment between RlmG and RsmCs from *E. coli* and *T. thermophilus* showed that their overall architectures look very similar (Fig. 5A), although they share only 22% amino acid identity. The DALI structure searches also showed that the structure most closely related to RlmG is RsmC-Tt with a Z-score of 33.6 and a RMSD of 2.5 Å for 331 C α atoms and RsmC-Ec with a Z-score of 31.4 and a RMSD of 2.8 Å for 319 C α atoms, respectively, in addition to MJ0882 (PDB entry 1dus) with a Z-score of 25.8 and a RMSD of 2.0 Å for 184 C α atoms, which has been predicted as the RsmC homolog (Sunita et al. 2007).

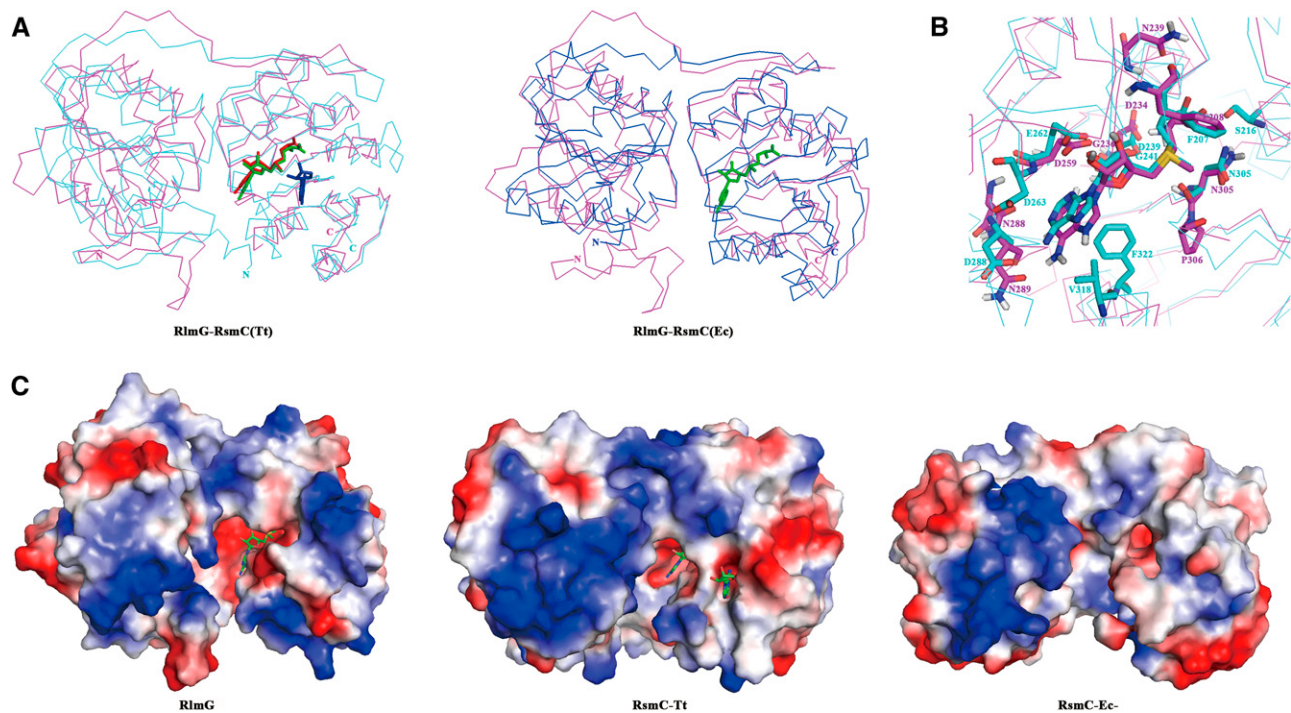


FIGURE 5. Structure comparisons between RlmG and RsmCs. (A) Tertiary structures alignment of RlmG-AdoMet shown in magenta (AdoMet was shown in green sticks), RsmC-AdoMet-guanosine from *T. thermophilus* shown in cyan (AdoMet and guanosine were shown in magenta and blue sticks, respectively, PDB entry 3DMH) and RsmC from *E. coli* shown in blue (PDB entry 2PJD). Their backbones are shown as C α trace. (B) Comparison of RlmG with RsmC-Tt in AdoMet binding region. (C) The molecular surface representation of RlmG (blue, +7KT; red, -7KT), RsmC-Tt (blue, +6KT; red, -6KT), and RsmC-Ec (blue, +7KT; red, -7KT), colored by their local electrostatic potential.

However, there are still many distinct differences, especially in the NTD, between RlmG and RsmCs. There are the positively charged protuberances both in the NTD and CTD seen from electrostatic potential mapped onto the RlmG surface, while the two RsmCs are almost uniformly negatively charged except that there is only a small positive patch on the conserved the NTD protuberance (Fig. 5C). The positively charged protuberance shows that differential conservation between RlmG and RsmCs (Fig. 1A; Supplemental Fig. S3) and in other MTase families of different specificity is likely closely related to the recognition and binding of their different rRNA substrates. Moreover, the catalytic pocket composed of negative charges located in the CTD of RlmG is larger than those of RsmCs, which may be used to accommodate a guanine ring and position its amino group next to the active site residues in the cavity. Conservation of the pocket suggests that the CTD of RsmC and RlmG is important for binding of the SAM cofactor and the catalysis of the methyl transfer reaction. In this context, it is interesting to note that the NTD of the RlmG structure places a positively charged surface area close to the active site and suggests a functional contribution of the NTD domain for substrate rRNA binding directly (Fig. 5C).

Moreover, the residues contributing to the charged surface patch are conserved among RlmG homologs, and this observation is consistent with the functional requirements (Fig. 1A).

In RsmCs structures, although there is also positively charged surface area close to the active site (Fig. 5C), the residues contributing to the charged surface patch are not conserved from different organisms (Supplemental Fig. S3). This difference may be due to their different substrate patterns and substrate recognition mechanism. It is interesting to note that the 29-nt loop in the leading NTD domain is a unique secondary structure element, where there is a unique helix $\alpha 1$ in RsmC-Tt predicted to contribute to the cofactor-binding site but missing in RsmC-Ec (Sunita et al. 2007; Demirci et al. 2008). In our RlmG structure, the flexible loop is conserved among RlmG homologs with positive charges (Fig. 1A) and may contribute to the RNA-binding site and create a better-defined binding pocket.

Not surprisingly, the comparison of RlmG with RsmC-Tt on the binding pattern of AdoMet showed that they share similar orientation in the conserved AdoMet-binding pocket of the CTD, and their interacting residues are also very similar. The AdoMet is located in conserved motifs between 305NPPF308 and 234DLGCGNG240 in RlmG corresponding to 305NPPF308 and 234DLGAGYG240 in RsmC-Tt (Fig. 5B), and 268NPPF271 and 202DVGCGAG208 in RsmC-Ec (Sunita et al. 2007; Demirci et al. 2008). Both hydrophobic F208 in RlmG and the corresponding F207 in RsmC-Tt approach the AdoMet sulfur atom and might contribute to the enzymatic function by destabilizing the cofactor; however,

the location of the corresponding RsmC-Ec F176 in the flexible loop cannot be modeled and is removed from the AdoMet-binding site. Mutation of key residues in the AdoMet-binding pocket of RsmC-Ec showed that the mutants D202A and D227A in the potential SAM-binding site in the CTD showed complete inability to bind the cofactor, while the N268A mutant in the predicted catalytic motif NPPF that coordinates interactions between SAM and the target guanosine showed almost fivefold reduction in the SAM-binding affinity (Sunita et al. 2007).

Implications for substrate RNA binding to RlmG in methylation of G1835

Comparisons of the structure of the RNA fragment (G1831–G1840) from the wild-type 50S subunit and that in the model showed that a significant conformational change occurs when the substrate is bound to RlmG (Fig. 4A,B) and an induced-fit mechanism may be required to form the active complex structure. RlmG may unfold its substrate RNA, which forms base pairs with neighboring nucleotides (the fragments C1902–C1905 and G1972–C1974) in 23S rRNA, into a single-stranded conformation. Then the target G1835, which has disengaged from its Watson-Crick pairing with C1905, will flip out of the helix to insert into the active site (Fig. 4D). Similar to RlmG, the target G1207 of RsmC-Tt has to disengage from the interaction with C1051 and flip out into the active site prior to its modification (Demirci et al. 2008). Such base flipping has been extensively reported in both RNA and DNA MTases. The single-stranded substrate is “refolded” on RsmA into a compact conformation with six key intra-RNA interactions, and a second base is “flipped out” from the core loop to stack against the adenine of AdoMet in addition to the target (Lee et al. 2005). The methylcytosine base at the hemi-methylated site is flipped out of the DNA helix in the mouse UHRF1 SRA-DNA complex and fits tightly into a protein pocket on the concave surface (Arita et al. 2008). Meanwhile, the 10-nt RNA fragment used in this model may be not long enough to represent the complete interaction between the substrate RNA and RlmG, and more interacting residues should be confirmed by mutagenesis.

An additional RNA binding domain in RlmG has been predicted to be necessary to reach substrate specificity (Sergiev et al. 2006). In RsmC-Tt structure, consistent with previous suggestions that RsmC can only recognize a subunit assembly intermediate (Tscherne et al. 1999), access to G1207 was occluded in the native structure by helix 18 and would require either rotation of the 30S head around the neck region or a substantial backward motion toward the solvent side of the subunit (Demirci et al. 2008). On the contrary, RlmG is supposed to directly accommodate the protein-free 23S rRNA substrate fragment to its positively charged cleft (Sergiev et al. 2006), which is confirmed by our RlmG structure and consistent with the location of the RNA fragment

in the model (Fig. 4C). Therefore, the different substrate specificity may lead to distinct structure and functional characterization especially in their substrate-recognizing domains.

In this study, we carried out structure and functional characterization of RlmG and provided evidence for its direct recognition and binding of protein-free rRNA. Based on these results, we propose the hairpin rRNA fragment containing G1835 is initially recognized by the NTD, and then the specific target will be bound to the active site in the CTD to trigger catalysis. Consequently, the catalysis process is carried out by coordinating roles of the NTD and CTD in vivo. The structure of RlmG provides the atomic-level picture of such RNA-modification enzymes of differential functional specialization in two domains apparently derived from a common ancestor.

MATERIALS AND METHODS

Protein purification and crystallization

The *rlmG* gene was PCR-amplified from *E. coli* K-12 genomic DNA and cloned into the expression vector pET21a with a non-cleavable C-terminal His₆ tag (Novagen) and expressed in BL21 (DE3) cells. The NTD variant was constructed by recloning the single domain into pET21a by removing the CTD in the PCR reaction. The CTD variant was constructed by recloning the CTD into the modified pET28a-SUMO (Novagen) in which the cleavable ubiquitin-like-specific protease 1 (Ulp1) fused with an N-terminal His₆ tag was introduced upstream of the BamHI site. These proteins were purified as previously described (Li et al. 2011). Prior to crystallization, RlmG was concentrated to 16 mg/mL.

The initial crystallization condition of RlmG-His was obtained under the crystallization conditions of 36# in Crystal Screen (Hampton Research) with the sitting drop vapor diffusion method at room temperature after 6 d. The crystals with typical size did not have good diffraction quality. In order to further improve the diffraction quality and resolution, several truncations of RlmG and the mixture of these truncations with cofactor AdoMet were applied for crystallization optimization. The mixture of RlmG and its truncations with AdoMet (Sigma) at a molar ratio of 1:10 were prepared and incubated on ice for 6 h before performing crystallization experiments. The best crystal was obtained from the mixture of RlmG (R9-K373) with AdoMet in solution 0.2 M Tris (pH 7.5), and 5% (w/v) PEG 8000 with the addition of 1% (w/v) protamine sulfate and 0.02 M HEPES sodium (pH 6.8) from Silver Bullets (Hampton Research) after 3–4 d.

Data collection, structure determination, and refinement

Single wavelength anomalous data from seleno-RlmG crystals were collected on the beamline station 3W1A of the Beijing Synchrotron Radiation Facility (BSRF). Before data collection, crystals were soaked for 5 sec in a cryoprotectant consisting of 20% (v/v) glycerol in the crystal mother liquid and then flash-frozen in liquid nitrogen. The temperature was held at 100 K by liquid nitrogen during data collection.

All data were processed using the program package HKL2000 (Otwinowski and Minor 1997), and collection statistics were summarized in Table 1. Selenium atoms in five methionines were identified using the program SOLVE (Terwilliger and Berendzen 1999). The initial phase was improved by maximum likelihood solvent-flattening density modification using the program RESOLVE (Terwilliger 2003), producing experimental electron density maps of excellent quality. Automated model building was performed with the program ARP/wARP (Perrakis et al. 1999). The structure was refined with the program COOT (Emsley and Cowtan 2004). The qualities of the final models were checked with the program PROCHECK (Laskowski et al. 1993). Refinement statistics and model parameters are given in Table 1. The program PyMOL (<http://www.pymol.sourceforge.net/>) was used to prepare structural figures.

Gel electrophoresis mobility shift assays

The 23S rRNA fragment (1827–1847) UGACGCCUGCCCCGU GCCGGA was chemically synthesized by Beijing AuGCT Biotechnology Co., Ltd. RNA samples (2 μ M final concentration) were annealed at 50°C for 10 min in the buffer (prepared by DEPC-sterilized water) containing 50 mM NaCl, 50 mM Tris (pH 8.0), 2 mM MgCl₂ and 1mM DTT, and then cooled at room temperature for another 10 min to form hairpin structure. RlmG-RNA complexes were prepared by adding RlmG (full-length, NTD, and CTD) at 0, 2, 6, 10, and 20 μ M final concentration and incubated for 30 min at room temperature. For each sample, free RNA and complexes were separated on a 10% acrylamide native gel run for 1 h at 200 V and 5 W at 4°C and visualized by Golden View (Biotium) staining.

Isothermal titration calorimetry

ITC was applied to quantitatively determine the binding affinity of full-length, NTD, and CTD of RlmG to the RNA fragment above. For the titration experiments, the protein was purified with the same method as above and was dialyzed against the buffer containing 50 mM Na-HEPES (pH 7.5), 0.15 M NaCl, 5% (v/v) glycerol, 2 mM MgCl₂, and 2 mM β -mercaptoethanol for 24 h. RNA was dissolved in the same buffer as above. The ITC experiments were carried out using a high-sensitivity iTC-200 microcalorimeter from Microcal (GE Healthcare) at 20°C using 100 μ M of RNA in the injector and 5–10 μ M of the protein in the sample cell. All samples were thoroughly degassed and then centrifuged to remove precipitates. Injection volumes of 2 μ L per injection were used for the different experiments, and for every experiment, the heat of dilution for each ligand was measured and subtracted from the calorimetric titration experimental runs for the protein. Consecutive injections were separated by 2 min to allow the peak to return to the baseline. Integrated heat data obtained for the ITCs were fitted in a single site model using a nonlinear least-squares minimization algorithm to a theoretical titration curve, using the MicroCal-Origin 7.0 software package.

Molecular modeling of RlmG-AdoMet-rRNA complex

The coordinates for the 21-nt hairpin rRNA were taken from the structure of the wild-type *E. coli* 50S subunit (PDB entry 2AW4), and the structure of RlmG-AdoMet was used as the receptor molecule and calculated with AutoDock 3.0.5 (Morris et al. 1998). Considering that some conformation changes of RlmG may take

place after being bound to substrate RNA, the receptor was treated as a flexible molecule (e.g., side chains of Phe208 and Asp214 are flexible regions), while the ligand (the RNA backbone) was kept rigidly oriented toward the cleft formed by the NTD and CTD of RlmG. In order to refine the complex model to satisfy a proposed catalytic mechanism, the N2 atom of G1835 was restrained to 3.6 Å of the methyl group on AdoMet. The grid size is 50 $\times$ 50 $\times$ 50 Å, and the grid step is 3 Å. Subsequently, both shape-only and shape-electrostatics correlation algorithms were used with a search radius of $n = 30$, and the top 10 docking solutions were inspected visually in Coot (Emsley and Cowtan 2004). Solutions from each round of docking were subsequently ranked, according to the proximity between the residues implicated in RNA binding to the G1835 nucleoside and affinity scores that describe clashes of the ligand with the receptor molecule, and best-scoring poses were regarded as the most likely models. However, no reasonable solutions were obtained when the 21-nt RNA was constrained in the vicinity of the activated methyl group of AdoMet. Then, the truncations of the RNA were tried and the best packed model (estimated free energy of binding = $+9.20 \times 10^4$ kcal/mol) was obtained by docking 10-nt

TABLE 1. Data collection and refinement statistics

Data collection	Peak
Wavelength (Å)	0.9794
Space group	P2 ₁
Unit-cell parameters	a = 67.19 Å, b = 42.13 Å, c = 72.43 Å, β = 104.5°
Resolution (Å)	2.30 (2.34–2.30) ^a
Number of unique reflections	17,821 (914)
Completeness (%)	99.9 (100)
Redundancy	7.2 (7.0)
Mean I/ σ (I)	39.71 (5.62)
Molecules in asymmetric unit	1
R _{merge} (%)	6.0 (53.0)
Structure refinement	
Resolution range (Å)	29.98–2.30
R _{work} /R _{free} (%)	20.3/25.2
Average B factor (Å ²)	
Main chain	27.09
Side chain	32.66
Ligand (SAM/PEG)	28.18/46.82
Waters	30.14
Ramachandran plot (%)	
Most favored	97.5
Allowed	2.2
Disallowed	0.3 (P17)
RMSDs	
Bond lengths (Å)	0.006
Bond angles (°)	0.978

^aThe values in parentheses are those for the highest resolution shell.

RNA (1831–1840) containing G1835, without any large conformational changes of the protein.

DATA DEPOSITION

The atomic coordinates and structure factors have been deposited in the Protein Data Bank (accession code 4DCM).

SUPPLEMENTAL MATERIAL

Supplemental material is available for this article.

ACKNOWLEDGMENTS

We thank the staff of the beamline station 3W1A of the Beijing Synchrotron Radiation Facility (BSRF) and the X33 beamline of the European Molecular Biology Laboratory (EMBL) for providing technical support and for many fruitful discussions. This work was supported by the grants from the National Natural Science Foundation of China (10979005) and the National Basic Research Program of China (2009CB918600).

Received March 21, 2012; accepted May 22, 2012.

REFERENCES

- Arita K, Ariyoshi M, Tochio H, Nakamura Y, Shirakawa M. 2008. Recognition of hemi-methylated DNA by the SRA protein UHRF1 by a base-flipping mechanism. *Nature* **455**: 818–821.
- Bujnicki JM, Rychlewski L. 2002. RNA:(guanine-N2) methyltransferases RsmC/RsmD and their homologs revisited—bioinformatic analysis and prediction of the active site based on the uncharacterized Mj0882 protein structure. *BMC Bioinformatics* **3**: 10. doi: 10.1186/1471-2105-3-10.
- Chow CS, Lamichhane TN, Mahto SK. 2007. Expanding the nucleotide repertoire of the ribosome with post-transcriptional modifications. *ACS Chem Biol* **2**: 610–619.
- Connolly K, Rife JP, Culver G. 2008. Mechanistic insight into the ribosome biogenesis functions of the ancient protein KsgA. *Mol Microbiol* **70**: 1062–1075.
- Decatur WA, Fournier MJ. 2002. rRNA modifications and ribosome function. *Trends Biochem Sci* **27**: 344–351.
- Demirci H, Gregory ST, Dahlberg AE, Jøgl G. 2008. Crystal structure of the *Thermus thermophilus* 16S rRNA methyltransferase RsmC in complex with cofactor and substrate guanosine. *J Biol Chem* **283**: 26548–26556.
- Demirci H, Larsen LH, Hansen T, Rasmussen A, Cadambi A, Gregory ST, Kirpekar F, Jøgl G. 2010. Multi-site-specific 16S rRNA methyltransferase RsmF from *Thermus thermophilus*. *RNA* **16**: 1584–1596.
- Douthwaite S, Fourmy D, Yoshizawa S. 2005. *Fine-tuning of RNA functions by modification and editing*. Springer, New York.
- Emsley P, Cowtan K. 2004. *Coot: Model-building tools for molecular graphics*. *Acta Crystallogr D Biol Crystallogr* **60**: 2126–2132.
- Grosjean H. 2005. Fine-tuning of RNA functions by modification and editing. In *Topics in current genetics* (ed. S Hohmann), Vol. 12, pp. 1–16. Springer, New York.
- Guymon R, Pomerantz SC, Crain PF, McCloskey JA. 2006. Influence of phylogeny on posttranscriptional modification of rRNA in thermophilic prokaryotes: The complete modification map of 16S rRNA of *Thermus thermophilus*. *Biochemistry* **45**: 4888–4899.
- Hoang C, Chen J, Vizthum CA, Kandel JM, Hamilton CS, Mueller EG, Ferré-D'Amaré AR. 2006. Crystal structure of pseudouridine synthase RluA: Indirect sequence readout through protein-induced RNA structure. *Mol Cell* **24**: 535–545.
- Kumar A, Saigal K, Malhotra K, Sinha KM, Taneja B. 2011. Structural and functional characterization of Rv2966c reveals an RsmD-like methyltransferase from *Mycobacterium tuberculosis* and the role of its N-terminal domain in target recognition. *J Biol Chem* **286**: 19652–19661.
- Laskowski RA, MacArthur MW, Moss DS, Thornton JM. 1993. PROCHECK: A program to check the stereochemical quality of protein structures. *J Appl Crystallogr* **26**: 283–291.
- Lee T, Agarwalla S, Stroud R. 2005. A unique RNA fold in the RsmA-RNA-cofactor ternary complex contributes to substrate selectivity and enzymatic function. *Cell* **120**: 599–611.
- Li LQ, Liu Y, Liu P, Dong YH. 2011. Solution conformation of the response regulator proteins from *Deinococcus radiodurans* studied by SAXS. *Chin Phys C* **35**: 974–977.
- Morris GM, Goodsell DS, Halliday RS, Huey R, Hart WE, Belew RK, Olson AJ. 1998. Automated docking using a Lamarckian genetic algorithm and empirical binding free energy function. *J Comput Chem* **19**: 1639–1662.
- Ofengand J, Del Campo M. 2004. Modified nucleosides in *Escherichia coli* ribosomal RNA. In *EcoSal—Escherichia coli and Salmonella: Cellular and molecular biology* (ed. R Curtiss), Module no. 4.6.1. ASM Press, Washington, DC. <http://www.ecosal.org>.
- Ofengand J, Fournier MJ. 1998. The pseudouridine residues of rRNAs: Number, location, biosynthesis and function. In *Modification and editing of RNA: The alteration of RNA structure and function* (ed. H Grosjean, R Benne), pp. 229–253. ASM Press, Washington, DC.
- Osterman IA, Sergiev PV, Tsvetkov PO, Makarov AA, Bogdanov AA, Dontsova OA. 2011. Methylated 23S rRNA nucleotide m²G1835 of *Escherichia coli* ribosome facilitates subunit association. *Biochimie* **93**: 725–729.
- Otwinowski Z, Minor W. 1997. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol* **276**: 307–326.
- Perrakis A, Morris R, Lamzin V. 1999. Automated protein model building combined with iterative structure refinement. *Nat Struct Biol* **6**: 458–463.
- Purta E, O'Connor M, Bujnicki JM, Douthwaite S. 2009. YgdE is the 2'-O-ribose methyltransferase RlmM specific for nucleotide C2498 in bacterial 23S rRNA. *Mol Microbiol* **72**: 1147–1158.
- Sergiev PV, Lesnyak DV, Bogdanov AA, Dontsova OA. 2006. Identification of *Escherichia coli* m²G methyltransferases: II. The *yjgO* gene encodes a methyltransferase specific for G1835 of the 23S rRNA. *J Mol Biol* **364**: 26–31.
- Sergiev PV, Bogdanov AA, Dontsova OA. 2007. Ribosomal RNA guanine-(N2)-methyltransferases and their targets. *Nucleic Acids Res* **35**: 2295–2301.
- Sunita S, Purta E, Durawa M, Tkaczuk KL, Swaathi J, Bujnicki JM, Sivaraman J. 2007. Functional specialization of domains tandemly duplicated within 16S rRNA methyltransferase RsmC. *Nucleic Acids Res* **35**: 4264–4274.
- Terwilliger TC. 2003. Automated main-chain model building by template matching and iterative fragment extension. *Acta Crystallogr D Biol Crystallogr* **59**: 38–44.
- Terwilliger TC, Berendzen J. 1999. Automated MAD and MIR structure solution. *Acta Crystallogr D Biol Crystallogr* **55**: 849–861.
- Tscherne JS, Nurse K, Popienick P, Ofengand J. 1999. Purification, cloning, and characterization of the 16S rRNA m²G1207 methyltransferase from *Escherichia coli*. *J Biol Chem* **274**: 924–929.
- Xu Z, O'Farrell HC, Rife JP, Culver GM. 2008. A conserved rRNA methyltransferase regulates ribosome biogenesis. *Nat Struct Mol Biol* **15**: 534–536.