Acta Crystallographica Section D
Biological Crystallography
ISSN 1399-0047

The structure of Rv2372c identifies an RsmE-like methyltransferase from *Mycobacterium tuberculosis*

Atul Kumar, Santosh Kumar and Bhupesh Taneja*

Structural Biology Unit, CSIR – Institute of Genomics and Integrative Biology, Mathura Road, New Delhi 110 020, India

Correspondence e-mail: btaneja@igib.res.in

Received 25 September 2013
Accepted 11 December 2013

PDB reference: Rv2372c, 4l69

U1498 of 16S rRNA plays an important role in translation fidelity as well as in antibiotic response. U1498 is present in a methylated form in the decoding centre of the ribosome. In this study, Rv2372c from *Mycobacterium tuberculosis* has been identified as an RsmE-like methyltransferase which specifically methylates U1498 of 16S rRNA at the N3 position and can complement RsmE-deleted *Escherichia coli*. The crystal structure of Rv2372c has been determined, and reveals that the protein belongs to a distinct class in the SPOUT superfamily and exists as a dimer. The deletion of critical residues at the C-terminus of Rv2372c leads to an inability of the protein to form stable dimers and to abolition of the methyltransferase activity. A ternary model of Rv2372c with its cofactor *S*-adenosylmethionine (SAM) and the 16S rRNA fragment ^{1487}16S rRNA^{1510} helps to identify binding pockets for SAM (in the deep trefoil knot) and substrate RNA (at the dimer interface) and suggests an S_N2 mechanism for the methylation of N3 of U1498 in 16S rRNA.

1. Introduction

Ten methylated nucleotides are present in 16S ribosomal RNA of *Escherichia coli*. Although specific roles for all methylated positions have not been well established, the majority of these methylated nucleotides are distributed around the decoding centre and are thought to play significant roles in ribosome function. Nucleotide methylations in 16S rRNA occur during different stages of ribosome maturation and assembly and are brought about by specific methyltransferases. These methyltransferases have been identified in *E. coli* as RsmA for $m_2^6$1518 and $m_2^6$1519, RsmB for m^5C967, RsmC for m^2G1207, RsmD for m^2G966, RsmE for m^3U1498, RsmF for m^5C1407, RsmG for m^7G527, RsmH and RsmI for m^4Cm1402, and RsmJ for m^2G1516 modifications. As many as six of these ten methylations, namely m^4Cm1402, m^5C1407, m^3U1498, m^2G1516, m_2^6A1518 and m_2^6A1519, are present in the terminal helices 44 and 45 of the 30S ribosomal subunit (Fig. 1).

The conserved positions 1400–1405 and 1496–1510 in the apical part of helix 44 are part of the A-site in the ribosomal decoding centre and play important roles in aminoacyl-tRNA selection and protein synthesis *in vitro* (Cunningham *et al.*, 1993; Berk *et al.*, 2006; Stanley *et al.*, 2010). Conformational changes and nucleotide modifications in this region modulate the ribosome structure and function by affecting intersubunit interactions, translation initiation (Qin *et al.*, 2012) and translocation (Frank & Agrawal, 2000). A direct role for at least two 16S rRNA nucleotides in this region, C1402 and U1498, has been shown in ribosome activity. A lack of methylation at C1402 affects the decoding fidelity of the

ribosome by increasing translation initiation at non-AUG codons as well as by decreasing the rate of UGA read-through (Kimura & Suzuki, 2010). A U1498G mutation, on the other hand, affects the formation of the first peptide bond (Ringquist *et al.*, 1993). The role of U1498 in ribosome function has been further highlighted by the direct contact of U1498 observed with the ribose phosphate backbone of mRNA in the codon–anticodon helix (Korostelev *et al.*, 2006).

In addition to its roles in ribosomal function and fidelity, U1498 also affects the response to antibiotics. The crystal structure of ribosome in complex with hygromycin B reveals a sequence-specific binding mode for hygromycin between the A and P sites, and the drug interacts with several conserved nucleotides including U1498 in helix 44 of 16S rRNA (Brodersen *et al.*, 2000). The role of U1498 in drug response is further emphasized by the identification of U1498C among three mutations that confer different levels of hygromycin resistance in *Mycobacterium smegmatis* (Pfister *et al.*, 2003). U1498 is present in a methylated form in the decoding centre of 16S rRNA. This methylation is brought about by the highly specific methyltransferase RsmE. The growth phenotype of Δ*rsmE E. coli*, which lacks the ability to methylate U1498, is defective when grown in competition with wild-type cells (Basturea *et al.*, 2006), highlighting the importance of this methylation and its potential role in modulating the ribosomal functions mediated by this nucleotide.

The emergence of drug-resistant strains is one of the impediments to curing tuberculosis (http://www.who.int/tb/publications/global_report/). Despite the role of methylations in ribosome function and drug resistance, *erm*, *tlyA*, *gidB* and *rsmD* are the only ribosomal methyltransferases of *M. tuberculosis* that have been characterized to date (Buriánková *et al.*, 2004; Johansen *et al.*, 2006; Okamoto *et al.*, 2007; Kumar *et al.*, 2011). Although the efforts of various structural genomics consortia have led to the availability of proteins with RsmE-like structures in the PDB, a biochemical role has only been described for *E. coli* RsmE to date (Basturea *et al.*, 2006; Basturea & Deutscher, 2007). The recent structure of *E. coli* RsmE (Zhang *et al.*, 2012) displays the core fold of a SPOUT family methyltransferase (SPOUT; COG1385; Tkaczuk *et al.*, 2007) with a deep trefoil knot in its C-terminal domain. The structure reveals a modified arrangement of the two subunits in the RsmE dimer in a perpendicular orientation, unlike the parallel orientation of TrmD subunits (Ahn *et al.*, 2003). This modified arrangement of the two subunits in RsmE leaves a single functional unit for methyltransferase activity and suggests that dimerization is critical for function (Zhang *et al.*, 2012).

Here, we have identified and characterized Rv2372c as an RsmE-like methyltransferase of *M. tuberculosis* by a genome database search and structural and biochemical analysis. We have solved the structure of Rv2372c at 2.9 Å resolution, which reveals it to be an RsmE orthologue that can complement RsmE-deleted *E. coli* cells. We have also identified the critical residues necessary for dimerization, deletion of which leads to loss of dimer formation and abolition of the methyltransferase activity. We explore the molecular mechanism for methyltransferase activity on the basis of a ternary model of Rv2372c with SAM and 16S rRNA.

2. Materials and methods

2.1. Cloning, expression and purification of recombinant Rv2372c constructs

The open reading frame encoding Rv2372c was PCR-amplified from H37Rv genomic DNA using forward and reverse primers. The forward primer introduced a *Bgl*II site to the start codon, while the reverse primer introduced an *Xho*I site 3′ to the stop codon. After digestion with *Bgl*II and *Xho*I, the PCR product was cloned into pET28-His$_{10}$-Smt3 vector digested with *Bam*HI and *Xho*I to give the expression plasmid pRsm-1. pRsm-1 encodes Rv2372c fused to an N-terminal His$_{10}$-Smt3 tag. The recombinant construct was transformed into *E. coli* BL21 (DE3) cells for expression and the tagged protein was purified as described previously (Kumar *et al.*, 2011). The His$_{10}$-Smt3 tag was cleaved by incubation of the eluted protein with the Smt3-specific protease Ulp1 (the Ulp1:protein ratio was 1:500) at 4°C overnight and the tag was removed by passage through an Ni–NTA column. The cleavage site leaves one additional residue (serine) at the N-terminus of Rv2372c. Recombinant Rv2372c was collected

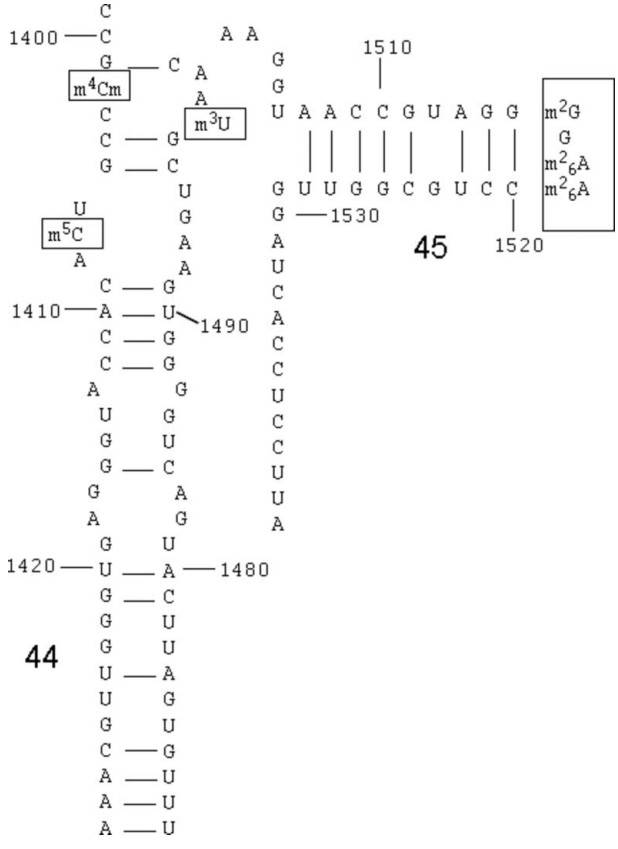


Figure 1
Schematic view of the secondary structure of helices 44 and 45 of *E. coli* 16S rRNA. Several methylated nucleotides of 16S rRNA are clustered together in helices 44 and 45. The methylated positions are highlighted in boxes, while helix numbers are indicated in bold.

in the flowthrough and the protein was further purified by cation-exchange chromatography on a Mono S column (Amersham). The purified recombinant Rv2372c was finally dialyzed against dialysis buffer [10 m*M* Tris buffer pH 8.0, 5 m*M* EDTA, 50 m*M* NaCl, 10%(*v*/*v*) glycerol, 0.1%(*v*/*v*) β-mercaptoethanol], concentrated using Vivaspin 3 kDa molecular-weight cutoff filter units and stored at 4°C until further use.

C-terminally truncated Rv2372c constructs were PCR-amplified from pRsm-1. Reverse primers were designed to terminate at the 246th and 249th amino-acid positions of the 262-residue Rv2372c, resulting in 16-residue and 13-residue deletions, respectively. The forward primer was the same as that used for full-length Rv2372c. PCR products were digested with *Bgl*II and *Xho*I and cloned between the *Bam*HI and *Xho*I sites of pET28-His$_{10}$-Smt3 to give the expression plasmids pRsm-13 and pRsm-16. The truncated proteins were expressed and purified from *E. coli* BL21 (DE3) following the same procedure as described above for full-length Rv2372c, resulting in the truncated proteins RSM-13 and RSM-16 lacking residues 249–262 or 246–262, respectively.

2.2. Protein crystallization, data collection and structure refinement

Purified Rv2372c was crystallized by the sitting-drop vapour-diffusion method by mixing 1 µl of 10 mg ml^{-1} protein solution and 1 µl reservoir solution (100 m*M* Tris–HCl pH 8.0, 1.5 *M* potassium acetate) at 24°C. Crystals were obtained in two to three weeks. The crystals were cryoprotected in a solution consisting of 20% ethylene glycol in reservoir solution and flash-cooled in liquid nitrogen before data collection.

Diffraction data were collected on the BM14 beamline at the ESRF. The data were processed using *MOSFLM* (Leslie & Powell, 2007) and were scaled using *SCALA* within the *CCP*4 package (Winn *et al.*, 2011). The crystals belonged to space group $P4_12_12$ and contained one molecule in the crystallographic asymmetric unit. Data-collection and refinement statistics are summarized in Table 1.

The structure of Rv2372c was solved by molecular replacement with the help of the *BALBES* server (Long *et al.*, 2008). The best solution was obtained with the polyalanine model of PDB entry 1vhk (hypothetical protein from *Bacillus subtilis*; Badger *et al.*, 2005); the two sequences had 33% sequence identity over 87% sequence coverage. The initial model contained 208 residues with an $R_{\rm work}$ of 0.369 and an $R_{\rm free}$ of 0.423. The initial model was further improved by iterative cycles of model building in *Coot* (Emsley *et al.*, 2010) and refinement using *REFMAC*5 (Murshudov *et al.*, 2011) and *BUSTER* (Bricogne *et al.*, 2011) with TLS refinement (Painter & Merritt, 2006; residues 2–248). The final coordinates have been deposited in the PDB with accession code 4l69.

2.3. Construct for determining the *in vivo* activity of full-length and truncated Rv2372c

The *in vivo* activities of full-length and truncated proteins were determined by expressing them under an arabinose promoter in the vector pBAD24 containing an ampicillin selection marker (Guzman *et al.*, 1995). Plasmids pRsm-1, pRsm-13 and pRsm-16 were digested with *Nco*I and *Xho*I and the fragments containing the genes along with the fused N-terminal His$_{10}$-Smt3 tag were cloned between the *Nco*I and *Xho*I sites of pBAD24, giving the expression constructs pRsm-1′, pRsm-13′ and pRsm-16′, respectively. Plasmid pBSmt2 lacking any insert was prepared in a previous study (Kumar *et al.*, 2011) and was also used as a negative control here.

Table 1
Data-collection and model-refinement statistics.

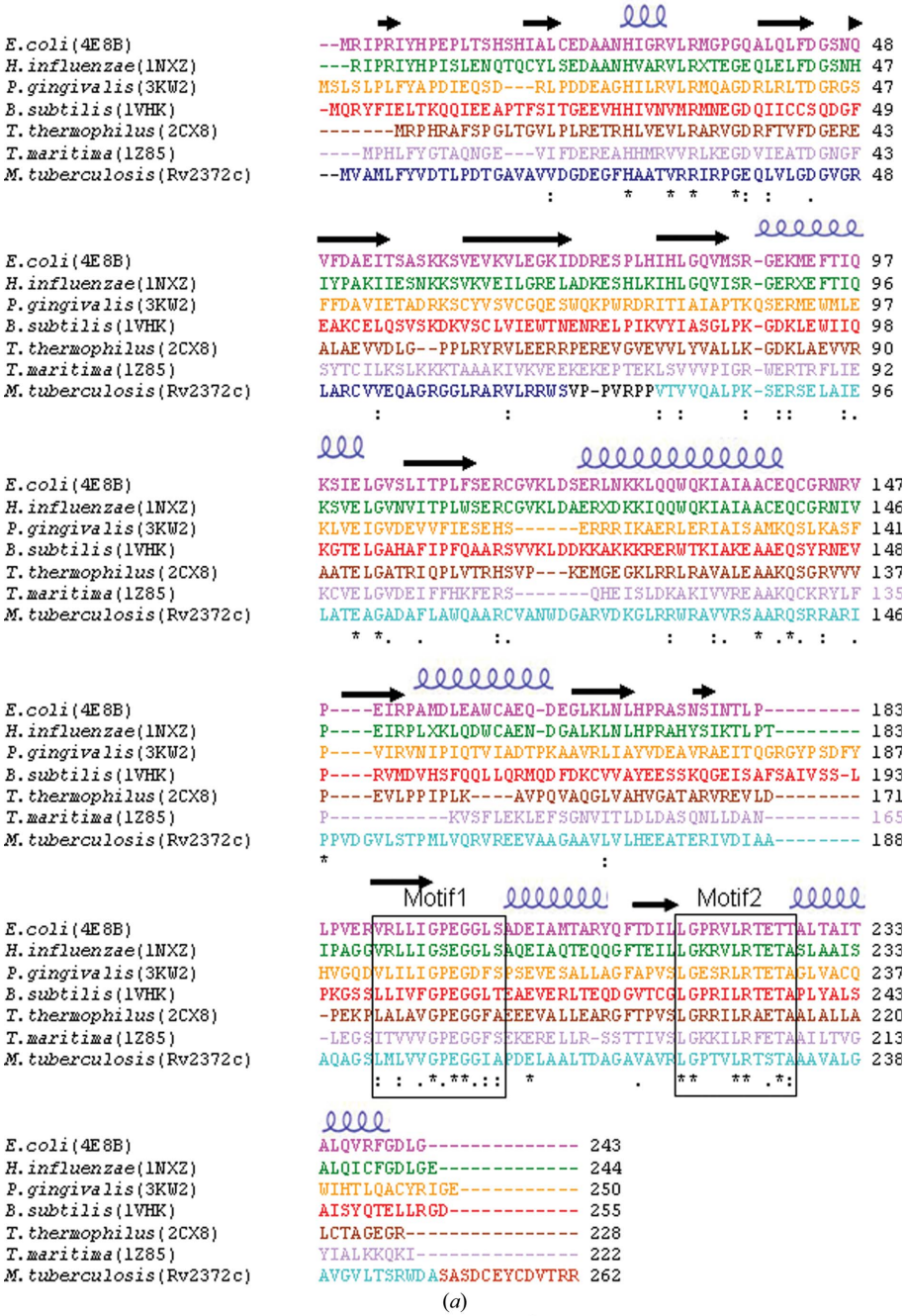
Values in parentheses are for the highest resolution shell.

Crystal and data-collection statistics	
Space group	$P4_12_12$
Unit-cell parameters (Å, °)	$a = b = 81.95$, $c = 115.136$, $\alpha = \beta = \gamma = 90$
Resolution (Å)	57.94–2.90 (3.06–2.90)
No. of unique reflections	8836 (1278)
Completeness (%)	97.0 (98.1)
Multiplicity	8.1 (8.2)
$R_{\rm meas}$† (%)	7.2 (57.4)
$\langle\langle I\rangle/\sigma(\langle I\rangle)\rangle$‡	16.5 (4.2)
Refinement statistics	
Resolution (Å)	51.75–2.90 (3.24–2.90)
No. of reflections (working/test)	8822/416
$R_{\rm work}$§	0.201 (0.235)
$R_{\rm free}$¶	0.239 (0.318)
Total protein residues/atoms	244/1752
Total water molecules	20
Wilson *B* factor (Å^2)	93.67
Average *B* factor (Å^2)	
Protein atoms	97.06
Water molecules	74.60
R.m.s. deviations from ideal††	
Bond lengths (Å)	0.008
Bond angles (°)	1.04
Ramachandran plot‡‡	
Most favoured regions (%)	95.8
Disallowed regions (%)	0.4
Rotamers‡‡	
Poor rotamers (%)	2.96

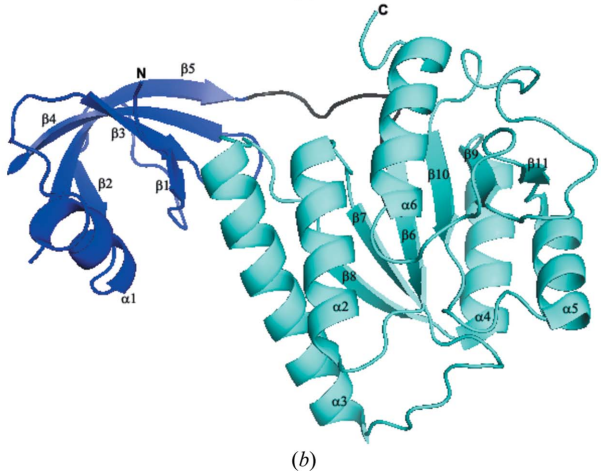
† $R_{\rm meas}$ as described in Diederichs & Karplus (1997). ‡ $\langle\langle I\rangle/\sigma(\langle I\rangle)\rangle$ is the mean $I(hkl)$ over the standard deviation of the mean $I(hkl)$ averaged over all reflections in a resolution shell. § $R_{\rm work} = \sum_{hkl} \big||F_{\rm obs}| - |F_{\rm calc}|\big| / \sum_{hkl} |F_{\rm obs}|$, where $|F_{\rm obs}|$ is the observed structure-factor amplitude and $|F_{\rm calc}|$ is the calculated structure-factor amplitude. ¶ $R_{\rm free}$ is the *R* factor based on 5% of the data which were excluded from refinement. †† As per the standard amino-acid dictionary of Engh & Huber (1991). ‡‡ From the *MolProbity* server (Chen *et al.*, 2010).

2.4. Complementation of *rsmE*-deleted *E. coli* cells

The m^{3}U1498 methyltransferase activity of Rv2372c and the truncated proteins was determined by transforming *rsmE*-deleted KL16 *E. coli* cells (KL16Δ*rsmE*; Das *et al.*, 2008) with pRsm-1′, pRsm-13′, pRsm-16′ or pBsmt2. Cells were grown at 37°C to $A_{600} \simeq 0.5$ in the presence of 100 µg ml^{-1} ampicillin and 50 µg ml^{-1} kanamycin. The protein was expressed by induction with 0.2% arabinose and incubation for 3 h at the same temperature, and the cells were harvested for RNA purification. Total RNA was isolated from the cells using an RNeasy kit (Qiagen). 5 µg of RNA was annealed to a ^{32}P-5′-end labelled primer complementary to sequence 1513–1531 of *E. coli* 16S rRNA. Primer-extension reaction was carried out using M-MuLV reverse transcriptase (Fermentas) as per



(*a*)



(*b*)

the manufacturer's protocol. The reaction was terminated with formamide/EDTA at 85°C and the products were analyzed on a 15% polyacrylamide (19:1)/7 *M* urea gel and visualized by phosphorimaging. For comparison of the size of the extension products, synthetic molecular-weight markers with lengths of 29–32 nucleotides designed in a previous study (Kumar *et al.*, 2011) were used.

2.5. Circular-dichroism spectroscopy measurements

Far-UV CD spectra were collected on a Jasco J815 CD spectrometer using a quartz cuvette with a path length of 0.1 cm at room temperature. Ellipticity data were collected in the range 250–195 nm. The concentration of the proteins (Rv2372c, RSM-13 and RSM-16) for CD measurements was kept at 0.15 mg ml^{-1} in 5 m*M* HEPES pH 7.4, 50 m*M* KCl. Each spectrum was recorded as an average of three scans. In all experiments, the contributions of the buffer to the spectra were subtracted and the mean residue ellipticities were determined before plotting the spectra.

2.6. Cross-linking with glutaraldehyde

To analyze the most stable oligomeric state of full-length Rv2372c and its C-terminally truncated forms, namely RSM-13 and RSM-16, the proteins were cross-linked with increasing concentrations of glutaraldehyde. The cross-linking reaction was stopped by adding 1 *M* Tris and 5× SDS–PAGE loading

Figure 2
Sequence alignment and overall structure of Rv2372c. (*a*) Structure-based sequence alignment of Rv2372c and its structural homologues as identified by *DALI*. The sequence of each homologue is shown in a different colour. The NTD and CTD of Rv2372c are shown in blue and cyan, respectively, the short linker connecting the two is shown in black and the C-terminal tail (disordered in the crystal structure) is shown in brick red. The two conserved sequence motifs of the RsmE family are indicated. (*b*) The overall structure of Rv2372c consists of two distinct domains, a PUA-like NTD (blue) and a SPOUT-like CTD (cyan) containing a deep trefoil knot, separated by a short linker region (black). The N- and C-termini are labelled.

dye. The samples were then loaded onto a 12% denaturing polyacrylamide gel and the oligomeric state was calculated by comparison with a standard molecular-weight protein marker (Fermentas).

2.7. Modelling of the ternary complex of Rv2372c with SAM and 16S rRNA

The potential SAM binding site was identified by superposing the coordinates of Rv2372c with those of PDB entry 2egv (rRNA methyltransferase from *Aquifex aeolicus* in complex with SAM; RIKEN Structural Genomics/Proteomics Initiative, unpublished work). To understand the possible mechanism of catalysis of Rv2372c, the final coordinates of Rv2372c were docked with a fragment of helix 44 of 16S ribosomal RNA (comprising nucleotide positions 1487–1510 of 16S rRNA) from the coordinates of PDB entry 3ofo (Dunkle *et al.*, 2010) using the *PatchDock* server (Schneidman-Duhovny *et al.*, 2003). The default parameters for the *PatchDock* server were used for the docking experiment. The position of U1498 in the docked complex with Rv2372c was manually curated in *Coot* (in the presence of SAM obtained as above) for analysis of the potential interactions between the two molecules.

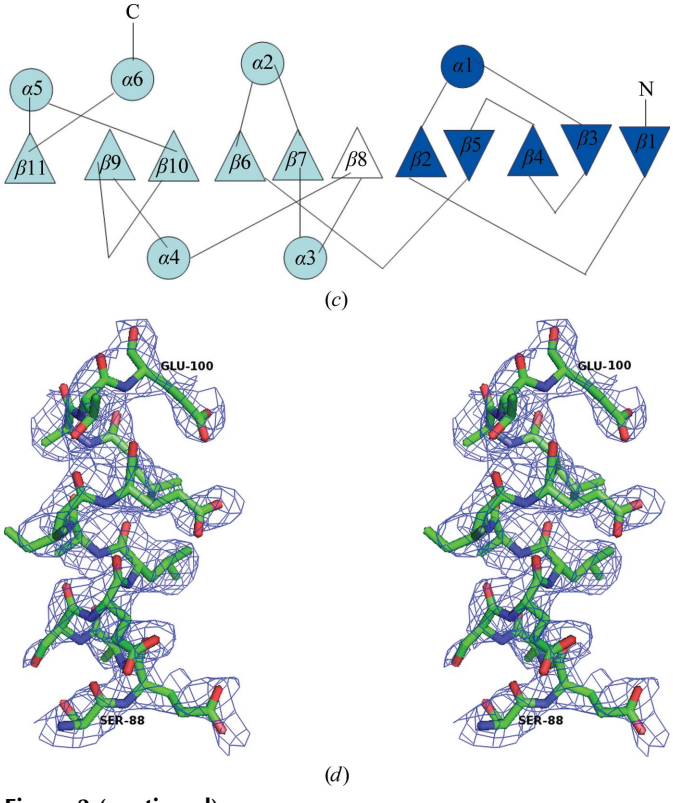


Figure 2 (continued)
(*c*) Topology diagram of Rv2372c, with α-helices and β-strands denoted as circles and triangles, respectively. The secondary-structure elements are numbered as in (*b*). The conserved structural elements of the NTD or CTD are shown in colour, whereas the variable strand β8 is shown in white. (*d*) Stereoview of the OMIT map of the interface helix at the 1.5σ level indicates a good fit.

3. Results

3.1. Identification and sequence analysis of Rv2372c of *M. tuberculosis*

A search for the homologue of RsmE in *M. tuberculosis* resulted in Rv2372c as the top hit (*E* value of 10^{-19}). Rv2372c consists of an open reading frame of 262 amino acids and has 25% sequence identity to *E. coli* RsmE, with 60 identical and 104 similar residues in the sequence alignment. A *BLAST* search of Rv2372c against nonredundant protein sequences gives top hits with uncharacterized proteins from Actinobacteria and further matches with several other bacterial sequences, with the *E. coli* protein as the only characterized member. A search for the presence of conserved domains shows that the region 1–242 of Rv2372c belongs to the RNA methyltransferase superfamily, while the further stretch of residues at the C-terminus did not show a match to any conserved domains. Examination of the sequence alignment reveals that Rv2372c of *M. tuberculosis* has a unique extension of 13 amino acids (Fig. 2*a*) that was not observed in any other sequences.

A PROSITE scan did not identify any conserved sequence motifs in Rv2372c. Multiple sequence alignments indicate several conserved residues that may have important functional roles. The C-terminal half of Rv2372c appears to be more conserved than the N-terminal region, with only four residues, namely His27, Arg33, Gly37 and Val31, being conserved in the N-terminal region. The N-terminal region shows structural similarities to RNA-binding proteins (described later) and the conserved residues may be involved in interactions with substrate nucleotides. There are two conserved regions towards the C-terminus of the sequence that have previously been identified as distinct motifs that distinguish *E. coli* RsmE from other SPOUT methyltransferases (MTases; Basturea *et al.*, 2006; Zhang *et al.*, 2012). Motif 1 in Rv2372c comprises residues ^{196}LVVGPEGGIA^{205} and motif 2 comprises residues ^{222}LGPTVLRTSTA^{232} (Fig. 2*a*). As described later in the text, residues in these conserved motifs take part in SAM binding and are functionally significant.

3.2. Crystallization and structure determination of Rv2372c

Rv2372c was purified to homogeneity and crystallized as described in §2. The structure of Rv2372c was determined to 2.9 Å resolution with final $R_{\rm work}$ and $R_{\rm free}$ values of 0.201 and 0.239, respectively. The final model of Rv2372c consists of a total of 244 residues. The first two residues at the N-terminus (including the N-terminal serine obtained as a cloning artefact), residues 57–59 and 14 residues at the C-terminus were disordered and could not be built in the final model. The presence of the intact protein in the crystal was confirmed by SDS–PAGE (Supplementary Fig. S1[1]).

The overall structure consists of two distinct domains, an N-terminal domain (NTD) comprising residues 2–71 and a C-terminal domain (CTD) comprising residues 79–248, sepa-

[1] Supporting information has been deposited in the IUCr electronic archive (Reference: RR5057).

rated by a short linker region (residues 72–78) (Figs. 2*b*, 2*c* and 2*d*). The NTD of Rv2372c exhibits a large amount of conformational flexibility as inferred from the high average *B* factor (overall) of 124.79 Å^2 in comparison to an average *B* factor (overall) of 87.04 Å^2 for the CTD.

The NTD of Rv2372c contains a twisted sheet composed of five strands (β1–β5), which are mostly antiparallel to each other, with a single helix (α1) on one side (Figs. 2*b* and 2*c*). A *DALI* (Holm & Sander, 1993) search for structural homologues of the NTD returned top hits with the N-terminal domains of previously uncharacterized proteins, proteins with an RNA methyltransferase fold (PDB entry 1nxz, *Z*-score 10.1, r.m.s.d. of 2.0 Å for 74 C$^\alpha$ atoms, Forouhar *et al.*, 2003; PDB entry 1vhy, *Z*-score 9.4, r.m.s.d. of 2.0 Å for 73 C$^\alpha$ atoms, Badger *et al.*, 2005; PDB entry 4e8b, *Z*-score 9.2, r.m.s.d. of 1.9 Å for 73 C$^\alpha$ atoms, Zhang *et al.*, 2012) as well as with the RNA-binding PUA (**p**seudo**u**ridine synthase and **a**rchaeosine transglycosylase) domain (Pérez-Arellano *et al.*, 2007): (PDB entry 2ane, *Z*-score 5.3, r.m.s.d. of 2.8 Å for 65 C$^\alpha$ atoms, Li *et al.*, 2005; PDB entry 2nwa, *Z*-score 5.3, r.m.s.d. of 3.4 Å for 61 C$^\alpha$ atoms, Northeast Structural Genomics Consortium, unpublished work; PDB entry 1ze2, *Z*-score 4.9, r.m.s.d. of 2.6 Å for 58 C$^\alpha$ atoms, Phannachet & Huang, 2004). Structural comparisons between the NTD of Rv2372c and PUA domains reveals that the NTD of Rv2372c has diverged significantly from the typical α/β architecture of the PUA fold comprising a six-stranded pseudo-β-barrel capped by an α-helix on each apical side of the pseudobarrel (Pérez-Arellano *et al.*, 2007; Supplementary Fig. S2). RsmE specifically methylates U1498 in 16S rRNA and is not known to methylate any other nucleotide in DNA or RNA. The divergence in the PUA-like domain at the N-terminus is likely to be an adaptation to recognize target RNA structures with high specificity.

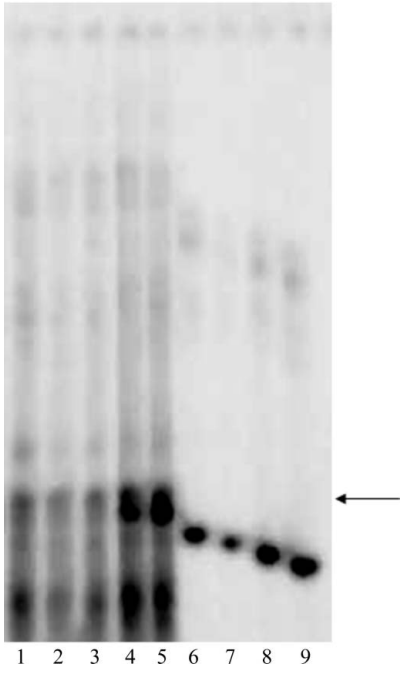


Figure 3
Primer-extension analysis for *in vivo* activity of Rv2372c. Primer-extension reactions were carried out using M-MuLV reverse transcriptase to monitor the *in vivo* methylation carried out by Rv2372c. Products were analyzed on a 15% urea denaturing polyacrylamide gel and visualized by phosphorimaging. The primer-extension products in the different lanes correspond to reactions carried out on total RNA isolated from *E. coli* KL16Δ*rsmE* (lane 1), *E. coli* KL16Δ*rsmE* containing pBSmt2 expressing only the His_{10}-Smt3 tag (lane 2), *in vitro* reaction on total RNA purified from *E. coli* KL16Δ*rsmE* incubated with purified Rv2372c (lane 3), *E. coli* KL16Δ*rsmE* complemented with pRsm-1′ (lane 4) and the parent *E. coli* KL16 strain (lane 5). Lanes 6–9 contain ^{32}P-end-labelled oligonucleotide markers ME1, ME2, ME3 and ME4 of lengths 32, 31, 30 or 29 nucleotides, based on our previous study (see text for further information) to ascertain the size of the extension product. The position of the expected product of the extension reaction when m^{3}U1498 is methylated is indicated.

The C-terminal domain is the catalytic domain of Rv2372c and contains conserved motifs harbouring residues involved in ligand (SAM) binding and in methyltransferase activity. The CTD belongs to the SPOUT superfamily of methyltransferases, with a central six-stranded parallel β-sheet (β6–β11) sandwiched between two layers of helices on either side (α3 and α4 on one side and α2, α5 and α6 on the other) and a deep trefoil knot formed by threading the β11–α6 segment of Rv2372c through the loop connecting β9 and β10 at its C-terminus (Figs. 2*b* and 2*c*). The CTD of Rv2372c contains an additional β-strand (β8) in addition to the conserved topology of SPOUT proteins (Fig. 2*c*). This additional strand appears to be present in a 'melted state' in *E. coli* RsmE. The presence of additional 'variable' secondary-structure elements in addition to the conserved topological elements has been observed in several SPOUT proteins (Tkaczuk *et al.*, 2007) and may be utilized in conferring additional structural stability or substrate specificity to enzymes of this family.

A *DALI* search with the CTD of Rv2372c shows significant similarity to the C-terminal domains of RNA methyltransferases with the SPOUT fold, with the hypothetical protein YqeU as the top hit (PDB entry 1vhk, *Z*-score 23.4, r.m.s.d. of 1.7 Å for 161 C$^\alpha$ atoms; Badger *et al.*, 2005). *E. coli* RsmE (PDB entry 4e8b, *Z*-score 21.2, r.m.s.d. of 1.7 Å for 164 C$^\alpha$ atoms; Zhang *et al.*, 2012), 2′-*O*-methyltransferase (PDB entry 1ipa, *Z*-score 13.7, r.m.s.d. of 2.7 Å for 135 C$^\alpha$ atoms; Nureki *et al.*, 2002), the tRNA methyltransferase TrmH (PDB entry 1v2x, *Z*-score 11.4, r.m.s.d. of 2.7 Å for 134 C$^\alpha$ atoms; Nureki *et al.*, 2004) and the tRNA methyltransferase TrmD (PDB entry 1p9p, *Z*-score 7.8, r.m.s.d. of 3.5 Å for 115 C$^\alpha$ atoms; Elkins *et al.*, 2003) are other proteins with the SPOUT fold with significant similarity, suggesting that C-terminal domain of Rv2372c is likely to share a similar catalytic mechanism with all of these proteins despite all of them acting on vastly different substrates in a highly specific manner (Supplementary Fig. S3).

3.3. *In vivo* activity of Rv2372c

To elucidate the *in vivo* activity of Rv2372c, *E. coli* KL16Δ*rsmE* (Das *et al.*, 2008) was transformed with pRsm-1′. The total RNA isolated from the induced cells was used for primer-extension analysis as described in §2. Extension of the primer stops at position 1499, one nucleotide before the modified nucleotide (m^{3}U1498), because of altered Watson–

Crick base pairing at the methylated nucleotide. *E. coli* KL16Δ*rsmE* cells were complemented by *rv2372c*, as seen by inhibition of the extension reaction in *E. coli* KL16Δ*rsmE* cells transformed with pRsm-1′ (Fig. 3, lane 4). A similar result was obtained in wild-type *E. coli* (Fig. 3, lane 5) but not in vector-only control (Fig. 3, lane 2) or in *E. coli* KL16Δ*rsmE* (Fig. 3, lane 1). *rv2372c* hence complements *E. coli* KL16Δ*rsmE* cells by methylating U1498 of 16S rRNA at the N3 position and inhibiting the primer extension at A1499.

The primer-extension reaction carried out on RNA purified from *E. coli* KL16Δ*rsmE* and incubated with Rv2372c *in vitro* resulted in longer products as Rv2372c failed to methylate naked unmodified RNA in the *in vitro* reaction (Fig. 3, lane 3). The absence of any methylation in the *in vitro* reaction on purified naked RNA (purified from *E. coli* KL16Δ*rsmE*) indicates that, like *E. coli* RsmE, Rv2372c also requires its substrate (16S rRNA) in the context of the 30S ribosome for m^3U1498 methytransferase activity.

3.4. The C-terminus of Rv2372c is important for protein dimerization

As described above, a search for the CTD carried out using the *DALI* server (Holm & Sander, 1993) revealed close matches with several putative methyltransferases, with *E. coli* RsmE being the only characterized member. A similar result was obtained in a search for overall structural homologues of Rv2372c. Several of the top hits have been crystallized by structural genomics consortia, but their functions have not been explored. All of these structural homologues of Rv2372c have been reported as dimers. Structural analysis of Rv2372c suggests that the functional state is formed by two subunits

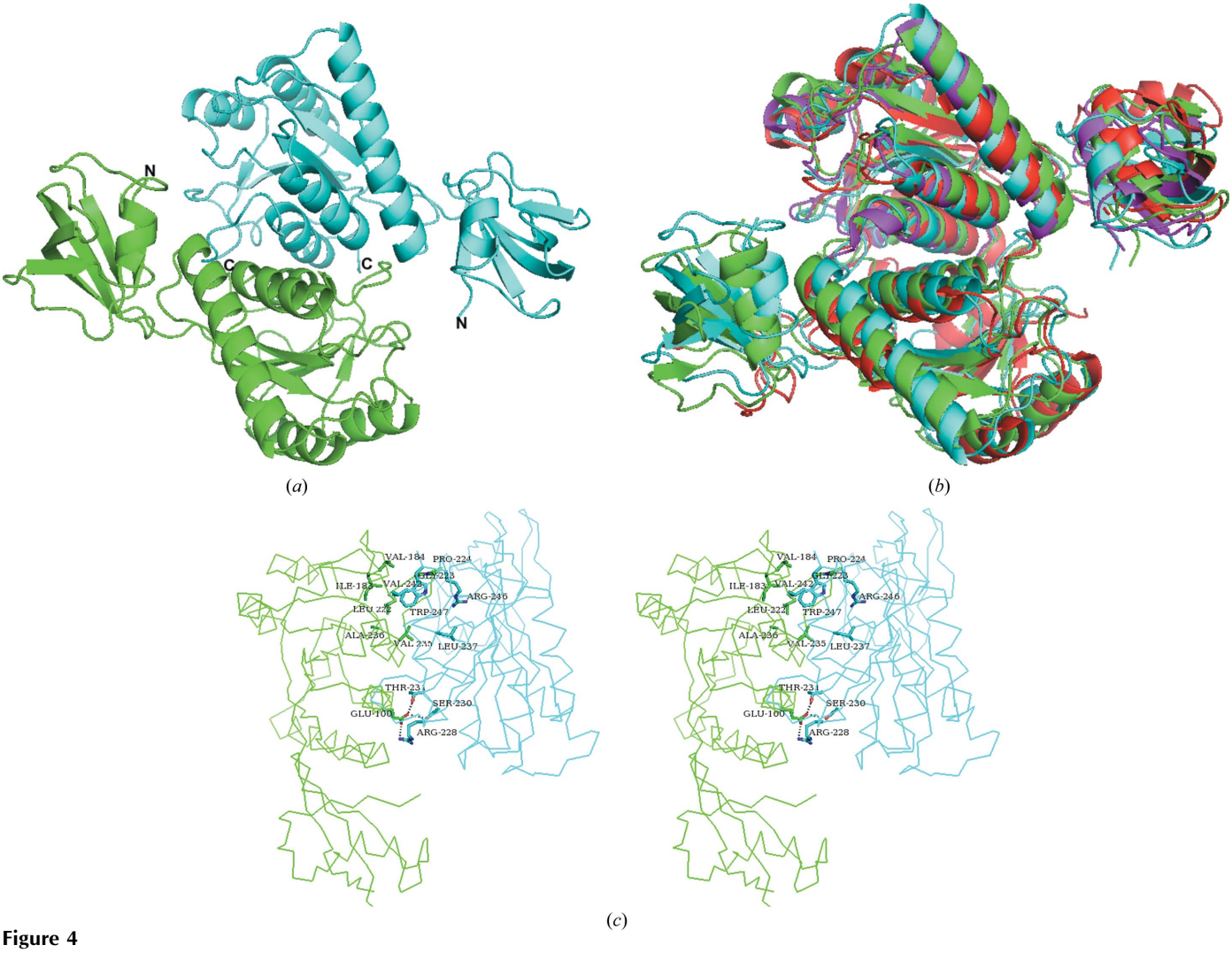


Figure 4
Rv2372c exists as a functional dimer. (*a*) Rv2372c exists as a dimer with the two subunits separated along the crystallographic twofold axes. The dimer structure of Rv2372c revealed by the *PISA* (*Protein Interface, Surfaces and Assemblies*) server is shown with the two subunits in cyan and green. The orientation of the Rv2372c subunits is identical in all subsequent figures for better visualization. (*b*) Superposition of the overall structure of Rv2372c in the dimeric state with its homologues: superposition of Rv2372c (cyan) with *E. coli* RsmE (PDB entry 4e8b, magenta), *H. influenzae* YggJ (PDB entry 1nxz, green) and *B. subtilis* hypothetical protein (PDB entry 1vhk, red) shows that overall structure is similar and the proteins have the same overall fold. (*c*) Stereo image of the residues involved in the dimer interface of Rv2372c. The C^{α} backbone of the two subunits is shown in cyan or green. The residues involved in interactions are shown in ball-and-stick representation.

separated along the crystallographic twofold axis, with the two subunits arranged perpendicular to each other (Figs. 4*a* and 4*b*). The dimer buries nearly 10.2% of the total area, corresponding to 2618 Å^2 of buried surface as analyzed by the *PISA* server (Krissinel & Henrick, 2007). The dimeric interface is stabilized by several salt bridges and hydrogen bonds at one end and by hydrophobic interactions at the other (Fig. 4*c*). Notably, Arg228 and Thr231 in motif 2 and Glu100 are conserved across various RsmE structural homologues (Fig. 2*a*), confirming the importance of these residues in the structural maintenance of a functional dimer.

Rv2372c has an extension of 13 amino acids at its C-terminus that was not observed in the sequence of any other homologues (Fig. 2*a*). Residues 249–262 at the C-terminus of Rv2372c were missing in the crystal structure, preventing unambiguous assignment of the role of this C-terminal tail in dimer formation. To monitor the potential role of this C-terminal tail in oligomerization, a C-terminal truncation of Rv2372c lacking the terminal tail of 13 residues (RSM-13) was constructed and showed no loss of dimerization (Supplementary Fig. S4*a*) or methyltransferase activity (Fig. 5, lane 2). The crystal structure of Rv2372c shows that the C-terminus is present in a highly hydrophobic region surrounded by several small apolar residues. Leu237, Val242 and Trp247 and the hydrophobic tail of Arg246 of one subunit, along with Ile183, Val184, Leu222, Gly223, Pro224, Val235 and Ala236 of the second subunit, are clustered to form a hydrophobic cage (Fig. 4*c*).

The C-terminus of several SPOUT family members including those from *Haemophilus influenzae* (PDB entry 1nxz; Forouhar *et al.*, 2003), *E. coli* (PDB entry 4e8b; Zhang *et al.*, 2012), *B. subtilis* (PDB entry 1vhk; Badger *et al.*, 2005), *Pseudomonas gingivalis* (PDB entry 3kw2; New York SGX Research Center for Structural Genomics, unpublished work) and *Thermotoga maritima* (PDB entry 1z85; Joint Center for Structural Genomics, unpublished work) show conservation of this hydrophobic cage in the subunit packing. The deletion of the terminal residues Arg246, Trp247 and Asp248 at the C-terminus of RSM-13 disrupts this hydrophobic cage, leading to structural perturbations and the inability of the resultant RSM-16 to form stable dimers (Supplementary Fig. S4*b*) and the loss of its methyltransferase activity (Fig. 5, lane 3), although the overall folded states of wild-type Rv2372c, RSM-13 and RSM-16 appear to be similar as determined by circular-dichorism measurements (Fig. 6).

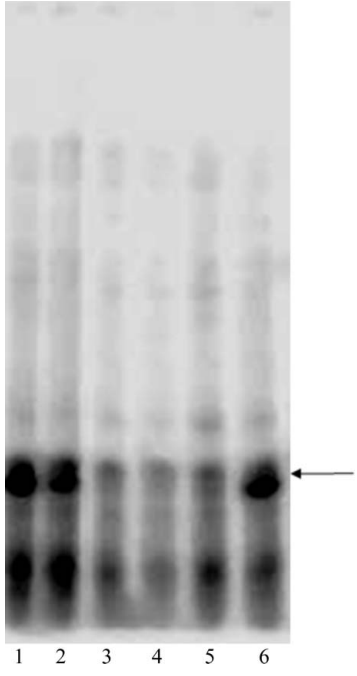


Figure 5
In vivo activity of truncated proteins. The *in vivo* activity of RSM-13 and RSM-16 was monitored by primer-extension analysis using M-MuLV reverse transcriptase; the reaction products were analyzed on a 15% urea denaturing polyacrylamide gel and were visualized by phosphorimaging. Reactions were carried out with RNA purified from *E. coli* KL16Δ*rsmE* transformed with pRsm-1′ which encodes full-length Rv2372c (lane 1), *E. coli* KL16Δ*rsmE* transformed with pRsm-13′ (lane 2), *E. coli* KL16Δ*rsmE* transformed with pRsm-16′ (lane 3), *E. coli* KL16Δ*rsmE* transformed with pBSmt2 expressing only the His_{10}-Smt3 tag (lane 4), RNA purified from *E. coli* KL16Δ*rsmE* (lane 5) and RNA purified from *E. coli* KL16 (lane 6). The position of the expected product of the extension reaction when m^3U1498 is methylated is indicated.

3.5. Mechanism of substrate binding to Rv2372c

Methyltransferases with the SPOUT fold exhibit binding of SAM to the C-terminal trefoil knot (Tkaczuk *et al.*, 2007). As described above, the structure of Rv2372c belongs to the SPOUT superfamily of methyltransferases and the molecular interactions involved in trefoil formation and stabilization of the dimer interface have been identified. In order to investigate the molecular interactions of Rv2372c with the SAM ligand, several efforts were made to obtain crystals of the Rv2372c–SAM complex. However, despite multiple attempts, no crystals were obtained of the complex. The coordinates of SAM were then modelled by superposing Rv2372c on another SPOUT-family rRNA methyltransferase in complex with SAM (PDB entry 2egv). The two structures are highly similar, with an overall *Z*-score of 23.7 and an r.m.s.d. of 2.2 Å for 227 residues, and hence are expected to have similar binding modes to the cofactor. SAM binds at the deep trefoil knot towards the C-terminal ends of the β9 and β10 strands of Rv2372c on one side and the loop connecting the knot of β11

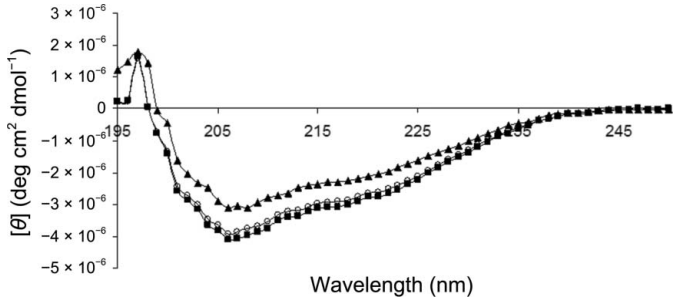


Figure 6
CD analysis of Rv2372c and truncated products. CD analysis of full-length Rv2372c (open circles), RSM-13 (filled squares) and RSM-16 (open triangles) indicates similar secondary structures suggestive of folded proteins.

and α6 on the other (Fig. 7*a*), suggesting that there are two available pockets for SAM binding. SAM binding in each subunit is stabilized through specific interactions with several residues of the two conserved motifs. The model seems robust as SAM is buried close to the hydrophobic pocket of the dimeric interface with no steric clashes and is stabilized by the formation of several potential hydrogen bonds to conserved motifs of the subunit. For instance, O3′ of SAM is located 2.8 Å from the main-chain NH of Gly199 in motif 1 in this model. Similarly, main-chain atoms of motif 2 (Leu222, Leu227 and Thr225) are present at hydrogen-bonding distances from N1, N6 and O4′ of the modelled SAM, respectively (Fig. 7*b*).

To understand the catalytic mechanism of Rv2372c, the crystal structure was docked against a stretch of 16S rRNA nucleotides flanking U1498 (^{1487}16S rRNA^{1510}) using the *PatchDock* server. The model of Rv2372c and rRNA shows that unlike SAM, which binds in the trefoil knot of individual subunits, ^{1487}16S rRNA^{1510} binds at the dimer interface near the SAM binding site. N3 of U1498 in the docked model is placed within hydrogen-bonding distance of Glu201 of Rv2372c in the current model (Fig. 8). Glu201 is part of the conserved motif 1 of RsmE (which corresponds to the conserved motif 2 of other SPOUT members) and is positioned to carry out nucleophilic attack. A conserved glutamate has previously been shown to be involved in catalysis in tRNA methyltransferases [for instance, Glu124 of *Thermus thermophilus* TrmH (PDB entry 1v2x) and Glu116 of *E. coli* TrmD (PDB entry 1p9p)]. Furthermore, mutation of this conserved glutamate has been shown to result in loss of activity in tRNA methyltransferases (Elkins *et al.*, 2003; Nureki *et al.*, 2004; Watanabe *et al.*, 2005, 2006).

In addition to the conserved glutamate in the model, conserved arginines (Arg90 and Arg228 of motif 2) are also present in the vicinity of Glu201 and may take part in the reaction through stabilization of the intermediate. The ternary model of Rv2372c–SAM–U1498 (of 16S rRNA) appears to be further stabilized by potential interactions between O4 of U1498 and Arg90 of Rv2372c and between O2 of U1498 and the carboxyl group of SAM. An arginine in motif 1 of tRNA methyltransferases of the SPOUT family has also previously been proposed to play an important role in catalysis (Watanabe *et al.*, 2005). Although motif 1 is not present in RsmE proteins, the positive charge at this position is conserved as an arginine or lysine at residue 90 in the sequence alignment (Fig. 2*a*). Alternately, the conserved Arg228, which is present in motif 2 and is at 3.1 Å from O4 of U1498 in the current model, may also take part in this catalytic step. Notably, the side chain of Arg90 is disordered in the crystal structure of Rv2372c, possibly owing to the absence of a bound substrate and/or ligand. Based on these results, we propose an S_N2 mechanism for the activity of Rv2372c mediated by conserved arginines (Arg90/228) and Glu201, similar to tRNA methyltransferases (Fig. 9).

An electrostatic surface for the ternary model of Rv2372c–SAM–^{1487}16S rRNA^{1510} was analyzed in order to further understand the binding modes of SAM and RNA to Rv2372c. The surface charge suggests that RNA binds at the dimer interface by utilizing the positive charge distributed over the surface of the two subunits during catalysis (Fig. 10*a*). The role of ribosomal proteins in modulating this interaction cannot be ruled out by our model. Interestingly, while one face of Rv2372c is positively charged (Fig. 10*a*), the opposite face is negatively charged (Fig. 10*b*). This arrangement not only plays a crucial role during the selection of helix 44 as a target for methylation, but also acts as a sink to direct the rRNA to the catalytic face of Rv2372c.

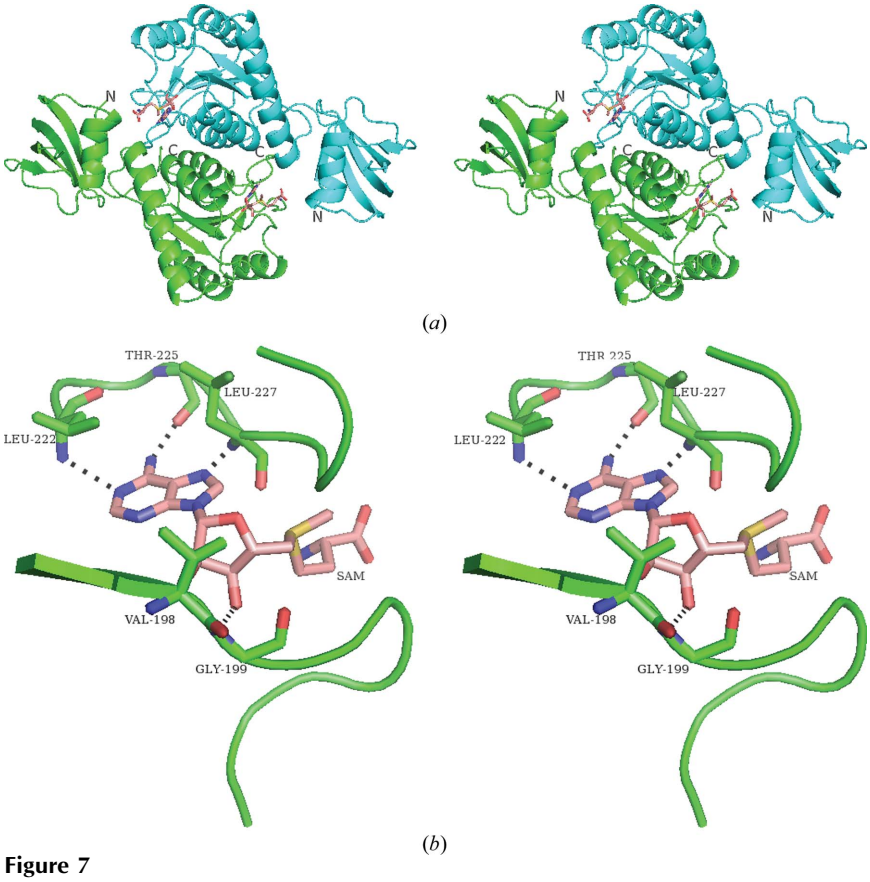


Figure 7
Interactions of SAM with Rv2372c. (*a*) Stereoview of SAM bound to Rv2372c. SAM (shown in sticks) binds close to the dimer interface in the deep trefoil pocket of each subunit. The two subunits are shown in different colours and the N- and C-termini of each is labelled. (*b*) Close-up view of the interactions of SAM with specific residues of the conserved motifs in Rv2372c (see text for further details).

4. Discussion

Both 23S and 16S ribosomal RNA consist of several base and nucleotide modifications. Most of these modifications are clustered

around important functional regions of the ribosome, such as the peptidyl-transfer centre in the large subunit or the decoding centre in the small subunit, highlighting their importance (Brimacombe *et al.*, 1993). Here, we have identified Rv2372c from *M. tuberculosis* as an orthologue of RsmE in *E. coli*. Rv2372c can complement RsmE in a mutant *E. coli* strain. Rv2372c requires the 30S ribosome subunit as substrate to methylate N3 of U1498 in 16S rRNA and is unable to methylate U1498 when incubated with naked ribosomal RNA purified from *rsmE*-deleted *E. coli* cells. The ability of Rv2372c to use *E. coli* ribosome subunits as substrate shows that despite significant sequence divergence from the *E. coli* RsmE, Rv2372c retains the conserved residues necessary for recognition of the tertiary structure of RNA in assembled ribosomes as well as the ribosomal components. This recognition is imparted by residues in the two conserved motifs of Rv2372c through specific intermolecular interactions.

The structure of Rv2372c is only the second structure to be determined of the RsmE family of proteins. The crystal structure of Rv2372c reveals that this protein also belongs to the SPOUT superfamily, with a deep trefoil knot in its catalytic C-terminal domain (SPOUT fold) and an additional PUA-like N-terminal domain involved in RNA binding. SpoU and TrmD are canonical members of the SPOUT superfamily and exist in a dimeric state, which is essential for catalysis of methyltransferase on tRNA. RsmE has previously been shown to form a distinct class of SPOUT methyltransferases with its own sequence motifs (Basturea *et al.*, 2006; Zhang *et al.*, 2012). However, critical residues in the sequence motifs of both SPOUT (Watanabe *et al.*, 2005) and RsmE-like proteins (Fig. 2*a*) are conserved across both families. These conserved residues play similar roles in function, namely Arg228 and Thr231 of Rv2372c (hydrogen bonding), Leu222, Gly223 and Val235 of Rv2372c (hydrophobic interactions in dimer formation), Gly199, Leu222 and Leu227 (SAM binding), and Glu201 and Arg228 (activity). Mutation of Arg223 in *E. coli* RsmE (equivalent to Arg228 in Rv2372c) leads to a >90% loss in activity (Zhang *et al.*, 2012). Our structural analysis indicates that loss of dimer formation owing to this mutation is the most likely cause of loss of activity. Mutation of Val139 in the hydrophobic pocket of YibK from *H. influenzae* has previously been shown to lead to a loss of dimerization in SpoU family proteins (Mallam & Jackson, 2007). Val235 in Rv2372c is structurally conserved at this position and packs against Trp247 in the hydrophobic pocket of the dimer interface. The presence of Trp247 in the sequence of Rv2372c is unique, as no other members of the RsmE family contain a bulky hydrophobic residue in this region. The loss of function upon deletion of Arg246, Trp247 and Asp248 (in addition to the extended tail) highlights the importance of dimerization in activity. However, a role for the flexible C-terminal tail remains to be determined.

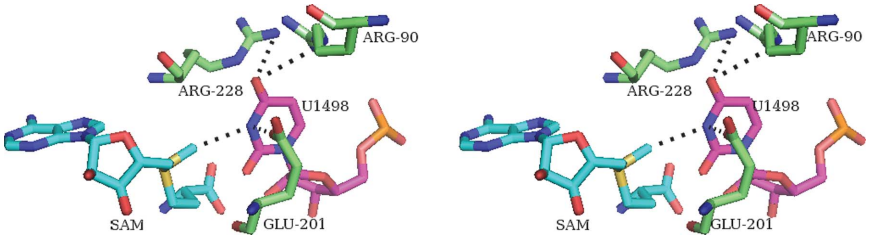


Figure 8
Ternary model of Rv2372c with RNA and SAM. Close-up view indicating the interactions of different residues of Rv2372c (green) with SAM (cyan) and U1498 (magenta) of the ^{1487}16S rRNA^{1510} fragment. Glu201 is the catalytic residue, while Arg90 or Arg228 are likely to stabilize the intermediate (see text for details).

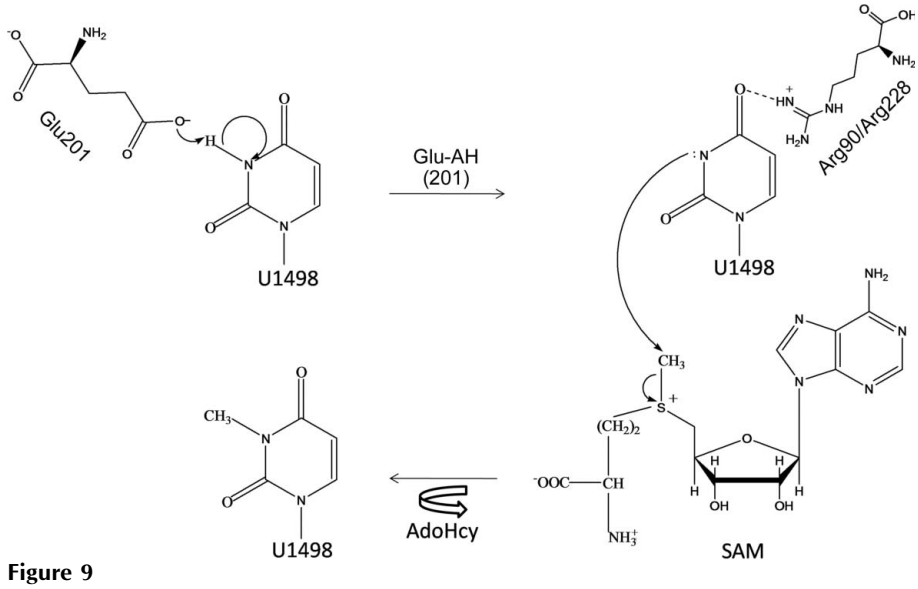


Figure 9
Proposed mechanism of catalysis by Rv2372c. The methyl transfer by Rv2372c is proposed to take place by an S_N2 mechanism. Methylation of U1498 on 16S rRNA is initiated by the deprotonation of N3 of U1498 by Glu201. This results in a shift of the electron pair to the O4 position, leaving protonated Glu201 (Glu-AH). Arg90 and Arg228 are positioned in the vicinity to stabilize this intermediate. The transfer of the methyl (CH_3) group to N3 of U1498 on 16S rRNA leaves the by-product *S*-adenosylhomocysteine (AdoHCy).

Rv2372c presents a positive charge distributed on one face of the dimer for essential electrostatic interactions with the ribosomal components of the 30S ribosome during the methylation of U1498. The importance of dimerization in catalysis is also supported by the ternary model suggesting potential interactions of U1498 with residues in the dimeric interface (Fig. 8). We hence propose that the absence of methyltransferase activity upon loss of dimer formation appears to be caused by loss of substrate binding.

TrmH is a prototypical 2′-*O*-methyltransferase of the SPOUT family. Asn35

of the conserved motif 1 of TrmH is required for binding to the 5′-phosphate, while the conserved Arg41 of the same motif is required for catalysis. RsmE-like proteins have lost the motif 1 of other SPOUT superfamily proteins. However, Rv2372c has conserved the catalytic Glu201 (in motif 2 of the SPOUT superfamily or motif 1 of Rv2372c) for activity. The ternary model of Rv2372c–SAM–^{1487}16S rRNA^{1510} shows that the conserved Arg90 or Arg228 interacts with O4 of U1498, whereas N3 of U1498 is at a hydrogen-bonding distance from the conserved Glu201 of motif 1 of Rv2372c. We hence propose an S_N2 mechanism for activity of Rv2372c, similar to the tRNA methyltransferases of the SPOUT superfamily (Christian *et al.*, 2010). Despite significant sequence divergence and the specific motifs of SPOUT and RsmE-like proteins, the RsmE family of methyltransferases and tRNA methyltransferase share a similar mechanism of catalysis towards different substrates. However, RsmE-like proteins appear to have evolved from other SPOUT superfamily members to methylate N3 of U1498 instead of having 2′-*O*-methyltransferase (TrmH/SpoU) or N1-methyltransferase (TrmD) activity. RsmE proteins hence belong to a distinct class of the SPOUT superfamily that has evolved from the prototypical TrmH for methylation of the specific substrate U1498 of 16S rRNA.

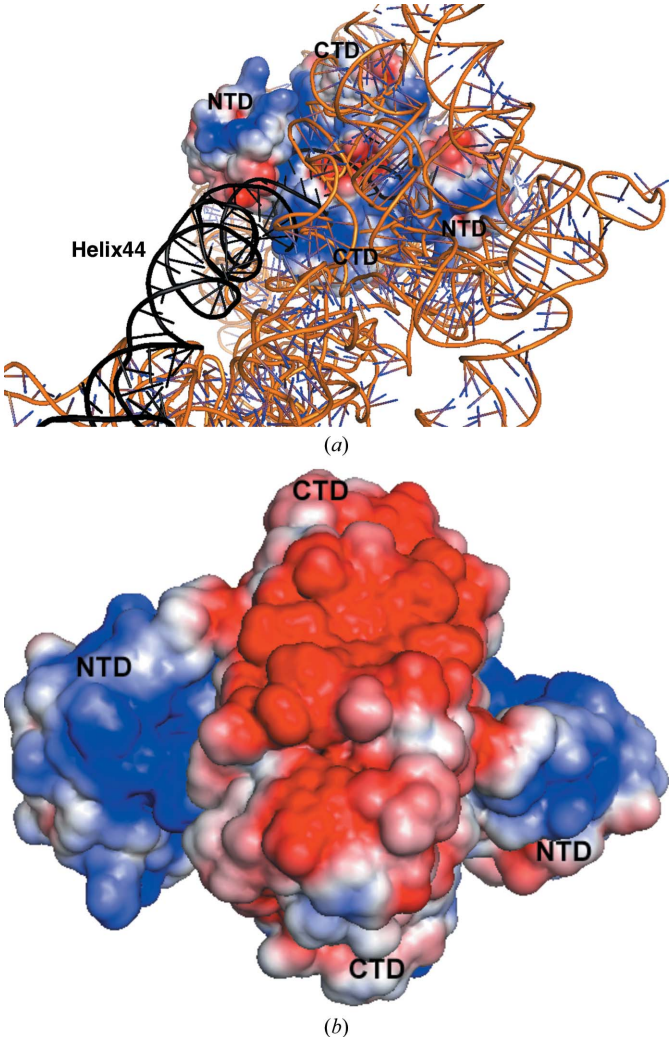


Figure 10
Electrostatic surface of Rv2372c. (*a*) The electrostatic surface potential of Rv2372c indicates positive charge on one face of Rv2372c that can bind 16S rRNA. There are several potential electrostatic interactions of Rv2372c with the negative phosphate backbone of the 16S rRNA. Helix 44, where U1498 lies, is shown in black. (*b*) The opposite face of Rv2372c, rotated 180° with respect to (*a*), has a strong distribution of negative charge on its surface. The surface in (*a*) and (*b*) is coloured with a blue to red gradient from $-2.5k_BT$/e to $+2.5k_BT$/e.

Rv2372c has been suggested as a potential 'high-confidence' drug target for *M. tuberculosis* (Raman *et al.*, 2008). Since U1498, the target of methylation by Rv2372c, is present in the crucial helix 44 of the ribosome, we wanted to investigate the effect of RsmE deletion in *E. coli* on different aminoglycosides or macrolides. *E. coli* KL16Δ*rsmE*, lacking methylation ability of U1498, was screened against several antibiotics. However, no significant effect of this single mutation was observed in drug sensitivity (Supplementary Table S1). Since U1498 is one of several methylated nucleotides in the hydrophobic cluster of helix 44, the lack of a single methylated group may not be sufficient to observe an effect on growth under our experimental conditions. Alternately, higher concentrations of investigated antibiotics may need to be tested. U1498 has been shown to be involved in direct interaction with hygromycin (Brodersen *et al.*, 2000): O4 of U1498 forms a hydrogen bond to O30 of ring IV of hygromycin (Supplementary Fig. S5). Although the absence of methylation at N3 does not directly affect this interaction, it is likely to affect local conformations in this region. U1498 has also been found to be in proximity to hygromycin in other crystal structures (Supplementary Fig. S5). The ribosome is one of the most common targets of antibiotics in the cell. These antibiotics bind to similar but not identical sites on the ribosome. This can be seen by the somewhat larger distances of ∼8 Å between U1498 and bound paromomycin or viomycin in the crystal structures (Supplementary Fig. S5). The role of ribosomal RNA methyltransferase and methylations in *M. tuberculosis* in growth and antibiotic resistance are beginning to emerge. The effect of the loss of methylation of specific rRNA nucleotides on antibiotic response hence needs to be explored in mycobacterial cells in greater detail.

In conclusion, these results confirm the identification of Rv2372c as an RsmE-like methyltransferase through a combination of structural, *in vivo* and *in vitro* biochemical studies and thereby unambiguously confirm the presence of nucleotide methylations in ribosomal RNA in mycobacteria. The role of U1498 in translation fidelity as well as antibiotic response lends greater significance to the work. These results may be exploited further in understanding ribosomal function during growth (or latency) as well as during drug-design studies.

AK acknowledges ICMR and SK acknowledges CSIR for fellowships. This work was supported by FAC-02 funding: a CSIR project to BT. We would like to thank Pallavi Shukla for help with purification, Dr Souvik Maiti and Sriyans Jain for help with CD and ITC experiments, Professor Umesh Varshney for his generous gift of *rsmE* knockout strains of

E. coli, Professor Stewart Shuman and Dr Krishna Murari Sinha for the pET28-His$_{10}$-Smt3 plasmid and Dr Hassan Behrali and Dr Babu Manjasetty at BM14, ESRF for data collection.

References

Ahn, H. J., Kim, H.-W., Yoon, H.-J., Lee, B. I., Suh, S. W. & Yang, J. K. (2003). *EMBO J.* **22**, 2593–2603.

Badger, J. *et al.* (2005). *Proteins*, **60**, 787–796.

Basturea, G. N. & Deutscher, M. P. (2007). *RNA*, **13**, 1969–1976.

Basturea, G. N., Rudd, K. E. & Deutscher, M. P. (2006). *RNA*, **12**, 426–434.

Berk, V., Zhang, W., Pai, R. D. & Cate, J. H. D. (2006). *Proc. Natl Acad. Sci. USA*, **103**, 15830–15834.

Bricogne, G., Blanc, E., Brandl, M., Flensburg, C., Keller, P., Paciorek, W., Roversi, P, Sharff, A., Smart, O. S., Vonrhein, C. & Womack, T. O. (2011). *BUSTER* v.2.10.0. Cambridge: Global Phasing Ltd.

Brimacombe, R., Mitchell, P., Osswald, M., Stade, K. & Bochkariov, D. (1993). *FASEB J.* **7**, 161–167.

Brodersen, D. E., Clemons, W. M. Jr, Carter, A. P., Morgan-Warren, R. J., Wimberly, B. T. & Ramakrishnan, V. (2000). *Cell*, **103**, 1143–1154.

Buriánková, K., Doucet-Populaire, F., Dorson, O., Gondran, A., Ghnassia, J.-C., Weiser, J. & Pernodet, J.-L. (2004). *Antimicrob. Agents Chemother.* **48**, 143–150.

Chen, V. B., Arendall, W. B., Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., Murray, L. W., Richardson, J. S. & Richardson, D. C. (2010). *Acta Cryst.* D**66**, 12–21.

Christian, T., Lahoud, G., Liu, C., Hoffmann, K., Perona, J. J. & Hou, Y.-M. (2010). *RNA*, **16**, 2484–2492.

Cunningham, P. R., Nurse, K., Weitzmann, C. J. & Ofengand, J. (1993). *Biochemistry*, **32**, 7172–7180.

Das, G., Thotala, D. K., Kapoor, S., Karunanithi, S., Thakur, S. S., Singh, N. S. & Varshney, U. (2008). *EMBO J*, **27**, 840–851.

Diederichs, K. & Karplus, P. A. (1997). *Nature Struct. Biol.* **4**, 269–275.

Dunkle, J. A., Xiong, L., Mankin, A. S. & Cate, J. H. D. (2010). *Proc. Natl. Acad. Sci. USA*, **107**, 17152–17157.

Elkins, P. A., Watts, J. M., Zalacain, M., van Thiel, A., Vitazka, P. R., Redlak, M., Andraos-Selim, C., Rastinejad, F. & Holmes, W. M. (2003). *J. Mol. Biol.* **333**, 931–949.

Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. (2010). *Acta Cryst.* D**66**, 486–501.

Engh, R. A. & Huber, R. (1991). *Acta Cryst.* A**47**, 392–400.

Forouhar, F., Shen, J., Xiao, R., Acton, T. B., Montelione, G. T. & Tong, L. (2003). *Proteins*, **53**, 329–332.

Frank, J. & Agrawal, R. K. (2000). *Nature (London)*, **406**, 318–322.

Guzman, L. M., Belin, D., Carson, M. J. & Beckwith, J. (1995). *J. Bacteriol.* **177**, 4121–4130.

Holm, L. & Sander, C. (1993). *J. Mol. Biol.* **233**, 123–138.

Johansen, S. K., Maus, C. E., Plikaytis, B. B. & Douthwaite, S. (2006). *Mol. Cell*, **23**, 173–182.

Kimura, S. & Suzuki, T. (2010). *Nucleic Acids Res.* **38**, 1341–1352.

Korostelev, A., Trakhanov, S., Laurberg, M. & Noller, H. F. (2006). *Cell*, **126**, 1065–1077.

Krissinel, E. & Henrick, K. (2007). *J. Mol. Biol.* **372**, 774–797.

Kumar, A., Saigal, K., Malhotra, K., Sinha, K. M. & Taneja, B. (2011). *J. Biol. Chem.* **286**, 19652–19661.

Leslie, A. G. W. & Powell, H. R. (2007). *Evolving Methods for Macromolecular Crystallography*, edited by R. J. Read & J. L. Sussman, pp. 41–51. Dordrecht: Springer.

Li, M., Rasulova, F., Melnikov, E. E., Rotanova, T. V., Gustchina, A., Maurizi, M. R. & Wlodawer, A. (2005). *Protein Sci.* **14**, 2895–2900.

Long, F., Vagin, A. A., Young, P. & Murshudov, G. N. (2008). *Acta Cryst.* D**64**, 125–132.

Mallam, A. L. & Jackson, S. E. (2007). *Structure*, **15**, 111–122.

Murshudov, G. N., Skubák, P., Lebedev, A. A., Pannu, N. S., Steiner, R. A., Nicholls, R. A., Winn, M. D., Long, F. & Vagin, A. A. (2011). *Acta Cryst.* D**67**, 355–367.

Nureki, O., Shirouzu, M., Hashimoto, K., Ishitani, R., Terada, T., Tamakoshi, M., Oshima, T., Chijimatsu, M., Takio, K., Vassylyev, D. G., Shibata, T., Inoue, Y., Kuramitsu, S. & Yokoyama, S. (2002). *Acta Cryst.* D**58**, 1129–1137.

Nureki, O., Watanabe, K., Fukai, S., Ishii, R., Endo, Y., Hori, H. & Yokoyama, S. (2004). *Structure*, **12**, 593–602.

Okamoto, S., Tamaru, A., Nakajima, C., Nishimura, K., Tanaka, Y., Tokuyama, S., Suzuki, Y. & Ochi, K. (2007). *Mol. Microbiol.* **63**, 1096–1106.

Painter, J. & Merritt, E. A. (2006). *Acta Cryst.* D**62**, 439–450.

Pérez-Arellano, I., Gallego, J. & Cervera, J. (2007). *FEBS J.* **274**, 4972–4984.

Pfister, P., Risch, M., Brodersen, D. E. & Böttger, E. C. (2003). *Antimicrob. Agents Chemother.* **47**, 1496–1502.

Phannachet, K. & Huang, R. H. (2004). *Nucleic Acids Res.* **32**, 1422–1429.

Qin, D., Liu, Q., Devaraj, A. & Fredrick, K. (2012). *RNA*, **18**, 485–495.

Raman, K., Yeturu, K. & Chandra, N. (2008). *BMC Syst. Biol.* **2**, 109.

Ringquist, S., Cunningham, P., Weitzmann, C., Formenoy, L., Pleij, C., Ofengand, J. & Gold, L. (1993). *J. Mol. Biol.* **234**, 14–27.

Schneidman-Duhovny, D., Inbar, Y., Polak, V., Shatsky, M., Halperin, I., Benyamini, H., Barzilai, A., Dror, O., Haspel, N., Nussinov, R. & Wolfson, H. J. (2003). *Proteins*, **52**, 107–112.

Stanley, R. E., Blaha, G., Grodzicki, R. L., Strickler, M. D. & Steitz, T. A. (2010). *Nature Struct. Mol. Biol.* **17**, 289–293.

Tkaczuk, K. L., Dunin-Horkawicz, S., Purta, E. & Bujnicki, J. M. (2007). *BMC Bioinformatics*, **8**, 73.

Watanabe, K., Nureki, O., Fukai, S., Endo, Y. & Hori, H. (2006). *J. Biol. Chem.* **281**, 34630–34639.

Watanabe, K., Nureki, O., Fukai, S., Ishii, R., Okamoto, H., Yokoyama, S., Endo, Y. & Hori, H. (2005). *J. Biol. Chem.* **280**, 10368–10377.

Winn, M. D. *et al.* (2011). *Acta Cryst.* D**67**, 235–242.

Zhang, H., Wan, H., Gao, Z.-Q., Wei, Y., Wang, W.-J., Liu, G.-F., Shtykova, E. V., Xu, J.-H. & Dong, Y.-H. (2012). *J. Mol. Biol.* **423**, 576–589.