Published online 22 February 2012 *Nucleic Acids Research, 2012, Vol. 40, No. 11* ***5149–5161***
doi:10.1093/nar/gks163

Crystal structures of the tRNA:m^{2}G6 methyltransferase Trm14/TrmN from two domains of life

Marcus Fislage[1,2], Martine Roovers[3], Irina Tuszynska[4], Janusz M. Bujnicki[4,5], Louis Droogmans[6] and Wim Versées[1,2,*]

[1]VIB Department of Structural Biology, [2]Structural Biology Brussels, Vrije Universiteit Brussel, Pleinlaan 2, 1050 Brussel, Belgium, [3]Institut de Recherches Microbiologiques Jean-Marie Wiame, Ave E. Gryson 1, 1070 Bruxelles, Belgium, [4]International Institute of Molecular and Cell Biology in Warsaw, Trojdena 4 St, 02-109 Warsaw, [5]Institute of Molecular Biology and Biotechnology, Faculty of Biology, Adam Mickiewicz University, 61-614 Poznan, Poland and [6]Laboratoire de Microbiologie, Université Libre de Bruxelles (ULB), Ave E. Gryson 1, 1070 Bruxelles, Belgium

Received September 15, 2011; Revised January 16, 2012; Accepted January 30, 2012

ABSTRACT

Methyltransferases (MTases) form a major class of tRNA-modifying enzymes needed for the proper functioning of tRNA. Recently, RNA MTases from the TrmN/Trm14 family that are present in Archaea, Bacteria and Eukaryota have been shown to specifically modify tRNA^{Phe} at guanosine 6 in the tRNA acceptor stem. Here, we report the first X-ray crystal structures of the tRNA m^{2}G6 (N^2-methylguanosine) MTase $_{TTC}$TrmN from *Thermus thermophilus* and its ortholog $_{Pf}$Trm14 from *Pyrococcus furiosus*. Structures of $_{Pf}$Trm14 were solved in complex with the methyl donor *S*-adenosyl-L-methionine (SAM or AdoMet), as well as the reaction product *S*-adenosyl-homocysteine (SAH or AdoHcy) and the inhibitor sinefungin. $_{TTC}$TrmN and $_{Pf}$Trm14 consist of an N-terminal THUMP domain fused to a catalytic Rossmann-fold MTase (RFM) domain. These results represent the first crystallographic structure analysis of proteins containing both THUMP and RFM domain, and hence provide further insight in the contribution of the THUMP domain in tRNA recognition and catalysis. Electrostatics and conservation calculations suggest a main tRNA binding surface in a groove between the THUMP domain and the MTase domain. This is further supported by a docking model of TrmN in complex with tRNA^{Phe} of *T. thermophilus* and via site-directed mutagenesis.

INTRODUCTION

RNA molecules are often highly post-transcriptionally modified, with over 100 different chemical modifications known to date. The majority and largest variety of modified nucleotides are found in transfer RNA (tRNA) (1). These modifications play different structural and functional roles and contribute to (i) the proper folding and stability of tRNA, (ii) the correct codon–anticodon recognition at the decoding center of the ribosome and (iii) the recognition of the tRNA by its cognate aminoacyl-transferase (2). Methylation constitutes by far the most abundant kind of nucleotide modification, and has been reported on the 2′-*O*-atom of ribose (3), and at various positions of the nucleotide bases on carbon and nitrogen atoms (4). These base methylations can be part of a biosynthetic pathway, leading to more complex hypermodifications that are often present in the anticodon stem–loop. Examples of such methyltransferases (MTases) that have attracted a lot of attention recently are the Trm9/Trm112 complex, involved in the biosynthesis of methoxycarbonylmethyl-5-uridine (mcm^{5}U) (5) and TYW5, involved in the synthesis of the hypermodified nucleoside wybutosine (6). On the other hand, the methyl group is often the end product of the base modification, as even simple methylations are known for their ability to stabilize the tertiary structure of tRNA (7,8).

One abundant type of methylation is the m^{2}G modification, which is well characterized at positions G10, G26 and G27 of various tRNAs (9,10). However, m^{2}G modifications are also known to exist at positions 6, 7, 9 and 18 (11). Trm11 was shown in yeast to facilitate the m^{2}G10

*To whom correspondence should be addressed. Tel: +32 26291849; Fax: +32 26291963; Email: wim.versees@vib-vub.be

modification in complex with the 'hub' protein Trm112, whereas in Archaea a single polypeptide TrmG10 is required for the same modification (12,13). Trm1 is involved in the production of m_2^2G26 (N^2,N^2-dimethylguanosine) in most eukaryotic and archaeal tRNAs and in the production of m_2^2G26 and m^2G27 in the bacterium *Aquifex aeolicus* (9,10,14).

Recently, Trm14 was identified in Archaea as the tRNA MTase that catalyzes the formation of the m^2G modification at position 6 in the acceptor stem of tRNA (15). In *Methanocaldococcus jannaschii*, it modifies $tRNA^{Cys}$ using *S*-adenosyl-L-methionine (SAM) as methyl donor. Independently, we identified this enzyme in Bacteria and showed that *Thermus thermophilus* TrmN (ortholog of the archaeal Trm14; nomenclature according to bacterial tRNA MTases) catalyzes formation of m^2G6 in $tRNA^{Phe}$ (16). Orthologs of Trm14 are also found in Eukaryota, but their activity has not been experimentally tested yet.

Known RNA MTases can be classified into four superfamilies, including Rossmann-fold (RFM), SPOUT (SpoU and TrmH), radical-SAM and FAD/NAD(p)-dependent MTases [reviewed in (17)]. RFM enzymes are the largest superfamily of MTases. They share a common structure with a seven-stranded mixed β-sheet that is a variation on the classical (di-)nucleotide binding Rossmann fold (18). Crystal structures of representatives of all four classes have been solved (19,20), while structures in complex with tRNA or a tRNA mimic are available only for representatives of the RFM and SPOUT superfamily (21,22).

While some MTases catalyze the methyl transfer reaction using a catalytic domain alone, others are fused to one of the various RNA binding domains (23). One of these domains is the THUMP domain (named after THioUridine synthase, MTase and Pseudouridine synthase), which was initially proposed to be an ancient RNA binding domain on the basis of bioinformatics analyses (24). It was proposed that the THUMP domain consists of a minimal core, which is often fused to a so-called N-terminal ferredoxin-like domain (NFLD domain) (25). Relatively, few structures of proteins containing a THUMP domain have been solved to date. The crystal structures of ThiI, which is involved in the s^4U modification, show the THUMP domain linked to a sulfur transfer catalytic domain (25), while in the cytidine deaminase CDAT8, it is fused to a deaminase domain (26). PUS10, on the other hand, has a core-THUMP domain linked to a Psi synthase domain, involved in the formation of pseudouridine (27). Examples of tRNA-modifying enzymes containing a THUMP domain fused to an RFM domain are described in literature, like Trm11 and PAB1283, involved in the m^2G10 and m_2^2G10 modification (12,13,28), and Trm14, involved in the formation of m^2G on position 6 of certain tRNAs (15). Although these proteins are well described, their structures and mode of interactions with the RNA substrates remain unknown. It has been proposed that the THUMP domain is mainly involved in the modification of nucleotides in the core of tRNA (13), but very little is known about its exact roles in (t)RNA binding and in catalysis of the modification reactions.

Here, we present the X-ray crystal structures of the bacterial tRNA:m^2G6 MTase $_{TTC}$TrmN from *Thermus thermophilus* (TTC1157) and its archaeal ortholog $_{Pf}$Trm14 from *Pyrococcus furiosus* (PF1002) in complex with the methyl donor SAM, the product *S*-adenosylhomocysteine (SAH) and the inhibitor SFG (sinefungin). This study provides the first detailed crystallographic structure analysis of tRNA-modifying enzymes consisting of an MTase domain fused to a THUMP domain. Analysis of electrostatics combined with conservation of the surface amino acids suggest a binding patch for the substrate tRNA. A docking model of tRNA with $_{TTC}$TrmN proposes the binding of tRNA in a groove between the two domains. This binding model is supported by site-directed mutagenesis in the proposed binding region.

MATERIALS AND METHODS

Protein expression and purification

$_{Pf}$Trm14 was cloned in a pET30 vector containing a C-terminal histidine tag and $_{TTC}$TrmN was cloned in a pET28 vector containing an N-terminal histidine tag, as described (29). The proteins were expressed in *Escherichia coli* Rosetta (DE3) ($_{Pf}$Trm14) or *E. coli* BL21 (DE3) cells ($_{TTC}$TrmN). Cells were initially grown at 310 K to an OD_{600} of 0.5. The strain expressing Se–Met-$_{Pf}$Trm14 was grown in minimal medium supplemented with selenomethionine (30). After induction with 0.1 mM isopropyl β-D-1-thiogalactopyranoside (IPTG), the cells grew at 288 K and were harvested the next day. Proteins were subsequently purified via nickel affinity chromatography as described earlier (29).

The $_{Pf}$Trm14 loop mutants were constructed by the use of four oligonucleotides (Supplementary Table S1), directing the desired deletion/insertion. Site-directed mutagenesis was performed according to the protocol described in the QuikChange site-directed mutagenesis kit (Stratagene). Incorparation of the mutations and integrity of the rest of the open reading frame was confirmed via sequencing. The used oligonucleotides are outlined in Supplementary Table S1.

Preparation of tRNA and *in vitro* assays

The *in vitro* transcript of $tRNA^{Phe}$ of *T. thermophilus* was generated according to a method described earlier (31). The tRNA MTase assay was based on the procedure described previously (32). The reaction mixture for the tRNA-MTase assay (400 µl) consisted of 50 mM Tris–HCl pH 8, 5 mM $MgCl_2$, 10^6 cpm of radioactive [α^{32}P] GTP $tRNA^{Phe}$ transcript, 500 µM SAM and variable amounts of purified protein. After 30 min of incubation at 60°C for $_{TTC}$TrmN or 70°C for $_{Pf}$Trm14, the reaction was stopped by phenol extraction and the tRNA was ethanol precipitated. The recovered radioactive tRNA was then digested completely by nuclease P1 (1 µg), in the presence of 5 µg total yeast tRNA as carrier. Conversion of pG to pm^2G was analyzed by 1D-thin layer chromatography (TLC) on cellulose plates (Merck) in solvent A (isobutyric acid/concentrated NH_4OH/water;

66/1/33; v/v/v). The migration pattern was visualized by autoradiography.

For band shift assays, 10^4 cpm of radioactive [α^{32}P] GTP $tRNA^{Phe}$ transcript from *T. thermophilus* was incubated in the presence of increasing amounts of enzyme in buffer B (20% glycerol, 50 mM Tris pH 8) at 60°C for $_{TTC}$TrmN or 70°C for $_{Pf}$Trm14 for 30 min. The binding reaction (total volume of 20 µl) was stopped by the addition of 2 µl of stop solution (0.05% bromophenol blue in 30% glycerol) and the mixture was separated by 6% PAGE (190 mm × 160 mm × 1.5 mm) in TB buffer at room temperature. The PAGE was subjected to a voltage of 180 V until the samples entered the gel and further at 150 V until the end of the run (~2 h).

Crystallization, data collection and structure determination

Wild-type unlabeled $_{Pf}$Trm14 was crystallized in a buffer consisting of 50 mM Tris–HCl pH 8, 10 mM $MgCl_2$, 500 mM NaCl, 280 mM imidazole, 1 mM DTT by 1 : 1 mixing with crystallization solution (100 mM Tris–acetate pH 8, 32% polyethylene glycol (PEG) 4000, 15% glycerol) in a hanging drop vapor diffusion setup. Wild-type unlabeled $_{TTC}$TrmN was crystallized in a buffer consisting of 50 mM Tris–HCl pH 8, 250 mM NaCl and 350 mM imidazole by 1 : 1 mixing with crystallization solution (100 mM citrate/phosphate pH 3.5, 15% PEG 6000, 200 mM NaCl, 100 mM sodium citrate) in a hanging drop vapor diffusion setup (29). For the phasing, selenomethionine (Se–Met)-derivatized $_{Pf}$Trm14 was crystallized in a crystallization solution containing 100 mM Tris–acetate pH 8, 32% PEG 4000, 15% glycerol, after streak seeding from a native $_{Pf}$Trm14 crystal. All crystals were flash frozen in liquid nitrogen using either the crystallization buffer ($_{Pf}$Trm14) or using crystallization buffer containing 20% glycerol as cryoprotectant ($_{TTC}$TrmN). Crystals of protein–ligand complexes of $_{Pf}$Trm14 were obtained by overnight soaking, using mother liquor containing 1 mM SFG or 1 mM SAM. After 24-h incubation crystals were flash frozen in liquid nitrogen using the mother liquor containing 20% glycerol as cryoprotectant.

The diffraction data of Se–Met-derivatized $_{Pf}$Trm14 were collected at 100 K at the ID14-4 beamline (ESRF, Grenoble) using a single wavelength of 0.97936 Å, corresponding to the selenium absorption peak. The statistics of the data collection and processing are summarized in Table 1. The raw data were processed and scaled using the XDS suite (33). The structure was solved using the SAS protocol of Auto-Rickshaw: the EMBL-Hamburg automated crystal structure determination platform (34). The input diffraction data were prepared and converted for use in Auto-Rickshaw using programs of the CCP4 suite (35). Heavy atom structure factor (FA) values were calculated using the program SHELXC (36). Twelve heavy atoms were found using the program SHELXD (37). The correct hand for the substructure was determined using the programs ABS (38) and SHELXE (39). Initial phases were then calculated after density modification using the program SHELXE (39). The initial phases were improved using density modification and phase extension using the program DM (40). The model was partially built using the program ARP/wARP (41,42). Further model building was done manually using Coot (43) and the structure refinement was carried out using refmac5 (44).

Diffraction data of native $_{Pf}$Trm14 either non-soaked (containing endogenous SAH) or soaked with SFG and SAM were collected at 100 K at the ID23-2 (ESRF, Grenoble) or PX-III (SLS, Villingen) beamlines, respectively. The diffraction data of the native $_{TTC}$TrmN were collected at the ID23-2 beamline (ESRF, Grenoble). The statistics of the data collection and processing are summarized in Table 1. The $_{Pf}$Trm14-SAH, -SFG, -SAM and $_{TTC}$TrmN structures were solved by molecular replacement using PhaserMR (45). In the case of the $_{TTC}$TrmN structure, the individual THUMP (without the residues 47–63), and RFM domains of the Se–Met $_{Pf}$Trm14 structure were used as search models in molecular replacement. In the case of the soaked $_{Pf}$Trm14 structures, the full-length Se–Met $_{Pf}$Trm14 structure was used as a search model. Model building and refinement were performed as for the Se–Met $_{Pf}$Trm14 structure. TLS refinement was implemented in the refinement protocol for $_{TTC}$TrmN, using four individual TLS groups determined by TLSMD (46,47). Stereochemical validation of all models was done using the Molprobity server (48). The figures were prepared using Pymol (http://www.pymol.org). The statistics of structure refinement and final structure models are summarized in Table 1.

Computational analysis, modeling of nucleotide binding and tRNA docking

Electrostatic surface calculations were prepared with PDB2PQR (49) using the PARSE force field and APBS (50). Amino acid conservation was calculated through the ConSurf server (51). Omit maps were generated using omit (52). Modeling of SAM into the active site pocket of $_{TTC}$TrmN and of guanosine into the $_{Pf}$Trm14-SAM and $_{TTC}$TrmN structures were done using the program Epitope Match (53). For the modeling of SAM, $_{Pf}$Trm14 was used as epitope model. For the modeling of guanosine into the $_{Pf}$Trm14-SAM structure and $_{TTC}$TrmN structure, the crystal structure of RsmC (PDB code 3DMH) was used as epitope model (54).

In order to predict the structure of the $_{TTC}$TrmN-$tRNA^{Phe}$ complex, we used the computational docking method. Since an unbound structure of *T. thermophilus* $tRNA^{Phe}$ is not available, we have built its homology model with ModeRNA (55) using a structure of *E. coli* $tRNA^{Phe}$ as a template (PDB code 3L0U). In a separate docking experiment, we have also used the coordinates of *T. thermophilus* $tRNA^{Phe}$ taken from the cocrystal structure with the cognate phenylalanyl-tRNA synthetase (PDB code 2IY5). For prediction of the complex structure, we used a combination of procedures described in the references (28,56). Briefly, a low-resolution method GRAMM (57) was used to generate 30 000 alternative models (decoys) with physically reasonable geometric compatibility between protein and RNA structures. We have filtered these decoys using FILTREST3D (58) to retain structures

Table 1. Data-collection, refinement and validation statistics of the structures of $_{Pf}$Trm14 and $_{TTC}$TrmN

Data set	$_{Pf}$Trm14				$_{TTC}$TrmN Apo form
	Se-SAD peak	SAH bound	SAM bound	SFG bound	
Data-collection					
X-ray source	ESRF ID 14-4	ESRF ID23-2	SLS PX-III	SLS PX-III	ESRF ID23-2
X-ray wavelength (Å)	0.97936	0.8726	1.00150	0.98000	0.8726
Temperature (K)	100	100	100	100	100
Space group	$P2_1$	$P2_1$	$P2_1$	$P2_1$	$P3_221$
Unit-cell parameter (Å, °)	a = 82.2, b = 45.0, c = 120.7, β = 91.3	a = 85.5, b = 45.3, c = 122.2, β = 91.2	a = 82.1, b = 45.2, c = 121.4, β = 91.8	a = 82.3, b = 45.1, c = 121.1, β = 91.8	$a = b$ = 65.9, c = 144.6
Resolution range (Å)	50–2.3 (2.4–2.3)	50–2.2 (2.3–2.2)	50–1.95 (2.0–1.95)	50–2.27 (2.4–2.27)	50–2.05 (2.16–2.05)
Total/unique reflections	581 156/77 726	162 685/46 452	462 285/65 283	306 409/79 652	502 157/23 534
R_{merge} (%)[a]	11.4 (61.2)	10.8 (57.7)	7.3 (54.4)	14.7 (64.3)	10.1 (52.5)
R_{meas} (%)[b]	12.3 (65.7)	12.7 (68.4)	7.8 (59.1)	17.1 (74.8)	10.4 (53.7)
Data completeness (%)	99.7 (99.3)	99.7 (99.7)	99.5 (98.9)	98.7 (92.3)	100 (100)
Average I/σ	14.0 (3.8)	9.3 (2.7)	18.3 (3.6)	8.6 (2.1)	22.4 (8.1)
Redundancy	7.4 (7.6)	3.5 (3.5)	7.1 (6.6)	3.8 (3.7)	21.3 (22.1)
Refinement					
R_{work}/R_{free}[c]		18.6/22.0	16.5/19.7	20.5/21.9	15.7/21.2
Total number					
Amino acid residues		728	728	735	332
Water molecules		373	383	253	196
Ligand atoms		56 (2 × SAH, 1 × ACT)	54 (2 × SAM)	54 (2 × SFG)	5 (PO4)
rmsd					
Bond length (Å)		0.015	0.015	0.015	0.016
Bond angles (°)		1.6	1.4	1.6	1.5
Ramachandran Plot (%)					
Favored regions		98.1	98.3	98.1	97.9
Allowed regions		1.9	1.7	1.9	1.8
Disallowed regions		0	0	0	0.3 (Glu88A)
PDB code		3TLJ	3TM4	3TM5	3TMA

Number in parentheses are statistics in the highest resolution shell.

[a] $R_{merge} = \sum\sum\left(\left|I_i(h)\right| - \left|\langle I(h)\rangle\right|\right)/\sum\sum I_i(h)$.

[b] $R_{meas} = \sum\sqrt{n_h/(n_h-1)}\sum\sum\left(\left|I_i(h)\right| - \left|\langle I(h)\rangle\right|\right)/\sum\sum I_i(h)$.

[c] $R_{work} = \sum \left|F(h)_o\right| - \left|F(h)_c\right| / \sum \left|F(h)_o\right|$, $F(h)_o$ and $F(h)_c$ are observed and calculated structure factor amplitudes, respectively. A random subset of data (5%) was used for the R_{free} calculation

with a distance between the target residue G6 in tRNA and the methyl group of SAM in the protein <20 Å. These selected decoys were scored with the Decoys As the Reference State (DARS-RNP) potential (56) and clustered according to their mutual similarity, to retain groups of very similar decoys. The overall best-scored complex, as well as a representative of the largest cluster of well-scored decoys, was selected for further consideration.

RESULTS AND DISCUSSION

Overall structure of *P. furiosus* Trm14 and *T. thermophilus* TrmN

The structure of $_{Pf}$Trm14 was solved by single wavelength anomalous dispersion (SAD) using selenomethionine-labeled protein and was refined to 2.3 Å resolution. With this initial structure, the structures of $_{Pf}$Trm14 in complex with SAH, SFG and SAM were solved by molecular replacement to a resolution of 2.2 Å, 1.95 Å and 2.27 Å, respectively. While the substrate SAM and the inhibitor SFG were soaked into the crystals, the product SAH co-purified with $_{Pf}$Trm14. $_{Pf}$Trm14 was crystallized in the space group $P2_1$ with two protein molecules per asymmetric unit related by a 2-fold non-crystallographic symmetry axis. The contact interface of the two molecules within the asymmetric unit only buries a solvent accessible surface area of 432 Å^2, indicating that these molecules do not correspond to a biological dimer (59). Additionally, the generation of all possible symmetry-related molecules in the crystal did not result in a biologically significant interface (biggest interface area: 515 Å^2). This matches with the results of a size exclusion chromatography experiment, which clearly indicates that the protein behaves as a monomer in solution (Supplementary Figure S1). Overall, the three $_{Pf}$Trm14 structures are highly similar with an rmsd of 0.4 Å (calculated over 730 backbone atoms) between the SAM- and SAH-bound structures and an rmsd of 1.2 Å or 1.3 Å (over 727 backbone atoms) between the SFG bound and SAH- or SAM-bound structures, respectively. Clear electron density is visible for most of the 367 amino acids, and some amino acids from the C-terminal his-tag are also visible. One loop spanning the residues 298–303 is highly flexible in

the $_{Pf}$Trm14-SAH and $_{Pf}$Trm14-SAM structures and could not be modeled. These residues are located in the RFM domain of $_{Pf}$Trm14, lining the SAM-binding pocket. Interestingly, this region became rigid when SFG was bound in the active site (see further).

The structure of $_{TTC}$TrmN was solved in the apo state by molecular replacement to a resolution of 2.05 Å. $_{TTC}$TrmN was crystallized in the space group *$P3_221$* with one monomer per asymmetric unit. Application of crystal symmetry operators yielded a symmetrical dimeric arrangement, with an accessible surface area buried in the interface of 799 Å^2 (about 5.1% of the total accessible surface area). This value is just below the threshold area used for biological relevant interfaces (59). Moreover, this interface is highly hydrated and made up of completely non-conserved residues, indicating that this interface is formed by crystal packing. Size exclusion experiments with the wt $_{TTC}$TrmN are not conclusive, since this protein interacts with the column matrix and hence elutes at a volume corresponding to a smaller molecular weight (Supplementary Figure S1). This can be partially circumvented by decreasing the high positive charge density of the protein and the K270E/R300E variant of $_{TTC}$TrmN (see further) elutes at a volume corresponding to a monomer. Together this suggests that $_{TTC}$TrmN, like $_{Pf}$Trm14, is a monomer in solution. In the $_{TTC}$TrmN structure, the same loop as in the $_{Pf}$Trm14 structure (amino acids 266–269) is disordered and did not show any electron density. Details of the data collection and refinement statistics are shown in Table 1.

Overall, the structures of $_{Pf}$Trm14 and $_{TTC}$TrmN are very similar (Figures 1 and 2) with an rmsd of 2.0 Å, calculated over all main chain atoms (286 residues). Globally, $_{Pf}$Trm14 and $_{TTC}$TrmN adopt a cylindrical shape with an approximate size of $66.0 \times 35 \times 35$ Å^3 (Figure 2). The structures consist of two globular domains. An N-terminal THUMP domain spanning approximately the residues 1–183 in $_{Pf}$Trm14 and 1–151 in $_{TTC}$TrmN, and a C-terminal RFM domain spanning approximately the residues 193–367 in $_{Pf}$Trm14 and the residues 160–335 in $_{TTC}$TrmN. Both domains are connected via a long linker containing several positively charged amino acids. The orientation of the THUMP and RFM domains toward each other is similar in $_{Pf}$Trm14 and $_{TTC}$TrmN.

Structural similarity searches with PDBeFold (http://www.ebi.ac.uk/msd-srv/ssm/) (60) revealed three protein structures that align over nearly the total length with the $_{TTC}$TrmN protein (Supplementary Table S2): a possible methylase from *C. difficile* (PDB code 3LDU, Tan,K., Wu,R., Buck,K. and Joachimiak, A., unpublished data), a predicted N6 adenine DNA methylase from *L. monocytogenes* (PDB code 3K0B, Nocek,B., Xu,X., Cui,H., Ng,J., Savchenko,A., Edwards,A. and Joachimiak,A., unpublished data) and the putative MTase SMU.422 from *Streptococcus mutans* (PDB code 3LDG, Wang,K.-T. and Su,X.-D., unpublished data). These three proteins belong to the 'putative RNA methylase family UPF0020' (PFAM nomenclature) and exhibit the same domain architecture and orientation as TrmN/Trm14. No publications describing these structures or their function was available at the time of the writing of this article. Moreover, the surface-exposed residues in the THUMP domain exhibit relatively low similarity to $_{TTC}$TrmN, suggesting that substrate specificity is not necessarily conserved. For these three structures, it should be noted that they contain an insertion of two α-helices in their MTase domain compared to the MTase domain of $_{TTC}$TrmN and $_{Pf}$Trm14, between α-helix 7 and β-strand 9 (corresponding to TrmN/Trm14 numbering).

The PDBeFold server was also used to search for structural similarities of the isolated THUMP and RFM domains of $_{Pf}$Trm14 and $_{TTC}$TrmN. The RFM domain of $_{TTC}$TrmN is not particularly similar to any experimentally characterized enzyme other than the three structures mentioned above. Instead, it exhibits highest similarity to various RFM domains of putative MTases with experimentally uncharacterized functions and unknown substrate specificities (Supplementary Table S2). These RFM domains are often fused to other domains, unrelated to THUMP. The THUMP domain is generally less conserved than the RFM domain, which is reflected in the relatively lower *Z*-scores of the top matches (Supplementary Table S2). The closest homologs of the THUMP domain of $_{Pf}$Trm14 and $_{TTC}$TrmN (with or without NFLD subdomain, see further) are proteins involved in different reactions and fused to other catalytic domains, like the ThiI enzyme involved in the formation of the s^4U modification in tRNA (25). Two RNA modification enzymes with known structures and possessing a THUMP domain, human pseudouridine synthase Pus10 (PDB code 2V9K) (27) and archaeal cytidine deaminase CDAT8 (PDB code 3G8Q) (26) were not reported in PDBeFold searches. CDAT8 possesses both the NFLD and core-THUMP subdomains of the THUMP domain (see further), while Pus10 lacks the NFLD subdomain. ThiI, Pus10 and CDAT8 possess all unrelated catalytic domains fused to the THUMP domain. The superposition of these structures onto $_{TTC}$TrmN using the core-THUMP subdomains reveals that the catalytic domains assume different orientations (Supplementary Figure S2). Only the unrelated catalytic domains of $_{TTC}$TrmN and ThiI assume relatively similar orientations with respect to the THUMP domain. At first sight, this might be related to the fact that $_{TTC}$TrmN and ThiI modify bases at similar positions in the tRNA (position 8 in ThiI versus position 6 in $_{TTC}$TrmN). However, also CDAT8 targets cytosine at position 8 but does not show the same domain orientation.

The THUMP domains of $_{Pf}$Trm14 and $_{TTC}$TrmN differ by the insertion of an anti-parallel β-sheet

The N-terminal THUMP domain of $_{TTC}$TrmN adopts a globular α/β structure with a total of seven β-strands and four α-helices (Figure 2B and D) of which helix 3 is kinked. The seven β-strands of this domain are arranged in two β-sheets that are connected by one long kinked β-strand that continues from one β-sheet to the other. The mixed β-sheet, consisting of the four C-terminal β-strands together with the two α-helices that are packed to one side of this β-sheet, corresponds to the core-THUMP domain as initially predicted by Aravind

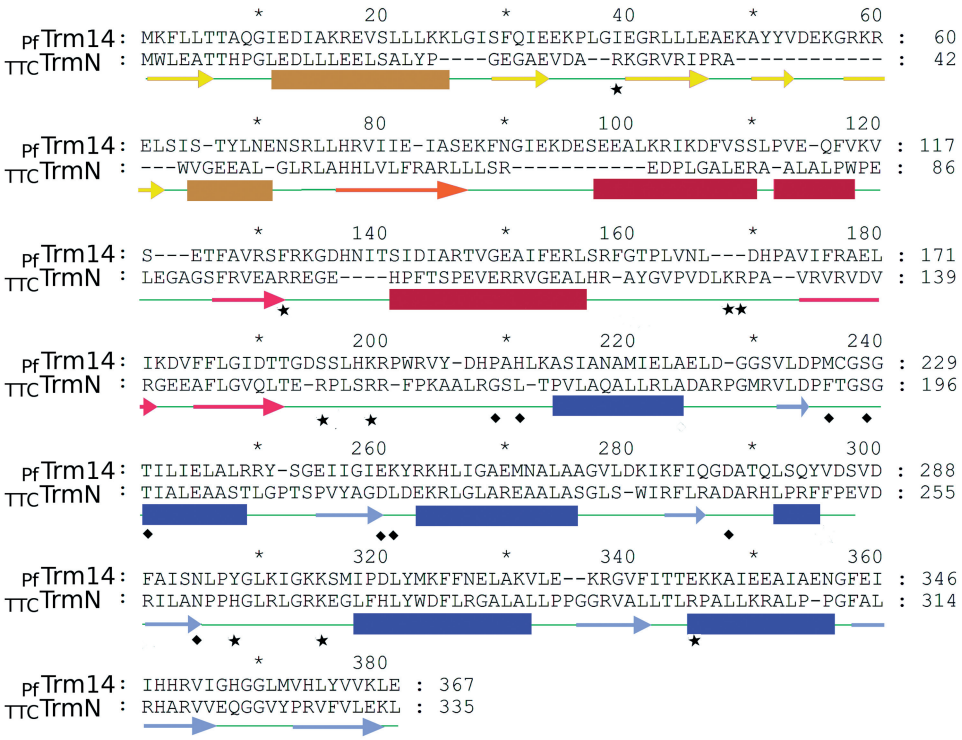

Figure 1. Sequence alignment of $_{Pf}$Trm14 and $_{TTC}$TrmN with the secondary structures, as deduced from the crystal structures, indicated below the sequence. Since the sequence identity of the two THUMP domains is rather low (22%), the sequence alignment was entirely based on the superposition of the two structures using CCP4 superpose. β-strands are indicated by lighter colored arrows, α-helices by darker colored bars. Secondary structures of the RFM domain are shown in blue. Secondary structures of the THUMP domain are shown in red for the core-THUMP subdomain and in yellow for the NFLD subdomain. The β-strand shared by these two subdomains is shown in orange. Squares depict residues in the active site of the methyltransferase domain interacting with the ligand. Residues that were mutated in this study are indicated by stars.

and Koonin (24). This core-THUMP domain is intimately associated with the N-terminal four-stranded anti-parallel β-sheet. This latter β-sheet, together with two α-helices, has previously been referred to as the NFLD domain (based on a topology similar to the ferredoxin-like fold) (25,26). However, from this and previous structures of the THUMP domain, it appears that both regions are subdomains of one globular domain fold and therefore both domains could also be regarded as one autonomously foldable THUMP domain (25,26,28).

The THUMP domains of $_{Pf}$Trm14 and $_{TTC}$TrmN are structurally very similar (backbone rmsd of 2.3 Å) and they are also very similar to the THUMP domains of ThiI (rmsd of 2.7 Å and 3.2 Å for superposition on $_{TTC}$TrmN and $_{Pf}$Trm14, respectively) and the cytidine deaminase CDAT8 (rmsd of 2.8 Å and 2.9 Å for superposition on $_{TTC}$TrmN and $_{Pf}$Trm14, respectively) (25,26).

One minor difference between the THUMP domains from Trm14 and TrmN, and also with the THUMP domains of ThiI and CDAT8, concerns the angle relating the β-sheets of the core-THUMP subdomain and the NFLD subdomain. These two β-sheets are somewhat more perpendicular to each other in $_{Pf}$Trm14 compared to $_{TTC}$TrmN and the THUMP domains of ThiI and CDAT8. Whether or not this change is relevant to the fine-tuning of the interaction with the tRNA substrate remains to be investigated.

Another obvious difference of the THUMP domain of $_{Pf}$Trm14 in comparison to $_{TTC}$TrmN and other known THUMP domains concerns an insertion of two anti-parallel β-strands between β-strand 3 and helix 2 of the NFLD subdomain of the THUMP domain of $_{Pf}$Trm14 (Figure 2A and C). This two-stranded anti-parallel β-sheet is extending away from the THUMP domain on the opposite side of the presumed tRNA binding surface and the SAM binding pocket of the RFM domain (see further). Sequence analysis using a BLAST search (61) showed that this insertion is conserved only in Archaea, with over 50 matches found. No comparable insertions were found in THUMP domains of Bacteria and Eukaryota. Interestingly, these two β-strands are made up mainly of conserved positively charged (five lysines or arginines) and aromatic residues (Y51, Y52), resulting in one side of the β-sheet having a positively charged surface (Figure 4A). To investigate a possible role of these two β-strands in tRNA binding and catalysis of the m^2G6 modification, we made two different deletion variants of $_{Pf}$Trm14. In a first variant, the peptide region spanning the inserted β-sheet (residues Y51 to E61) was replaced by two glycines that should be sufficient to span the distance between β-strand 3 and helix 2 of the THUMP domain ($_{Pf}$Trm14_2G variant). In a second variant, this region was replaced by the corresponding loop region of $_{TTC}$TrmN (R41 to G45)

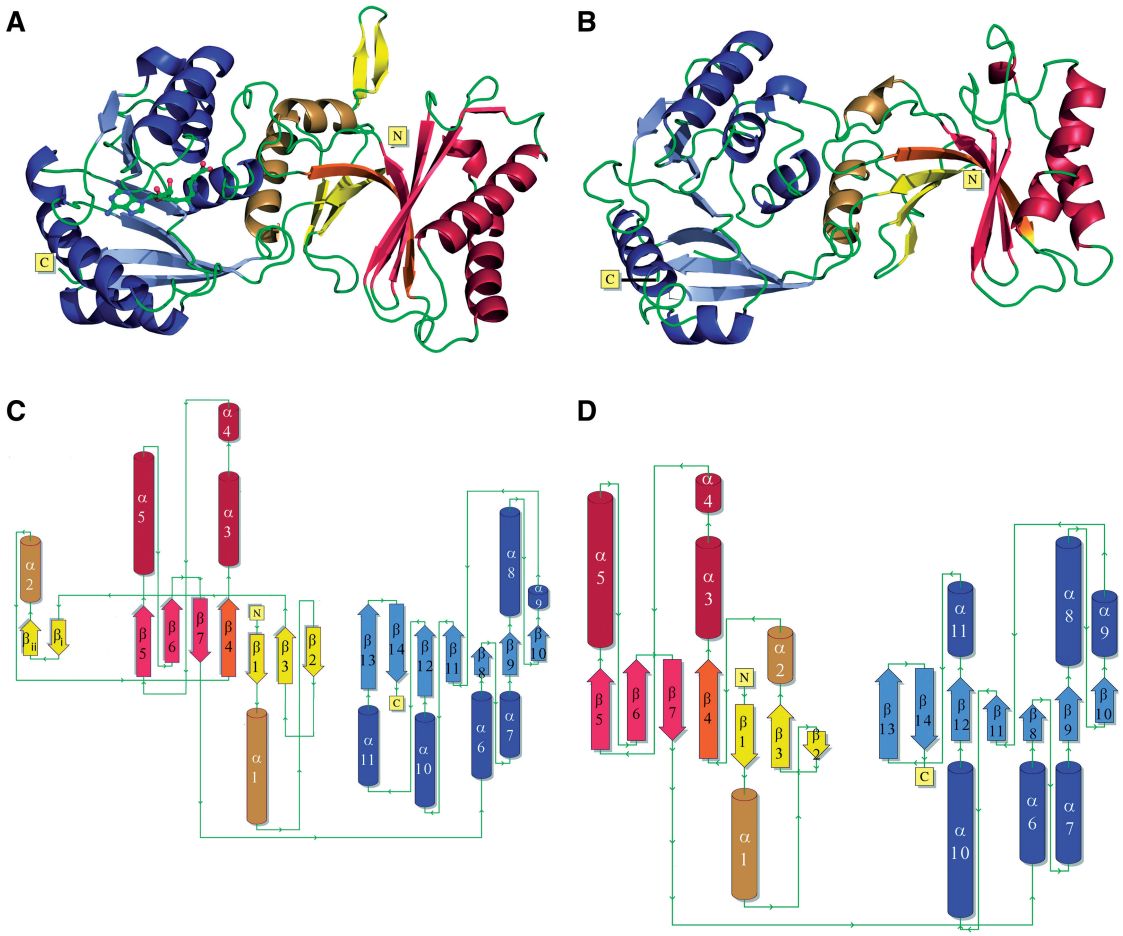

Figure 2. Crystal structures of $_{Pf}$Trm14 and $_{TTC}$TrmN. (**A** and **C**) show the structure of $_{Pf}$Trm14 in cartoon representation (A) and as a topology diagram (C); (**B** and **D**) show the structure of $_{TTC}$TrmN in cartoon representation (B) and as a topology diagram (D). β-strands are shown in red, yellow and blue for the core-THUMP, NFLD and RFM (sub)domains, respectively. α-Helices are shown in dark red, yellow and blue. The β-strand shared by the core-THUMP and NFLD subdomains is shown in orange. The two additional β-strands in the $_{Pf}$Trm14 structure are named β_i and β_{ii}. The structure of $_{Pf}$Trm14 contains sinefungin bound in the active site, represented in ball and stick.

($_{Pf}$Trm14_TTC variant). The G6 MTase activity and tRNA binding affinity of both variants were compared with the wild-type $_{Pf}$Trm14 (Supplementary Figure S3A and B). Interestingly, only small changes in affinity towards tRNA and in the catalytic rate were observed *in vitro* upon deletion. This might suggest that the β-strands are a conserved ornament of the basic NFLD fold or a relic of a previous activity. Alternatively, this element could be involved in another, yet undiscovered, function of archaeal Trm14.

Substrate binding and ligand-induced conformational changes in the RFM domain

The C-terminal MTase domains of $_{Pf}$Trm14 and $_{TTC}$TrmN both adopt a topology that is typical for the RFM superfamily of MTases (Figure 2C and D). Here, the C-terminal (7th) β-strand of a central seven-stranded β-sheets is inserted in an anti-parallel way between strands 5 and 6 of the sheet, and the β-sheet is packed on both sides by a total of six α-helices. The RFM domains of $_{TTC}$TrmN and $_{Pf}$Trm14 are very similar and superpose with an rmsd of 1.3 Å (using the main chain atoms of 155 residues).

While the $_{TTC}$TrmN structure was solved in the apo form, the $_{Pf}$Trm14 was solved with the RFM domain either bound to SAH ($_{Pf}$Trm14-SAH), SAM ($_{Pf}$Trm14-SAM) or SFG ($_{Pf}$Trm14-SFG). One loop region, connecting β-strand 4 and helix 5, is highly flexible and could not be observed in the electron density of the crystal structures of $_{TTC}$TrmN and $_{Pf}$Trm14-SAH and $_{Pf}$Trm14-SAM. Only when SFG was bound to $_{Pf}$Trm14 the loop was structured and became visible in the density map. Since both SAM and SFG were soaked into pre-existing crystals of $_{Pf}$Trm14, this loop movement is probably not caused by crystal packing interactions, but concerns a ligand-induced conformational change. Although SAM and SFG have the same net charge (Supplementary Figure S4), only SFG seems to evoke this conformational change. This might be due to the position of the positive charge, which is located on the S^{γ} in SAM and on the N^{ε} in SFG (62). Assuming a dissociative S_N2 reaction mechanism, one could speculate that the positive charge on the ε position would mimic an intermediate along the methyl

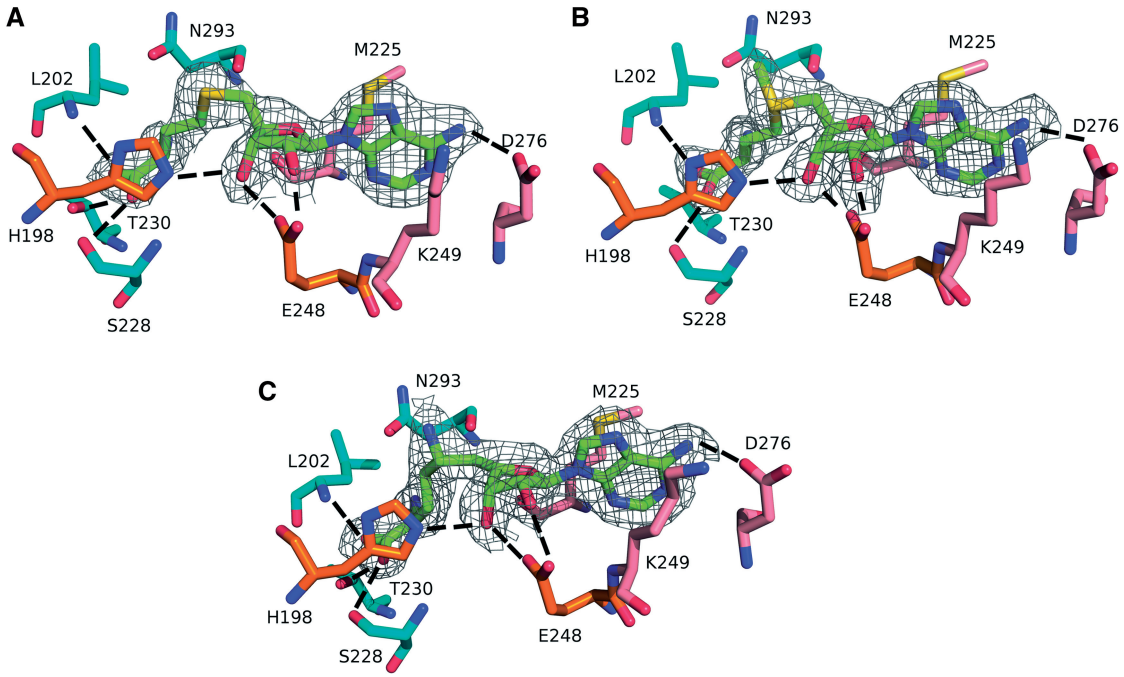

Figure 3. Active site of $_{Pf}$Trm14 bound to (**A**) SAH, (**B**) SAM and (**C**) SFG. The $2F_o - F_c$ omit density map is shown for the ligands. The map is contoured at 2 σ within 1.6 Å of the ligand. The ligands are shown with their C atoms colored green, the amino acid residues are shown with their C atoms colored pink (interacting with adenine), orange (interacting with ribose) or cyan (interacting with the homocysteine part). Hydrogen bonds are indicated by dashed lines.

transfer reaction pathway where the S^{γ}-C^{ε} bond in SAM is partially broken. Also the large loop region, connecting the THUMP domain to the first α-helix of the RFM domain, interacts with the methionine moiety of the bound co-factor and this loop undergoes some structural rearrangements in the SFG-bound form, moving the backbone 1.5 Å away from the ligand.

The adenosine moieties of the ligands SAH, SAM and SFG are bound to $_{Pf}$Trm14 in a pocket located at the first β–α–β–α–β motif of the Rossmann-fold and the cross-over towards the second β–α–β motif, as is often seen in other nucleotide binding proteins (63). The pocket for the methionine moiety is formed by the loop region connecting the THUMP domain to the RFM domain together with the first two α-helices of the RFM domain. Omit maps of the binding pocket of $_{Pf}$Trm14 (Figure 3) show that the ligand SAM is bound in an extended conformation (dihedral C4′, C5′, S, Cγ = −70.7°) which is common for SAM-dependent MTases (4,62). The reaction product SAH is also bound in an extended conformation but with a dihedral angle of −34.6°. The dihedral angle of the bound inhibitor SFG is −101.1°.

The SAM-binding cleft is mainly lined by residues belonging to the conserved motifs I, II, III, IV, V and X that exist in related N-MTases (9,62,64). The three ligands SAH, SAM and SFG are engaged in a similar set of hydrogen-bonding and van der Waals interactions with $_{Pf}$Trm14. Since $_{TTC}$TrmN was solved in the apo state only, modeling of SAM into the $_{TTC}$TrmN structure was performed using Epitope Match (rmsd 1.7 Å, 26 residues in the epitope) (Supplementary Figure S5). The model shows that the binding pocket is conserved in $_{TTC}$TrmN and chemically related residues are involved in the binding of SAM.

SAH is bound to the RFM domain by a total of 10 hydrogen bonds and several van der Waals interactions (Figure 3A). The adenine moiety of the co-factor is stacked between the side chains of Met225 and Lys249 of $_{Pf}$Trm14 using van der Waals interactions. The N6 exocyclic amine of the adenine base forms a hydrogen bond to Asp276 (Asp243 in $_{TTC}$TrmN), while its N1 is within hydrogen bonding distance to the main chain amine of Ala277. Asp276 is highly conserved and is part of the MTase motif III (P/G,QTXD$_{276}$AXXC/C/L), while Met225 is part of motif I (M$_{225}$XGXG).

The ribose moiety of the co-factor is hydrogen bonded with both its 2′- and 3′ hydroxyls to the side chain of the highly conserved Glu248 of motif II (Asp216 in $_{TTC}$TrmN). The 3′-OH is also involved in a hydrogen bond with His198.

The homocysteine/methionine moieties of SAH and SAM interact via their α-NH_3^+ with Asn293 (part of motif IV) and via their α-COO^- with Ser228 (side and main chain) and Thr230 which are located in or next to motif I, and with the main chain of Leu202, located in the proximity of motif X.

Motif IV is part of a possible guanosine binding pocket

The typical motif IV of RFM enzymes acting on exocyclic amine groups consists of the consensus sequence D/N-P/L-P-Y/F/H (65). In $_{Pf}$Trm14 and $_{TTC}$TrmN, this corresponds to N$_{293}$LPY and N$_{260}$PPH, respectively.

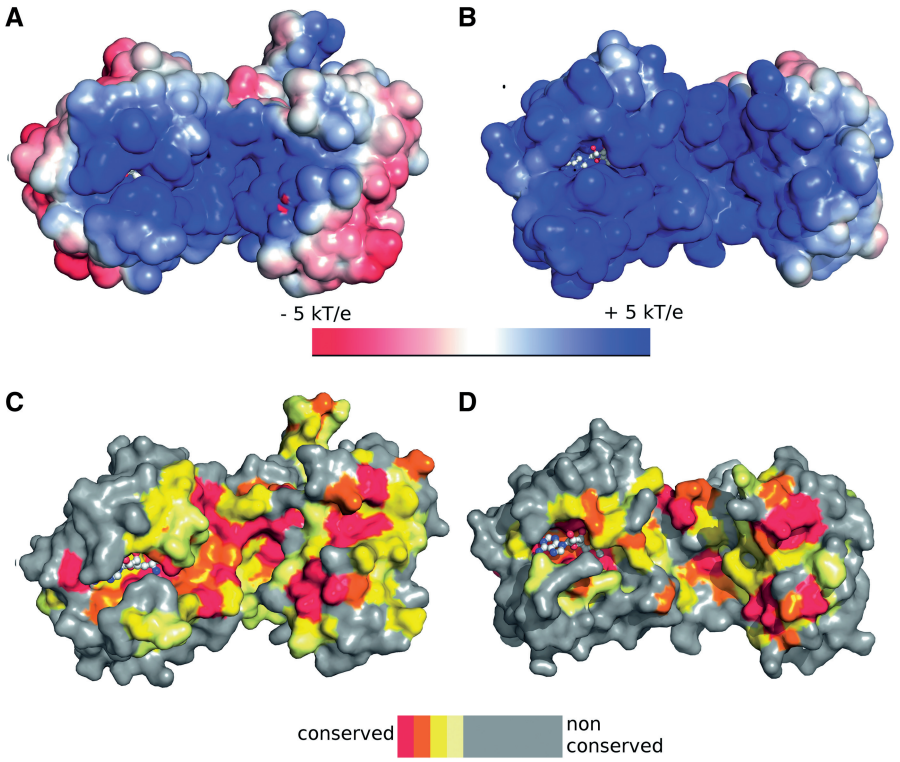

Figure 4. (**A** and **B**) Electrostatic surface potential mapped on the solvent accessible surface of $_{Pf}$Trm14 (A) and $_{TTC}$TrmN (B). k_b = Boltzmann's constant, T = temperature, e_c = charge of an electron. (**C** and **D**) Mapping of conservation scores of amino acids on the surface of $_{Pf}$Trm14 (C) and $_{TTC}$TrmN (D). The color code of the conservation score is indicated in the legend. The bound SFG ($_{Pf}$Trm14) and modeled SAM ($_{TTC}$TrmN) in the active site of the RFM domain are shown in ball and stick representation.

These residues are located right next to the flexible loop that is missing in our structures and line a pocket that leads towards the $^{\varepsilon}$CH3 of SAM. Asn293 is located in close proximity to the methyl group of SAM (3.9 Å). This pocket will likely accommodate the G6 of the tRNA that is modified by $_{Pf}$Trm14 and $_{TTC}$TrmN and provide the catalytic residues for the methyl transfer reaction. A crystal structure of the 16 S rRNA MTase RsmC in complex with SAM and guanine indeed shows that the residues of motif IV interact with the guanine substrate (54). In the latter structure, Phe308 (of the N_{305}PPF motif) interacts via base stacking with guanine, while Asn305 forms a hydrogen bond with the exocyclic amine group on position 2.

In order to confirm that essentially the same binding pocket for the G6 nucleoside exists in $_{Pf}$Trm14 and $_{TTC}$TrmN, an epitope match was performed using the program Epitope Match (53) and using the guanosine binding epitope of the RsmC-guanosine crystal structure (54). For $_{Pf}$Trm14 as well as $_{TTC}$TrmN, a possible binding pocket was found that could accommodate a guanosine (corresponding to G6) with only minor structural rearrangements (Supplementary Figure S6). In the case of $_{TTC}$TrmN, the epitope contained 13 residues with an rmsd of 2.7 Å. As expected, His263 of motif IV is in our model stacked with the guanosine base, although a small movement of this residue upon tRNA binding is necessary for perfect accommodation of guanosine. Such movement would be possible, since His263 is located next to the highly flexible loop of $_{Pf}$Trm14. The catalytic Asn263 residue (of the same motif IV) is in our model located in between the N2 atom of guanosine and the methyl group of SAM, bridging a distance of 4 Å between methyl donor and acceptor. In $_{Pf}$Trm14, the epitope contained 13 residues with an rmsd of 2.8 Å. In this case, Tyr296 of motif IV was involved in base stacking to the ribose, as proposed for the corresponding His of $_{TTC}$TrmN. The catalytic Asn293 is again located between the methyl donor (SAM) and the methyl acceptor (N2 of G6), bridging a distance of 4.9 Å between the donor and the acceptor.

To confirm the catalytic role of this motif in Trm14/TrmN, His263 of $_{TTC}$TrmN was substituted by an alanine. In agreement with the proposed importance of the His263 residue, the tRNA MTase activity of the $_{TTC}$TrmN(H263A) variant was decreased to 25% of the wild-type activity. The decrease in activity must be related to the accommodation of G6 in the catalytic pocket, since the affinity towards tRNA was unchanged upon introduction of the point substitution (Supplementary Figure S3C and D).

The tRNA substrate binds on the interface of the RFM and THUMP domains

The surface electrostatics of $_{TTC}$TrmN and $_{Pf}$Trm14 show a very large patch of positively charged residues leading

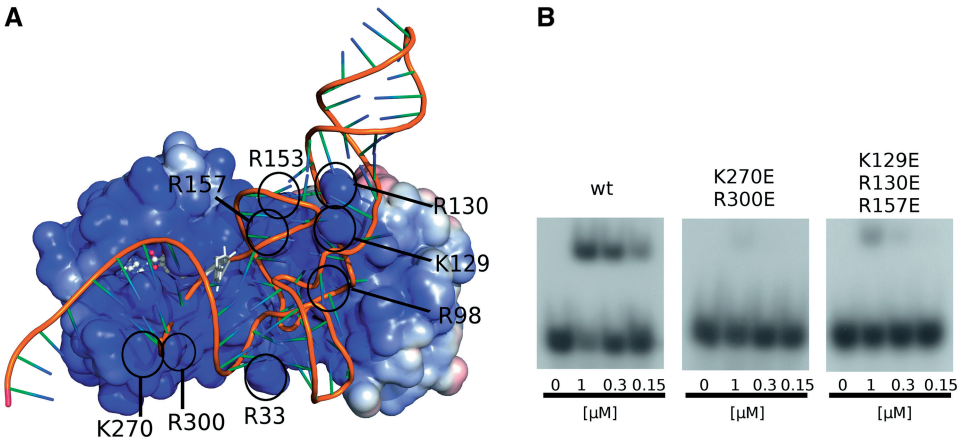

Figure 5. (**A**) Docking model of $tRNA^{Phe}$ of *T. thermophilus* onto $_{TTC}$TrmN. $_{TTC}$TrmN is shown with the electrostatic potential mapped on its solvent accessibility surface, and the tRNA (a homology model based on the orthologous $tRNA^{Phe}$ from *E. coli*) is shown in cartoon style. The guanosine on position 6 of the tRNA molecule and SAM, bound to the active site of the RFM domain, are shown in ball-and-stick and stick representation, respectively. Black circles indicate the position of residues that were mutated in this study. (**B**) Band shift assay showing binding of [α-^{32}P] GTP-labeled *T. thermophilus* $tRNA^{Phe}$ to increasing amounts of wt $_{TTC}$TrmN and the K270E/R300E and K129E/R130E/R157E variants. Band shift assays for other $_{TTC}$TrmN variants are shown in Supplementary Figure S7C. The docking model is available for download from: ftp://ftp.genesilico.pl/iamb/models/RNA.MTases/TrmN/.

into the SAM binding pocket of the RFM domain. The location of this positively charged surface overlaps in both proteins, although it is more pronounced in $_{TTC}$TrmN (Figure 4A and B). Such a positively charged surface matches with the tRNA binding and modification role of both proteins. Moreover, the positively charged residues overlap with the evolutionarily conserved region, suggesting a functional role in substrate tRNA binding (Figure 4C and D). The positively charged residues that form the proposed tRNA binding site are located on both the THUMP and RFM domains. A region of the solvent accessible surface with particularly strong positive electrostatics, due to the presence of several positively charged residues, is located around the interface of the THUMP domain and the RFM domain, building a groove suitable for tRNA accommodation. A significant contribution to the positive charge in this region is provided by the long peptide linker connecting the two domains, suggesting also a functional role for this region. A second region with positive electrostatics is centered around the SAM binding pocket, where the acceptor stem of the tRNA containing the target nucleotide G6 might bind. In $_{Pf}$Trm14 an additional patch of conserved, positively charged residues is present on the 2-stranded β-sheet extension of the THUMP domain, but a functional role in the m^2G6 MTase activity *in vitro* was ruled out (see before).

The THUMP domain has previously been suggested to be an RNA binding domain implicated in the modification of bases in the core of the tRNA molecules (13). However, previous studies with the isolated THUMP domain of the m^2G10 MTase PAB1283 have already shown that the isolated THUMP domain is not able to bind the tRNA substrate (28). This is in agreement with our suggestion that the tRNA binding surface of Trm14/TrmN is formed by both the RFM and THUMP domain, and that also the linker region contributes to a large extent. Consequently, it was found that in the case of $_{Pf}$Trm14 neither the isolated RFM nor the THUMP domain is able to bind tRNA or catalyze the methyltransfer reaction. Also an attempt to reconstitute the activity by mixing the individual domains was not successful, suggesting that both domains indeed synergistically collaborate, together with the linker, in the binding of the substrate (16).

To obtain a more detailed image of the binding mode of tRNA to Trm14/TrmN, a docking model was generated for the binding of $tRNA^{Phe}$ of *T. thermophilus* onto $_{TTC}$TrmN. An unbound structure of *T. thermophilus* $tRNA^{Phe}$ is not available. Therefore, we used two variants of *T. thermophilus* $tRNA^{Phe}$: a homology model of the unbound conformation based on the coordinates of a homolog from *E. coli* crystallized in the apo form (PDB code 3L0U) and an experimental model in the bound conformation taken from the cocrystal structure with phenylalanyl-tRNA synthetase (PDB code 2IY5). The unbound conformation corresponds to the unmodified substrate RNA as would be encountered by $_{TTC}$TrmN in solution. The bound tRNA conformation is altered, in particular in the anticodon loop, by interactions with the tRNA synthetase. However, in both cases docking yields the same two preferred orientations of the $tRNA^{Phe}$ towards $_{TTC}$TrmN. The best-scored docking model for the unbound tRNA conformation (which is similar to one of the top-scoring and largest clusters obtained for the bound conformation) is shown in Figure 5A and Supplementary Figure S7A. This model seems mechanistically most relevant, as the target nucleoside residue is positioned in such a way that it can be easily rotated out of the tRNA acceptor stem into the catalytic site, like observed in numerous other MTases acting on nucleic acids (66). In an alternative model, that represents the largest cluster for the unbound conformation, the tRNA is rotated by ~90°(shown in Supplementary Figure S7B). Here, the target nucleoside is positioned unfavorably to

enter the active site, suggesting that this orientation of tRNA may not be compatible with the catalysis of the methyl transfer.

In the preferred model (Figure 5A), $tRNA^{Phe}$ binds with its core, formed by the D and T stem and loops, in the groove formed between the RFM and THUMP domain, while a part of the anticodon stem interacts with the THUMP domain. Consistent with a role of Trm14/TrmN in G6 methylation, the tRNA is binding with its acceptor stem directed towards the RFM domain. In this model, the distance between the methyl group of SAM and the N^2 exocyclic amine group of G6 is 19.3 Å. The position of the G6 suggests that a flip-out mechanism could bring the guanine base in closer contact to the methyl donor for the reaction to occur. In this model, the tRNA is also binding in the vicinity of the flexible loop, adjacent to motif IV, that harbors the positively charged residues Arg266, Arg271 and Lys272, suggesting some induced fit upon tRNA binding.

To further validate the docking model, a number of positively charged residues in the proposed interaction surface were substituted by alanine and glutamate, and the ability of the resulting protein variants to bind the tRNA substrate was determined by band-shift assays (Supplementary Figures S5 and S7). The corresponding alanine and glutamate variants show a same tendency of reduced affinity towards $tRNA^{Phe}$. Yet, the effect of the glutamate substitutions is stronger due to the inversion of charge from positive (Arg or Lys) to negative (Supplementary Figure S7). Apart from direct interactions such as hydrogen bonds and salt bridges, also long-range electrostatics could affect the binding mode and affinity of the tRNA substrate. In the case of a highly positively charged molecule, such as $_{TTC}$TrmN, the effect of single alanine substitution on these long range interactions could be easily shielded by the surrounding charges.

Arg33 and Arg98 are located on the THUMP domain (respectively on the NFLD and core-THUMP subdomains) and line the highly positively charged groove between the THUMP and RFM domains (Supplementary Figure 5A). Substituting either of these residues by glutamate reduces the affinity for tRNA by at least a factor of 2 (Supplementary Figure S7C). The charged and conserved residues Arg129 and Arg130 are located in the THUMP domain of $_{TTC}$TrmN and are in the docking model in close proximity to the region between the D-arm and the anticodon arm of the tRNA substrate. R157 is part of the linker region between the THUMP and RFM domain, containing highly conserved positively charged residues, and is involved in the formation of the binding groove. In agreement with the model, a variant with a double substitution K129E/R130E exhibits a reduced affinity towards $tRNA^{Phe}$ by more than 2-fold (Supplementary Figure S7C). Interestingly, the triple substitution K129E/R130E/R157E invokes an even larger effect on affinity and nearly no binding is observed at the highest protein concentration used (~10-fold effect on K_D). These point variants clearly support the docking model and underscore the role of the groove between the THUMP and RFM domains in tRNA binding.

Lys270 and Arg300 are located on the RFM domain close to the SAM binding pocket (Supplementary Figure S5A). The K270E and R300E variants show a clear (~10-fold) decrease in their affinity to $tRNA^{Phe}$. In the K270E/R300E variant, no binding could be observed at the highest protein concentration used. These two residues are only in proximity with the tRNA molecule in the preferred docking model and not in the alternative docking model (compare Supplementary Figure S7A with S7B). Consequently, these results further support the docking model in Figure 5A as the biological relevant interaction mode between $_{TTC}$TrmN and $tRNA^{Phe}$.

AVAILABILITY

The docking models of $_{TTC}$TrmN and $tRNA^{Phe}$ discussed in this article are available for download from ftp://ftp.genesilico.pl/iamb/models/RNA.MTases/TrmN/

ACCESSION NUMBERS

Coordinates have been submitted to the RCSB Protein Data Bank with accession codes: 3TLJ for $_{Pf}$Trm14-SAH, 3TM4 for $_{Pf}$Trm14-SAM, 3TM5 for $_{Pf}$Trm14-SFG and 3TMA for $_{TTC}$TrmN.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online: Supplementary Figures 1–7; Supplementary Tables 1–2.

ACKNOWLEDGEMENTS

The authors would like to thank the staff at the beamline ID23-2 and ID14-4 of the ESRF in France and the staff of the beamline PX-III of the SLS in Switzerland for assistance during data collection. Authors would also like to thank Stefan Münnich for help during data processing.

FUNDING

Fonds Wetenschappelijk Onderzoek (G025909N to W.V.); FWO post-doctoral grant to W.V.; IWT pre-doctoral grant to M.F.; European Social Fund PhD fellowship to I.T.; Polish Ministry of Science and Higher Education (188/N-DFG/2008/0 to J.M.B.); European Research Council (StG grant RNA + P = 123D to J.M.B.); "Ideas for Poland" fellowship from the Foundation for Polish science to J.M.B.; Fonds pour le Recherche Fondamentale Collective (2.4.520.05F to L.D.); Fonds J. Brachet and Fonds D. et A. Van Buuren to L.D. Funding for open access charge: Fonds Wetenschappelijk Onderzoek (G025909N).

Conflict of interest statement. None declared.

REFERENCES

1. Cantara,W.A., Crain,P.F., Rozenski,J., McCloskey,J.A., Harris,K.A., Zhang,X., Vendeix,F.A., Fabris,D. and Agris,P.F. (2011) The RNA Modification Database, RNAMDB: 2011 update. *Nucleic Acids Res.*, **39**, D195–D201.

2. Hopper,A.K. and Phizicky,E.M. (2003) tRNA transfers to the limelight. *Genes Dev.*, **17**, 162–180.
3. Gray,M.W. (1975) Analysis of O2′-methylnucleoside 5′-phosphates in snake venom hydrolysates of RNA: identification of O2′-methyl-5-carboxymethyluridine as a constituent of yeast transfer RNA. *Can. J. Biochem.*, **53**, 735–746.
4. Hou,Y.M. and Perona,J.J. (2010) Stereochemical mechanisms of tRNA methyltransferases. *FEBS Lett.*, **584**, 278–286.
5. Kalhor,H.R. and Clarke,S. (2003) Novel methyltransferase for modified uridine residues at the wobble position of tRNA. *Mol. Cell Biol.*, **23**, 9283–9292.
6. Kato,M., Araiso,Y., Noma,A., Nagao,A., Suzuki,T., Ishitani,R. and Nureki,O. (2011) Crystal structure of a novel JmjC-domain-containing protein, TYW5, involved in tRNA modification. *Nucleic Acids Res.*, **39**, 1576–1585.
7. Helm,M., Giegé,R. and Florentz,C. (1999) A Watson-Crick base-pair-disrupting methyl group (m1A9) is sufficient for cloverleaf folding of human mitochondrial tRNALys. *Biochemistry*, **38**, 13338–13346.
8. Sakurai,M., Ohtsuki,T. and Watanabe,K. (2005) Modification at position 9 with 1-methyladenosine is crucial for structure and function of nematode mitochondrial tRNAs lacking the entire T-arm. *Nucleic Acids Res.*, **33**, 1653–1661.
9. Nishimoto,M., Higashijima,K., Shirouzu,M., Grosjean,H., Bessho,Y. and Yokoyama,S. (2008) Crystal structure of tRNA N2,N2-guanosine dimethyltransferase Trm1 from Pyrococcus horikoshii. *J. Mol. Biol.*, **383**, 871–884.
10. Awai,T., Kimura,S., Tomikawa,C., Ochi,A., Bessho,Y., Yokoyama,S., Ohno,S., Nishikawa,K., Yokogawa,T. *et al.* (2009) Aquifex aeolicus tRNA (N2,N2-guanine)-dimethyltransferase (Trm1) catalyzes transfer of methyl groups not only to guanine 26 but also to guanine 27 in tRNA. *J. Biol. Chem.*, **284**, 20467–20478.
11. Jühling,F., Mörl,M., Hartmann,R.K., Sprinzl,M., Stadler,P.F. and Putz,J. (2009) tRNAdb 2009: compilation of tRNA sequences and tRNA genes. *Nucleic Acids Res.*, **37**, D159–D162.
12. Purushothaman,S.K., Bujnicki,J.M., Grosjean,H. and Lapeyre,B. (2005) Trm11p and Trm112p are both required for the formation of 2-methylguanosine at position 10 in yeast tRNA. *Mol. Cell Biol.*, **25**, 4359–4370.
13. Armengaud,J., Urbonavicius,J., Fernandez,B., Chaussinand,G., Bujnicki,J.M. and Grosjean,H. (2004) N2-methylation of guanosine at position 10 in tRNA is catalyzed by a THUMP domain-containing, *S*-adenosylmethionine-dependent methyltransferase, conserved in Archaea and Eukaryota. *J. Biol. Chem.*, **279**, 37142–37152.
14. Awai,T., Ochi,A., Sengoku,T., Hirata,A., Bessho,Y., Yokoyama,S. and Hori,H. (2011) Substrate tRNA recognition mechanism of a multisite-specific tRNA methyltransferase, *Aquifex aeolicus* Trm1, based on the X-ray crystal structure. *J. Biol. Chem.*, **286**, 35236–46.
15. Menezes,S., Gaston,K.W., Krivos,K.L., Apolinario,E.E., Reich,N.O., Sowers,K.R., Limbach,P.A. and Perona,J.J. (2011) Formation of m2G6 in *Methanocaldococcus jannaschii* tRNA catalyzed by the novel methyltransferase Trm14. *Nucleic Acids Res.*, **39**, 7641–7655.
16. Roovers,M., Oudjama,Y., Fislage,M., Bujnicki,J.M., Versées,W. and Droogmans,L. (2012) The open reading frame TTC1157 of Thermus thermophilus HB27 encodes the methyltransferase forming N2-methylguanosine at position 6 in tRNA. *RNA*, **18**, 815–824.
17. Czerwoniec,A., Kasprzak,J.M., Kaminska,K.H., Rother,K., Purta,E. and Bujnicki,J.M. (2009) Folds and functions of domains in RNA modification enzymes. In: Grosjean,H. (ed.), *DNA and RNA Modification Enzymes: Structure, Mechanism, Function and Evolution.* Landes Bioscience, Austin, Texas, USA, pp. 289–302.
18. Bujnicki,J.M. (1999) Comparison of protein structures reveals monophyletic origin of the AdoMet-dependent methyltransferase family and mechanistic convergence rather than recent differentiation of N4-cytosine and N6-adenine DNA methylation. *In Silico Biol.*, **1**, 175–182.
19. Nishimasu,H., Ishitani,R., Yamashita,K., Iwashita,C., Hirata,A., Hori,H. and Nureki,O. (2009) Atomic structure of a folate/FAD-dependent tRNA T54 methyltransferase. *Proc. Natl Acad. Sci. USA*, **106**, 8180–8185.
20. Arragain,S., Garcia-Serres,R., Blondin,G., Douki,T., Clemancey,M., Latour,J.M., Forouhar,F., Neely,H., Montelione,G.T., Hunt,J.F. *et al.* (2010) Post-translational modification of ribosomal proteins: structural and functional characterization of RimO from *Thermotoga maritima*, a radical *S*-adenosylmethionine methylthiotransferase. *J. Biol. Chem.*, **285**, 5792–5801.
21. Ahn,H.J., Kim,H.W., Yoon,H.J., Lee,B.I., Suh,S.W. and Yang,J.K. (2003) Crystal structure of tRNA(m1G37)methyltransferase: insights into tRNA recognition. *EMBO J.*, **22**, 2593–2603.
22. Goto-Ito,S., Ito,T., Ishii,R., Muto,Y., Bessho,Y. and Yokoyama,S. (2008) Crystal structure of archaeal tRNA(m(1)G37)methyltransferase aTrm5. *Proteins*, **72**, 1274–1289.
23. Anantharaman,V., Koonin,E.V. and Aravind,L. (2002) Comparative genomics and evolution of proteins involved in RNA metabolism. *Nucleic Acids Res.*, **30**, 1427–1464.
24. Aravind,L. and Koonin,E.V. (2001) THUMP—a predicted RNA-binding domain shared by 4-thiouridine, pseudouridine synthases and RNA methylases. *Trends Biochem. Sci.*, **26**, 215–217.
25. Waterman,D.G., Ortiz-Lombardía,M., Fogg,M.J., Koonin,E.V. and Antson,A.A. (2006) Crystal structure of Bacillus anthracis ThiI, a tRNA-modifying enzyme containing the predicted RNA-binding THUMP domain. *J. Mol. Biol.*, **356**, 97–110.
26. Randau,L., Stanley,B.J., Kohlway,A., Mechta,S., Xiong,Y. and Söll,D. (2009) A cytidine deaminase edits C to U in transfer RNAs in Archaea. *Science*, **324**, 657–659.
27. McCleverty,C.J., Hornsby,M., Spraggon,G. and Kreusch,A. (2007) Crystal structure of human Pus10, a novel pseudouridine synthase. *J. Mol. Biol.*, **373**, 1243–1254.
28. Gabant,G., Auxilien,S., Tuszynska,I., Locard,M., Gajda,M.J., Chaussinand,G., Fernandez,B., Dedieu,A., Grosjean,H., Golinelli-Pimpaneau,B. *et al.* (2006) THUMP from archaeal tRNA:m22G10 methyltransferase, a genuine autonomously folding domain. *Nucleic Acids Res.*, **34**, 2483–2494.
29. Fislage,M., Roovers,M., Münnich,S., Droogmans,L. and Versées,W. (2011) Crystallization and preliminary X-ray crystallographic analysis of putative tRNA-modification enzymes from *Pyrococcus furiosus* and *Thermus thermophilus*. *Acta Crystallogr. Sect. F Struct. Biol. Cryst. Commun.*, **67**, 1432–1435.
30. Doublié,S. (1997) Preparation of selenomethionyl proteins for phase determination. *Meth. Enzymol.*, **276**, 523–530.
31. Reyes,V.M. and Abelson,J. (1987) A synthetic substrate for tRNA splicing. *Anal. Biochem.*, **166**, 90–106.
32. Droogmans,L. and Grosjean,H. (1991) 2′-O-methylation and inosine formation in the wobble position of anticodon-substituted tRNA-Phe in a homologous yeast in vitro system. *Biochimie*, **73**, 1021–1025.
33. Kabsch,W. (2010) XDS. *Acta Crystallogr. D Biol. Crystallogr.*, **66**, 125–132.
34. Panjikar,S., Parthasarathy,V., Lamzin,V.S., Weiss,M.S. and Tucker,P.A. (2005) Auto-Rickshaw: an automated crystal structure determination platform as an efficient tool for the validation of an X-ray diffraction experiment. *Acta Crystallogr. D Biol. Crystallogr.*, **61**, 449–457.
35. (1994) The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D Biol. Crystallogr.*, **50**, 760–763.
36. Sheldrick,G.M., Hauptman,H.A., Weeks,C.M., Miller,R. and Uson,I. *International tables for crystallography: Crystallography of biological macromolecules*, Vol. F. Kluwer Academic Publishers, Dordrecht, pp. 333–345.
37. Schneider,T.R. and Sheldrick,G.M. (2002) Substructure solution with SHELXD. *Acta Crystallogr. D Biol. Crystallogr.*, **58**, 1772–1779.
38. Hao,Q. (2004) ABS: a program to determine absolute configuration and evaluate anomalous scatterer substructure. *J. Appl. Crystallogr.*, **37**, 498–499.
39. Sheldrick,G.M. (2002) Macromolecular phasing with SHELXE. *Zeitschrift für Kristallographie*, **217**, 644–650.

40. Cowtan,K. *Joint CCP4 and ESF-EACBM Newsletter on protein Crystallography*., **31**, 34–38.
41. Perrakis,A., Morris,R. and Lamzin,V.S. (1999) Automated protein model building combined with iterative structure refinement. *Nat. Struct. Biol.*, **6**, 458–463.
42. Morris,R.J., Zwart,P.H., Cohen,S., Fernandez,F.J., Kakaris,M., Kirillova,O., Vonrhein,C., Perrakis,A. and Lamzin,V.S. (2004) Breaking good resolutions with ARP/wARP. *J. Synchrotron Radiat.*, **11**, 56–59.
43. Emsley,P., Lohkamp,B., Scott,W.G. and Cowtan,K. (2010) Features and development of Coot. *Acta Crystallogr. D Biol. Crystallogr.*, **66**, 486–501.
44. Murshudov,G.N., Vagin,A.A. and Dodson,E.J. (1997) Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. D Biol. Crystallogr.*, **53**, 240–255.
45. McCoy,A.J., Grosse-Kunstleve,R.W., Adams,P.D., Winn,M.D., Storoni,L.C. and Read,R.J. (2007) Phaser crystallographic software. *J. Appl. Crystallogr.*, **40**, 658–674.
46. Winn,M.D., Isupov,M.N. and Murshudov,G.N. (2001) Use of TLS parameters to model anisotropic displacements in macromolecular refinement. *Acta Crystallogr. D Biol. Crystallogr.*, **57**, 122–133.
47. Painter,J. and Merritt,E.A. (2006) Optimal description of a protein structure in terms of multiple groups undergoing TLS motion. *Acta Crystallogr. D Biol. Crystallogr.*, **62**, 439–450.
48. Chen,V.B., Arendall,W.B. III, Headd,J.J., Keedy,D.A., Immormino,R.M., Kapral,G.J., Murray,L.W., Richardson,J.S. and Richardson,D.C. (2010) MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D Biol. Crystallogr.*, **66**, 12–21.
49. Dolinsky,T.J., Nielsen,J.E., McCammon,J.A. and Baker,N.A. (2004) PDB2PQR: an automated pipeline for the setup of Poisson-Boltzmann electrostatics calculations. *Nucleic Acids Res.*, **32**, W665–W667.
50. Baker,N.A., Sept,D., Joseph,S., Holst,M.J. and McCammon,J.A. (2001) Electrostatics of nanosystems: application to microtubules and the ribosome. *Proc. Natl Acad. Sci. USA*, **98**, 10037–10041.
51. Ashkenazy,H., Erez,E., Martz,E., Pupko,T. and Ben-Tal,N. (2010) ConSurf 2010: calculating evolutionary conservation in sequence and structure of proteins and nucleic acids. *Nucleic Acids Res.*, **38**, W529–W533.
52. Vellieux,F.M.D. and Dijkstra,B.W. (1997) Computation of Bhat's OMIT maps with different coefficients. *J. Appl. Crystallogr.*, **30**, 396–399.
53. Jakuschev,S. and Hoffmann,D. (2009) A novel algorithm for macromolecular epitope matching. *Algorithms*, **2**, 498–517.
54. Demirci,H., Gregory,S.T., Dahlberg,A.E. and Jogl,G. (2008) Crystal structure of the *Thermus thermophilus* 16 S rRNA methyltransferase RsmC in complex with cofactor and substrate guanosine. *J. Biol. Chem.*, **283**, 26548–26556.
55. Rother,M., Rother,K., Puton,T. and Bujnicki,J.M. (2011) ModeRNA: a tool for comparative modeling of RNA 3D structure. *Nucleic Acids Res.*, **39**, 4007–4022.
56. Tuszynska,I. and Bujnicki,J.M. (2011) DARS-RNP and QUASI-RNP: new statistical potentials for protein-RNA docking. *BMC Bioinformatics*, **12**, 348.
57. Vakser,I.A. and Aflalo,C. (1994) Hydrophobic docking: a proposed enhancement to molecular recognition techniques. *Proteins*, **20**, 320–329.
58. Gajda,M.J., Tuszynska,I., Kaczor,M., Bakulina,A.Y. and Bujnicki,J.M. (2010) FILTREST3D: discrimination of structural models using restraints from experimental data. *Bioinformatics*, **26**, 2986–2987.
59. Ponstingl,H., Henrick,K. and Thornton,J.M. (2000) Discriminating between homodimeric and monomeric proteins in the crystalline state. *Proteins*, **41**, 47–57.
60. Krissinel,E. and Henrick,K. (2004) Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. *Acta Crystallogr. D Biol. Crystallogr.*, **60**, 2256–2268.
61. Altschul,S.F., Madden,T.L., Schäffer,A.A., Zhang,J., Zhang,Z., Miller,W. and Lipman,D.J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.*, **25**, 3389–3402.
62. Schluckebier,G., Kozak,M., Bleimling,N., Weinhold,E. and Saenger,W. (1997) Differential binding of *S*-adenosylmethionine *S*-adenosylhomocysteine and Sinefungin to the adenine-specific DNA methyltransferase M.TaqI. *J. Mol. Biol.*, **265**, 56–67.
63. Versées,W., Decanniere,K., Pellé,R., Depoorter,J., Brosens,E., Parkin,D.W. and Steyaert,J. (2001) Structure and function of a novel purine specific nucleoside hydrolase from *Trypanosoma vivax*. *J. Mol. Biol.*, **307**, 1363–1379.
64. Malone,T., Blumenthal,R.M. and Cheng,X. (1995) Structure-guided analysis reveals nine sequence motifs conserved among DNA amino-methyltransferases, and suggests a catalytic mechanism for these enzymes. *J. Mol. Biol.*, **253**, 618–632.
65. Bujnicki,J.M. (2000) Phylogenomic analysis of 16S rRNA:(guanine-N2) methyltransferases suggests new family members and reveals highly conserved motifs and a domain structure similar to other nucleic acid amino-methyltransferases. *FASEB J.*, **14**, 2365–2368.
66. Cheng,X. and Roberts,R.J. (2001) AdoMet-dependent methylation, DNA methyltransferases and base flipping. *Nucleic Acids Res.*, **29**, 3784–3795.