Structural basis for m^{7}G-cap hypermethylation of small nuclear, small nucleolar and telomerase RNA by the dimethyltransferase TGS1

Thomas Monecke, Achim Dickmanns and Ralf Ficner*

Abteilung für Molekulare Strukturbiologie, Institut für Mikrobiologie und Genetik, Georg-August-Universität Göttingen, 37077 Göttingen, Germany

ABSTRACT

The 5′-cap of spliceosomal small nuclear RNAs, some small nucleolar RNAs and of telomerase RNA was found to be hypermethylated in vivo. The Trimethylguanosine Synthase 1 (TGS1) mediates this conversion of the 7-methylguanosine-cap to the 2,2,7-trimethylguanosine (m$_3$G)-cap during maturation of the RNPs. For mammalian UsnRNAs the generated m^{2,2,7}G-cap is one part of a bipartite import signal mediating the transport of the UsnRNP-core complex into the nucleus. In order to understand the structural organization of human TGS1 as well as substrate binding and recognition we solved the crystal structure of the active TGS1 methyltransferase domain containing both, the minimal substrate m^{7}GTP and the reaction product S-adenosyl-L-homocysteine (AdoHcy). The methyltransferase of human TGS1 harbors the canonical class 1 methyltransferase fold as well as an unique N-terminal, α-helical domain of 40 amino acids, which is essential for m^{7}G-cap binding and catalysis. The crystal structure of the substrate bound methyltransferase domain as well as mutagenesis studies provide insight into the catalytic mechanism of TGS1.

INTRODUCTION

The m^{7}G-cap of small nuclear RNAs (snRNAs), small nucleolar RNAs (snoRNAs) as well as the telomerase RNA TLC1 is known to be hypermethylated by the Trimethylguanosine Synthase 1 (TGS1) during maturation of the respective ribonucleoprotein particles (RNPs). Especially the biogenesis of the human uridyl-rich snRNPs U1, U2, U4, U5 and U6, the major components of the spliceosome, is well characterized (1,2). Except for U6snRNP, in vertebrates the biogenesis of UsnRNPs involves a nucleocytoplasmic transport cycle. After transcription by RNA-polymerase II, the snRNAs U1, U2, U4 and U5 acquire an m^{7}G-cap and are subsequently exported to the cytoplasm in a cap dependent manner (3,4). In the cytoplasm seven Sm-proteins (B/B′, D1, D2, D3, E, F and G) are assembled in a doughnut-like structure around the highly conserved Sm-site (5′-PuA(U)$_n$GPu-3′) of the UsnRNA (5–7). The resulting Sm-core UsnRNP was shown to be a prerequisite for m^{7}G-cap hypermethylation (7,8) by a cap dimethyltransferase, catalyzing the addition of two methyl groups to the exocyclic NH$_2$ group of the N7-methylguanine (m^{7}G) (7,9,10). The m$_3$G-cap and the Sm-core serve as bipartite import signal, which mediates the import of the assembled UsnRNP-core into the nucleus (11–17). The import occurs by the concerted action of snurportin1 (SPN1), the survival of motor neuron (SMN)-complex and importinβ (Impβ). SPN1 binds the hypermethylated UsnRNA cap-guanine and the first nucleotide of the RNA in a stacked conformation and bridges the interaction to the import factor Impβ (13,18). After translocation Impβ dissociates from the complex due to the binding of Ran-GTP. A subcomplex containing at least the assembled UsnRNP and SPN1, which is bound to the m$_3$G-cap, remains (19) and its disassembly is not yet understood. The released UsnRNPs individually associate with specific sets of proteins, resulting in the particular, mature UsnRNPs U1, U2, U4 and U5 (20). However, the U6snRNP is generated by a different pathway that does not include such cytoplasmic maturation steps. In contrast to all other UsnRNPs it is transcribed by RNA polymerase III (21–23) and acquires a γ-monomethyl-phosphate-cap. Together with the other UsnRNPs it accumulates subsequent to maturation in nuclear (interchromatin) speckles and in cajal bodies (24,25).

The enzyme responsible for hypermethylation of the UsnRNA-m^{7}G-cap in yeast was first identified by Mouaikel and co-workers in 2002 and it was named TGS1 in accordance to its function (26). As a bona fide dimethyltransferase it catalyzes two successive methyl

*To whom correspondence should be addressed. Tel: +49 551 39 14071; Fax: +49 551 39 14082; Email: rficner@uni-goettingen.de

group transfers, each from one *S*-adenosyl-ʟ-methionine (AdoMet) to the exocyclic nitrogen N2 of the cap-guanine, generating two *S*-adenosyl-ʟ-homocysteine molecules (AdoHcy) and the modified $m^{2,2,7}G$ (m_3G)-cap (27). The two Sm-proteins SmB and SmD1 of the UsnRNP-core bind to TGS1 in yeast and in human cells suggesting an additional interaction between the dimethyltransferase and the RNP besides the cap (10,26). Although the yeast TGS1 (yTGS1) is not essential for cell viability its deletion results in a cold-sensitive phenotype, as the mutants show reduced growth at lower temperatures as well as missing m_3G-caps on certain UsnRNAs and snoRNAs (26). As a consequence of this deletion a splicing defect at restrictive temperatures is observed and linked to the appearance of nucleolar retention of U1snRNAs. Furthermore, the nucleolar morphology is disturbed and the cells show impairment in ribosome biogenesis. However, it turned out that it is the TGS1 protein *per se*, rather than its catalytic activity that is required for efficient ribosome synthesis (28).

In contrast to yTGS1, the orthologs from *Xenopus laevis*, *Drosophila melanogaster*, *Caenorhabditis elegans*, *Mus musculus* and *Homo sapiens* additionally contain large N-terminal domains.

The human ortholog of yTGS1 was originally designated PIMT (PRIP-interacting protein with methyltransferase domain) because of its binding capability to PRIP (PPAR-interacting protein) and the presence of the C-terminal methyltransferase domain (29). As shown recently, the purified C-terminal methyltransferase domain of hTGS1 containing the amino acids 576–853 is able to complement the cold-sensitive phenotype of a yeast Δtgs1-strain demonstrating that the human TGS1 methyltransferase domain is a true functional ortholog of the yeast enzyme (30). Hence, *H. sapiens* PIMT is referred to as hTGS1 with respect to its function and relatedness to yTGS1. It was found that hTGS1 is not only able to bind via its large N-terminus to PRIP, but also to enhance its action. PRIP was shown to interact with the peroxisome proliferator-activated receptor (PPAR). This suggests that the human TGS1 is a component of the nuclear receptor signal transduction cascade acting through PRIP, and at least in human cells is involved in transcriptional regulation (29). Besides that, hTGS1 was shown to interact with the transcriptional coactivators CREB-binding protein (CBP), p300 and PPAR-binding protein (PBP) *in vitro* and *in vivo* (31). hTGS1 exists in two isoforms, a full-length cytoplasmic and an N-terminally shortened nuclear form, which might be involved in snoRNP maturation (32,33). In addition it was found *in vivo* and *in vitro*, that the C-terminal part of the human cap dimethyltransferase interacts directly with the survival of motor neuron protein (SMNp) (34). This interaction is greatly impaired *in vivo* as well as *in vitro* by a SMN mutation (SMNΔEx7) that is most commonly present in patients who are adversely affected by the neuromuscular disease spinal muscular atrophy (SMA). Both proteins are localized in the cytoplasm as well as in the nuclear cajal bodies, where the assembled UsnRNPs and snoRNPs reside (34).

Contrary to the comparative mild effects upon deletion of yTGS1, the deletion of its counterpart in *D. melanogaster* is lethal to the larvae in the early pupal stages indicating an essential function in development (35). The fruit fly ortholog was named drosophila-tat-like (DTL) because of its capability to bind the TAR-RNA of the human immunodeficiency virus (HIV). Furthermore, by means of mutagenesis studies it could be shown that this lethal effect is causally linked to the observed absence of m_3G-caps on UsnRNPs, which is a direct result of the TGS1-knockdown (35).

The m_3G-cap does not solely occur on spliceosomal UsnRNPs, but also on a subset of small nucleolar RNAs (snoRNAs). The modification of these snoRNAs such as box C/D (U3) and box H/ACA (snR10/30), which are, at least in yeast, also hypermethylated by TGS1 is less well understood (36,37). The cap dimethyltransferase interacts with specific proteins of particular snoRNAs such as Cbf5 and Nop58, thus giving a possibility to distinguish between UsnRNA and different snoRNAs with respect to their state of methylation (26). As for yeast UsnRNAs, it was shown for the snoRNAs that they do not cycle through the cytoplasmic compartment during maturation (38). Interestingly, it is known that yTGS1 is localized exclusively in the nucleus and that this organism lacks the orthologs for important proteins required for m^7G-cap dependent UsnRNA export (phosphorylated adapter for RNA export, PHAX) and m_3G-dependent UsnRNP import (SPN1). Therefore it is likely that, in contrast to vertebrates, the maturation of UsnRNPs and snoRNPs in *Saccharomyces cerevisiae* occurs exclusively in the nucleus. This assumption is further confirmed by the fact, that the absence of m_3G-caps in yeast is not lethal and instead produces only a cold-sensitive phenotype and a mild splicing defect (26).

Quite recently, it could be demonstrated that in yeast the m_3G-cap at the 5′-end of the telomerase RNA TLC1 results from the catalytic activity of TGS1 as well (39). TLC1 shares several features with UsnRNAs and snoRNAs, among them the transcription by RNA-polymerase II, a high content of uridines, the ability to bind Sm-proteins and the presence of a 5′-m_3G-cap. Deletion of TGS1 in yeast cells influences the length, structure and function of telomeres and they show premature aging (39).

Although there are numerous biochemical data characterizing TGS1 from *H. sapiens*, *S. cerevisiae*, *D. melanogaster*, *Schizosaccharomyces pombe*, *Giardia lamblia* and *Trypanosoma brucei* (27,28,40–43) the three-dimensional structure of an active form of this enzyme is unknown so far. However, the crystal structure of a truncated, but catalytically inactive fragment of the human TGS1 was recently determined. This fragment contains the canonical methyltransferase domain, which on its own is not sufficient for m^7G-cap hypermethylation (44).

Here we present the 2 Å crystal structure of the catalytically active methyltransferase domain of hTGS1. The results provide insight into the structural organization and flexibility as well as a detailed view on substrate binding and recognition by TGS1. Furthermore, on the basis of the crystal structure and biochemical data including

mutagenesis studies, a mechanism for the dimethylation reaction catalyzed by TGS1 is proposed.

MATERIALS AND METHODS

Protein expression and purification

The human TGS1 fragment comprising the residues 618–853 ($hTGS1_{618–853}$) was subcloned from pGEX-6P-1 full-length TGS1 (residues 1–853; accession number Q96RS0) into BamHI/XhoI-sites of pGEX-6P-1 (GE Healthcare, Germany) and verified by sequencing. The GST-fusion constructs were expressed in *Escherichia coli* BL21(DE3) (Invitrogen, USA) at 16°C in ampicillin containing 2YT-medium, which was supplemented with 2% (w/v) α-D-glucose in order to suppress basal transcription. Expression of GST-$hTGS1_{618–853}$ was induced at $OD_{600} = 0.8$, adding IPTG to a final concentration of 0.5 mM. Directly after induction, 2% (v/v) ethanol and 50 mM K_2HPO_4 were added to the growing culture. The cells were harvested after 18 h of expression ($5000 \times g$, 20 min, 4°C) and resuspended in lysis buffer containing 50 mM Tris/HCl pH 7.5, 500 mM NaCl, 2 mM EDTA and 2 mM DTT. All subsequent steps were carried out at 4°C unless stated otherwise. Cells were disrupted using a microfluidizer 110S (Microfluidics, USA). The clarified lysate ($30\,000 \times g$, 30 min, 4°C) was subsequently loaded onto a GSH-Sepharose column (GE Healthcare, Germany) equilibrated with lysis buffer. Unbound proteins were removed by washing with 2 column volumes (CV) of lysis buffer. In order to eliminate RNA contaminations the loaded column was washed with 1 CV of a high salt buffer containing 50 mM Tris/HCl pH 7.5, 1 M NaCl and 2 mM DTT. After re-equilibration in lysis buffer the bound fusion protein was eluted with lysis buffer containing additionally 25 mM reduced glutathione. For cleavage the fusion protein (GST-$hTGS1_{618–853}$) was incubated with PreScission protease (GE Healthcare, Germany) at 4°C overnight in a 1:100 molar ratio of protease: fusion protein. Further purification was achieved using a Superdex S75 (26/60) gel filtration column (GE Healthcare, Germany) in a buffer containing 20 mM Tris/HCl pH 7.5, 200 mM NaCl and 2 mM DTT. In order to remove small amounts of remaining GST, the pooled gel filtration fractions were finally purified using a second GSH-sepharose equilibrated in gel filtration buffer. The pure protein was concentrated to 8 mg/ml using vivaspin concentrators with a molecular weight cut-off of 10 000 Da (Sartorius, Germany) and aliquots were frozen in liquid nitrogen and stored at −80°C.

Site-directed mutagenesis

Single amino-acid mutants of $hTGS1_{618–853}$ (Ser763Ala, Ser763Asp, Trp766Ala, Asp696Ala, Asp696Asn and Phe804Lys) were generated from the wild-type clone using the QuikChange® Site-Directed Mutagenesis Kit (Stratagene, USA) following the manufacturer's protocol. All mutations were confirmed by DNA sequencing and the mutants were expressed and purified in analogy to the procedures for the wild-type TGS1-fragment.

HPLC-based activity assay

In order to analyze the activity of the purified hTGS1 fragment as well as of the described mutants, an HPLC-based activity assay was applied (44). Total 0.25 nmol of purified $hTGS1_{618–853}$ were mixed with 5 nmol of the minimal substrate m^7GTP (Sigma-Aldrich, Germany) and 20 nmoles of the cofactor AdoMet (Sigma-Aldrich, Germany) in $1\times$ PBS. The mixture with a total volume of 10 µl was incubated at 37°C, the reaction was stopped by the addition of 1 µl 1 M $HClO_4$ followed by subsequent incubation on ice for 1 min. The solution was neutralized adding 20 µl 2 M Na-acetate. Precipitated protein was pelleted by centrifugation ($16\,000 \times g$, 10 min, 20°C) and the supernatant was loaded onto a reversed phase HPLC-column (Prontosil C18-AQ, Bischoff Chromatography, Germany), which was equilibrated in buffer A, containing 100 mM K_2HPO_4/KH_2PO_4 pH 6.5. The substrates and products of the reaction were eluted from the column applying a linear gradient from 0% to 60% of buffer B, containing buffer A and 50% (v/v) acetonitrile. Commercially available m^7GTP, AdoMet, AdoHcy (Sigma-Aldrich, Germany) and $m^{2,2,7}_{(3)}GTP$ (KEDAR, Poland) served as reference for column calibration.

Crystallization and structure determination

$hTGS1_{618–853}$ comprising the C-terminal 235 residues was crystallized by the vapor diffusion method in sitting drop 24-well ChrysChem-plates (Hampton Research, USA) and in presence of an 8-fold molar excess of both, the minimal substrate m^7GTP and the reaction product *S*-adenosyl-L-homocysteine (AdoHcy). The appropriate amount of AdoHcy was desiccated, mixed with the protein solution and incubated for 10 min at 20°C. Afterwards the cap-analog m^7GTP was added followed by an additional incubation for 10 min at 20°C. One microliter of a reservoir solution containing 13–16% PEG 8000 and 0.1 M MES pH 6.2–6.8 was mixed with 1 µl of the prepared protein-substrate solution (4–6 mg/ml). Rhombohedral-shaped single crystals with dimensions of $70 \times 70 \times 60\,\mu m^3$ grew at 20°C within 2 weeks. Crystals belong to the space group R3 with the cell dimensions $a = b = 156.2\,Å$, $c = 100.3\,Å$ and angles $\alpha = \beta = 90°$ and $\gamma = 120°$. The crystals were flash frozen in liquid nitrogen after soaking in reservoir solution containing an additional 15% (v/v) 1,2-propanediol as cryo protectant. An X-ray diffraction dataset of a crystal containing $hTGS1_{618–853}$ as well as m^7GTP and AdoHcy was collected at beamline BL14.2 of the electron synchrotron in Berlin (BESSY). Data were integrated, scaled and reduced with the HKL2000 suite (HKL Research, USA) and the structure was solved by means of molecular replacement. Therefore a starting model was designed, containing the canonical methyltransferase fold of the previously determined structure of SeMet-containing inactive hTGS1 (44). The structure of SeMet-hTGS1 comprising the residues Pro674-Arg848 served as search model in PHASER (45), which placed three molecules in the asymmetric unit. Missing residues, as well as 415 water molecules were added manually in COOT (46) and the structure was refined using REFMAC5 (47) to a reasonable R_{work} of

Table 1. Statistics of the data set of hTGS1 methyltransferase domain (aa 618–853) bound to m^7GTP and AdoHcy

Crystal	hTGS1$_{618–853}$ + m^7GTP + AdoHcy
Data collection	
Space group	R3
Cell dimensions	
a, b, c (Å)	156.2, 156.2, 100.3
α, β, γ (°)	90, 90, 120
Wavelength (Å)	0.9184
X-ray source	BL14.2, BESSY (Berlin)
Resolution range (Å)	30.00–2.00 (2.07–2.00)
Number of reflections	61489
Completeness (%)	99.8 (100.0)
R_{merge}[a] (%)	5.4 (35.9)
Average I/σ	16.5 (2.6)
Redundancy	2.4 (2.4)
Mosaicity (°)	0.20
Refinement	
Resolution (Å)	30.0–2.00
Molecules per AU	3
Number of atoms	
Protein	4973
Ligand	202
Waters	415
R_{work}[b] (%)	18.0
R_{free}[c] (%)	21.3
Figure of merit	0.86
Average B factors (Å^2)	
Protein	33.0
Ligand	32.9
Waters	38.9
RMS deviations	
Bond lengths (Å)	0.005
Bond angles (°)	0.914
Ramachandran statistics (%)	
Most	92.8
Allowed	7.2
Generous	0.0
Disallowed	0.0

Values in parentheses indicate the specific values in the particular highest resolution shell.
[a] $R_{merge} = \sum_{hkl}\sum_i |I_i(hkl) - <I_i(hkl)>|/\sum_{hkl}\sum_i I_i<(hkl)>$, where the sum i is over all separate measurements of the unique reflection hkl.
[b] $R_{work} = \sum_{hkl}||F_{obs}| - |F_{calc}||/\sum_{hkl}|F_{obs}|$.
[c] R_{free} as R_{work}, but summed over a 5% test set of reflections.

18.0% and a R_{free}-value of 21.3%. In the final model, 92.8% of the residues are located within the most favoured regions and 7.2% in the additionally allowed ones, whereas no residues lie in the generously allowed or disallowed regions of the Ramachandran plot (see Table 1). Figures were generated using PyMOL [DeLano, W.L. The PyMOL Molecular Graphics System (2002), DeLanoScientific, USA].

RESULTS

The globular core of yTGS1 (aa 58–266) was predicted to represent the catalytic domain, which corresponds to the C-terminal residues 653–845 in the human TGS1 (48) (sequence alignment in Supplementary Data). Recently, the crystal structure of this canonical methyltransferase domain of human TGS1 (aa 653–853) was solved, but this domain lacks catalytic activity suggesting an incomplete cap binding pocket (44). Interestingly, a fragment containing only 17 additional amino acids N-terminally

and therefore comprising the residues 636–853 gains catalytic activity (44). Since all attempts to crystallize the active TGS1$_{636–853}$ failed, several further elongated fragments of hTGS1 were tested in crystallization. Finally, the crystallization of a fragment starting at residue 618 (hTGS1$_{618–853}$) in the presence of substrate-analogs succeeded.

Purification and characterization of hTGS1$_{618–853}$

hTGS1$_{618–853}$ was cloned and purified as described in 'Materials and Methods' section. An HPLC-based activity assay was applied to determine the catalytic activity of this hTGS1 fragment (Figure 1). After the incubation of purified hTGS1$_{618–853}$ in the presence of m^7GTP and AdoMet, their significant decrease and conversely increasing amounts of reaction products $m^{2,2,7}GTP$ and AdoHcy indicate the catalytic activity of the analyzed fragment (Figure 1B). The additional peak, for which no commercially available standard exists, elutes at 8.5 ml and is most likely the reaction intermediate $m^{2,7}GTP$ that has previously been reported (27,30,42).
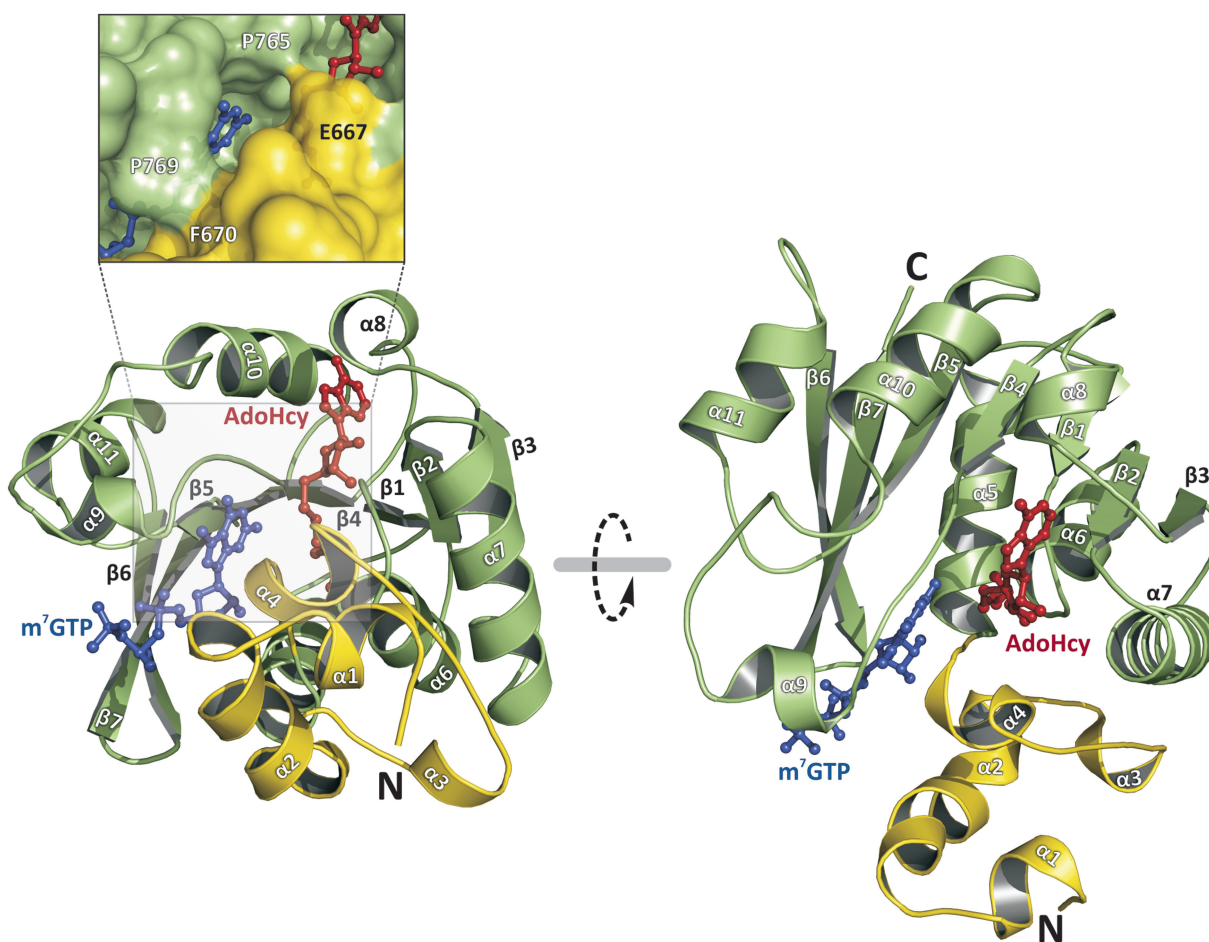
Overall structure of the active hTGS1 methyltransferase domain

The active fragment (hTGS1$_{618–853}$) was crystallized in the presence of m^7GTP and *S*-adenosyl-L-homocysteine (AdoHcy), the crystal structure was solved by means of molecular replacement and refined at a resolution of 2.0 Å. The statistics of X-ray diffraction data and structure refinement are summarized in Table 1. Crystals belong to the space group R3 with cell dimensions of $a = b = 156.2$ Å, $c = 100.3$ Å and $\alpha = \beta = 90°$ and $\gamma = 120°$, containing three TGS1 monomers in the asymmetric unit. Residues 634–844 of monomer 1 are well defined in the electron density map, whereas for monomers 2 and 3 only the residues 637–847 and 641–847 could be placed.

All three monomers superpose well with root mean square deviations of 0.47 Å, 0.60 Å and 0.28 Å for all three possible combinations (molecule 1/2, 2/3 and 1/3) and with respect to all common C_α atoms. Due to the high similarity of the three monomers only the structure of monomer 1 is discussed hereafter. One monomer consists of 11 α-helices and seven β-strands (Figure 2). It is divided into the core domain (residues Glu675-Asp844) forming the classical class I methyltransferase fold (green part) and an α-helical N-terminal extension (NTE) encompassing the residues Leu634-Ser671 (yellow part). Both domains are separated by a short loop comprising the three amino acids Val672, Thr673 and Pro674. The core domain of the monomer consists of a central seven-stranded β-sheet, which is flanked by α-helices on both sides, leading to the typical αβα sandwich and resembles the structure of the classical Rossmann-fold AdoMet-dependent methyltransferase superfamily. This fold is generally referred to as α/β twist with the first strand located in the middle of the β-sheet whereupon the following strands are placed consecutively outward to one edge. Thereafter, the chain returns to the middle of the β-sheet and continues out to the other edge. Thus, the β-sheet has a β6↑β7↓β5↑β4↑β1↑β2↑β3↑ topology with six strands
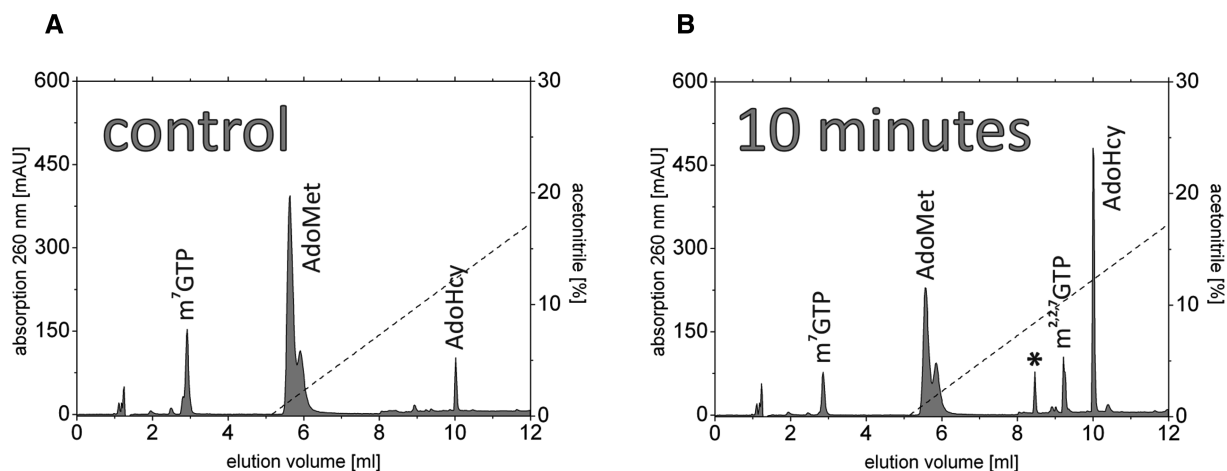
Figure 1. HPLC-based methyltransferase assay showing the catalytic activity of the crystallized hTGS1$_{618-853}$. (**A**) Chromatogram of the control reaction containing the substrates but no purified human hTGS1$_{618-853}$ and (**B**) chromatogram after 10 min of incubation at 37°C including the purified enzyme. The elution volume is plotted against the absorption at 260 nm (left *y*-axis) and the acetonitrile concentration (right *y*-axis), respectively. The corresponding peaks are labeled and the putative reaction intermediate m^{2,7}GTP is marked by an asterisk.

Figure 2. Overall structure of the human TGS1 methyltransferase domain. The crystallized methyltransferase domain comprising the residues Leu634-Asp844 is shown in cartoon representation and the bound substrate-analogs *S*-adenosyl-L-homocysteine (AdoHcy, red) and m^{7}GTP (blue) are depicted in ball-and-stick mode. The methyltransferase core (Glu675-Asp844) is colored in light green, whereas the N-terminal, α-helical extension (NTE; Leu634-Ser671) is shown in yellow. Both views are rotated by 65° with respect to each other around the indicated rotation axis. The secondary structure elements are labeled. The upper panel shows a surface representation of the active site cleft highlighting the enclosure of both ligands by the enzyme.

in a parallel and the last one (β-strand 7) in an anti-parallel orientation but positioned in between β-strand 5 and 6 (Figure 2). The β-sheet is twisted, thus the two outermost strands lie in an almost perpendicular orientation with respect to each other. Between two consecutive β-strands one or more α-helices build a right-handed crossover connection. Thus altogether three α-helices (α5, α6, α7) pack against one side of the β-sheet and another three helices (α8, α10, α11) on the opposing side. The α-helices adjacent to the β-sheet are oriented almost parallel to the strands, following the rotation impaired by this β-sheet. Using this conserved methyltransferase core as search model, the program DALI (49) found nearly 600 methyltransferases in the protein databank (PDB) with Z-scores greater than 12, indicating a closely related fold as discussed below.

The remaining four N-terminal α-helices of the structure (α1, α2, α3, α4) form a separate overall globular subdomain (Figure 2, *yellow part*) with a small

hydrophobic core. This small NTE packs against the core domain (Figure 2) and is involved in binding and recognition of both ligands. Interestingly, the residues Pro769 and Phe670 as well as Pro765 and Glu667 are in close proximity to each other and hence close the binding clefts on the top over both substrates as shown in the surface representation of Figure 2. A DALI-search using the NTE (aa 634–674) revealed that there are no structural related proteins or domains present in the protein data bank (PDB).

Substrate binding sites

The methyltransferase domain of human TGS1 harbors both, m^{7}GTP and AdoHcy bound in a pocket and they are held in close proximity to each other by an intricate pattern of hydrogen bonds. Either substrate is well defined in the $|F_o|-|F_c|$ electron density omit map as shown in Figure 3A.
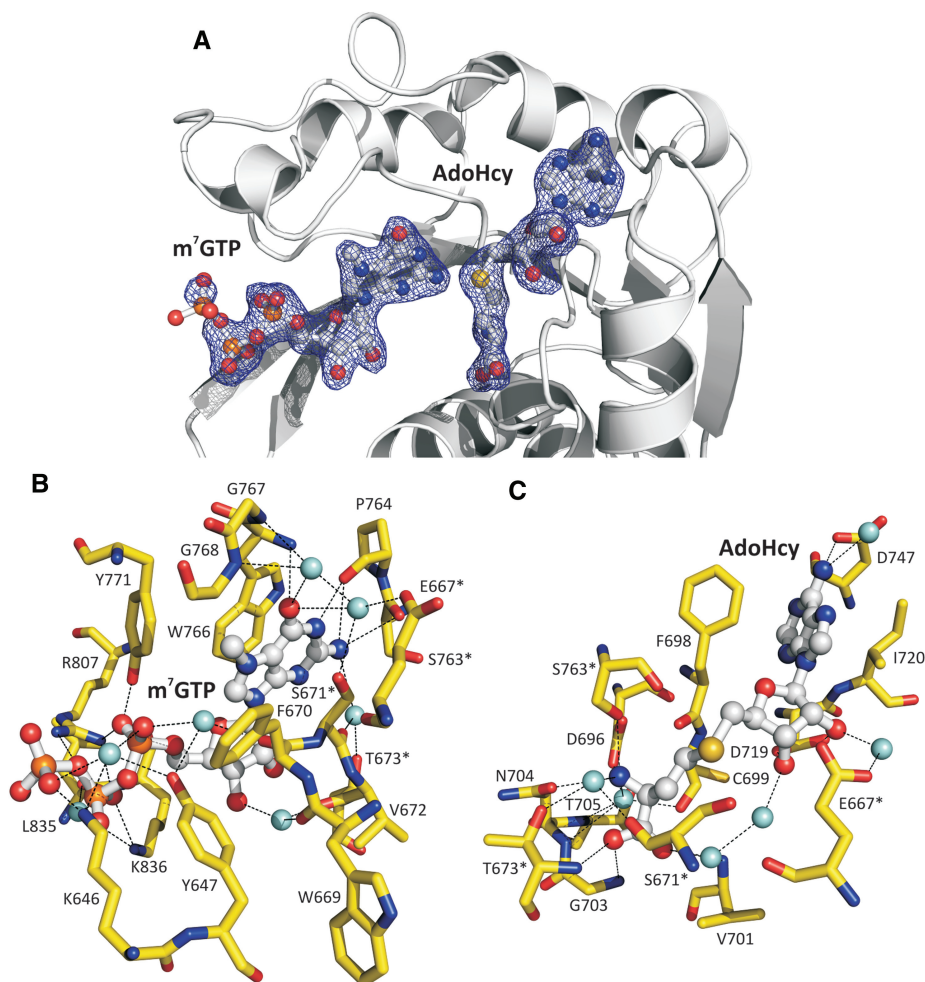
Figure 3. Binding of the minimal substrate 7-methylguanosine-triphosphate (m^{7}GTP) and the cofactor-analog *S*-adenosyl-L-homocysteine (AdoHcy) by hTGS1$_{618–853}$. (**A**) Overall binding of m^{7}GTP and AdoHcy (ball-and-stick mode) and relative orientation of the substrates to each other. The substrate molecules are surrounded by a $|F_o|-|F_c|$ electron density omit map contoured at 2.9σ (m^{7}GTP and AdoHcy were omitted). The helices α1–α4 were removed in this view for clarity reasons. (**B**) Detailed view of the binding-pockets for m^{7}GTP and (**C**) for AdoHcy. Protein residues involved in binding are labeled and drawn in stick mode (yellow) whereas the substrate-analogs are shown in ball-and-stick representation. Interactions of protein residues with water molecules (light blue) or substrate atoms (e.g. hydrogen bonds) are indicated by dashed lines. Amino acids participating in the binding of both ligands are marked by an asterisk.

The binding site for AdoHcy is mainly built up by residues belonging to the loops connecting β-strand 1 with α-helix 6, β-strand 2 with α-helix 7 and β-strand 3 with α-helix 8, respectively (Figures 2, 3A and C). The loop joining the methyltransferase core and the α-helix 4 of the NTE, as well as residues from the helix itself make additional contacts to AdoHcy. The adenine-moiety of AdoHcy is directly bound between the two side chains of Phe698 and Ile720 generating a tight hydrophobic interaction (Figure 3C). Furthermore, N6 of this adenine forms a hydrogen bond with Asp747 and a water molecule, respectively. The ribose of AdoHcy adopts a C1′-exo conformation in the crystal structure and its hydroxyls are bound by two hydrogen bonds each. One hydrogen bond is formed from both hydroxyls to each of the two oxygen atoms of Asp719. The other hydrogen bonds are mediated by two water molecules, while the water bound to the 2′-OH is further coordinated by Glu667. The homocysteine-moiety of AdoHcy, more precisely the amino and the carboxyl group, are also bound by an extended network of hydrogen bonds to Ser671, Thr673, Asp696, Cys699, Val701, Gly703, Asn704, Thr705 and Ser763, respectively (Figure 3C). Three water molecules additionally contribute to this interaction. Within this array of interactions, the AdoHcy is tightly bound in its binding pocket and suggests that the site of the methyl group present in AdoMet is oriented towards the N2 of the bound cap-guanine. The distance between the AdoHcy-sulfur and the methyl group acceptor N2 of the cap-guanine averages 4 Å, representing an ideal spacing for a nucleophilic attack of N2 on the AdoMet-methyl group.

The cap-binding pocket is located next to the AdoHcy binding site and in contrast to AdoHcy, the m^{7}GTP-ribose adopts a clear 2′-endo conformation (Figure 3A and B). The guanine of m^{7}GTP, which is located next to the homocysteine-moiety of AdoHcy, is perfectly stacked on Trp766. Both aromatic rings are oriented almost coplanar with a distance of 3.2 Å leading to a tight π–π interaction between them (Figure 3B). This base stacking is further stabilized by a cation–π interaction, due to the net positive electric charge of the guanine caused by the methylation of N7. The side chain of Trp766 is fixed by the side chain of Pro806 leading to a higher rigidity of its aromatic side chain (data not shown). On the opposing side of the m^{7}guanine the polar side chain of Ser671 limits the binding pocket. Further important contacts between the protein and the cap-analog involve all three phosphates α, β and γ, as well as the guanine nitrogens N2, N1 and the oxygen O6, generating the specificity for a guanine base. The carbonyl oxygen of Pro764 forms hydrogen bonds with N1 and N2, while the latter is further coordinated by a water molecule and the main chain carbonyl of Ser763. The O6 of m^{7}GTP is bound by the main chain amide of the stacking Trp766 as well as by two water molecules, which are held in position by a complex interaction network with Ser671, Gly767 and Gly768. Moreover, the γ-phosphate of the cap-analog is bound by Lys646 and a water molecule, whereas the β-phosphate is fixed by the same lysine residue and two additional interactions to atoms NH1 and NH2 of Arg807 (Figure 3B). The most pronounced interactions occur

with the α-phosphate, which is positioned by Tyr771, Arg807 and two water molecules, respectively.

Besides the two substrate-analogs, two molecules of the cryo protectant 1,2-propanediol are bound to the cap-analog m^{7}GTP and to the surface of the protein, respectively (data not shown). The first propylene glycol molecule binds via its hydroxyls to the β-phosphate and the 3′OH of the cap-nucleotide of all three molecules, whereas the second one is bound on the surface of the protein molecules 1 and 3 involving the residues Asp739 and Ala738, respectively. However, the interaction of 1,2-propanediol with m^{7}GTP does not interfere with its interaction to the protein. The two hydroxyls of propylene glycol forming the hydrogen bonds are expected to be replaced by water molecules in an aqueous solution lacking the cryo protectant.

Mutagenesis studies

The impact of residues involved in substrate binding was analyzed for the crystallized fragment (hTGS1$_{618-853}$). The residues Trp766, Ser763, Asp696 and Phe804, which participate in substrate binding or facilitate catalysis, were mutated and their influence on catalytic activity was measured using the HPLC-based activity assay. The effect of individual residues on catalytic activity of TGS1 is summarized in Table 2. For this purpose, the area of the m^{7}GTP-peak after 20 min of incubation with the accordant mutant protein was compared to the corresponding peak areas of a control reaction without enzyme (0% activity) and in the presence of the wild type enzyme (100% activity), respectively.

As expected, substitution of the stacking tryptophan by alanine (Trp766Ala) results in a complete loss of methyltransferase activity (<1% of wild-type activity). This observation underlines the ultimate importance of the aromatic indole ring being the exclusive cation–π and π–π stacking partner for the m^{7}G-cap in the binding pocket. Further on, replacement of serine 763, which was postulated to be a key residue in catalysis by alanine (Ser763Ala) roughly reduces the catalytic activity of the enzyme by a factor of five. A mutant containing an aspartate instead of serine 763 (Ser763Asp), as present in some other methyltransferases and the *G. lamblia* TGS2, shows almost wild-type activity confirming the hypothesis that a proton acceptor in this position enforces catalysis. When we tried to purify the mutant containing an alanine instead of aspartate 696 (Asp696Ala), the protein immediately precipitated subsequent to elution from the first

Table 2. Effect of selected amino acid mutations on catalytic activity of hTGS1$_{618-853}$

Mutation	Activity
wt (hTGS1$_{618-853}$)	100%
W766A	<1%
S763A	18%
S763D	97%
D696A	n/a
D696N	n/a
F804A	n/a
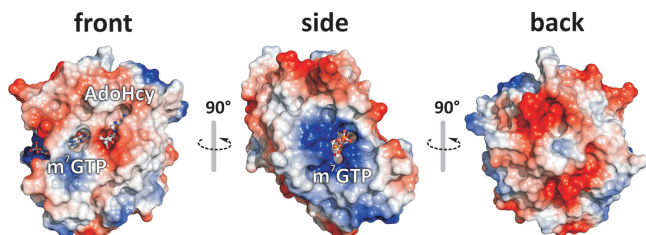
Figure 4. Charge distribution on the hTGS1 surface. Basic patches are colored in blue, while acidic ones are depicted in red. The three views are rotated by 90° with respect to each other. The front view has the same orientation as the left panel in Figure 2, while the side view shows the putative binding region for the ongoing UsnRNA backbone (blue patches). The back view reveals two acidic regions, representing potential interaction sites for the Sm-proteins B and D1 or other proteins.

GSH-sepharose-affinity column. Nevertheless, in order to test the effect of the lacking carboxyl group on the side chain of Asp696 without losing the structural integrity, we generated the more conservative mutant Asp696Asn. Surprisingly, this mutation showed the same effect on protein stability as the exchange against alanine, revealing the strict requirement of aspartate 696. The obvious structural importance therefore prevents the detailed analysis of the catalytic impact of Asp696. The analysis of replacement of Phe804 by an alanine or by a positively charged residue (e.g. lysine) was impeded by a related problem. Phe804 in combination with other residues is thought to form a hydrophobic pocket harboring the first transferred methyl group and therefore supporting the second methylation reaction (see below). However, replacement of this residue results in a significant reduction of bacterial cell growth during expression and mostly insoluble protein after cell lysis.

Potential protein and RNA interaction sites

In order to find potential binding sites of the hTGS1-methyltransferase domain for RNA nucleotides adjacent to the 5′-cap as well as for interacting proteins of the cognate RNPs, an electrostatic surface potential of the protein was calculated (Figure 4). The front view (left hand side) shows TGS1 in the same orientation as in Figure 2 (left-hand side) with the bound substrates. In the side view (middle) basic patches in close proximity to the m^7G-cap-binding pocket indicate the binding site for further RNA nucleotides. In fact, the first some ten nucleotides of the U1snRNP are accessible for interaction partners as demonstrated by RNAse H digestion experiments (13). The back view (right-hand side) reveals acidic patches of hTGS1 methyltransferase domain, which are likely to bind the positively charged C-terminal tails of Sm-proteins B and D1. The interaction of RNP specific proteins and TGS1 in human and yeast was investigated *in extenso* (26,34,48).

DISCUSSION

The RNA 5′-cap dimethyltransferases from diverse organisms have been biochemically characterized, however no crystal structure of an active enzyme has been available
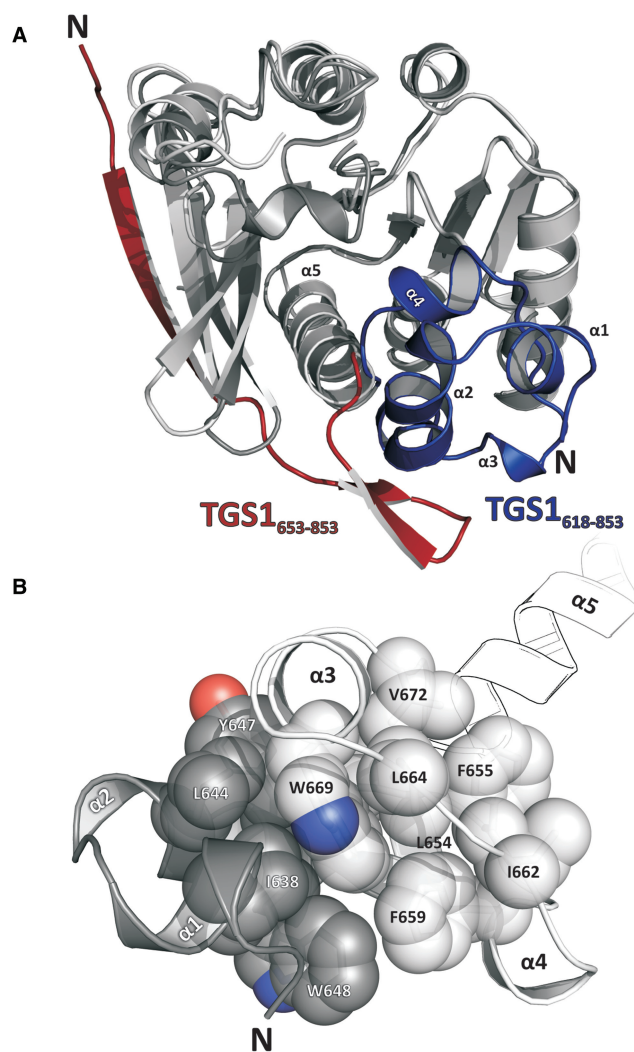
Figure 5. Superposition of the active (aa 618–853; NTE in blue) and inactive (aa 653–853; NTE in red; PDB ID 3EGI) hTGS1 methyltransferase domain (**A**) and detailed view on the hydrophobic core of the NTE of the active conformation (**B**). Hydrophobic residues of the NTE are drawn in space-filling spheres to clarify the core. Residues, that are not present in the inactive structure (Leu634-Arg651) and which contribute to the hydrophobic core are drawn in dark grey and labeled in white, while amino acids from 652 on are colored in white and labeled in black.

allowing an investigation of the structure-function relationship. In the present study we report the crystal structure of the active C-terminal methyltransferase domain of human m^7G-cap-specific dimethyltransferase TGS1 bound to the minimal substrate m^7GTP as well as the reaction product S-adenosyl-L-homocysteine (AdoHcy).

Recently, it was reported that the canonical and structurally conserved methyltransferase domain of hTGS1 is not sufficient for catalytic activity, but requires 17 additional residues to gain catalytic activity (44). Figure 5A shows the superposition of the active (hTGS1$_{618–853}$) and inactive (hTGS1$_{653–853}$) form of the hTGS1 methyltransferase domain. The structural reason for the loss of function of the shorter hTGS1 fragment is a significant change in the conformation of the N-terminal extension (NTE),

which forms a small α-helical subdomain in the active state (blue part in Figure 5), while it contains three β-strands in the inactive state (red part). The NTE is essential for m^{7}G-cap binding and as it completes the substrate binding pocket. Lys646, Tyr647, Glu667, Phe670, Ser671, Val672 and Thr673 of this NTE are involved in m^{7}G-cap binding (Figure 3). The NTE is mainly stabilized by its hydrophobic core consisting of the residues Ile638, Leu644, Trp648, Leu654, Phe655, Leu664, Gly668 and Trp669 (Figure 5B). The truncated NTE of hTGS1$_{653-853}$ adopts a different fold, since important residues of the hydrophobic core are missing, resulting in the formation of three β-strands instead of α-helices present in the NTE of the active form.

m^{7}G-cap-binding mode

To date several three-dimensional structures of m^{7}G-cap-binding proteins are known, clearly defining their binding characteristics for the positively charged, N7-methylated guanine of the 5′-cap structure. Among them are the eukaryotic initiation factor 4E (eIF4E) (50) the viral protein 39 (VP39) (51) and the small subunit of the cap-binding complex CBP20 (52) all of which show a tight stacking of the cap-N7-guanine between two aromatic side chains (eIF4E, Trp56/Trp102; VP39, Tyr22/Phe180; CBP20, Tyr20/Tyr43) (Figure 6). Differing thereof, the scavenger decapping enzyme DcpS (53) and the influenza virus polymerase subunit PB2 (54) stack the m^{7}G-cap only on one side by an aromate and by a non-aromatic side chain on the other (DcpS, Trp175/Leu206; PB2, Phe404/His357). Recently, it was shown that in the poly(A)-specific ribonuclease (PARN) a second stacking residue is completely missing and the N7-methylguanine is mainly bound by the aromate Trp475 (Figure 6) (55–57).

In order to gain insight into the cap binding mode of TGS1 the structural information of the protein Mj0882 (PDB ID 1DUS), which was shown to be a protein related to a family comprised of bacterial rRNA methyltransferases (RsmC/RsmD) was used to generate a homology model of yTGS1 (48). Based on this homology model of yTGS1, it was suggested that the m^{7}G-cap is stacked between the two side chains of Trp178 and Trp75 (which correspond to Trp766 and Trp669 in hTGS1). However, this cap binding mode remained questionable, as mutation of Trp75 in yTGS1 to an alanine did not decrease catalytic activity (48). Indeed the crystal structure

of hTGS1$_{618-853}$ reveals that only Trp766 (yTGS1 Trp178) stacks the m^{7}G-cap, while Trp669 (yTGS1 Trp75) only makes a water-mediated contact with the 3′OH of m^{7}GTP via the main chain carbonyl (Figure 3B). Moreover, the side chain of Trp669 is part of the hydrophobic core of the NTE, interacting for example with Leu664 and Ile638 (Figure 5B). The common second aromatic side chain sandwiching the m^{7}G-cap is replaced by the side chain of Ser671 that limits the cap-binding pocket spatially on the side opposing Trp766, allowing for some flexibility of the NTE (Figures 3B and 6). The structural data now explain the mutagenesis studies mentioned above.

Mutagenesis studies

The functional relevance of residues that are involved in substrate binding was analyzed by studying several single amino-acid mutants. As expected, the replacement of the stacking Trp766 (Trp766Ala) results in a complete loss of catalytic activity, whereas the exchange of another putative key residue in the active site, Ser763, has a milder effect. The mutation Ser763Ala decreases the activity to 18%. Contrary to our results the yTGS1 mutant Ser175Ala (corresponding to Ser763 in human TGS1) showed no defect in activity as shown indirectly by the immunoprecipitation of cellular RNA (48), while the corresponding exchange in the fruit fly TGS1 ortholog (DTL) was lethal to the larvae (35). The mutant Asp696Ala, as well as a more conservative mutation to an asparagine (Asp696Asn) resulted in insoluble and precipitating hTGS1, demonstrating the importance of this residue to maintain a structurally stable enzyme and an intact catalytic pocket. In contrast, Hausmann *et al.* reported the purification of this mutant and its activity to be reduced to below 1% (30). In this previous study additional six residues (Phe655, Thr673, Asn704, Asp719, Asn731, Arg807 and Asn808) predicted to be important for activity were mutated to alanine and characterized using a radioactive methylation assay (30). In summary, the results from this mutational analysis can be explained by the structure analysis of hTGS1$_{618-853}$. The residues Thr673 and Asn704 establish water mediated hydrogen bonds to the reaction product AdoHcy in the structure and their mutation to alanine reduces the activity to 13% and 5%, respectively. Asp719, which makes contacts to both ribose hydroxyls and therefore significantly contributes to
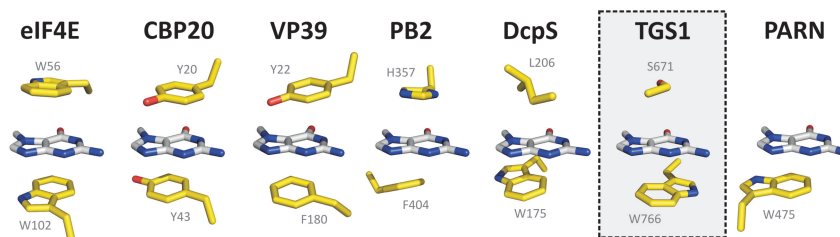

Figure 6. Comparison of the cap binding pockets of several cap binding proteins. The positively charged, N7-methylated guanine-moiety (light grey) is stacked in between or flanked by the individual, labeled protein residues (yellow). The interaction of m^{7}guanine with the eukaryotic initiation factor 4E (eIF4E; PDB ID 1L8B), the small subunit of the cap binding complex (CBP20; PDB ID 1H2T), the vaccinia virus protein 39 (VP39; PDB ID 1AV6), the influenza virus polymerase subunit PB2 (PDB ID 2VQZ), the scavenger decapping enzyme (DcpS; PDB ID 1ST0), hTGS1$_{618-853}$ and the poly(A)-specific ribonuclease (PARN; PDB ID 3CTR) is shown.

a binding of AdoHcy, reduces the catalytic activity to <1% when mutated to alanine. Both amines NH1 and NH2 of the guanidinium group of Arg807 form hydrogen bonds to the α- and the β-phosphate of the cap-analog, thus the mutation Arg807Ala results in a decrease of activity to 8% compared to the wild type protein. Replacement of Phe655, which is part of the hydrophobic core (Figure 5B), with alanine reduced the measured activity to less than 1% indicating its importance in the stabilization of the NTE and thus its connection to the methyltransferase core. Another important residue Asn731 is located in α-helix 7 and makes two contacts to the main chain carbonyl and amine of Lys663, which belongs to the NTE. The mutation Asn731Ala reduces the activity to 4% compared to the wild type protein. Located in the cleft between both domains it contributes significantly to the stability of their interaction and thus promotes the active conformation of the enzyme. A related function is fulfilled by the side chain Nδ of Asn808, which forms two hydrogen bonds to the main chain carbonyls of Ala774 (2.8 Å) and Tyr771 (3.2 Å) the latter of which is directly positioned in α-helix 9. Mutation of this asparagine to an alanine decreases the TGS1 activity to about 11%. This region encompassing the residues (Gly767 to Thr773) is not visible in the inactive structure of the $hTGS1_{653-853}$, which does not contain a correctly bound substrate. Thus the interactions between Asn808 and the mentioned carbonyls help to fix this region in a defined conformation competent to bind the cap-analog m^7GTP.

Catalytic mechanism

Catalysis by most methyltransferases is known to proceed by an S_N2 substitution reaction (58). On the basis of the presented structural and functional characterization of the human TGS1 methyltransferase domain in complex with substrate-analogs the following model for the catalytic mechanism of dimethylation by TGS1 is proposed. The exocyclic N2 of the m^7G performs a nucleophilic attack on the activated methyl group of the AdoMet. The distance of 4 Å between the AdoHcy sulfur and the cap N2 atoms observed in our crystal structure is consistent with that mechanism (59). In order to increase the nucleophilicity of the N2 its hybridization state has to be changed from sp^2 to sp^3. This transition might be stabilized by the main chain carbonyl oxygens of Ser763 and Pro764, which are in reasonable distance and a productive orientation (Figure 7). Such a mechanism was previously proposed for N6-adenine DNA methyltransferase M*Taq*I (60). Accordingly, the carbonyl oxygens of Pro764 and Ser763, which are both located out of the planar purine ring pull the hydrogens of N2 out of the conjugated system. In TGS1 the positive charge of the methylated guanine leads to an increased electron pull on N2 facilitating its deprotonation. This generates a lone pair of electrons that is no longer part of the conjugated system leading to an enhanced nucleophilicity of N2, which is required to attack the methyl group.

During this process the exocyclic NH_2 has to release a proton which most likely is accepted by a protein residue. A putative proton acceptor is the hydroxyl group of
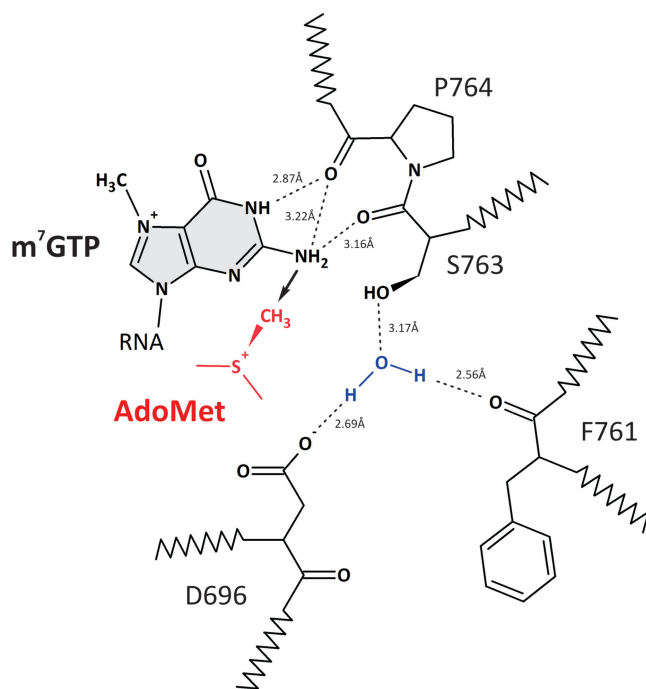
Figure 7. Proposed catalytic mechanism of methyltransfer by hTGS1. The methylation proceeds by an S_N2 substitution reaction. The exocyclic N2-amino group forms hydrogen bonds with the carbonyls of Ser763 and Pro764 of the motif $^{763}SPPW^{766}$ and the release of the hydrogen is further promoted by the net positive charge distributed over the whole methylated purine ring. Once shifted towards the tetrahedral state (sp^3) the N2 attacks the methyl group. After exchange of AdoHcy with a new AdoMet molecule the N2 methyl group is located outside of the ring plane and stabilized by a hydrophobic pocket formed by Trp766, Pro764 and Phe804. The second methylation then applies to the same mechanism as the first one with an additional activation of N2 by the methyl group. The proton release is facilitated further by the side chain of Ser763 connected to a water molecule, which on its own forms a hydrogen bond with the side chain carboxyl of Asp696 and the Phe761 carbonyl, respectively.

Ser763, which is, however, in a distance of 4.2 Å. Interestingly, a different rotamer of this side chain would reduce the distance to 2.5 Å, which would allow the direct proton transfer. In the observed conformation the serine hydroxyl group forms a hydrogen bond to a water molecule, which in turn is additionally bound to the carboxyl group of Asp696 and the main chain carbonyl group of Phe761. Therefore it appears likely that the proton is transferred via the serine side chain to the water molecule. However, this role of Ser763 is not essential, since its mutation to alanine reduces the activity to 18%. Furthermore, the exchange of Ser763 with an aspartate residue nearly retains catalytic activity compared to the wild type enzyme (97%). This mutation is observed in some other methyltransferases like in the *G. lamblia* TGS2 and a mimivirus methyltransferase (mimiTGS), altering the main catalytic motif from $^{763}SPPW^{766}$ to DPPW (42,61). In fact, for the tRNA N2, N2-guanosine dimethyltransferase Trm1 it was postulated that the negatively charged carboxyl group of the aspartate in this motif is the general base deprotonating the guanine N2 (62).

After the first methylation the reaction products AdoHcy and $m^{2,7}$G-cap dissociate from the active site, allowing the binding of two new substrates. The assumption that the dimethylation is not a processive reaction is supported by fact that *S. pombe* TGS1 under m^7GDP-excess conditions converts almost exclusively m^7GDP to $m^{2,7}$GDP. Only when the educt m^7GDP was limiting, the enzyme converted the reaction intermediate further to the end product $m^{2,2,7}_{(3)}$GDP in a pulse-chase experiment (27). After binding of the intermediate $m^{2,7}$GTP, the second methylation presumably applies to the same mechanism. The N2 is activated additionally by the first bound methyl group located outside of the plane of the guanine ring. This methyl group would fit in an adjacent hydrophobic pocket formed by Trp766, Pro764 and Phe804 (data not shown). The activated N2 attacks the newly bound AdoMet-methyl group, leading to the second methylation reaction, which results in the fully modified $m^{2,2,7}$G-cap.

Comparison with structurally and functionally related methyltransferases

The three-dimensional structures of several other RNA-modifying methyltransferases targeting exocyclic amino groups of nucleobases are known. Among them are the rRNA-adenine dimethyltransferase KsgA (63), the rRNA guanine methyltransferase RsmC (64) and the tRNA-guanine-N2 dimethyltransferase Trm1 (62). The methyltransferase core domains of KsgA, RsmC, Trm1 and hTGS1, encompassing the canonical αβα sandwich, are structurally very similar, whereas the remaining parts of these enzymes differ significantly.

RsmC contains two separate methyltransferase domains, with only the C-terminal one binding the substrate-guanosine and AdoMet in the crystal structure (64). Unlike hTGS1, RsmC contains the four residues ^{305}NPPF^{308} in both the N-terminal and the C-terminal methyltransferase domains. RsmC binds the guanosine mainly by its stacking interaction with the aromatic side chain of Phe308. However, the distance of 4.7Å between the AdoMet methyl group and the exocyclic target amino group suggests that the substrates are in a non-productive orientation.

Interestingly, the *Pyrococcus horikoshii* Trm1 contains a C-terminal extension, which is proposed to be involved in tRNA-binding and composed of six α-helices and six β-strands forming the three subdomains C1–C3 in addition to the conserved αβα-sandwich. The N-terminus contains two additional β-strands generating a small β-sheet enclosing α-helix 1, which represents the first helix of the methyltransferase domain and a two β-strand insertion between α-helix 3 and β-strand 7 elongating the present β-sheet to nine strands.

The four related methyltransferases KsgA, RsmC, Trm1 and hTGS1 bind the cofactor AdoMet or its analog AdoHcy in a structurally conserved binding site, but the residues that contribute to the binding of AdoMet are not strictly conserved. The common catalytic motif of KsgA, RsmC and hTGS1, which has the sequence ^{1}(N/D/S)-^{2}(P/L)-^{3}(P)-^{4}(W/F/Y), is altered in Trm1 as it lacks the second residue. The first residue of this motif is involved in the activation of the methyl group acceptor, which is held in position by the two main chain carbonyl oxygens of the first and second residues. The aromatic side chain of the last motif residue is generally believed to bind the methyl-acceptor nucleobase by stacking interaction as experimentally demonstrated for RsmC and hTGS1.

In summary, our structural studies on hTGS1 revealed the necessity of an unexpected N-terminal and helical extension required for m^7G-cap binding and catalytic activity in addition to the canonical methyltransferase domain.

ACCESSION NUMBER

Accession number: PDB ID: 3GDH.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

FUNDING

REFERENCES

1. Dickmanns,A. (2009) Import and Export of snRNPs. In Kehlenbach,R. (ed.), *Nuclear Transport*. Landes Bioscience, Austin, USA.
2. Dickmanns,A. and Ficner,R. (2005) Role of the 5′-cap in the biogenesis of spliceosomal snRNPs. *Top. Curr. Genet.*, **12**, 179–204.
3. Cougot,N., van Dijk,E., Babajko,S. and Seraphin,B. (2004) Cap-tabolism. *Trends Biochem. Sci.*, **29**, 436–444.
4. Hernandez,N. (2001) Small nuclear RNA genes: a model system to study fundamental mechanisms of transcription. *J. Biol. Chem.*, **276**, 26733–26736.
5. Kambach,C., Walke,S. and Nagai,K. (1999) Structure and assembly of the spliceosomal small nuclear ribonucleoprotein particles. *Curr. Opin. Struct. Biol.*, **9**, 222–230.
6. Kambach,C., Walke,S., Young,R., Avis,J.M., de la Fortelle,E., Raker,V.A., Luhrmann,R., Li,J. and Nagai,K. (1999) Crystal structures of two Sm protein complexes and their implications for the assembly of the spliceosomal snRNPs. *Cell*, **96**, 375–387.

7. Plessel,G., Fischer,U. and Luhrmann,R. (1994) m3G cap hyper-methylation of U1 small nuclear ribonucleoprotein (snRNP) in vitro: evidence that the U1 small nuclear RNA-(guanosine-N2)-methyltransferase is a non-snRNP cytoplasmic protein that requires a binding site on the Sm core domain. *Mol. Cell Biol.*, **14**, 4160–4172.

8. Mattaj,I.W. (1986) Cap trimethylation of U snRNA is cytoplasmic and dependent on U snRNP protein binding. *Cell*, **46**, 905–911.

9. Massenet,S., Pellizzoni,L., Paushkin,S., Mattaj,I.W. and Dreyfuss,G. (2002) The SMN complex is associated with snRNPs throughout their cytoplasmic assembly pathway. *Mol. Cell Biol.*, **22**, 6533–6541.

10. Raker,V.A., Plessel,G. and Luhrmann,R. (1996) The snRNP core assembly pathway: identification of stable core protein heteromeric complexes and an snRNP subcore particle in vitro. *EMBO J.*, **15**, 2256–2269.

11. Fischer,U., Sumpter,V., Sekine,M., Satoh,T. and Luhrmann,R. (1993) Nucleo-cytoplasmic transport of U snRNPs: definition of a nuclear location signal in the Sm core domain that binds a trans-port receptor independently of the m3G cap. *EMBO J.*, **12**, 573–583.

12. Hamm,J., Darzynkiewicz,E., Tahara,S.M. and Mattaj,I.W. (1990) The trimethylguanosine cap structure of U1 snRNA is a component of a bipartite nuclear targeting signal. *Cell*, **62**, 569–577.

13. Huber,J., Cronshagen,U., Kadokura,M., Marshallsay,C., Wada,T., Sekine,M. and Luhrmann,R. (1998) Snurportin1, an m3G-cap-specific nuclear import receptor with a novel domain structure. *EMBO J.*, **17**, 4114–4126.

14. Mattaj,I.W. and De Robertis,E.M. (1985) Nuclear segregation of U2 snRNA requires binding of specific snRNP proteins. *Cell*, **40**, 111–118.

15. Narayanan,U., Achsel,T., Luhrmann,R. and Matera,A.G. (2004) Coupled in vitro import of U snRNPs and SMN, the spinal muscular atrophy protein. *Mol. Cell*, **16**, 223–234.

16. Narayanan,U., Ospina,J.K., Frey,M.R., Hebert,M.D. and Matera,A.G. (2002) SMN, the spinal muscular atrophy protein, forms a pre-import snRNP complex with snurportin1 and importin beta. *Hum. Mol. Genet.*, **11**, 1785–1795.

17. Palacios,I., Hetzer,M., Adam,S.A. and Mattaj,I.W. (1997) Nuclear import of U snRNPs requires importin beta. *EMBO J.*, **16**, 6783–6792.

18. Strasser,A., Dickmanns,A., Luhrmann,R. and Ficner,R. (2005) Structural basis for m3G-cap-mediated nuclear import of spliceosomal UsnRNPs by snurportin1. *EMBO J.*, **24**, 2235–2243.

19. Huber,J., Dickmanns,A. and Luhrmann,R. (2002) The importin-beta binding domain of snurportin1 is responsible for the Ran- and energy-independent nuclear import of spliceosomal U snRNPs in vitro. *J. Cell Biol.*, **156**, 467–479.

20. Will,C.L. and Luhrmann,R. (2001) Spliceosomal UsnRNP biogenesis, structure and function. *Curr. Opin. Cell Biol.*, **13**, 290–301.

21. Krol,A., Carbon,P., Ebel,J.P. and Appel,B. (1987) Xenopus tropicalis U6 snRNA genes transcribed by Pol III contain the upstream promoter elements used by Pol II dependent U snRNA genes. *Nucleic Acids Res.*, **15**, 2463–2478.

22. Kunkel,G.R., Maser,R.L., Calvet,J.P. and Pederson,T. (1986) U6 small nuclear RNA is transcribed by RNA polymerase III. *Proc. Natl Acad. Sci. USA*, **83**, 8575–8579.

23. Reddy,R., Henning,D., Das,G., Harless,M. and Wright,D. (1987) The capped U6 small nuclear RNA is transcribed by RNA polymerase III. *J. Biol. Chem.*, **262**, 75–81.

24. Sleeman,J.E., Ajuh,P. and Lamond,A.I. (2001) snRNP protein expression enhances the formation of Cajal bodies containing p80-coilin and SMN. *J. Cell Sci.*, **114**, 4407–4419.

25. Sleeman,J.E. and Lamond,A.I. (1999) Newly assembled snRNPs associate with coiled bodies before speckles, suggesting a nuclear snRNP maturation pathway. *Curr. Biol.*, **9**, 1065–1074.

26. Mouaikel,J., Verheggen,C., Bertrand,E., Tazi,J. and Bordonne,R. (2002) Hypermethylation of the cap structure of both yeast snRNAs and snoRNAs requires a conserved methyltransferase that is localized to the nucleolus. *Mol. Cell*, **9**, 891–901.

27. Hausmann,S. and Shuman,S. (2005) Specificity and mechanism of RNA cap guanine-N2 methyltransferase (Tgs1). *J. Biol. Chem.*, **280**, 4021–4024.

28. Colau,G., Thiry,M., Leduc,V., Bordonne,R. and Lafontaine,D.L. (2004) The small nucle(ol)ar RNA cap trimethyltransferase is required for ribosome synthesis and intact nucleolar morphology. *Mol. Cell Biol.*, **24**, 7976–7986.

29. Zhu,Y., Qi,C., Cao,W.Q., Yeldandi,A.V., Rao,M.S. and Reddy,J.K. (2001) Cloning and characterization of PIMT, a protein with a methyltransferase domain, which interacts with and enhances nuclear receptor coactivator PRIP function. *Proc. Natl Acad. Sci. USA*, **98**, 10380–10385.

30. Hausmann,S., Zheng,S., Costanzo,M., Brost,R.L., Garcin,D., Boone,C., Shuman,S. and Schwer,B. (2008) Genetic and biochemical analysis of yeast and human cap trimethylguanosine synthase: Functional overlap of TMG caps, snRNP components, pre-mRNA splicing factors, and RNA decay pathways. *J. Biol. Chem.*, **283**, 31706–31718.

31. Misra,P., Qi,C., Yu,S., Shah,S.H., Cao,W.Q., Rao,M.S., Thimmapaya,B., Zhu,Y. and Reddy,J.K. (2002) Interaction of PIMT with transcriptional coactivators CBP, p300, and PBP differential role in transcriptional regulation. *J. Biol. Chem.*, **277**, 20011–20019.

32. Enunlu,I., Papai,G., Cserpan,I., Udvardy,A., Jeang,K.T. and Boros,I. (2003) Different isoforms of PRIP-interacting protein with methyltransferase domain/trimethylguanosine synthase localizes to the cytoplasm and nucleus. *Biochem. Biophys. Res. Commun.*, **309**, 44–51.

33. Girard,C., Verheggen,C., Neel,H., Cammas,A., Vagner,S., Soret,J., Bertrand,E. and Bordonne,R. (2008) Characterization of a short isoform of human Tgs1 hypermethylase associating with small nucleolar ribonucleoprotein core proteins and produced by limited proteolytic processing. *J. Biol. Chem.*, **283**, 2060–2069.

34. Mouaikel,J., Narayanan,U., Verheggen,C., Matera,A.G., Bertrand,E., Tazi,J. and Bordonne,R. (2003) Interaction between the small-nuclear-RNA cap hypermethylase and the spinal muscular atrophy protein, survival of motor neuron. *EMBO Rep.*, **4**, 616–622.

35. Komonyi,O., Papai,G., Enunlu,I., Muratoglu,S., Pankotai,T., Kopitova,D., Maroy,P., Udvardy,A. and Boros,I. (2005) DTL, the Drosophila homolog of PIMT/Tgs1 nuclear receptor coactivator-interacting protein/RNA methyltransferase, has an essential role in development. *J. Biol. Chem.*, **280**, 12397–12404.

36. Terns,M.P. and Dahlberg,J.E. (1994) Retention and 5′ cap trimethylation of U3 snRNA in the nucleus. *Science*, **264**, 959–961.

37. Watkins,N.J., Lemm,I., Ingelfinger,D., Schneider,C., Hossbach,M., Urlaub,H. and Luhrmann,R. (2004) Assembly and maturation of the U3 snoRNP in the nucleoplasm in a large dynamic multiprotein complex. *Mol. Cell*, **16**, 789–798.

38. Terns,M.P., Grimm,C., Lund,E. and Dahlberg,J.E. (1995) A common maturation pathway for small nucleolar RNAs. *EMBO J.*, **14**, 4860–4871.

39. Franke,J., Gehlen,J. and Ehrenhofer-Murray,A.E. (2008) Hypermethylation of yeast telomerase RNA by the snRNA and snoRNA methyltransferase Tgs1. *J. Cell Sci.*, **121**, 3553–3560.

40. Gunzl,A., Bindereif,A., Ullu,E. and Tschudi,C. (2000) Determinants for cap trimethylation of the U2 small nuclear RNA are not conserved between Trypanosoma brucei and higher eukaryotic organisms. *Nucleic Acids Res.*, **28**, 3702–3709.

41. Hausmann,S., Ramirez,A., Schneider,S., Schwer,B. and Shuman,S. (2007) Biochemical and genetic analysis of RNA cap guanine-N2 methyltransferases from Giardia lamblia and Schizosaccharomyces pombe. *Nucleic Acids Res.*, **35**, 1411–1420.

42. Hausmann,S. and Shuman,S. (2005) Giardia lamblia RNA cap guanine-N2 methyltransferase (Tgs2). *J. Biol. Chem.*, **280**, 32101–32106.

43. Ruan,J.P., Ullu,E. and Tschudi,C. (2007) Characterization of the Trypanosoma brucei cap hypermethylase Tgs1. *Mol. Biochem. Parasitol.*, **155**, 66–69.

44. Monecke,T., Dickmanns,A., Strasser,A. and Ficner,R. (2009) Structure analysis of the conserved methyltransferase domain of human trimethylguanosine synthase TGS1. *Acta Crystallogr. D Biol. Crystallogr.*, **65**, 332–338.

45. McCoy,A.J. (2007) Solving structures of protein complexes by molecular replacement with Phaser. *Acta Crystallogr. D Biol. Crystallogr.*, **63**, 32–41.

46. Emsley,P. and Cowtan,K. (2004) Coot: model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.*, **60**, 2126–2132.
47. 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.
48. Mouaikel,J., Bujnicki,J.M., Tazi,J. and Bordonne,R. (2003) Sequence-structure-function relationships of Tgs1, the yeast snRNA/snoRNA cap hypermethylase. *Nucleic Acids Res.*, **31**, 4899–4909.
49. Holm,L., Kaariainen,S., Rosenstrom,P. and Schenkel,A. (2008) Searching protein structure databases with DaliLite v.3. *Bioinformatics*, **24**, 2780–2781.
50. Niedzwiecka,A., Marcotrigiano,J., Stepinski,J., Jankowska-Anyszka,M., Wyslouch-Cieszynska,A., Dadlez,M., Gingras,A.C., Mak,P., Darzynkiewicz,E., Sonenberg,N. *et al.* (2002) Biophysical studies of eIF4E cap-binding protein: recognition of mRNA 5′ cap structure and synthetic fragments of eIF4G and 4E-BP1 proteins. *J. Mol. Biol.*, **319**, 615–635.
51. Hodel,A.E., Gershon,P.D. and Quiocho,F.A. (1998) Structural basis for sequence-nonspecific recognition of 5′-capped mRNA by a cap-modifying enzyme. *Mol. Cell*, **1**, 443–447.
52. Mazza,C., Segref,A., Mattaj,I.W. and Cusack,S. (2002) Large-scale induced fit recognition of an m(7)GpppG cap analogue by the human nuclear cap-binding complex. *EMBO J.*, **21**, 5548–5557.
53. Gu,M., Fabrega,C., Liu,S.W., Liu,H., Kiledjian,M. and Lima,C.D. (2004) Insights into the structure, mechanism, and regulation of scavenger mRNA decapping activity. *Mol. Cell*, **14**, 67–80.
54. Guilligay,D., Tarendeau,F., Resa-Infante,P., Coloma,R., Crepin,T., Sehr,P., Lewis,J., Ruigrok,R.W., Ortin,J., Hart,D.J. *et al.* (2008) The structural basis for cap binding by influenza virus polymerase subunit PB2. *Nat. Struct. Mol. Biol.*, **15**, 500–506.
55. Monecke,T., Schell,S., Dickmanns,A. and Ficner,R. (2008) Crystal structure of the RRM domain of poly(A)-specific ribonuclease reveals a novel m(7)G-cap-binding mode. *J. Mol. Biol.*, **382**, 827–834.
56. Nagata,T., Suzuki,S., Endo,R., Shirouzu,M., Terada,T., Inoue,M., Kigawa,T., Kobayashi,N., Guntert,P., Tanaka,A. *et al.* (2008) The RRM domain of poly(A)-specific ribonuclease has a noncanonical binding site for mRNA cap analog recognition. *Nucleic Acids Res.*, **36**, 4754–4767.
57. Wu,M., Nilsson,P., Henriksson,N., Niedzwiecka,A., Lim,M.K., Cheng,Z., Kokkoris,K., Virtanen,A. and Song,H. (2009) Structural basis of m(7)GpppG binding to poly(A)-specific ribonuclease. *Structure*, **17**, 276–286.
58. Schubert,H.L., Blumenthal,R.M. and Cheng,X. (2003) Many paths to methyltransfer: a chronicle of convergence. *Trends Biochem. Sci.*, **28**, 329–335.
59. Zheng,Y.J. and Bruice,T.C. (1997) A theoretical examination of the factors controlling the catalytic efficiency of a transmethylation enzyme: catechol O-methyltransferase. *J. Am. Chem. Soc.*, **119**, 8137–8145.
60. Goedecke,K., Pignot,M., Goody,R.S., Scheidig,A.J. and Weinhold,E. (2001) Structure of the N6-adenine DNA methyltransferase M.TaqI in complex with DNA and a cofactor analog. *Nat. Struct. Mol. Biol.*, **8**, 121–125.
61. Benarroch,D., Qiu,Z.R., Schwer,B. and Shuman,S. (2009) Characterization of a mimivirus RNA cap guanine-N2 methyltransferase. *RNA*, **15**, 666–674.
62. Ihsanawati, 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.
63. O'Farrell,H.C., Scarsdale,J.N. and Rife,J.P. (2004) Crystal structure of KsgA, a universally conserved rRNA adenine dimethyltransferase in Escherichia coli. *J. Mol. Biol.*, **339**, 337–353.
64. 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.