RNA methyltransferases involved in 5′ cap biosynthesis

Magdalena Byszewska¹, Miroslaw Śmietański¹, Elżbieta Purta¹, and Janusz M Bujnicki¹,²,*

¹International Institute of Molecular and Cell Biology; Warsaw, Poland; ²Institute of Molecular Biology and Biotechnology; Faculty of Biology; Adam Mickiewicz University; Poznan, Poland

*These authors contributed equally to the work.

In eukaryotes and viruses that infect them, the 5′ end of mRNA molecules, and also many other functionally important RNAs, are modified to form a so-called cap structure that is important for interactions of these RNAs with many nuclear and cytoplasmic proteins. The RNA cap has multiple roles in gene expression, including enhancement of RNA stability, splicing, nucleocytoplasmic transport, and translation initiation. Apart from guanosine addition to the 5′ end in the most typical cap structure common to transcripts produced by RNA polymerase II (in particular mRNA), essentially all cap modifications are due to methylation. The complexity of the cap structure and its formation can range from just a single methylation of the unprocessed 5′ end of the primary transcript, as in mammalian U6 and 7SK, mouse B2, and plant U3 RNAs, to an elaborate m⁷Gppm⁶⁶AmpAmpCmpm⁶Um structure at the 5′ end of processed RNA in trypanosomases, which are formed by as many as 8 methylation reactions. While all enzymes responsible for methylation of the cap structure characterized to date were found to belong to the same evolutionarily related and structurally similar Rossmann Fold Methyltransferase superfamily, that uses the same methyl group donor, S-adenosylmethionine; the enzymes also exhibit interesting differences that are responsible for their distinct functions. This review focuses on the evolutionary classification of enzymes responsible for cap methylation in RNA, with a focus on the sequence relationships and structural similarities and dissimilarities that provide the basis for understanding the mechanism of biosynthesis of different caps in cellular and viral RNAs. Particular attention is paid to the similarities and differences between methyltransferases from human cells and from human pathogens that may be helpful in the development of antiviral and antiparasitic drugs.

Introduction

Nascent transcripts produced by RNA polymerases universally carry a 5′ triphosphate (5′ppp). Processed RNA molecules, such as rRNAs and tRNAs, generated from precursors whose 5′ segments were removed by nucleolytic cleavage, carry a 5′ monophosphate (5′p). In several types of cellular and viral RNAs, the 5′ end is further modified enzymatically, by a variety of modification enzymes, to introduce various chemical structures that are collectively dubbed as the “5′ caps”. This cap is absent in bacterial and archaeal transcripts.

The most typical and widely studied cap modification comprises the addition of an N⁷-methylguanosine (m⁷G) linked via an inverted 5′–5′ triphosphate bridge to the 5′-terminal nucleotide of the transcript.¹ This structure termed cap0 is a characteristic feature of transcripts that are produced by RNA polymerase II, such as messenger RNAs (mRNAs) of all eukaryotic organisms and many viral RNAs. It is typically introduced in sequential steps: (1) hydrolysis of 5′ γ-phosphate of a nascent pre-mRNA to generate a 5′ diphosphate mRNA end; (2) transfer of a guanine monophosphate nucleoside; and (3) methylation of the guanine at the N⁷ position. The cap0 structure was shown to be essential for cell growth of Saccharomyces cerevisiae² and survival of mammalian cells;³ it is critical for mRNA interactions with many nuclear and cytoplasmic proteins and has multiple important roles in gene expression, including enhancement of RNA stability, splicing, nucleocytoplasmic transport, and translation initiation.⁴⁵ Enzymes responsible for cap0 formation have been well characterized in many organisms and viruses.

In many instances, m⁷G-capped RNAs are modified further, in particular by additional methylation steps at the cap0 guanosine or methylation of the first few transcribed nucleoside residues. For instance, the cap0 guanosine is modified by addition of 2 methyl groups at the N⁷ position, yielding a trimethylguanosine (m²,⁷,²⁷G or TMG) cap, in some small nuclear RNAs (snRNAs) and nucleolar RNAs ( snoRNAs) required for pre-mRNA splicing (e.g., U1, U2, U4, and U5), pre-rRNA processing (U3 and U8), and telomere addition (telomerase RNA), as well as in several selenoprotein mRNAs.⁶

In higher eukaryotes, the 5′ ends of mRNA and snRNA are modified further by ribose 2′-O-methylation on the first and
second transcribed nucleosides, yielding cap1 and cap2 modifications, respectively. In humans, cap0 and cap1 methylations are found on all mRNA molecules, while about half of the capped and polyadenylated RNA molecules contain a 2'-O-methylated residue at the second transcribed position. The U1, U2, U4, and U5 snRNAs are methylated at both the first 2 positions. Cap1 and cap2 methylations in U2 snRNA are required for the formation of spliceosomal E-complex and, as a consequence, for efficient pre-mRNA splicing. In some organisms, such as in Trypanosomes, as many as 4 first residues of the nascent transcript undergo ribose methylation, to generate the cap4 structure. These additional methylation steps are often important for RNA processing, translation and stability, although their role has not been fully elucidated.

Alternative capping pathways have been invented by certain viruses. For instance, in Alphaviruses, the precursor of cap0 is first methylated to a m7G triphosphate and only then connected to the 5' end of the RNA. Nonsegmented negative-sense (nsNS) RNA viruses have evolved a different mechanism for mRNA cap formation in that the guanylyltransferase transfers GDP rather than GMP onto the 5' end of the RNA and the resulting cap structure is first monomethylated on the ribose of the first transcribed residue (yielding GpppAm structure), and only later the guanosine is methylated to m7G.

In addition to ribose 2'-O-methylation, base moieties of the first transcribed nucleosides may be methylated, thereby increasing the catalog of 5' cap structures. In particular, the first adenine nucleoside of the transcript is often methylated at the N6 position. In Trypanosomes, the fourth uridine residue is also methylated at the N6 position. The role of these base modifications is unclear, and the enzymes responsible for these modifications remain to be characterized.

Some small RNAs, including mammalian U6 and 7SK, mouse B2, and plant U3, present a completely different 5' structure, which is chemically minimalistic compared to the elaborate guanosine cap. This alternative cap is generated by methylation of a γ-phosphate oxygen at the unprocessed 5' end of the primary transcript.

It is clear that apart from guanosine addition to the 5' end, essentially all cap modifications are due to methylations. The cap structure of mRNAs in trypanosomes, m7Gpppm6AmppAmpCmpm3Um, is formed with as many as 8 methylation steps. In all cases that have been experimentally characterized to date, methylaitions of caps in all organisms and viruses are catalyzed by S-adenosyl-l-methionine (SAM)-dependent methyltransferases (Table 1). For the most common types of methylation reactions implicated in cap modification, the crystal structures of representative proteins have been determined (Table 2). All cap methyltransferases characterized structurally belong to the Rossmann Fold Methyltransferase (RFM) superfamily. The topology of the RFM fold is very similar to the typical Rossmann fold (β6-β5-α4-β1-α2-β3), with an additional, 7th β-strand inserted into the sheet in an antiparallel manner (β6-α1-β5-α4-β1-α2-β3) (Fig. 1). The methyl group donor (SAM) binding site is formed by loops following strands 1, 2, and 3, while the substrate to be methylated is typically bound by loops following strands 4, 5, and 6. Various families of RFM enzymes exhibit fusions with other domains, extensions of termini, and insertions within the conserved RFM domain, in particular following strand 5. These elaborations of the common fold are often involved in substrate binding or in oligomerization.

In this review, we discuss cellular and viral methyltransferases involved in 5' cap RNA biosynthesis, with emphasis on the sequence structure relationships in the light of the experimentally determined structures of enzymes complexed with their ligands. We focus on comparison of enzymes with similar activities that generate products with chemically similar structures.

Throughout the article, we follow the nomenclature of cap modifying enzymes and their products commonly used in the literature. We use terms “capX methylation” and “capX methyltransferase” (where X is a number) to refer to some enzymatic activity or its product at a particular position X. On the other hand, the term “capX structure” is used to refer to a fully modified cap structure. For instance “cap2 structure” indicates a cap methylated on the inverted guanosine and the first 2 ribose sugars of the nucleotide sequence; i.e., m7GpppNm2Nm3.

**Cap-specific m7G methyltransferases**

Cellular and viral RNA cap guanine-N7-methyltransferases methylate RNA with the GpppN cap0 structure. Many viruses, however, replicate in the cytoplasm of their eukaryotic host, and the cellular capping machinery is not accessible for their RNAs; hence, these viruses have evolved their own capping enzymes to form a cap structure that can be recognized by the cellular translation machinery for gene expression. Examples include Flaviviridae, Nidovirales, Mononegavirales and Poxviridae (reviewed in ref. 12). While the cellular and viral mRNA capping apparatus is functionally similar, the enzyme organization differs greatly across evolution.

The Abd1 protein from *Saccharomyces cerevisiae* is a monofunctional cap0 methyltransferase, and biochemically has been one of the best studied methyltransferases involved in cap structure biosynthesis. Its enzymatic activity is critical for yeast cell growth and the gene *ABD1* that encodes the Abd1 protein is essential. Abd1 is a founding member of a protein family that is strongly conserved in eukaryotes as well as in viruses. The crystal structure of *S. cerevisiae* Abd1 itself could not be determined, but eventually it was solved for its homolog, the Ecm1 protein from *E. cuniculi* (Fig. 2). Purified Ecm1 is a monomeric protein that catalyzes methyl transfer to GpppRNA to form cap0, but also to free mononucleotides GTP, GDP or dGTP (deoxy-GTP). The methyltransferase domain in Ecm1, and by inference also in other homologous cap0 methyltransferases, exhibits the RFM fold with a characteristic insertion that forms a characteristic β-meander structure involved in the formation of the cap- binding site. This insertion is common between methyltransferases that methylate G to m7G in the RNA cap, and
methyltransferases that N-methylate the amino acid glycine. This relationship between 2 different types of methyltransferases, as well as the cap0 methyltransferase structure, were correctly predicted using bioinformatics before the first structure of the cap0 methyltransferase was determined.

In the human capping system, the cap0 methyltransferase (RNMT) consists of a catalytic subunit related to Abd1 and an obligate activating subunit, RAM (RNMT-activating miniprotein). The C-terminal catalytic domain of RNMT has essentially the same structure as Abd1. RNMT also has an N-terminal domain that is conserved in mammals, but not required for catalytic activity. However, it contains 2 nuclear localization signal motifs and the nuclear localization of RNMT is essential for cell viability. The cap0 methyltransferases, members of the above-mentioned family, were also identified and characterized in other eukaryotes, including TbCmt1 in Trypanosoma brucei, for example.

As mentioned above, the viral cap0 methyltransferases possess a catalytic domain that is closely related to the eukaryotic cap0 methyltransferases, but it often functions in the context of other domains. For instance, the vaccinia virus possesses an enzyme that is composed of D1 and D12 polypeptides that execute all 3 steps in cap0 biosynthesis. The D1 subunit contains triphosphatase and guanylyltransferase activities in the N-terminal domain, and a cap0 methyltransferase domain that forms a heterodimer with the D12 subunit. The methyltransferase active site is located entirely in the D1 subunit and has a weak cap0 modification activity that is stimulated allosterically by D12. Interestingly, the D12 structure resembles a degenerate cap'0-ribose methyltransferase domain (see below), but it lacks a proper SAM binding site and does not show any methyltransferase activity on its own.

In the SARS-coronavirus, a nonstructural protein 14 (nsP14) was initially identified as an exoribonuclease (and termed ExoN). Later, it was shown that it also exhibits cap0 methyltransferase activity. Analysis of protein variants with substitutions of conserved residues in the ExoN (N-terminal) and methyltransferase (C-terminal) domains revealed that both active sites are functionally distinct; however, the integrity of the ExoN domain turned out to be essential for the function of the cap0 methyltransferase domain. Nsp14 shows little sequence similarity to known methyltransferases; however, its structure has not been determined experimentally, hence its phylogenetic relationships to other enzymes remain unclear.

Cap-specific 2'-O-ribose methyltransferases

A poxvirus cap1-forming enzyme (VP39 protein from vaccinia virus), was the first methyltransferase involved in the cap structure formation, for which a crystal structure was determined and also the first one for which a structure of a ternary complex of an enzyme with the cofactor and RNA substrate was determined. It has become one of the best studied members of a large family of methyltransferases that act on the 2'-OH-ribose group in RNA, which includes also enzymes such as RmrJ and fibrillar. Although they share little sequence identity with each other, these 2'-O-ribose methyltransferases are characterized by the presence of a conserved tertiary fold characteristic for all RFM enzymes and a conserved K-D-K catalytic triad between the methyl group donor binding site, and the cap binding site. VP39 is a single-domain protein with additional structural elements at both the N- and C-termini, which wrap around the RFM core and form a binding pocket for the cap. In the ternary complex, the m7G base of the cap is bound sandwiched between 2 aromatic side chains, and oriented in such a way that the Hoogsteen edge modified by addition of the methyl group on N7 faces the protein, thus explaining the ability of VP39 to sense the methylation status of the substrate, which is the basis of its preference for substrates that already have an N7-methylated cap.

In humans, cap1 formation is catalyzed by the CMTr1 enzyme. It is composed of several domains, including the N-terminal catalytic RFM domain with a conserved K-D-K triad characteristic for 2'-O-ribose methyltransferases and a guanylyltransferase-like domain that lacks catalytic residues. The N-terminal domain of CMTr1 shares a global architecture with the VP39 protein and is sufficient for cap1 activity in vitro. Interestingly, while the cofactor-binding sites, active sites, and the sites of binding of the nascent RNA chain exhibits similarities with the VP39 and CMTr1 enzymes (and likewise the conformations

| Species | Cellular enzymes | Viral enzymes |
|---------|-----------------|--------------|
| Homo sapiens | RNMT | TGS |
| Saccharomyces cerevisiae | Abd1 | Tgs1 |
| Encephalitozoon cuniculi | Ecm1 | Tgs1 |
| Giardia lamblia | D1/D12 | Tgs1 |
| Trypanosoma brucei | TbCmt1 | Tgs1 |
| Vaccinia virus | D12 | Tgs1 |
| Flavivirus | NS5 | Tgs1 |
| Vesicular stomatitis virus | L protein | Tgs1 |
| Reovirus | lambda 2 | Tgs1 |
| Blu etongue virus | VP4 | Tgs1 |
| SARS Coronavirus | nsp14 | Tgs1 |

| Species | Cap-position | base | 2'-O-ribose | other |
|---------|--------------|------|-------------|-------|
| Homo sapiens | RNMT | cap0 | TMG | cap1 |
| Saccharomyces cerevisiae | Abd1 | cap0 | TMG | cap1 |
| Encephalitozoon cuniculi | Ecm1 | cap0 | TMG | cap1 |
| Giardia lamblia | D1/D12 | cap0 | TMG | cap1 |
| Trypanosoma brucei | TbCmt1 | cap0 | TMG | cap1 |
| Vaccinia virus | D12 | cap0 | TMG | cap1 |
| Flavivirus | NS5 | cap0 | TMG | cap1 |
| Vesicular stomatitis virus | L protein | cap0 | TMG | cap1 |
| Reovirus | lambda 2 | cap0 | TMG | cap1 |
| Blu etongue virus | VP4 | cap0 | TMG | cap1 |
| SARS Coronavirus | nsp14 | cap0 | TMG | cap1 |
Table 2. Experimentally determined structures of cap-specific methyltransferases. "cap0 + cap1" indicates 2 activities encoded in separate domains in one polypeptide, while "cap0/cap1" indicates 2 activities associated with one domain. — indicates structures available in the PDB, for which no corresponding articles are available in the literature.

| MTase type | organism / virus | protein | ligand1 | ligand2 | PDB |
|------------|-----------------|---------|---------|---------|-----|
| cap0 (m7G) | Homo sapiens    | RNMT    | sinefungin | —       | 3eppa |
| cap0 (m7G) | Homo sapiens    | RNMT    | SAH     | —       | 3bgva |
| cap0 (m7G) | Encephalitozoon cuniculi | RNMT | sinefungin | —       | 2hv96 |
| cap0 (m7G) | Encephalitozoon cuniculi | RNMT | AzosAM | —       | 1Z3b |
| cap0 (m7G) | Encephalitozoon cuniculi | RNMT | SAH | m7GpppG | 1ri1 |
| cap0 (m7G) | Encephalitozoon cuniculi | RNMT | — | — | 1ri2 |
| cap0 (m7G) | Encephalitozoon cuniculi | RNMT | — | — | 1ri3 |
| cap0 (m7G) | vaccinia virus | D1, D12 | SAH | — | 2vdw33 |
| cap0 (m7G) | vaccinia virus | D1, D12 | SAH | — | 4cke88 |
| cap0 (m7G) | vaccinia virus | D1, D12 | SAH | — | 4ckb88 |
| cap0 (m7G) | vaccinia virus | D1, D12 | SAH | — | 4ckc88 |
| cap1 (XpppNm) | Homo sapiens | CMTtr1 | SAM | m7GpppG | 4n48 |
| cap1 (XpppNm) | Homo sapiens | CMTtr1 | SAM | m7GpppG | 4n49 |
| cap1 (XpppNm) | — | — | — | — | 4na9 |
| cap1 (XpppNm) | vaccinia virus | VP39 | SAH | m7GpppGAA | 1av6 |
| cap1 (XpppNm) | vaccinia virus | VP39-dC26 | — | m7Ade | 3mag9 |
| cap1 (XpppNm) | vaccinia virus | VP39-dC26 | — | m7Ade | 1b42 |
| cap1 (XpppNm) | vaccinia virus | VP39-dC26 | — | m7Cyt | 3mct9 |
| cap1 (XpppNm) | vaccinia virus | VP39-dC26 | — | m7Cyt | 1bky9 |
| cap1 (XpppNm) | vaccinia virus | VP39-D182A | — | m7G | 4d9g9 |
| cap1 (XpppNm) | vaccinia virus | VP39-E233Q | — | m7G | 1eqa9 |
| cap1 (XpppNm) | vaccinia virus | VP39-E233A | — | m7G | 1e9m9 |
| cap1 (XpppNm) | vaccinia virus | VP39 | — | — | 1vp3 |
| cap1 (XpppNm) | vaccinia virus | VP39-dC26 | — | — | 1vp9 |
| cap1 (XpppNm) | vaccinia virus | VP39-dC26 | — | — | 1v39 |
| cap1 (XpppNm) | vaccinia virus | VP39-dC26 | — | — | 1jte9 |
| cap1 (XpppNm) | vaccinia virus | VP39 | — | — | 1ktl |
| cap1 (XpppNm) | vaccinia virus | VP39 | — | — | 1jtf9 |
| cap1 (XpppNm) | vaccinia virus | VP39-dC26 | — | — | 1jte9 |
| cap1 (XpppNm) | vaccinia virus | VP39-dC26 | — | — | 1ktl |
| cap1 (XpppNm) | vaccinia virus | VP39-dC26 | — | — | 1jtf9 |
| cap1 (XpppNm) | SARS virus | ns10-ns16 | SAM | — | 3m24 |
| cap0 + cap1 | reovirus | lambda2 | — | — | 1e96 |
| cap0 + cap1 | bluetongue virus | VP4 | — | — | 2iha1 |
| cap0 + cap1 | bluetongue virus | VP4 | — | — | 2ihp1 |
| cap0 + cap1 | bluetongue virus | VP4 | — | — | 2jh8 |
| cap0 + cap1 | bluetongue virus | VP4 | — | — | 2jh9 |
| cap0 + cap1 | bluetongue virus | VP4 | — | — | 2jhc |
| cap0/cap1 | West Nile virus | NS5 | SAH | — | 2oy0 |
| cap0/cap1 | Wesselsbron virus | wv-MTase | SAM | m7GpppG | 3emb51 |
| cap0/cap1 | Wesselsbron virus | wv-MTase | SAM | m7GpppG | 3elw51 |
| cap0/cap1 | Wesselsbron virus | wv-MTase | SAM | m7GpppG | 3elu51 |
| cap0/cap1 | Wesselsbron virus | wv-MTase | SAM | m7GpppG | 3elv51 |
| cap0/cap1 | Wesselsbron virus | wv-MTase | sinefungin | — | 3elb51 |
| cap0/cap1 | Wesselsbron virus | wv-MTase | sinefungin | m7GpppG | 3emd51 |
| cap0/cap1 | Meaban virus | mVMTase | SAH | — | 2oct52 |
| cap0/cap1 | Murray Valley enc. virus | NS5 | SAH | — | 2px2 |
| cap0/cap1 | Murray Valley enc. virus | NS5 | SAH | — | 2px4 |
| cap0/cap1 | Murray Valley enc. virus | NS5 | SAH | — | 2px5 |
| cap0/cap1 | Murray Valley enc. virus | NS5 | m7GTP | — | 2px8 |
| cap0/cap1 | Murray Valley enc. virus | NS5 | GpppG | — | 2pxa53 |
| cap0/cap1 | Murray Valley enc. virus | NS5 | SAM | — | 2pxb53 |
| cap0/cap1 | Dengue virus | NS5 | — | m7GpppA | 2px2 |
| cap0/cap1 | Dengue virus | NS5 | — | m7GpppG | 2px4 |
| cap0/cap1 | Dengue virus | NS5 | — | m7GpppG | 2px5 |
| cap0/cap1 | Dengue virus | NS5 | — | m7GpppG | 2px8 |
| cap0/cap1 | Dengue virus | NS5 | — | m7GpppG | 2pxa53 |
| cap0/cap1 | Dengue virus | NS5 | — | m7GpppG | 2pxb53 |
| cap0/cap1 | Dengue virus | NS5 | — | m7GpppG | 2pxc53 |
| cap0/cap1 | Dengue virus | NS5 | — | m7GpppG | 2pxd53 |
| cap0/cap1 | Dengue virus | NS5 | — | m7GpppG | 2pxe53 |
| cap0/cap1 | Dengue virus | NS5 | — | m7GpppG | 2pxf53 |
| cap0/cap1 | Dengue virus | NS5 | — | m7GpppG | 2pxg53 |

(Continued on next page)
of the respective ligands), their cap-binding sites exhibit large differences in the shape of the m7G-binding pocket. As a result, CMTr1 binds m7G in a different way, in which the sugar edge of the cap guanosine faces the protein, and the methyl group on N7 faces the solvent (Fig. 3). These structural differences explain why CMTr1 is relatively insensitive to the absence of cap0 methylation and therefore is able to act, at least in vitro, on substrates with unmethylated guanosine.

Proteins with cap1 methyltransferase activities were also characterized in the alfalfa looper moth Autographa californica nucleopolyhedrovirus (orf6940) and in T. brucei (TbMTr144,45). Both of these enzymes are relatively closely related to the human CMTr1 enzyme.

In the SARS virus, cap1 methylation is catalyzed by a complex comprised of 2 partners: the nsp16 protein that is clearly related to the above-mentioned cap1 methyltransferases, but is inactive on its own, and a small regulatory protein nsp10 that is required for nsp16 to bind both the SAM methyl group donor and the RNA substrate. The crystal structure of the snp10-nsp16 complex showed that, in nsp16, the SAM-binding region is partially degenerated compared to “partner-independent” ribose methyltransferases, and nsp10 stabilizes the SAM binding pocket and extends the RNA-binding groove of nsp16.43

Apart from the enzymes responsible for cap1 methylation, methyltransferases have been characterized that act on additional residues in the nascent RNA chain. Many eukaryotic organisms possess a 2'-O-ribose methyltransferase that methylates the 2nd residue in mRNA and in other RNA molecules. The cap2 methyltransferase has been characterized in T. brucei (TbMTr244,45) and in humans (CMTr246). Interestingly, while CMTr2 appears to be closely related to its human paralog CMTr1 as well as to TbMTr1, TbMTr2 is more closely related to the vaccinia virus cap1 methyltransferase.44 In trypanosomes, a third 2'-O-ribose cap methyltransferase was identified and termed TbMTr3, which is responsible for the methylation of the third residue of the cap and is required for the methylation of the fourth residue.47,48

Table 2. Experimentally determined structures of cap-specific methyltransferases. “cap0 + cap1” indicates 2 activities encoded in separate domains in one polypeptide, while “cap0/cap1” indicates 2 activities associated with one domain. a — indicates structures available in the PDB, for which no corresponding articles are available in the literature. (Continued)

| MTase type | organism / virus | protein | ligand1 | ligand2 | PDB |
|-----------|-----------------|---------|---------|---------|-----|
| cap0/cap1 | Dengue virus    | NS5     | —       | GpppG   | 2p3t |
| cap0/cap1 | Dengue virus type 2 | NS5     | SAH     | SAH     | 1l9k |
| cap0/cap1 | Dengue virus type 2 | NS5     | SAH     | GMP     | 2p1d |
| cap0/cap1 | Dengue virus type 2 | NS5     | SAH     | ribavirin | 1r6a |
| TMG (m2,2,7G) | Homo sapiens | TGS1    | SAH     | m2GpppG | 3gdh |
| TMG (m2,2,7G) | Homo sapiens | TGS1    | —       | m2GpppA | 3ge9 |
| mpppN    | Homo sapiens    | BCDIN3  | SAM     | —       | 3g07 |

Figure 1. Schematic representation of the conserved core of Rossmann-fold Methyltransferase (RFM) catalytic domains. The β-sheet is composed of 7 β-strands (gray arrows) surrounded by 6 α-helices (semi-transparent tubes) forms the fold that is typical for SAM-dependent methyltransferases. All secondary structure elements of the conserved core are labeled as α1, β1, etc. The stars indicate points of most frequent insertions and terminal fusions with other domains.

Figure 2. Crystal structure of the cap0 methyltransferase from Encephalitozoon cuniculi. A stick representation of the ligands bound to cap0 methyltransferase. The guanosine cap analog position was defined based on the structure deposited as 1RI2 in the PDB, and the methyl group donor position was depicted based on the structure deposited as 1R14 in the PDB. Secondary structure elements that correspond to elements of the conserved RFM core are labeled. Secondary structure elements outside of the conserved core are not labeled.
Proteins with cap0 and cap1 methyltransferase activities

A number of viral proteins were reported to possess both cap0 and cap1 methyltransferase activities. In most of them, this is due to the presence of multiple domains. For instance, in the human reovirus (a virus with a dsRNA genome), the cap structure formation is catalyzed by a large multidomain protein lambda 2, which in turn is a part of the reovirus core: an assembly with a relative molecular mass of 52 MDa that synthesizes, modifies and exports viral mRNA. The structure of the human reovirus core has been solved at low resolution, revealing a series of domains that include a putative guanylyltransferase domain and 2 putative methyltransferase (RFM) domains. It has been suggested that the order of the domains in the lambda 2 protein corresponds to the order of the capping reactions: guanosine transfer followed by cap0 and cap1 methylation. However, comparison of domain structures suggested that the functional assignments may be different, as the RFM domain 1 shared a putative active site with the corresponding structurally characterized 2'-O-ribose methyltransferases, including the cap1 methyltransferase, whereas the RFM domain 2 exhibited structural similarity to the cap0 methyltransferases. It should be noted that the putative cap1 methyltransferase domain of reovirus exhibits a similar cap-binding platform formed by N- and C-terminal extensions, as in VP39 and human CMTr1 enzymes; however, its putative m'G-binding site is more open.

In bluetongue virus, another member of the reoviruses, the structure of the VP4 protein revealed a multi-domain protein with an N-terminal guanylyltransferase domain and 2 RFM domains, of which one was inserted into another. The inserted RFM domain exhibited clear similarities to the cap1 methyltransferases and in 3 crystal forms had GpppG, m'GDP, or GTP bound in the position of the cap-binding site, while the other RFM domain exhibited low but significant similarity to known cap0 methyltransferase structures.

In the non-segmented, negative-sense single-stranded RNA viruses [order Mononegavirales (MNV)] that include pathogens such as respiratory syncytial virus, measles, mumps, rabies, parainfluenza, vesicular stomatitis virus (VSV), and Marburg and Ebola viruses, one of the common components of the viral ribonucleoprotein core is the large (L) protein, which encodes multiple functions such as the RNA-dependent RNA polymerase and activities responsible for mRNA capping, cap0 and cap1 methylation, poly(A) polymerase and protein kinase. Using bioinformatics methods, we and others predicted that the C-terminal region of that protein (conserved region VI) encodes a domain homologous to 2'-O-ribose methyltransferases and is likely to function as a cap1 methyltransferase. Later it was found that, in VSV, this region is essential not only for cap1, but also for cap0 methyltransferase activity and that the same SAM-binding site and part of the K-D-K triad is used for both reactions. The structural basis of this phenomenon remains to be determined.

In Flaviviruses (positive-sense, single-stranded RNA viruses), an RFM domain with a similar dual methyltransferase function was identified. In a non-structural protein 5, the N-terminus was first unambiguously characterized as a cap1 (2'-O-ribose) methyltransferase. Later, it was shown that this domain takes part also in cap0 (m'G) methylation using the same SAM-binding site during cap synthesis. Interestingly, in these viruses, the order of methylation is different than in Mononegavirales, as cap0 methylation precedes cap1 methylation. Several structures were determined for the flavivirus cap methyltransferases known or...
predicted to be bifunctional, including Dengue, West Nile, Wesselbron, Meaban, and Murray Valley encephalitis viruses and they all revealed high similarity to the cap1 methyltransferases, and little if any similarity to the classical cap0 methyltransferases. It should be noted that these methyltransferases share a similar cap-binding platform structure with VP39 and human CMTr1 enzymes (a platform formed by N- and C-terminal extensions); however, the orientation of the bound guanosine residue suggests that their mode of cap-recognition is different from both poxvirus and human enzymes (Fig. 3).

Tgs1/Tgs2
The enzyme responsible for the trimethylguanosine (m2,2,7G, TMG) synthesis was first identified in yeast and named yTgs1. The Tgs enzymes of budding and fission yeast and Giardia are relatively small polypeptides (239–315 amino acids) consisting of little more than an RFM methyltransferase catalytic domain (Fig. 4), whereas metazoan Tgs1 proteins are much larger, because they include an N-terminal extension not found in lower eukaryotes. Tgs1 activity is strictly dependent on prior cap0 (m7G) methylation, thereby restricting its activity to RNAs that were already methylated by cap0 methyltransferase. Similar substrate requirements are characteristic for the Giardia Tgs2 enzyme. Interestingly, in contrast to Tgs1 methyltransferases able to catalyze 2 sequential N2 methylation steps leading to TMG cap formation, Tgs2 activity is apparently limited to a single round of N2 methylation, resulting in the synthesis of a 2,7-dimethylguanosine (m2,7G) product. Bioinformatics analyses predicted that the Tgs enzymes are related to a large group of RFM enzymes that act on exocyclic amine groups in nucleic acid bases, including m4A, m4C, and m2G and have a characteristic NPPY-like motif at the active site. The crystal structure of the active C-terminal methyltransferase domain of the human TGS enzyme bound to a minimal substrate m7GTP as well as the reaction product SAH has been reported, confirming these predictions and revealing the atomic details of these protein-ligand interactions.

Other methyltransferases involved in cap-specific base modifications
Studies of cap composition of human mRNAs conducted in mid-70s revealed that when the first nucleotide of the transcript is an adenosine, this base can be methylated to m6A. The enzyme that catalyzes the conversion of mGpppAm ends of mRNA to mGpppm6Am has been isolated from a cytoplasmic fraction of HeLa cells. The isolated enzyme showed no activity toward internal adenosines. Recently, Schwartz and coworkers studied the m6A mRNA methylome following depletion of multiprotein methyltransferase complex components METTL3, METTL14, KIAA1429, and WTAP, and implicated the involvement of the METTL3, METTL14, and KIAA1429 proteins in m6A formation at the internal sites but not at the 5' sites. The full characterization of the cap-specific m6A methyltransferase activity requires further studies in vitro.

In trypanosomes, the first adenine of the hypermethylated cap4 structure is not only methylated at the ribose, but also dimethylated at the N6 position, forming m6,6Am. The methyltransferase responsible for the latter reaction remains unknown. Further, the fourth uracil in that structure is modified to m3U, and the enzyme responsible for this modification also remains unknown.

Bin3/γ-methyltransferases
The γ-methylphosphate cap structure is unique in that it is an alternative to the guanosine-containing cap. It is formed by a single methyltransfer reaction to a γ-phosphate oxygen at the 5' end of the primary transcripts of certain small RNA molecules such as mammalian U6 and 7SK, mouse B2 and plant U3. The enzyme responsible for this reaction, Bicoid-interacting protein 3 (Bin3), is a methyltransferase conserved in eukaryotes. It is, however, absent from S. cerevisiae. A structure of the human Bin3 homolog (BCDIN3) was determined, revealing a conserved RFM core (Fig. 5). An enzyme-substrate complex is not yet available, and the details of protein-RNA recognition and the mechanism of discrimination between Bin3 substrates and non-substrates remain to be determined.

Christopher F. Sullivan, Fengting Zhang, and Li-Ping Zheng
Department of Chemistry, University of Kansas, Lawrence, Kansas 66045-7501, USA
Conclusions and Future Perspectives

In recent years significant progress has been made in understanding the mechanism of formation of different RNA cap structures. This progress has been driven in particular by the identification and characterization of novel methyltransferases that take part in cap biosynthesis, and by the determination of their crystal structures. This knowledge also has a practical dimension, as the capping process is essential for eukaryotic cells as well as for the life cycle of viruses that infect them. In this context, the difference between the structures of the human enzymes and the enzymes from human pathogens could be exploited to develop new drugs. In particular, viruses that evolved alternative enzymes to synthesize the same cap structures as are synthesized by human cellular machinery are attractive targets for the development of inhibitors that could specifically block viral methyltransferases as close homologs of the vaccinia virus cap1 methyltransferase. While the analysis of protein-RNA interactions and search for potential regulatory molecules (e.g., inhibitors) could be guided by homology models developed so far, experimental determination of high resolution structures for cap methyltransferases in trypanosomes would be definitely useful.

A complete understanding of RNA cap biosynthesis requires not only structure determination of the enzymes that are well characterized biochemically, but also the identification of the genes and proteins that encode the cap methyltransferation machinery. Some of the prominent enzymatic activities known to exist that are still awaiting unequivocal identification of the corresponding genes and proteins include m6A methylation of the first transcribed nucleoside of capped RNAs in humans, and m6,6A and m3U methylation of the first and the fourth residues in the cap4 structure in capped RNAs in trypanosomes. A comprehensive biochemical and structural characterization of these enzymes could further contribute to the possibility of developing new drugs against trypanosomal parasites and new tools to study RNA metabolism in human cells.

Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed.

Funding

This analysis was supported in part by the Foundation for Polish Science (FNP, grant TEAM/2009-4/2 to J.M.B.). M.B. and E.P. were additionally supported by National Science Center (NCN, 2011/03/D/NZ1/03247 to E.P.). M.S. was additionally supported by the START Fellowship from the FNP. J.M.B. was additionally supported by the European Research Council (ERC, StG grant RNA5P = 123D) and by the “Ideas for Poland” fellowship from the FNP. Research on structural aspects of protein-RNA interactions in the Bujnicki laboratory has been also supported by grants from the European Union’s Seventh Framework Program (REGPOT grant FishMed; grant agreement no 316125), the Polish Ministry of Science and Higher Education (MNNiSW, grant POIG.02.03.00-00-003/09) and by the infrastructure financed by the European Union—the European Regional Development Fund within the Operational Program “Innovative economy” for 2007–2013 (CePT, grant POIG.02.02.01-14-024/08-00).
58. Ray D, Shah A, Tilgner M, Guo Y, Zhao Y, Dong H, 1606 Volume 11 Issue 12
RNA Biology
59. Bollati M, Milani M, Mastrangelo E, Ricagno S, Tede-
55. Ferron F, Longhi S, Henrissat B, Canard B. Viral
deas TS, Zhou Y, Li H, Shi PY. West Nile virus 5'-cap
51. Sutton G, Grimes JM, Stuart DI, Roy P. Bluetongue
methylation by the large polymerase protein. J Virol
O methylation of the vesicular stomatitis virus mRNA
50. Bujnicki JM, Rychlewski L. Reassignment of specific-
structural protein of ssRNA negative-strand viruses.
56. Hausmann S, Ramirez A, Schneider S, Schwer B, Shu-
mRNA cap guanine-N2 methyltransferase (Tgs1). J Biol
71. Wei CM, Gershowitz A, Moss B. Methylated nucleoti-
date prediction and phylogenetic analysis of the 2'-
mRNA cap guanine-N2 methyltransferases from Giardia lamblia
52. Keith JM, Muthukrishnan S, Moss B. Effect of methyl-
5. Sutton G, Grimes JM, Stuart DI, Roy P. Bluetongue
mRNA capping machinery reveal the identity of the 7SK capping enzyme. Mol Cell 2007; 27:262-74; PMID:17643375; http://dx.doi.org/10.1016/j.molcel.2007.06.027
54. Jeronimo C, Forget D, Bouchard A, Li Q, Chua G,
53. Jeronimo C, Forget D, Bouchard A, Li Q, Chua G,
52. Keith JM, Muthukrishnan S, Moss B. Effect of methyl-
51. Sutton G, Grimes JM, Stuart DI, Roy P. Bluetongue
methylation by the large polymerase protein. J Virol
50. Bujnicki JM, Rychlewski L. Reassignment of specific-
structural protein of ssRNA negative-strand viruses.
56. Hausmann S, Ramirez A, Schneider S, Schwer B, Shu-
mRNA cap guanine-N2 methyltransferase (Tgs1). J Biol
71. Wei CM, Gershowitz A, Moss B. Methylated nucleoti-
date prediction and phylogenetic analysis of the 2'-
mRNA cap guanine-N2 methyltransferases from Giardia lamblia
52. Keith JM, Muthukrishnan S, Moss B. Effect of methyl-
5. Sutton G, Grimes JM, Stuart DI, Roy P. Bluetongue
mRNA capping machinery reveal the identity of the 7SK capping enzyme. Mol Cell 2007; 27:262-74; PMID:17643375; http://dx.doi.org/10.1016/j.molcel.2007.06.027
54. Jeronimo C, Forget D, Bouchard A, Li Q, Chua G,
53. Jeronimo C, Forget D, Bouchard A, Li Q, Chua G,
52. Keith JM, Muthukrishnan S, Moss B. Effect of methyl-
51. Sutton G, Grimes JM, Stuart DI, Roy P. Bluetongue
methylation by the large polymerase protein. J Virol
O methylation of the vesicular stomatitis virus mRNA
50. Bujnicki JM, Rychlewski L. Reassignment of specific-
structural protein of ssRNA negative-strand viruses.
56. Hausmann S, Ramirez A, Schneider S, Schwer B, Shu-
mRNA cap guanine-N2 methyltransferase (Tgs1). J Biol
71. Wei CM, Gershowitz A, Moss B. Methylated nucleoti-
date prediction and phylogenetic analysis of the 2'-
mRNA cap guanine-N2 methyltransferases from Giardia lamblia
52. Keith JM, Muthukrishnan S, Moss B. Effect of methyl-
5. Sutton G, Grimes JM, Stuart DI, Roy P. Bluetongue
mRNA capping machinery reveal the identity of the 7SK capping enzyme. Mol Cell 2007; 27:262-74; PMID:17643375; http://dx.doi.org/10.1016/j.molcel.2007.06.027
54. Jeronimo C, Forget D, Bouchard A, Li Q, Chua G,
53. Jeronimo C, Forget D, Bouchard A, Li Q, Chua G,
89. Hu G, Gershon PD, Hodel AE, Quiocho FA. mRNA cap recognition: dominant role of enhanced stacking interactions between methylated bases and protein aromatic side chains. Proc Natl Acad Sci U S A 1999; 96:7149-54; PMID:10377383; http://dx.doi.org/10.1073/pnas.96.13.7149

90. Hodel AE, Gershon PD, Shi X, Wang SM, Quiocho FA. Specific protein recognition of an mRNA cap through its alkylated base. Nat Struct Biol 1997; 4:350-4; PMID:9145102; http://dx.doi.org/10.1038/nstb0350

91. Hu G, Oguro A, Li C, Gershon PD, Quiocho FA. The “cap-binding slot” of an mRNA cap-binding protein: quantitative effects of aromatic side chain choice in the double-stacking sandwich with cap. Biochemistry 2002; 41:7677-87; PMID:12056899; http://dx.doi.org/10.1021/bi0201926

92. Benarroch D, Egloff MP, Mulard L, Guerreiro C, Romette JL, Canard B. A structural basis for the inhibition of the NS5 dengue virus mRNA 2'-O-methyltransferase domain by ribavirin 5'-triphosphate. J Biol Chem 2004; 279:35638-43; PMID:15152003; http://dx.doi.org/10.1074/jbc.M400460200