CMTr mediated 2`-*O*-ribose methylation status of cap adjacent-nucleotides across animals

THOMAS C. DIX[1,2*], IRMGARD U. HAUSSMANN[3*], SARAH BRIVIO[4], MOHANNAKARTHIK P. NALLASIVAN[1,2], YAVOR HADZHIEV[2,5], FERENC MÜLLER[2,5], BERNDT MÜLLER[4], JONATHAN PETTITT[4] AND MATTHIAS SOLLER[1,2,6]

[1]School of Biosciences, College of Life and Environmental Sciences, University of Birmingham, Edgbaston, Birmingham, B15 2TT, United Kingdom

[2]Birmingham Centre for Genome Biology, University of Birmingham, Edgbaston, Birmingham, B15 2TT, United Kingdom

[3]Department of Life Science, Faculty of Health, Education and Life Sciences, Birmingham City University, Birmingham, B15 3TN, United Kingdom

[4]Institute of Cancer and Genomic Sciences, College of Medical and Dental Sciences, University of Birmingham, Edgbaston, Birmingham, B15 2TT, United Kingdom

[5]School of Medicine, Medical Sciences and Nutrition, Institute of Medical Sciences, University of Aberdeen, Aberdeen, AB25 2ZD, United Kingdom

[6]Corresponding author: m.soller@bham.ac.uk

Phone: +44 121 414 5905

Running title: Cap 2'-*O*-ribose methylation status in animals

Key Words: mRNA methylation, 2'-*O*-ribose methylation

Abstract

Cap methyltransferases (CMTrs) *O*-methylate the 2` position of the ribose (cOMe) of cap-adjacent nucleotides of animal, protist and viral mRNAs. Animals generally have two CMTrs, while trypanosomes have three and many viruses encode one in their genome. In the splice leader of mRNAs in trypanosomes the first four nucleotides contain cOMe, but little is known about the status of cOMe in animals. Here, we show that cOMe is prominently present on the first two cap-adjacent nucleotides with species- and tissue-specific variations in *C. elegans*, honeybees, zebrafish, mouse and human cell lines. In contrast, *Drosophila* contains cOMe primarly on the first cap-adjacent nucleotide. De novo RoseTTA modelling of CMTrs reveals close similarities of the overall structure and near identity for the catalytic tetrad, and for cap and co-factor binding for human, *Drosophila* and *C. elegans* CMTrs. While viral CMTrs maintain the overall structure and catalytic tetrad, they have diverged in cap and co-factor binding. Consistent with the structural similarity, both CMTrs from *Drosophila* and humans methylate the first cap-adjacent nucleotide of an AGU consensus start. Since the second nucleotide is also methylated upon heat stress in *Drosophila*, these findings argue for regulated cOMe important for gene expression regulation.

Introduction

Eukaryotic mRNAs contain a cap at their beginning that can be followed by variably methylated nucleotides. The main function of the cap is to protect mRNAs from degradation and to recruit translation initiation factors, but also to promote splicing and 3` end processing (Topisirovic et al. 2011; Gonatopoulos-Pournatzis and Cowling 2014).

Caps are added co-transcriptionally in two steps shortly after transcription initiation of RNA polymerase II. First, a guanosine is added by RNGTT (RNA guanylyltransferase and 5′ phosphatase capping enzyme) in a characteristic 5`-5` linkage to the first nucleotide of the mRNA and this guanosine is then methylated at the *N7* position by RNMT (RNA guanine-7 methyltransferase). Subsequently, the first two cap adjacent nucleotides can then be methylated at the ribose (2`-*O*-ribose methylation, cOMe) to various degrees between tissues and transcripts (Perry and Kelley 1974; Furuichi et al. 1975; Wei et al. 1975; Furuichi and Shatkin 2000; Galloway and Cowling 2019).

The cOMe mRNA modifications are introduced by dedicated cap methyltransferases (CMTrs) (Langberg and Moss 1981; Belanger et al. 2010; Werner et al. 2011). Most animals, including *C. elegans*, *Drosophila* and mice model organisms as well as humans have two *CMTr* genes (*CMTr1* and *CMTr2*), while trypanosomes have three *CMTr* genes (Bangs et al. 1992). Moreover, many viruses including corona virus have their own *CMTr* gene (Ferron et al. 2012; Netzband and Pager 2020). Analysis of the first cap adjacent nucleotide has revealed variations in methylation between tissues and transcripts (Kruse et al. 2011; Mauer et al. 2017). Analysis of the second and following cap-adjacent nucleotides remains technically challenging (Anreiter et al. 2021). Initial studies on human CMTr activities had suggested that CMTr1 2`-*O*-ribose methylates the first and CMTr2 the second cap-adjacent nucleotide on a capped polyA substrate RNA or a transcript starting with three guanosines (Langberg and Moss 1981; Werner et al. 2011), but CMTr2 has also been shown to 2`-*O*-ribose methylate the second nucleotide of U1 or U2 snRNA starting with m7GpppAmpUpC (Werner et

al. 2011). Mainly through the analysis of cOMe in trypanosomes, which is found on the first four nucleotides of the unique splice-leader trans-spliced onto each mRNA, it has been possible to analyse the contribution of individual CMTrs through knock-outs (Arhin et al. 2006a; Arhin et al. 2006b; Zamudio et al. 2006; Zamudio et al. 2007). These studies suggested that CMTr1 would methylate the first nucleotide and CMTrs 2 and 3 methylate the following nucleotides of the unique splice leader added to mRNAs by trans-splicing, and these findings have been extrapolated to human CMTr1 and CMTr2 (Langberg and Moss 1981; Werner et al. 2011). However, recent findings in *Drosophila* using a novel recapping assay revealed that both CMTrs introduce cOMe on the first nucleotide redundantly. In addition, vaccinia CMTr VP39 can methylate up to the first three nucleotides of a trypanosome capped splice leader substrate *in vitro* (Haussmann et al. 2022).

In *Drosophila*, CMTr1 is the main enzyme and accounts for the majority of cOME (Haussmann et al. 2022). CMTr1 in *Drosophila* and humans is nuclear, while CMTr2 localizes predominantly to the cytoplasm, but is also present in the nucleus and at cell membranes (Werner et al. 2011; Haussmann et al. 2022). Human CMTr1 interacts with the CTD of Pol II (Haline-Vaz et al. 2008). In *Drosophila* CMTr1 was shown to globally localize to sites of transcription, while CMTr2 only localizes to a subset of transcription sites suggesting divergent regulatory roles rather than constitutive functions in adding cOMe at a specific position in the beginning of the mRNA (Haussmann et al. 2022).

In addition to methylation introduced by CMTrs, when the first nucleotide of the mRNA is an adenosine and carries cOMe, it can also be methylated by PCIF1 in vertebrates and some other organisms. However, the mechanism for cap adenosine *N6*-methylation is different from internal methylation of adenosine (Keith et al. 1978; Kruse et al. 2011; Mauer et al. 2017; Akichika et al. 2019; Balacco and Soller 2019; Boulias et al. 2019; Sendinc et al. 2019; Sun et al. 2019; Pandey et al. 2020).

A high-resolution structure has been determined for the methyltransferase domain of human CMTr1 bound to a capped oligonucleotide and the methyl donor S-adenosylmethionine (SAM) (Smietanski et al. 2014). This structure has then served to model the methyltransferase domain of human CMTr2 bound to a capped oligonucleotide and SAM. In addition, structures of various viral CMTrs haven been determined (Hodel et al. 1998; Malet et al. 2007; Zhao et al. 2015; Ferrero et al. 2019; Viswanathan et al. 2020). These structures reveal deep pockets for binding the cap and SAM. Because contacts between CMTr1 and cap-adjacent nucleotides are restricted to the RNA backbone, it is thought the 2`-*O*-ribose methylation occurs sequence independent.

Here, we applied a novel sensitive assay to analyse 2`-*O*-ribose methylation of cap-adjacent nucleotides beyond the first nucleotide to determine cOMe in various model organisms including *C. elegans*, *Drosophila*, honey bees, zebrafish and mouse as well as human cell lines. This analysis reveals species- and tissue-specific differences in cOMe. In *Drosophila* cOMe is mainly found on the first nucleotide. In contrast, in *C. elegans*, honey bees, zebrafish, mouse and human cell lines, cOME is found on the first and second nucleotide. De novo structural modelling using RoseTTA structure prediction from human CMTr1 and viral CMTrs reveals overlapping configurations in the catalytic tetrad and binding of the cap structure, the RNA backbone and the methyl donor S-adenosylmethione (SAM) in animal CMTrs and trypanosome TbMTr1, but minor differences to viral CMTrs and trypanosome TbMTr2 and 3. Consistent with the structural similarity, both CMTrs from *C. elegans*, *Drosophila* and humans methylate the first cap-adjacent nucleotide of a capped substrate RNA starting with an AGU consensus observed for the majority of transcription start sites (Haussmann et al. 2022). Moreover, *Drosophila CMTr* double mutants are sensitive to heat stress, which can induce cOMe at the second position in *Drosophila* indicating that CMTrs can dynamically introduce cOMe at multiple positions in the first few nucleotides of an mRNA.

Results

Cap-adjacent nucleotides are 2`-O-ribose methylated in a species and tissue specific manner

An alignment of CMTr1 and CMTr 2 orthologues of model animals and trypanososmes shows strong conservation of the overall domain structure and in the catalytic domain, but increasing differences outside these domains as expected from their phylogenetic divergence (Supplemental Figure S1 and S2) (Werner et al. 2011). It has so far not been possible to directly and unambiguously determine the extent of 2`-*O*-methylation of cap-adjacent nucleotides in small amounts of mRNA. To close this technical gap, we have developed an assay based on recapping with ^{32}P-alphaGTP followed by digestion with RNAse I, which does not cleave in the presence of cOMe to allow for detection of consecutive cOMe on the first nucleotides of an mRNA by comparison to markers (Note, in vitro transcribed RNA by bacterial polymerases requires a G or A as first nucleotide, which will be a 5' tri-phosphate and thus can be capped) (Haussmann et al. 2022).

In the trypanosome *T. brucei*, we find a six nucleotide fragment consisting of the cap GTP, four 2`-*O*-ribose methylated nucleotides and one non-methylated nucleotide as previously reported by mass spectrometry analysis of the unique splice leader sequence, which is added to all mRNAs (Figure 1A) (Bangs et al. 1992; Arhin et al. 2006b; Zamudio et al. 2006). In *C. elegans*, we most prominently find cOMe on the first two cap adjacent nucleotides of mRNA at similar levels. *Drosophila* carries cOMe as previously reported only prominently on the first cap-adjacent nucleotide of mRNA and cOMe is absent in *CMTr1/2* double mutants. In honey bees, zebrafish inner organs and brain, mouse brain and in human HEK293T cells, cOME is also found prominently on the first two cap adjacent nucleotides at about equal amounts (Figure 1A).

To specifically analyze methylation of the first nucleotide in polyA mRNA, we decapped and dephosphorylated polyA mRNA and labelled the first nucleotide by ^{32}P-

gammaATP. After digestion into individual nucleotides, they were separated on 2D thin layer chromatography (TLC) (Figure 1B). In trypanosomes we detected *N6*-dimethylated adenosine that is also 2`-*O*-ribose methylated ($pm^6{}_2Am$, and to a lesser extent, *N6*-dimethylated adenosine ($pm^6{}_2A$, Figure 1C) as previously reported (Bangs et al. 1992). In *C. elegans*, we detected cOMe prominently on adenosine (pAm) (Figure 1D), but could not detect pGm present in about 70% of trans-spliced mRNAs starting with G in the splice leader (Allen et al. 2011; Pettitt et al. 2014), because pGm runs at the same position as pC (Haussmann et al. 2022). In *C. elegans*, pm^6Am was absent in accordance with the absence of a Pcif1 orthologue (Pandey et al. 2020), but pm^6Am was also absent in *Drosophila* and honey bees (Figure 1D-F). In Drosophila, the core catalytic sequence differs that could explain the absence, but this is not the case in honey bees (Supplemental Figure S3).

In zebrafish embryos and inner organs we detected pCm, pAm and m^6Am (Figure 1G and H). In the mouse brain, pm^6Am is most prominent, followed by pCm and pAm (Figure 1I). In human HCT116 cells, pCm is most prominent, followed by pm^6Am and pAm (Figure 1J).

Eukaryotic and viral core CMTrs structures align in the catalytic centre to bind the cap and SAM

A key feature of methyltransferases is the characteristic Rossmann-like fold (Figure 2A) (Medvedev et al. 2021). To obtain insights into conserved features and the molecular mechanisms directing 2`-*O*-methylation of cap-adjacent nucleotides in diverse species we aligned the catalytic core from animal, protist and viral CMTrs (Figure 2B). This alignment reveals strict conservation of the four amino acids KDKE forming the catalytic tetrad (Lys, Asp, Lys and Glu) and two major clades of an extended consensus surrounding the catalytic centre for animal CMTrs including trypanosome TbMTr1 and viral CMTrs alongside similar trypanosome TbMTr2 and 3, respectively (Bujnicki and Rychlewski 2001; Feder et al. 2003; Zamudio et al. 2007).

Although the structure for human CMTr1 had been determined, extrapolating functional relationships from sequence alignments is limited because the sequence in more distantly related species such as *Drosophila* and *C. elegans* has diverged too much. Therefore, we used *de novo* structural prediction based on recently developed machine learning algorithms RoseTTA to compare the structures of CMTrs across model organisms, protists and viruses (Baek et al. 2021; Jumper et al. 2021). In particular, we built theoretical models of the methyltransferase (MTase) domain of both CMTr1 and CMTr2 from humans, mice, zebrafish, *Drosophila* and *C. elegans*, as well as from the trypanosome *T. brucei* TbMTr1-3. In addition, we also modelled the MTase domain of viral CMTr non-structural protein 16 (nsp16) from SARS-CoV-2, gill-associated virus (GAV), infectious bronchitis virus (IBV) and Nam Dihn virus (NDiV). Of note, the non-structural protein 5 (ns5) from vaccinia (VP39), Zika, Dengue and West Nile viruses combines the entire capping process into one protein (Sutton et al. 2007).

To evaluate the accuracy of the predicted structures, we superimposed all models with corresponding structures available from the Protein Data Bank (PDB) for CMTr1 (4n48), SARS-CoV-2 nsp16 (6wks) and vaccinia VP39 (1av6). These structure-based pair-wise alignment confirmed that they completely superimpose with Root-Mean-Square Deviation (RMSD) values, Template Modeling (TM) and Global Distance Test (GDT) scores of 1.9/0.92/0.79 for CMTr1, 1.1/0.85/0.82 for VP39 and 1.1/96/0.92 for nsp16 (DALI server, Supplemental Figure S4) (Holm 2020).

To quantify the three-dimensional similarity of modelled CMTr structures, we performed an “all against all” comparative analysis via the DALI server of the MTase domain (Holm 2020). As an out-group control, we performed the same “all against all” analysis against the X-ray crystal structure of the human METTL3 (Methyltransferase-like 3, 5il0) MTase domain, which also contains Rossmann-like fold and catalyses methylation of internal adenosine residues at the *N6* position (Wang et al. 2016; Balacco and Soller 2019).

As expected, the MTase domain of orthologous proteins framed in black clustered together with high structural similarity scores (z= 40-60, Figure 2C) reminiscent of phylogenetic analysis from primary sequence alignment (Werner et al. 2011). Interestingly, trypanosome TbMTr1 is highly similar to eukaryotic CMTr1 (z=39-44), whereas TbMTr2 and 3 are more closely related to vaccinia virus CMTr VP39 (z=24, Figure 2C). In contrast, the MTAse domain of CMTrs compared to the METTL3 MTase domain had a very low structural similarity score (z<5, Figure 2C).

Superimposing the MTase domain of CMTr1 or CMTr2 from human, *Drosophila* and *C. elegans* revealed that they are highly similar in the core catalytic centre and only diverged in peripheral parts by extended loops (Figure 2 D and E). In particular, the spatial position of the four amino acids forming the catalytic tetrad is near identical in the three animal species differing less than 1.5 Å (Supplemental Figure S5A and B). Interestingly, CMTr2 contains an extended loop that occludes the SAM-binding pocket. Such a loop has also been found in SARS-CoV-2 nsp16, which opens upon association with nsp10 to allow binding of SAM (Vithani et al. 2021).

For the viral MTAse domains from SARS-CoV-2 nsp 16, vaccinia VP39 and zika ns5, the core catalytic centre also showed a high similarity, but peripheral parts are more diverged compared to animal CMTrs (Figure 2F). Likewise, the four amino acids in the catalytic centre adopt a similar position as in animal CMTrs differing less than 4 Å, but the first Lys of the catalytic tetrad shows more positional flexibility (Supplemental Figure S4C).

Detailed analysis of the amino acids involved in cap and SAM binding in human, *Drosophila* and *C. elegans* CMTR1 reveals an essentially identical configuration (Supplemental Figure S4D). Only one amino acid recognizing the amine group at position 2 of the cap guanosine changed from Asp to Ser in *C. elegans*. Other minor alterations include a change of Asn to Asp in recognising the ribose of SAM in *Drosophila*. In CMTr2, although the amino acids involved in both cap and SAM binding have changed considerably between human, *Drosophila* and *C. elegans*, the contact

positions are strictly conserved (Supplemental Figure S4E). Likewise, viral CMTrs have substantially diverged to a non-overlapping binding mode between vaccinia, Zika and SARS-CoV-2. In addition, the cap is bound differently between animal CMTrs and viral CMTrs resulting in altered solvent exposure of the cap guanosine (Supplemental Figure S4F).

Human CMTr1 and CMTr2 structures align with trypanosome TbMTr1, and vaccinia VP39 with trypanosome TbMTr2 and 3 in the catalytic centre to bind the cap and SAM

Superimposing the MTase domains of human CMTr1 and CMTr2 with the trypanososme TbMTr1 MTase domain shows that the overall structures are very similar, particularly in the core catalytic centre and only diverged in peripheral parts (Figure 3A). However, in CMTr2 an extended alpha helix protrudes into the catalytic site and SAM binding pocket suggesting structural rearrangement upon binding of either SAM or the cap. Again, the spatial position of the four amino acids forming the catalytic tetrad is near identical in human CMTr1 and 2, and trypanosome TbMTr1 differing less than 0.9 Å (Figure 3B).

Detailed analysis of the amino acids involved in cap and SAM binding in human CMTr1 and CMTr2 compared to trypanosome TbMTr1 reveals essentially an identical configuration of how the cap and SAM are bound, although with variations in the contacting amino acids (Figure 3C). Intriguingly, some amino acids altered in human CMTr2 compared to CMTr1 are the same in *Drosophila* CMTr2 compared to human CMTr1 such as the Asp contacting the amine group at position 2 of the cap guanosine and Arg contacting the gamma phosphate in the 5`-5` cap linkage. In the SAM binding pocket CMTr1 has two additional contacts at the carboxy group of methionine and at the ribose (Figure 3C).

Vaccinia VP39 preferentially 2`-*O*-ribose methylates the first cap-adjacent nucleotide, but can extend the methylation to the second and even third nucleotide (Haussmann

et al. 2022). Superimposing the MTase domains of vaccinia VP39 with trypanosome TbMTr2 and TbMTr3 MTase domains reveals principally an overall structures that is very similar, particularly in the core catalytic centre, but TbMTr2 and TbMTr3 contain additional sequences in protruding loops above the substrate binding site (Supplemental Figure S6A). Again, the spatial position of the four amino acids forming the catalytic tetrad is near identical in vaccinia VP39 compared to trypanosome TbMTr2 and TbMTr3 MTase domains differing less than <1.5 Å (Supplemental Figure S6B). Likewise, the amino acids involved in cap and SAM binding in vaccinia VP39 and, trypanosome TbMTr2 and TbMTr3 MTase domains are identical for side chain interactions (Supplemental Figure S6C).

Taken together, the high similarity of the different CMTr structures suggests that they will have very similar properties in adding cOMe to cap-adjacent nucleotides.

Both human CMTrs 2`-*O*-ribose methylate the first and second cap-adjacent nucleotide

Consistent with the highly similar structural models for CMTr1 and CMTr2, both can add cOMe to the first nucleotide in *Drosophila* (Haussmann et al. 2022). To test the activity of human CMTrs, we expressed them in human HEK293T cells (Figure 4A) and incubated them with a capped AGU consensus starting substrate RNA.

In this assay, human CMTr1 primarily methylates the first, and to a low level also the second cap-adjacent nucleotide at the 2` ribose position (Figure 4B lane 6), but no activity in the cell extract was detected (Fig 4B, lane 5). In contrast, the main products detected after incubation of the substrate RNA with human CMTr2 *in vitro* contain 2`-*O*-ribose methylation on the second and third cap-adjacent nucleotide (Figure 4B lane 7), which is also evident for the vaccinia CMTr (Figure 1A, lane 2 and Figure 4B, lane 2). However, after the CMTr2 incubation 2`-*O*-ribose methylation of the first cap-adjacent nucleotide is not seen indicating that CMTr2 directly methylates the second cap-adjacent nucleotide once the first is methylated (Figure 4B lane 7). When human

CMTr1 and CMTr2 were incubated together, cOMe on the second and third nucleotide increased (Figure 4B, lane 8), but methylation with CMTr2 is slower than with CMTr1 since now also methylation on the first cap-adjacent nucleotide is detected (Figure 4B, lane 8 compared to lane 7).

Taken together, both human CMTrs methylate the first cap-adjacent nucleotide as found in *Drosophila*, but CMTr2 more efficiently adds cOMe to the second and third cap-adjacent nucleotide preferably on a substrate containing cOMe on the first cap-adjacent nucleotide.

C. elegans CMTr2 is redundant for 2`-_O_-ribose methylation of the second cap-adjacent nucleotide

To further evaluate the activity of CMTr1 *in vivo*, we used *C. elegans* because both the first and second nucleotide carry cOMe. CMTr1 is well conserved in *C. elegans* (Supplemental Fig S1), has the same methyltransferase domain structure, configuration of the catalytic tetrad, cap and SAM binding as human and *Drosophila* CMTrs (Figure 2 and 3A).

A *CMTr2* null mutant in *C. elegans* harbouring a small deletion inducing a frameshift in the catalytic domain is viable (Figure 5A). In these mutants, the second cap-adjacent nucleotide still carries cOMe indicating that CMTr1 and 2 can also act redundant in adding cOMe to the second nucleotide *in vivo* (Figure 5B).

Elevated temperature reduces viability of _Drosophila CMTr1/2_ double mutants and induces 2`-_O_-ribose methylation of the second cap-adjacent nucleotide

Since methylation of mRNA has been linked to provide robustness to gene expression (Dezi et al. 2016; Haussmann et al. 2016; Roignant and Soller 2017), we wondered, whether loss of cOMe would reduce heat-tolerance of flies. When mutant flies of individual *CMTr1* and *CMTr2* genes were reared at 29ºC, which is above the preferred temperature of 24º (Sayeed and Benzer 1996), *CMTr1/2*^{null} double, but not single

mutant flies were less tolerant to elevated temperature and showed significantly reduced survival ($p<0.001$, Figure 6A). These results again demonstrate redundancy of CMTrs in *Drosophila* (Haussmann et al. 2022).

We then tested whether acute increase in temperature for two days in adult flies would induce cOMe. Surprisingly, we now detected cOMe at the second cap-adjacent nucleotide in *Drosophila* in response to elevated temperature (Figure 6B). However, when we reared *C. elegans* or *Drosophila* at low or high temperatures (15ºC and 25ºC for *C.elegans* and 18ºC and 29ºC for *Drosophila*) we did not detect differences in the extent of cOMe indicating no role of cOMe in temperature adaptation per se, but rather as a short-term response to stress (Supplemental Figure S7).

Discussion

The most prominent and variable methylation of mRNA in animals and some of their parasites including trypanosomes and viruses such as SARS-CoV-2 is found at cap-adjacent nucleotides (Kruse et al. 2011; Mauer et al. 2017; Galloway and Cowling 2019). Intriguingly, both CMTrs in *Drosophila* add cOMe to the first nucleotide *in vivo* as shown by knock-outs, as well as *in vitro* methylation assays (Haussmann et al. 2022). Here, we find that the ribose is variably 2`-*O*-methylated (cOMe) by CMTrs in a species and tissue-specific manner. In particular, the first nucleotide mostly contains cOMe, while cOMe on the second is variable among species and tissues with very little present in *Drosophila* and up to about half in other species. Likewise, if the first nucleotide is an adenosine, it is also variably methylated at the *N6* position (m^6Am) by PCIF1 in vertebrates (Akichika et al. 2019). In *C. elegans*, the entire Mettl3 writer complex directing internal m^6A and its YTH reader proteins are absent, and also PCIF1 is absent (Dezi et al. 2016; Balacco and Soller 2019), hence we did not detect m^6Am on the first nucleotide in this species. In *Drosophila*, PCIF1 has a phenylalanine to histidine amino acid substitution in the NPPF catalytic centre motif (Pandey et al. 2020), that could explain the absence of m^6Am. In bees, we also did not detect m6Am,

but the change of phenylalanine to tyrosine in the NPPF catalytic centre motif unlikely explains the absence.

On the first nucleotide, cOMe is found most prominently at high levels of about 80% or more in all species using a re-capping assay, but is lower when using a decapping and re-labelling assay (Haussmann et al. 2022). Several reasons could account for this differences. First, when preparing polyA RNA, non-cOMe containing RNA seems to co-purify even after two rounds of polyA or ribo-minus selection. Also, incubation with the 5'-3' exonuclease Xrn-1, that uses 5' phosphorylated substrate only marginally reduces this contaminant (Haussmann et al. 2022). Such RNA could be of various sources beyond degradation products and include un-capped sense and antisense RNAs and snoRNAs with complementarity to polyadenylated mRNA. Intriguingly, however, in both the TLC assay to analyse the first nucleotide and CAGEseq, adenosine was found to be most frequent first nucleotide in *Drosophila* mRNA with little difference between the two methods (Haussmann et al. 2022). Second, using different decapping enzymes including tobacco acid pyrophosphatase, bacterial RppH or mammalian decapping enzyme, we did not observe differences in the level of cOMe by labelling with T4 polynucleotide kinase indicating robustness of the assay, but also that these enzymes are not sensitive to mRNA secondary structure often found in the 5'UTR. Whether the recently commercially available yeast yDcpS decapping enzyme is sensitive to secondary structure has not been extensively tested, but can be evaluated in the future when more CAGEseq data is becoming available from using yDcpS for library preparation (Wulf et al. 2019; Yan et al. 2022). Recently, also mass spectrometric analysis of caps in polyA mRNA has become possible and these frequencies are similar to the recapping data (Akichika et al. 2019; Wang et al. 2019; Galloway et al. 2020).

In contrast to mass spectrometric analysis of caps in polyA mRNA, our recapping assay and separation on 22% denaturing acrylamide gels requires much less input material. Although this method is not able to identify the sequence of cap-adjacent

nucleotides, it can accurately identify how many nucleotides adjacent to the cap are 2`-*O*-ribose methylated in mRNAs from wild type or CMTr mutants, or in synthetic substrate RNAs by comparison to markers. The capped di-nucleotide of endogenous mRNA runs broader with about three identifiable bands as a result of sequence heterogeneity, while longer capped oligomers are merging into one band. We also noticed differences in ribose phosphate configurations, but the phosphatase activity of T4 PNK could not resolve this issue (Das and Shuman 2013).

Initial experiments characterizing CMTr activity using a capped polyA substrate and DEAE paper elecrtrophoresis suggested preferential activity of CMTr1 for the first nucleotide, and CMTr2 for the second nucleotide (Langberg and Moss 1981). Similar results were also obtained with an other unnatural capped RNA substrate starting with three guanosines (Werner et al. 2011). Although CMTr2 adds cOMe to U1 and U2 snRNA starting with AUA and AUC, whether it can also methylate the first nucleotide has not been tested (Werner et al. 2011). The cOMe status of mRNA has also been analysed in trypanosome mutants for CMTrs using an RNAse T2 digestion essay measuring cOMe by cleavage protection, because RNAses require a non-methylated 2`-hydroxyl group for catalysis (Motorin and Marchand 2018) and mRNAs contain a unique splice leader sequence (Bangs et al. 1992). Indeed, results from TbMTr2 and 3 mutants support a model whereby TbMTr1 would 2`-*O*-methylate the ribose of the first nucleotide and TbMTr2 and 3 the ribose of the three additional nucleotides (Arhin et al. 2006a; Arhin et al. 2006b; Zamudio et al. 2006; Zamudio et al. 2007). Possibly, this model specifically applies to trypanosomes, because of the unique sequence of the splice leader in combination with sequence specificity of TbMTrs (Mittra et al. 2008). Intriguingly, however, reinterpretation of the T2 digestion assay from in TbMTr1 mutants suggest that an AmAmC fragment is present indicating that TbMTr2 and 3 can add cOMe to the first and second nucleotide (Arhin et al. 2006a; Arhin et al. 2006b; Zamudio et al. 2006; Zamudio et al. 2007). Likewise, the vaccinia viral CMTr VP39, to which TbMTr2 and 3 are most closely related, adds cOMe to the first nucleotide, but

410 can also methylate additional nucleotides (Haussmann et al. 2022). In addition, our
411 previous genetic and biochemical studies of cOMe deposition in *Drosophila* showed
412 that both CMTr1 and CMTr2 can methylate the first cap-adjacent nucleotide and that
413 they act redundantly (Haussmann et al. 2022). Here, we further substantiate this view
414 and show that also human CMTr2 can add cOMe to the first cap-adjacent nucleotide
415 of a capped AGU consensus starting RNA substrate. Moreover, CMTr1 can also
416 methylate the second nucleotide in a CMTr2 knock-out in *C. elegans*, which are viable.
417 In *Drosophila*, CMTr1 is the main enzyme as its absence reduces cOMe by about 80%.
418 In addition, CMTr1 co-localizes with RNA Pol II to most if not all sites of transcription
419 (Haussmann et al. 2022). Given the physical interaction of CMTr1 with RNA Pol II
420 (Haline-Vaz et al. 2008), CMTr1 is likely also the main enzyme in other organisms, but
421 since its loss is lethal in *C. elegans* and mice (Lee et al. 2020), this is more difficult to
422 test.
423 From a structural point, the models of CMTr1 and 2 from humans, *Drosophila* and *C.*
424 *elegans* are very similar in their overall methyltransferase structure, the catalytic tetrad
425 and how the cap and SAM are bound, which is consistent with the capacity of both
426 enzymes to add cOMe on both the first and second nucleotide. Although both enzymes
427 only recognize the RNA backbone, but not the first nucleotide, there could be
428 preference for adding cOMe to the second nucleotide based on sequence context. In
429 fact, such a scenario is suggested from the differences in the methylation pattern of
430 vaccinia VP39 to a consensus AGU starting mRNA and the AACUAA starting
431 trypanosome splice leader, which becomes more extensively 2`-*O*-ribose methylated
432 (Haussmann et al. 2022). In this context, it is interesting that cOMe of the second
433 nucleotide is most variable between different species and tissues. Hence, CMTr1 and
434 2 could differ in their activity in adding cOMe to the nucleotides after the first and impact
435 differentially on gene expression this way. Moreover, developmental and cell-type
436 specific roles are further suggested from the distinct expression profile of CMTrs in
437 *Drosophila* (Haussmann et al. 2022). Likewise, CMTr1 is mostly nuclear while CMTr2

shows prominent cytoplasmic localization in both *Drosophila* and humans (Werner et al. 2011; Haussmann et al. 2022). In fact, a developmental role in *Drosophila* trachea development has specifically been attributed to CMTr2 suggesting sequence specificity, which has been found for trypanosome TbMTr1 (Englund et al. 1999; Mittra et al. 2008). Progress in the analysis of cOMe has been hampered by many technical challenges, but the combination of now available knock-out models and the development of effective in vitro assay systems will allow to address substrate specificity in more detail in the future.

Taken together, our analysis of CMTrs across animals reveals redundant roles in adding cOMe to the first and second cap-adjacent nucleotide and potentially also to the third. Since transcription start sites are heterogeneous (Ohler et al. 2002) and such variability mediates a translational response to heat stress (Tamarkin-Ben-Harush et al. 2017), alterations in the methylation pattern of cap-adjacent nucleotides could essentially impact on gene expression regulating stability, localization and translation.

Materials and Methods

Animal husbandry and cell culture

C. elegans worms were kept on *E. coli* coated agar. The *CMTr2* mutant strain was generated by the National Bioresource Project, Tokyo, Japan, which is part of the International *C. elegans* Gene Knockout Consortium and validated with PCR primers tm4453F (ATGATTTTGCCAGAAACCCGCG) and tm4453R (TGGTGCTTCCATCTGCAGTAAC).

Flies were kept on standard cornmeal-agar food (1% industrial-grade agar, 2.1% dried yeast, 8.6 % dextrose, 9.7 % cornmeal and 0.25 % Nipagin, all in (w/v)) in a 12 L : 12D cycle (Haussmann et al. 2013). CMTr mutants were as described (Haussmann et al. 2022). For the analysis of *Drosophila* survival at 29º C, flies were allowed to lay eggs on agar plates containing 1% grape juice and live yeast on top for one day at 25º C

(Ustaoglu et al. 2019). After 48 hours, larvae were washed off and collected on a fine mesh and the temperature shifted to 29º C. Groups of 30 larvae were transferred to food vials and flies were reared at designated temperatures. Honey bees were obtained and maintained as described (Decio et al. 2019; Decio et al. 2021).

Zebrafish were maintained in designated facility (according to UK Home Office regulations) in a recirculating system (ZebTEC, Tecniplast) at 26 °C in a 10-hour dark, 14-hour light photoperiod and fed three times daily. Animal work presented in this study was carried out under the project licences 40/3681 and P51AB7F76 assigned to the University of Birmingham, UK. Inner organs were spleen, liver and heart.

Mouse tissues were obtained from the Biomedical Service Unit of the University of Birmingham. Human HCT116 and HEK293T cells (ATCC) were cultured in McCoy's 5A and DMEM medium (Lonza) with 10% heat-inactivated FBS and 1% penicillin/streptomycin, respectively.

Statistical analysis of behavioral data

Behavioral data was analyzed using GraphPad Prism. One-way ANOVA followed by a Tukey`s post-hoc test was used for comparing multiple groups.

Analysis of cap-adjacent 2`-*O*-ribose methylation

Total RNA was extracted with Trizol (Invitrogen) and polyA mRNA was prepared by oligo dT selection according to the manufacturer (Promega). For the analysis of 5' cap structures, 100 ng of polyA mRNA were decapped by yDcpS in 20 µl for 1 h at 37º C according to the manufacturer's instruction (NEB). Then the RNA was extracted by phenol/$CHCl_3$ and ethanol precipitated in the presence of glycogen. The RNA was then labelled in a total volume of 20 µl containing 2 µl capping buffer (NEB), 1 µl SAM (2 mM), 0.25 µl ^{32}P-alphaGTP (3000 Ci/mmol, 6.6 µM; Hartmann Analytics, Germany), 0.5 µl RNAse Protector (Roche) and 0.5 µl capping enzyme (NEB) by incubation for 1 h at 37º C. The volume was then increased to 30 µl and 54 µl AMPure XP magnetic

beads (Beckman Coulter) were added and the labelled mRNA purified according to the manufacturer's instructions. The RNA was then eluted in 10 µl DEPC treated water. An 2.5 µl aliquot was digested by adding 0.3 µl NEB buffer 2 and 0.3 µl RNAse I for 2 h and then 10 µl gel loading buffer was added (98% deionized formamide, 10 mM EDTA, 0.025 % xylene cyanol FF and 0.025% bromphenol blue) products were analysed on 22% denaturing polyacrylamide gels (National Diagnostics) pre-run for 2 h. Gels were soaked in 20% PEG400, 10% acetic acid and 40 % methanol for 10 min and then dried on a Whatman 3MM paper. Dried gels were then exposed to a storage phosphor screen (BioRad) and scanned by a Molecular Imager FX in combination with QuantityOne software (BioRad). As a marker, a 31 nt *in vitro* transcript of the *per* gene made from a T7 2.5A promoter was capped and processed as described (Haussmann et al. 2022).

For the analysis of the first nucleotide in mRNA, polyA mRNA was purified from 30 µg of total RNA by oligo dT selection (Promega). Then, 50 ng of polyA mRNA was incubated with terminator nuclease (Epicenter) according to the manufacturer's instructions to remove rRNA followed by phenol/$CHCl_3$ and ethanol precipitation in the presence of glycogen (Roche). The mRNA was then decapped using RppH (NEB) and dephosphorylated by Antarctic phosphatase (NEB) in NEB buffer 2 supplemented with 0.1 mM $ZnCl_2$ in 20 µl. Then the RNA was extracted by phenol/$CHCl_3$ and precipitated in the presence of glycogen. The 5`-end of dephosphorylated mRNAs were then labeled using 10 units of T4 PNK (NEB) and 0.25 µl ^{32}P-gammaATP (6000 Ci/mmol, 25 µM; Perkin-Elmer). The labeled RNA was precipitated, and resuspended in 10 µl of 50 mM sodium acetate buffer (pH 5.5) and digested with nuclease P1 (SIGMA) for 1 h at 37º C. Two microliters of each sample was loaded on cellulose F TLC plates (20x20 cm; Merck) and run in a solvent system of isobutyric acid:0.5 M NH_4OH (5:3, v/v), as first dimension, and isopropanol:HCl:water (70:15:15, v/v/v), as the second dimension. TLCs were repeated from biological replicates. The identity of the nucleotide spots was determined as described (Keith 1995; Kruse et al. 2011). TLCs were exposed to a

storage phosphor screen (BioRad) and scanned by a Molecular Imager FX in combination with QuantityOne software (BioRad).

Human CMTrs were cloned into a pSp plasmid containing a cytomegalovirus promoter using primers hCMTr1 F1 (CTGAAATCACTTTTTTTCAGGTTGGACCGGTGCCACCATGAAGAGGAGAACTGACCCAGAATGCACTGCC) and hCMTr1 R1 (GTCATCGTCATCCTTGTAATCGCTCGAGGCCCTGTGCATCTGGATGAAGGAGAGGAC), and hCMTr2 F1 (CTGAAATCACTTTTTTTCAGGTTGGACCGGTGCCACCATGAGTAAGTGCAGAAAGACACCAGTTCAG) and hCMTr2 R1 (GTCATCGTCATCCTTGTAATCGCTCGAGTTTTGTAACTGAAGGCTGTTGATAATTTCTTC), after reamplification with primer FLAG Strep R (GTTAGCAGACTTCCTCTGCCCTCGCTAGCCTTCTCGAACTGCGGGTGGGACCAGGCGCTCTTGTCATCGTCATCCTTGTAATC) to add a FLAG and a Strep tag to the C-term before the 2A peptide followed by a puromycin resistance gene. Proteins were then expressed in HEK293T cells after transfection with Transit2020 (Mirrus) reagent modified from the manufacturer's instructions. Briefly, 2 µl plasmid DNA (1µg/µl) were mixed with 4 µl transfection regent and 600 µl serum-free DMEM media added and put on 1 Mio cells, grown in one well of a 6-well plate, for 2.5 hours before media was changed. After one day, puromycin (1:200, 10 mg/ml in water) was added and cells grown for two days. Cells were then lysed in 100 µl protein extraction buffer (20 mM Tris-HCL (pH 7.4), 137 mM NaCl, 1 mM EDTA, 25% glycerol (v/v), 1% NP-40 (v/v) 1 mM DTT, 1 mM PMSF and protease inhibitor cocktail (Roche).

To determine CMTr activity, 10 µl protein extract of CMTr1 was mixed with 10 µl reaction mix (20 mM Tris-HCL (pH 7.4), 0.2 µM SAM, 5 mM $MgCl_2$, 2 mM DTT, 1 µl RNAse Protector (Roche) and 1 µmole ^{32}P-alphaGTP capped substrate RNA) and incubated for 30 min at room temperature. For CMTr2, 40 µl extract were mixed with 200 µl (20 mM Tris-HCL (pH 7.4), 137 mM NaCl) and 10 µl MAGStrep magnetic beads

(type3 XT beads, IBA) and incubated for 2 hours at 4º C. For the assay, 10 µl reaction mixture was added to the beads and incubated as described above. After phenol/$CHCl_3$ extraction and ethanol precipitation in the presence of glycogen (Roche), the RNA was taken up in 10 µl of water. Then, 0.3 µl NEB buffer 2 and 0.3 µl RNAse I (NEB) was added to 2.5 µl RNA and digested for 2 hours. Then 10 µl gel loading solution was added and analysed on 22% gels as described above.

As substrate RNA, a 600 nt RNA starting with 31 nts from the *Drosophila per* gene containing a AGU consensus start was used (Haussmann et al. 2022), which was transcribed with T7 polymerase from a EcoRI and Spe I linearized plasmid in a 100 µl reaction using the Ambion MEGAscript kit (Ustaoglu et al. 2021). Six µg RNA (3 µl 11 µM) was then labelled with 3 µl ^{32}P-alphaGTP in 20 µl as described above and 1 µl was used per CMTr labeling reaction.

Structural modeling

Alignments of CMTrs was done by ClustalW and conservative substitution determined by BLOSUM-62. Full-length amino acid sequences were used for theoretical modelling via the Robetta protein structure prediction service using the RoseTTA deep learning server (Baek et al. 2021). The theoretical models of CMTrs were 'trimmed' to the minimal methyltransferase catalytically active unit and written to a new PDB file. This region is indicated on the CMTr1 and 2 alignments (Supplemental Figure S1 and S2).

Quality assessment of theoretical protein structures

The quality of the theoretical models was determined by error-coverage plots determined by the RoseTTA algorithm on the Robetta server (Baek et al. 2021). In addition, theoretical models were superimposed with their corresponding experimentally determined protein structures hMTr1 (4n48), SARS-CoV-2 nsp16 (6wks) and Vaccinia virus VP39 (1av6) (Hodel et al. 1998; Smietanski et al. 2014; Viswanathan et al. 2020). For quantitative comparison, the theoretical model and

corresponding X-ray crystal structure were subject to DALI server pairwise analysis to determine average RMSDs for all common residues and a structural similarity z-score (Holm 2020). Moreover, the Template Modelling (TM) and Global Distance Test (GDT) score was determined via the TM-score server (Xu and Zhang 2010). The DALI server was used for "all against all" protein structure comparison (Holm 2020). Protein structure visualisation, superimposition and determination of Cartesian co-ordinates were performed using the UCSF Chimera package downloaded from the Resource for Biocomputing, Visualization, and Informatics at the University of California, San Francisco (Pettersen et al. 2004).

Acknowledgments

We thank Mark Carrington and Nancy Standard for *T. brucei* total RNA, Jane Nimmo for honeybees, Caroline Chadwick for mouse tissue, Pawel Grzechnik for HEK293T cells, Roland Arnold for HCT116 cells, Rupert Fray for plasmids and the National Bioresource Project, Tokyo, Japan for the *C. elegans CMTr2* strain.

Funding

MS acknowledges funding from the Leverhulme Trust and BBSRC (BB/R002932/1), JP and BM from BBSRC (BB/T002859/1) and FM from the Wellcome Trust (106955/Z/15/Z).

Author contributions

IUH and MS performed molecular biology experiments, TD performed structural modelling, MPN performed cell culture experiments, JP and SB performed *C. elegans* experiments and YH zebrafish experiments. IUH and MS conceived the project and wrote the original draft of the manuscript. MS, IUH, TD and all other authors reviewed and edited. MS, BM, JP and FM supervised and acquired funding.

606 **Competing interests**

607 The authors declare no competing interests.

608

609 **References**

610 Akichika S, Hirano S, Shichino Y, Suzuki T, Nishimasu H, Ishitani R, Sugita A, Hirose
611 Y, Iwasaki S, Nureki O et al. 2019. Cap-specific terminal N (6)-methylation of
612 RNA by an RNA polymerase II-associated methyltransferase. *Science* **363**:
613 eaav0080.

614 Allen MA, Hillier LW, Waterston RH, Blumenthal T. 2011. A global analysis of C.
615 elegans trans-splicing. *Genome Res* **21**: 255-264.

616 Anreiter I, Mir Q, Simpson JT, Janga SC, Soller M. 2021. New Twists in Detecting
617 mRNA Modification Dynamics. *Trends Biotechnol* **39**: 72-89.

618 Arhin GK, Li H, Ullu E, Tschudi C. 2006a. A protein related to the vaccinia virus cap-
619 specific methyltransferase VP39 is involved in cap 4 modification in
620 Trypanosoma brucei. *Rna* **12**: 53-62.

621 Arhin GK, Ullu E, Tschudi C. 2006b. 2'-O-methylation of position 2 of the
622 trypanosome spliced leader cap 4 is mediated by a 48 kDa protein related to
623 vaccinia virus VP39. *Molecular and biochemical parasitology* **147**: 137-139.

624 Baek M, DiMaio F, Anishchenko I, Dauparas J, Ovchinnikov S, Lee GR, Wang J, Cong
625 Q, Kinch LN, Schaeffer RD et al. 2021. Accurate prediction of protein
626 structures and interactions using a three-track neural network. *Science* **373**: 871-
627 876.

628 Balacco DL, Soller M. 2019. The m(6)A Writer: Rise of a Machine for Growing Tasks.
629 *Biochemistry* **58**: 363-378.

630 Bangs JD, Crain PF, Hashizume T, McCloskey JA, Boothroyd JC. 1992. Mass
631 spectrometry of mRNA cap 4 from trypanosomatids reveals two novel
632 nucleosides. *J Biol Chem* **267**: 9805-9815.
633 Belanger F, Stepinski J, Darzynkiewicz E, Pelletier J. 2010. Characterization of hMTr1,
634 a human Cap1 2'-O-ribose methyltransferase. *J Biol Chem* **285**: 33037-33044.
635 Boulias K, Toczydlowska-Socha D, Hawley BR, Liberman N, Takashima K, Zaccara
636 S, Guez T, Vasseur JJ, Debart F, Aravind L et al. 2019. Identification of the
637 m(6)Am Methyltransferase PCIF1 Reveals the Location and Functions of
638 m(6)Am in the Transcriptome. *Mol Cell* **75**: 631-643 e638.
639 Bujnicki JM, Rychlewski L. 2001. Reassignment of specificities of two cap
640 methyltransferase domains in the reovirus lambda 2 protein. *Genome Biol* **2**:
641 Research0038.
642 Das U, Shuman S. 2013. Mechanism of RNA 2',3'-cyclic phosphate end healing by T4
643 polynucleotide kinase-phosphatase. *Nucleic Acids Res* **41**: 355-365.
644 Decio P, Ustaoglu P, Derecka K, Hardy ICW, Roat TC, Malaspina O, Mongan N,
645 Stoger R, Soller M. 2021. Thiamethoxam exposure deregulates short ORF gene
646 expression in the honey bee and compromises immune response to bacteria. *Sci*
647 *Rep* **11**: 1489.
648 Decio P, Ustaoglu P, Roat TC, Malaspina O, Devaud JM, Stoger R, Soller M. 2019.
649 Acute thiamethoxam toxicity in honeybees is not enhanced by common
650 fungicide and herbicide and lacks stress-induced changes in mRNA splicing.
651 *Sci Rep* **9**: 19196.
652 Dezi V, Ivanov C, Haussmann IU, Soller M. 2016. Nucleotide modifications in
653 messenger RNA and their role in development and disease. *Biochem Soc Trans*
654 **44**: 1385-1393.

655 Englund C, Uv AE, Cantera R, Mathies LD, Krasnow MA, Samakovlis C. 1999. adrift,
656 a novel bnl-induced Drosophila gene, required for tracheal pathfinding into the
657 CNS. *Development* **126**: 1505-1514.
658 Feder M, Pas J, Wyrwicz LS, Bujnicki JM. 2003. Molecular phylogenetics of the
659 RrmJ/fibrillarin superfamily of ribose 2'-O-methyltransferases. *Gene* **302**: 129-
660 138.
661 Ferrero DS, Ruiz-Arroyo VM, Soler N, Usón I, Guarné A, Verdaguer N. 2019.
662 Supramolecular arrangement of the full-length Zika virus NS5. *PLoS pathogens*
663 **15**: e1007656.
664 Ferron F, Decroly E, Selisko B, Canard B. 2012. The viral RNA capping machinery as
665 a target for antiviral drugs. *Antiviral Res* **96**: 21-31.
666 Furuichi Y, Morgan M, Shatkin AJ, Jelinek W, Salditt-Georgieff M, Darnell JE. 1975.
667 Methylated, blocked 5 termini in HeLa cell mRNA. *Proc Natl Acad Sci U S A*
668 **72**: 1904-1908.
669 Furuichi Y, Shatkin AJ. 2000. Viral and cellular mRNA capping: past and prospects.
670 *Adv Virus Res* **55**: 135-184.
671 Galloway A, Atrih A, Grzela R, Darzynkiewicz E, Ferguson MAJ, Cowling VH. 2020.
672 CAP-MAP: cap analysis protocol with minimal analyte processing, a rapid and
673 sensitive approach to analysing mRNA cap structures. *Open Biol* **10**: 190306.
674 Galloway A, Cowling VH. 2019. mRNA cap regulation in mammalian cell function
675 and fate. *Biochim Biophys Acta Gene Regul Mech* **1862**: 270-279.
676 Gonatopoulos-Pournatzis T, Cowling VH. 2014. Cap-binding complex (CBC).
677 *Biochem J* **457**: 231-242.

678 Haline-Vaz T, Silva TC, Zanchin NI. 2008. The human interferon-regulated ISG95
679 protein interacts with RNA polymerase II and shows methyltransferase activity.
680 *Biochem Biophys Res Commun* **372**: 719-724.
681 Haussmann IU, Bodi Z, Sanchez-Moran E, Mongan NP, Archer N, Fray RG, Soller M.
682 2016. m6A potentiates Sxl alternative pre-mRNA splicing for robust
683 Drosophila sex determination. *Nature* **540**: 301-304.
684 Haussmann IU, Hemani Y, Wijesekera T, Dauwalder B, Soller M. 2013. Multiple
685 pathways mediate the sex-peptide-regulated switch in female Drosophila
686 reproductive behaviours. *Proc Biol Sci* **280**: 20131938.
687 Haussmann IU, Wu Y, Nallasivan MP, Archer N, Bodi Z, Hebenstreit D, Waddell S,
688 Fray R, Soller M. 2022. CMTr cap-adjacent 2'-O-ribose mRNA
689 methyltransferases are required for reward learning and mRNA localization to
690 synapses. *Nat Commun* **13**: 1209.
691 Hodel AE, Gershon PD, Quiocho FA. 1998. Structural basis for sequence-nonspecific
692 recognition of 5'-capped mRNA by a cap-modifying enzyme. *Mol Cell* **1**: 443-
693 447.
694 Holm L. 2020. Using Dali for Protein Structure Comparison. *Methods in molecular*
695 *biology (Clifton, NJ)* **2112**: 29-42.
696 Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, Tunyasuvunakool
697 K, Bates R, Žídek A, Potapenko A et al. 2021. Highly accurate protein structure
698 prediction with AlphaFold. *Nature* **596**: 583-589.
699 Keith G. 1995. Mobilities of modified ribonucleotides on two-dimensional cellulose
700 thin-layer chromatography. *Biochimie* **77**: 142-144.

Keith JM, Ensinger MJ, Mose B. 1978. HeLa cell RNA (2'-O-methyladenosine-N6-)-methyltransferase specific for the capped 5'-end of messenger RNA. *J Biol Chem* **253**: 5033-5039.

Kruse S, Zhong S, Bodi Z, Button J, Alcocer MJ, Hayes CJ, Fray R. 2011. A novel synthesis and detection method for cap-associated adenosine modifications in mouse mRNA. *Sci Rep* **1**: 126.

Langberg SR, Moss B. 1981. Post-transcriptional modifications of mRNA. Purification and characterization of cap I and cap II RNA (nucleoside-2'-)-methyltransferases from HeLa cells. *J Biol Chem* **256**: 10054-10060.

Lee YL, Kung FC, Lin CH, Huang YS. 2020. CMTR1-Catalyzed 2'-O-Ribose Methylation Controls Neuronal Development by Regulating Camk2alpha Expression Independent of RIG-I Signaling. *Cell Rep* **33**: 108269.

Malet H, Egloff MP, Selisko B, Butcher RE, Wright PJ, Roberts M, Gruez A, Sulzenbacher G, Vonrhein C, Bricogne G et al. 2007. Crystal structure of the RNA polymerase domain of the West Nile virus non-structural protein 5. *J Biol Chem* **282**: 10678-10689.

Mauer J, Luo X, Blanjoie A, Jiao X, Grozhik AV, Patil DP, Linder B, Pickering BF, Vasseur JJ, Chen Q et al. 2017. Reversible methylation of m6Am in the 5' cap controls mRNA stability. *Nature* **541**: 371-375.

Medvedev KE, Kinch LN, Dustin Schaeffer R, Pei J, Grishin NV. 2021. A Fifth of the Protein World: Rossmann-like Proteins as an Evolutionarily Successful Structural unit. *J Mol Biol* **433**: 166788.

Mittra B, Zamudio JR, Bujnicki JM, Stepinski J, Darzynkiewicz E, Campbell DA, Sturm NR. 2008. The TbMTr1 spliced leader RNA cap 1 2'-O-ribose

methyltransferase from Trypanosoma brucei acts with substrate specificity. *J Biol Chem* **283**: 3161-3172.

Motorin Y, Marchand V. 2018. Detection and Analysis of RNA Ribose 2'-O-Methylations: Challenges and Solutions. *Genes (Basel)* **9**.

Netzband R, Pager CT. 2020. Epitranscriptomic marks: Emerging modulators of RNA virus gene expression. *Wiley interdisciplinary reviews RNA* **11**: e1576.

Ohler U, Liao GC, Niemann H, Rubin GM. 2002. Computational analysis of core promoters in the Drosophila genome. *Genome Biol* **3**: RESEARCH0087.

Pandey RR, Delfino E, Homolka D, Roithova A, Chen KM, Li L, Franco G, Vagbo CB, Taillebourg E, Fauvarque MO et al. 2020. The Mammalian Cap-Specific m(6)Am RNA Methyltransferase PCIF1 Regulates Transcript Levels in Mouse Tissues. *Cell Rep* **32**: 108038.

Perry RP, Kelley DE. 1974. Existence of methylated messenger RNA in mouse L cells. *Cell* **1**: 37-42.

Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE. 2004. UCSF Chimera--a visualization system for exploratory research and analysis. *Journal of computational chemistry* **25**: 1605-1612.

Pettitt J, Philippe L, Sarkar D, Johnston C, Gothe HJ, Massie D, Connolly B, Müller B. 2014. Operons are a conserved feature of nematode genomes. *Genetics* **197**: 1201-1211.

Roignant JY, Soller M. 2017. m6A in mRNA: An Ancient Mechanism for Fine-Tuning Gene Expression. *Trends Genet* **33**: 380-390.

Sayeed O, Benzer S. 1996. Behavioral genetics of thermosensation and hygrosensation in Drosophila. *Proc Natl Acad Sci U S A* **93**: 6079-6084.

749 Sendinc E, Valle-Garcia D, Dhall A, Chen H, Henriques T, Navarrete-Perea J, Sheng
750 W, Gygi SP, Adelman K, Shi Y. 2019. PCIF1 Catalyzes m6Am mRNA
751 Methylation to Regulate Gene Expression. *Mol Cell* **75**: 620-630 e629.
752 Smietanski M, Werner M, Purta E, Kaminska KH, Stepinski J, Darzynkiewicz E,
753 Nowotny M, Bujnicki JM. 2014. Structural analysis of human 2'-O-ribose
754 methyltransferases involved in mRNA cap structure formation. *Nat Commun* **5**:
755 3004.
756 Sun H, Zhang M, Li K, Bai D, Yi C. 2019. Cap-specific, terminal N(6)-methylation by
757 a mammalian m(6)Am methyltransferase. *Cell Res* **29**: 80-82.
758 Sutton G, Grimes JM, Stuart DI, Roy P. 2007. Bluetongue virus VP4 is an RNA-
759 capping assembly line. *Nat Struct Mol Biol* **14**: 449-451.
760 Tamarkin-Ben-Harush A, Vasseur JJ, Debart F, Ulitsky I, Dikstein R. 2017. Cap-
761 proximal nucleotides via differential eIF4E binding and alternative promoter
762 usage mediate translational response to energy stress. *Elife* **6**.
763 Topisirovic I, Svitkin YV, Sonenberg N, Shatkin AJ. 2011. Cap and cap-binding
764 proteins in the control of gene expression. *Wiley interdisciplinary reviews RNA*
765 **2**: 277-298.
766 Ustaoglu P, Gill JK, Doubovetzky N, Haussmann IU, Dix TC, Arnold R, Devaud JM,
767 Soller M. 2021. Dynamically expressed single ELAV/Hu orthologue elavl2 of
768 bees is required for learning and memory. *Commun Biol* **4**: 1234.
769 Ustaoglu P, Haussmann IU, Liao H, Torres-Mendez A, Arnold R, Irimia M, Soller M.
770 2019. Srrm234, but not canonical SR and hnRNP proteins, drive inclusion of
771 Dscam exon 9 variable exons. *RNA* **25**: 1353-1365.

772 Viswanathan T, Arya S, Chan SH, Qi S, Dai N, Misra A, Park JG, Oladunni F,
773 Kovalskyy D, Hromas RA et al. 2020. Structural basis of RNA cap modification
774 by SARS-CoV-2. *Nat Commun* **11**: 3718.
775 Vithani N, Ward MD, Zimmerman MI, Novak B, Borowsky JH, Singh S, Bowman GR.
776 2021. SARS-CoV-2 Nsp16 activation mechanism and a cryptic pocket with
777 pan-coronavirus antiviral potential. *Biophysical journal* **120**: 2880-2889.
778 Wang J, Alvin Chew BL, Lai Y, Dong H, Xu L, Balamkundu S, Cai WM, Cui L, Liu
779 CF, Fu XY et al. 2019. Quantifying the RNA cap epitranscriptome reveals novel
780 caps in cellular and viral RNA. *Nucleic Acids Res* **47**: e130.
781 Wang X, Feng J, Xue Y, Guan Z, Zhang D, Liu Z, Gong Z, Wang Q, Huang J, Tang C
782 et al. 2016. Structural basis of N(6)-adenosine methylation by the METTL3-
783 METTL14 complex. *Nature* **534**: 575-578.
784 Wei CM, Gershowitz A, Moss B. 1975. Methylated nucleotides block 5' terminus of
785 HeLa cell messenger RNA. *Cell* **4**: 379-386.
786 Werner M, Purta E, Kaminska KH, Cymerman IA, Campbell DA, Mittra B, Zamudio
787 JR, Sturm NR, Jaworski J, Bujnicki JM. 2011. 2'-O-ribose methylation of cap2
788 in human: function and evolution in a horizontally mobile family. *Nucleic Acids*
789 *Res* **39**: 4756-4768.
790 Wulf MG, Buswell J, Chan SH, Dai N, Marks K, Martin ER, Tzertzinis G, Whipple
791 JM, Corrêa IR, Jr., Schildkraut I. 2019. The yeast scavenger decapping enzyme
792 DcpS and its application for in vitro RNA recapping. *Sci Rep* **9**: 8594.
793 Xu J, Zhang Y. 2010. How significant is a protein structure similarity with TM-score =
794 0.5? *Bioinformatics* **26**: 889-895.

795 Yan B, Tzertzinis G, Schildkraut I, Ettwiller L. 2022. Comprehensive determination of
796 transcription start sites derived from all RNA polymerases using ReCappable-
797 seq. *Genome Res* **32**: 162-174.
798 Zamudio JR, Mittra B, Foldynová-Trantírková S, Zeiner GM, Lukes J, Bujnicki JM,
799 Sturm NR, Campbell DA. 2007. The 2'-O-ribose methyltransferase for cap 1 of
800 spliced leader RNA and U1 small nuclear RNA in Trypanosoma brucei. *Mol*
801 *Cell Biol* **27**: 6084-6092.
802 Zamudio JR, Mittra B, Zeiner GM, Feder M, Bujnicki JM, Sturm NR, Campbell DA.
803 2006. Complete cap 4 formation is not required for viability in Trypanosoma
804 brucei. *Eukaryotic cell* **5**: 905-915.
805 Zhao Y, Soh TS, Lim SP, Chung KY, Swaminathan K, Vasudevan SG, Shi PY, Lescar
806 J, Luo D. 2015. Molecular basis for specific viral RNA recognition and 2'-O-
807 ribose methylation by the dengue virus nonstructural protein 5 (NS5). *Proc Natl*
808 *Acad Sci U S A* **112**: 14834-14839.
809

810 **Figure legends**

811 Figure 1. Analysis of mRNA cap 2`-*O*-ribose methylation in various species.
812 (A) Recapping of mRNA with ^{32}P-alphaGTP from Trypanosomes (*T. brucei*), adult *C.*
813 *elegans*, adult *Drosophila* wild type and CMTr1/2 double flies, worker honey bees,
814 zebrafish inner organs and brain, mouse brain and human HEK293T cells. 5'cap
815 structures were separated on a 22% denaturing polyacrylamide gels after digestion
816 with RNAse I (lanes 4-11, right) Markers – M1: RNAse I digested ^{32}P-alphaGTP
817 capped *in vitro* transcript starting with AGU. M2: RNAse I digested ^{32}P-alphaGTP
818 capped *in vitro* transcript starting with AGU and 2'-*O*-ribose methylated with vaccinia
819 CMTr. Sequences of markers are shown on the left and of cap structures from different
820 species are shown on the right except for the sequence from Trypanososmes, which

821 is shown on the left. L: Alkaline hydrolysis of a 5' ^{32}P-labeled RNA oligonucleotide with
822 the nucleotide number indicated in white. (B) Schematic diagram of a 2D thin layer
823 chromatography (TLC) depicting standard and 2`-*O*-ribose methylated nucleotides.
824 For orientation, pA is indicated with an asterisk and pAm with an arrow. (C-J) TLCs
825 showing modifications of the first cap-adjacent nucleotides of *T. brucei* (C), *C. elegans*
826 (D), *Drosophila* (E), honey bees (F), zebrafish embryos and inner organs (G and H),
827 mouse brain (I) and human HCT116 cells (J).
828
829 **Figure 2. Structural comparisons of animal, protist and viral CMTrs.**
830 (A) Schematic depiction of the Rossmann-like fold. (B) Structure-based multiple
831 alignment of the sequence context surrounding the four key catalytic residues (yellow).
832 The consensus for animal and viral CMTrs is shown on top and bottom, respectively.
833 (C) Three-dimensional similarity of modelled CMTr structures determined by the DALI
834 server depicting the structural similarity dendrogram on top and "all against all" multiple
835 comparison at the bottom with orthologous proteins framed in black. Human Mettl3
836 (red star) was used as an outgroup indicating no similarity by near zero z-score (grey).
837 (D,E) Three-dimensional structural alignment of CMTr1 (D) and CMTr2 (E) from
838 humans (red), *Drosophila* (grey) and *C. elegans* (blue). The black star denotes the
839 position of alpha 11 helix which is absent in *C. elegans*. The position of the cap
840 analogue (light blue) and the SAM (yellow) were inserted by superimposition from the
841 published structure (PDB 4n48). (F) Three-dimensional structural alignment of CMTrs
842 from vaccinia (red), SARS-CoV-2 (grey) and zika virus (blue). The position of the cap
843 analogue (light blue) and the SAM (yellow) were inserted by superimposition from the
844 published structure (PDB 1av6).
845
846 **Figure 3. Comparison of human CMTr1 with human CMTr2 and trypanosome**
847 **TbMTr1.**

(A) Superimposition of human CMTr1 (red), modelled human CMTr2 (grey) and trypanosome TbMTr1 (blue). Magenta circles indicate the base which has been removed for clarity. The position of the cap analogue (light blue) and the SAM (yellow) were inserted by superimposition from the published structure (PDB 4n48).
(B) Configuration of the four amino acids forming the catalytic centre of human CMTr1 (red), modelled human CMTr2 (grey) and trypanosome TbMTr1 (blue). Magenta circles indicate the base which has been removed for clarity.
(C) Substrate and co-factor recognition of human CMTr1 (red), modelled human CMTr2 (grey) and trypanosome TbMTr1 (blue). Amino acids contacting the cap binding are shown on top and amino acids contacting the co-factor SAM (yellow) are shown at the bottom. The catalytic centre is indicated by a dashed circle and the turquoise circle indicates solvent facing. Contacts of amino acids via side chains are indicated by or purple circles and via the backbone in green circles. Green and blue thin dashed lines indicate direct hydrogen bonds and hydrogen bonds via a water molecule, respectively. Magenta thick dashed lines indicate aromatic stacking. Methyl-groups are shown in yellow circles.

Figure 4. Analysis of mRNA cap 2`-*O*-ribose methylation in *C. elegans CMTr2* null mutants.
(A) Representative Western-blot from two replicates of human CMTr1 and CMTr2 tagged with FLAGStrep expressed in human HEK293T cells detected with anti-FLAG antibodies. Molecular weight markers in kDa are shown on the left. The asterisk on the right denotes an unspecific band.
(B) 2`-*O*-ribose methylation of a $^{32}$PalphaGTP capped *in vitro* transcript starting with AGU by human CMTR1 and CMTr2 expressed in human HEK293T cells, vaccinia CMTr (vCMTr) or extract from untransfected cells (-). 5`cap structures were separated on a 22% denaturing polyacrylamide gels after digestion with RNAse I (lanes 6-8, right). Markers – M1: RNAse I digested ^{32}P-alphaGTP capped *in vitro* transcript starting

with AGU. M2: RNAse I digested ^{32}P-alphaGTP capped *in vitro* transcript starting with AGU and 2'-*O*-ribose methylated with vaccinia CMTr. Sequences of cap structures are shown on the sides. L: Single nucleotide ladder with nucleotide number indicated in white.

Figure 5. Analysis of mRNA cap 2`-*O*-ribose methylation in *C. elegans CMTr2* null mutants.

(A) Genomic organization of the *C. elegans CMTr2* locus depicting the exon intron structure is shown is shown on the left. The methyltransferase domain including key amino acid residues involved in catalysis (K117, D235 and K275 in yellow) is shown in dark blue and the catalytically inactive methyltransferase-like domain in light blue. The deletion in the *CMTr2* null allele (tm4453) leads to a frameshift and is validated by PCR products separated on an agarose gel on the right. M: 100 base pair ladder. (B) Recapping of mRNA with ^{32}P-alphaGTP from adult *C. elegans* wild type and *CMTr2* null mutants. 5'cap structures were separated on a 22% denaturing polyacrylamide gels after digestion with RNAse I (lanes 4 and 5, right) Markers – M1: RNAse I digested ^{32}P-alphaGTP capped *in vitro* transcript starting with AGU. M2: RNAse I digested ^{32}P-alphaGTP capped *in vitro* transcript starting with AGU and 2'-*O*-ribose methylated with vaccinia CMTr. Sequences of markers are shown on the left and of cap structures from *C. elegans* are shown on the right. L: Alkaline hydrolysis of a 5' ^{32}P-labeled RNA oligonucleotide with the nucleotide number indicated in white.

Figure 6. Impact of heat stress on *CMTr* double mutant viability and mRNA cap 2`-*O*-ribose methylation in *Drosophila*.

(A) Viability of *CMTr1*^{null} and *CMTr2*^{null} single and double mutant flies at 25º and 29º C shown as mean±SE (*n*=3, except for *CMTr1*/*2*^{null} at 25ºC n=4, p<0.001). (B) Recapping of mRNA with ^{32}P-alphaGTP from adult *Drosophila* wild type and *CMTr* double mutants at 25ºC and adult *Drosophila* wild type flies kept at 29ºC for two days. 5'cap structures

were separated on a 22% denaturing polyacrylamide gels after digestion with RNAse I (lanes 4 and 5, right) Markers –M1: RNAse I digested ^{32}P--alphaGTP capped *in vitro* transcript starting with AGU and 2'-*O*-ribose methylated with vaccinia CMTr. M2: RNAse I digested ^{32}P-alphaGTP capped *in vitro* transcript starting with AGU. Sequences of markers are shown on the left and of cap structures from *Drosophila* are shown on the right. L: Alkaline hydrolysis of a 5' ^{32}P-labeled RNA oligonucleotide with the nucleotide number indicated in white.

Figure 1

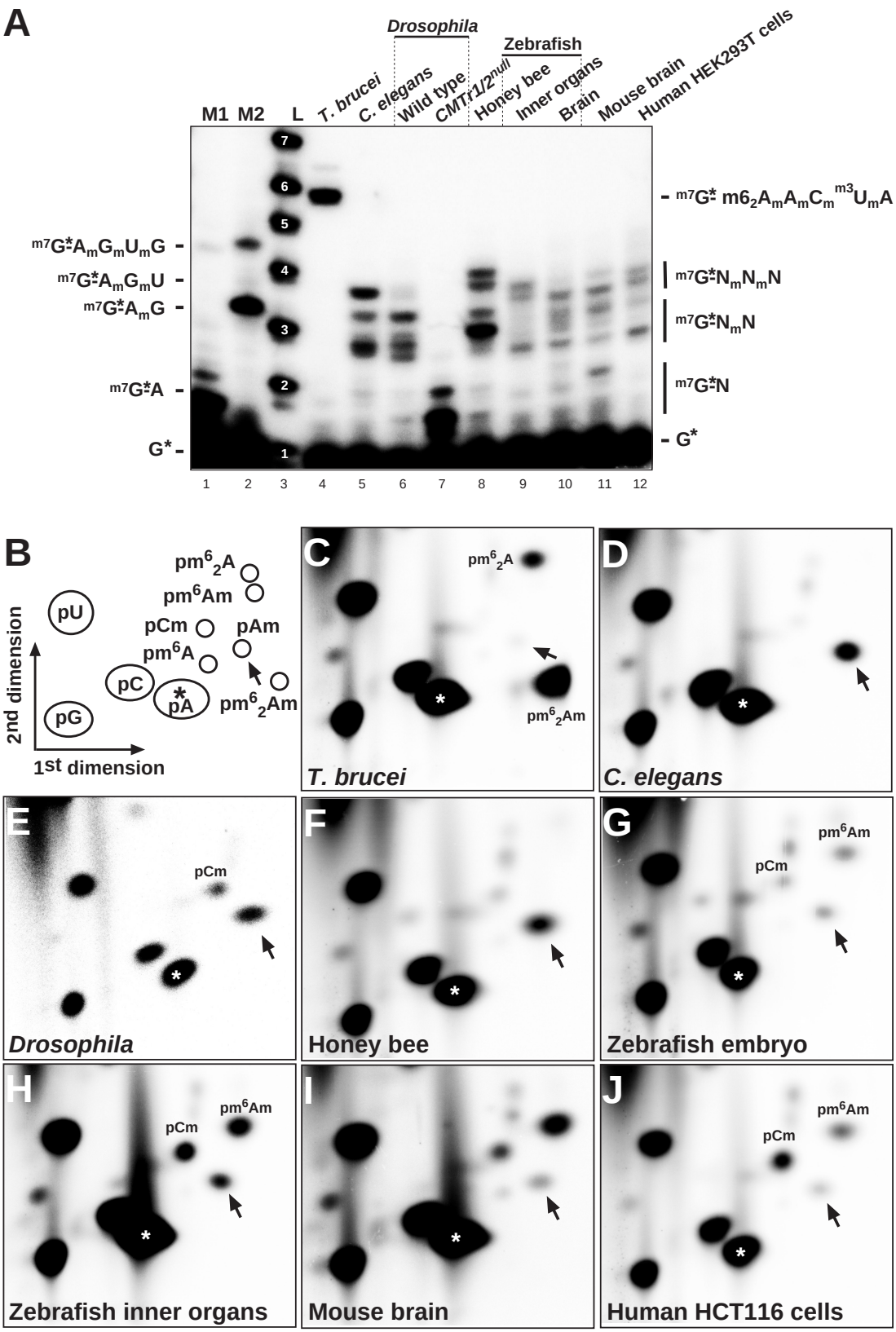


Figure 2

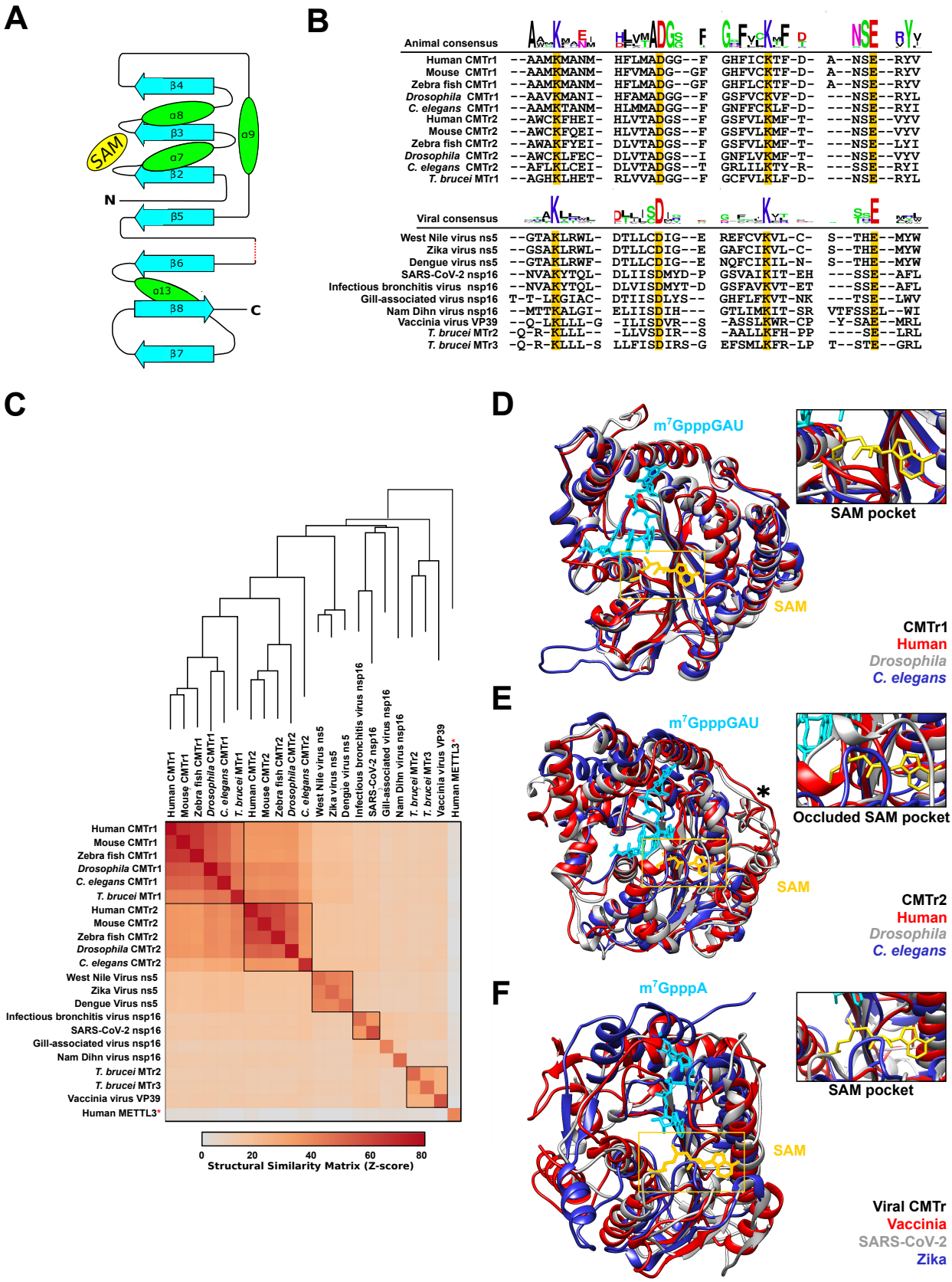


Figure 3

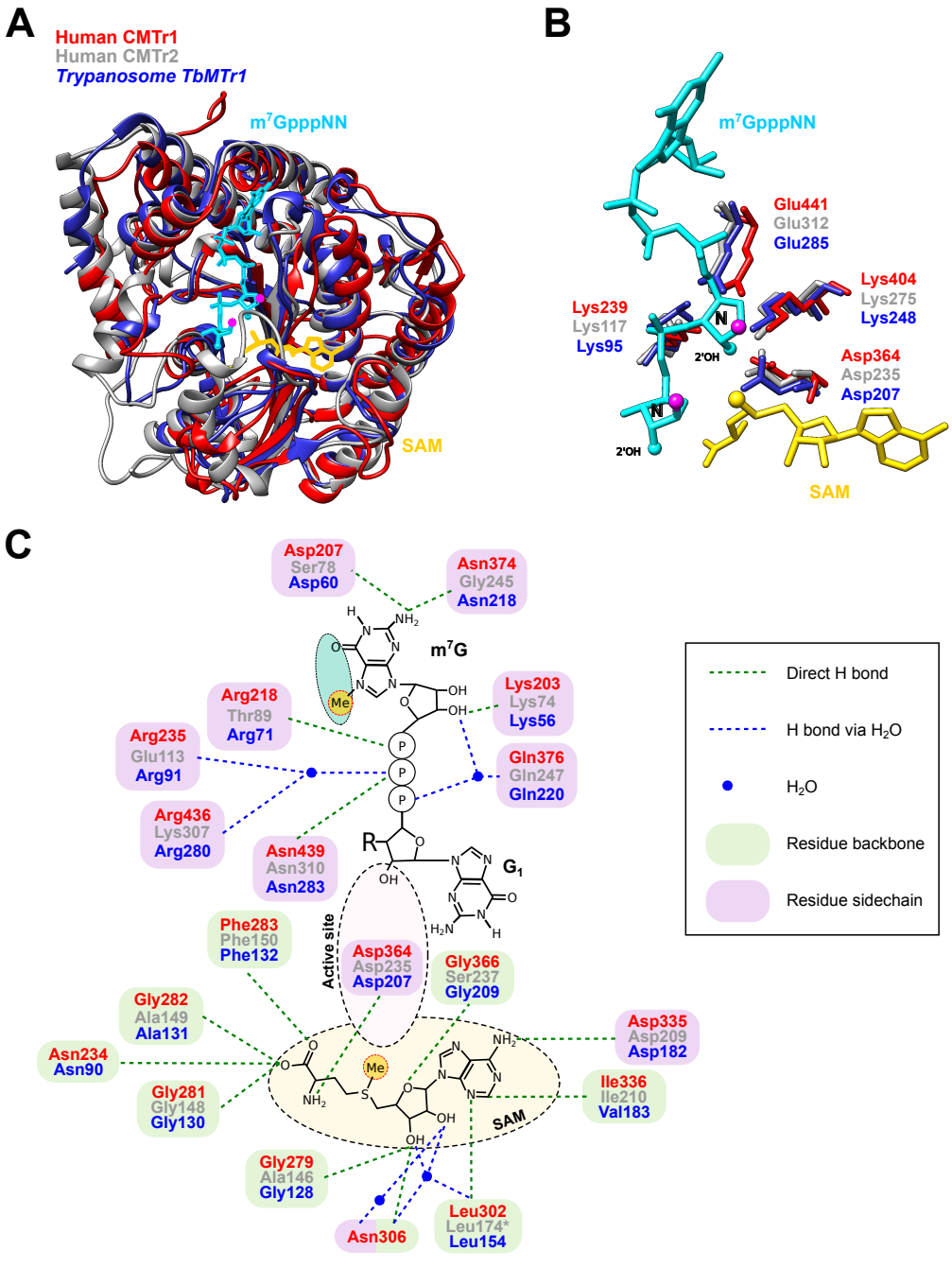


Figure 4

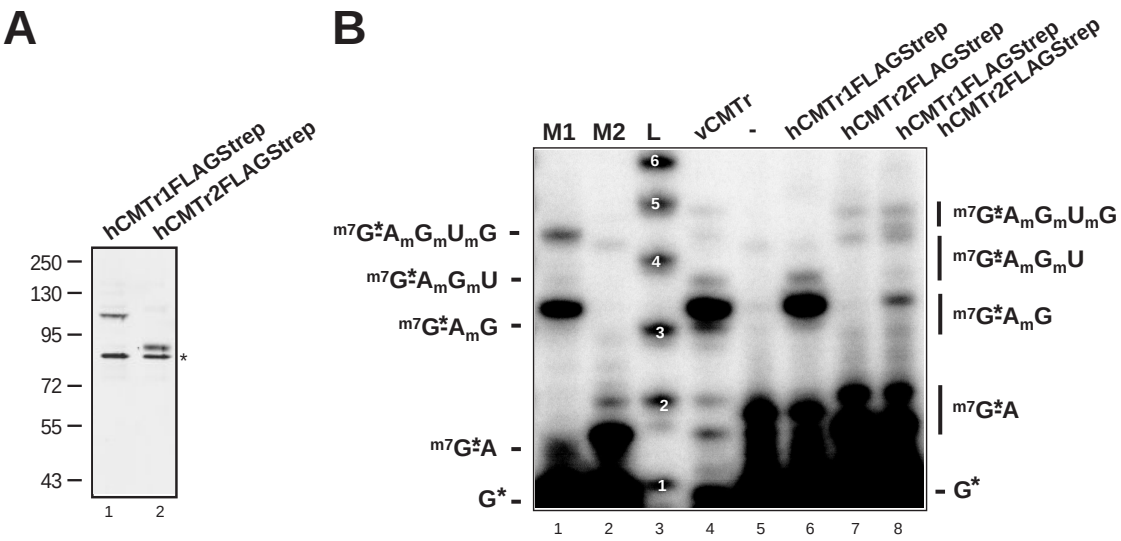


Figure 5

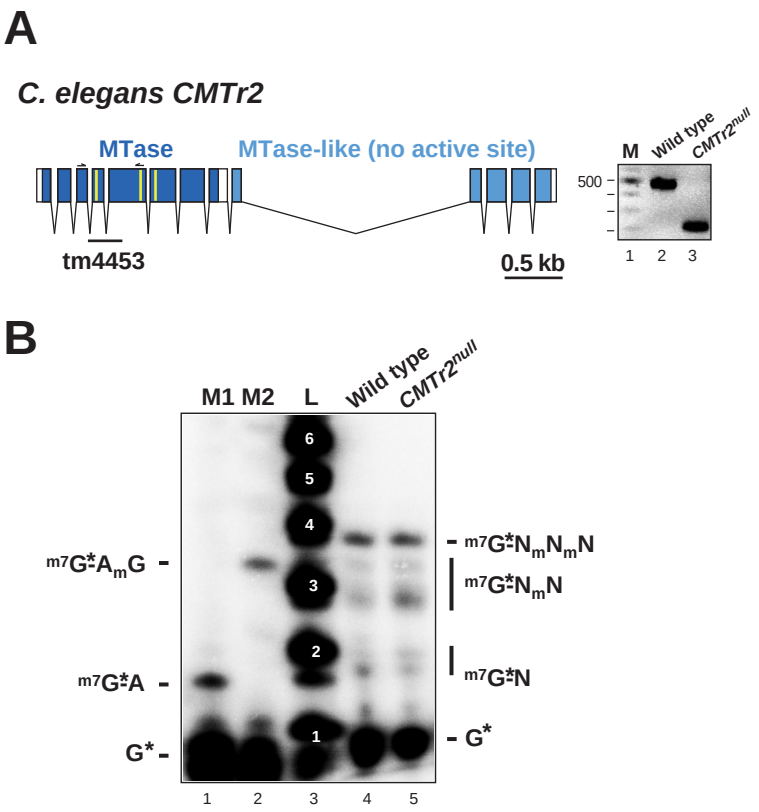
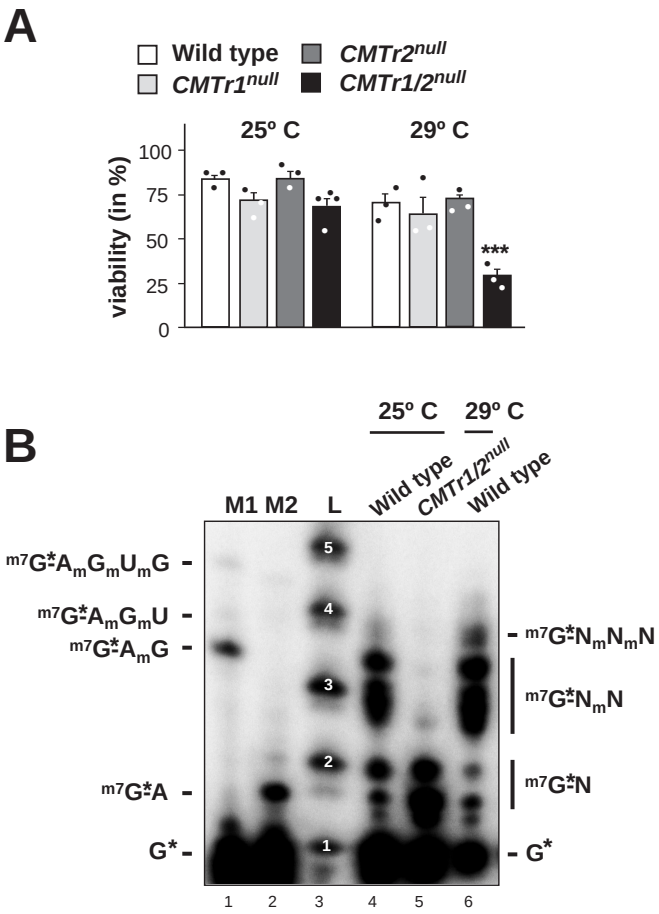


Figure 6



CMTr mediated 2`-O-ribose methylation status of cap adjacent-nucleotides across animals

Thomas Dix, Irmgard Haussmann, Sarah Brivio, et al.

RNA published online August 15, 2022