

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

THOMAS C. DIX^{1,2*}, IRMGARD U. HAUSSMANN^{3*}, SARAH BRIVIO⁴, MOHANNAKARTHIK P. NALLASIVAN^{1,2}, YAVOR HADZHIEV^{2,5}, FERENC MÜLLER^{2,5}, BERNDT MÜLLER⁴, JONATHAN PETTITT⁴ AND MATTHIAS SOLLER^{1,2,6}

Figure legends for Supplemental Figures

Supplemental Figure S1. Alignment of CMTr1 from human, mouse, zebrafish, *Drosophila*, *C. elegans* and trypanosome TbMTr1. The domain structure is indicated on top of the sequence. Identical and conserved amino acids are indicated in red and orange, respectively. Blue arrows indicate the amino acids of the catalytic tetrad and red arrows the boundary for the methyltransferase domain used for modelling.

Supplemental Figure S2. Alignment of CMTr2 from human, mouse, zebrafish, *Drosophila* and *C. elegans*. The domain structure is indicated on top of the sequence. Identical and conserved amino acids are indicated in red and orange, respectively. Blue arrows indicate the amino acids of the catalytic tetrad and red arrows the boundary for the methyltransferase domain used for modelling.

Supplemental Figure S3. Alignment of PCIF1 from human, mouse, zebrafish, *Drosophila*, *C. elegans* and *T. brucei*. The domain structure is indicated on top of the sequence. Identical and conserved amino acids are indicated in red and orange, respectively. Pink arrows indicate amino acids in the catalytic centre important for catalytic activity and green arrows amino acids involved in cap binding.

Supplemental Figure S4. Validation of modelled structures. Superimposed structures for human CMTr1 (A), vaccinia virus VP39 (B) and SARS-CoV-2 nsp16 (C) are shown on the left with the X-ray structure in orange, the modelled structure in magenta, the cap analogue in light blue and SAM in yellow. On the right side the error estimate (in Angstrom) is plotted for each residue with the methyltransferase domain indicated in yellow. SARS-CoV-2 nsp16 was co-crystallized with its bound co-factor nsp10.

Supplemental Figure S5. Structural comparison of animal CMTr1, CMTr2 and viral CMTr tetrad configuration, cap and SAM binding.

(A-C) Configuration of the four amino acids forming the catalytic centre of CMTr1 and CMTr2 for human (red), *Drosophila* (grey) and *C. elegans* (blue) and viral CMTr for Vaccinia VP39 (red), Zika ns5 (grey) and SARS-CoV-2 nsp16 (blue), respectively. SAH: S-adenosylhomocystein (orange). The position of the cap analogue (light blue) and the SAM (yellow) or SAH (orange) were inserted by superimposition from the published structure (PDB 4n48 in A and B, and PDB 6wks for nsp16 and PDB 1av6 for VP39 and ns5 with the cap analogue in white for VP39 in C).

(D-F) Substrate and co-factor recognition of CMTr1 and CMTr2 for human (red), *Drosophila* (grey) and *C. elegans* (blue) and viral CMTr for Vaccinia VP39 (red), Zika ns5 (grey) and SARS-CoV-2 nsp16 (blue), respectively. 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 orange 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.

Supplemental Figure S6. Comparison of vaccinia VP39 structure with trypanosome TbMTr2 and TbMTR3.

(A) Superimposition of vaccinia VP39 (red), modelled trypanosome TbMTr2 (grey) and TbMTr3 (blue). The position of the cap analogue (light blue) and the SAH (yellow) were inserted by superimposition from the published structure (PDB 1av6).

(B) Configuration of the four amino acids forming the catalytic centre of vaccinia VP39 (red), modelled trypanosome TbMTr2 (grey) and TbMTr3 (blue). SAH: S-adenosylhomocystein.

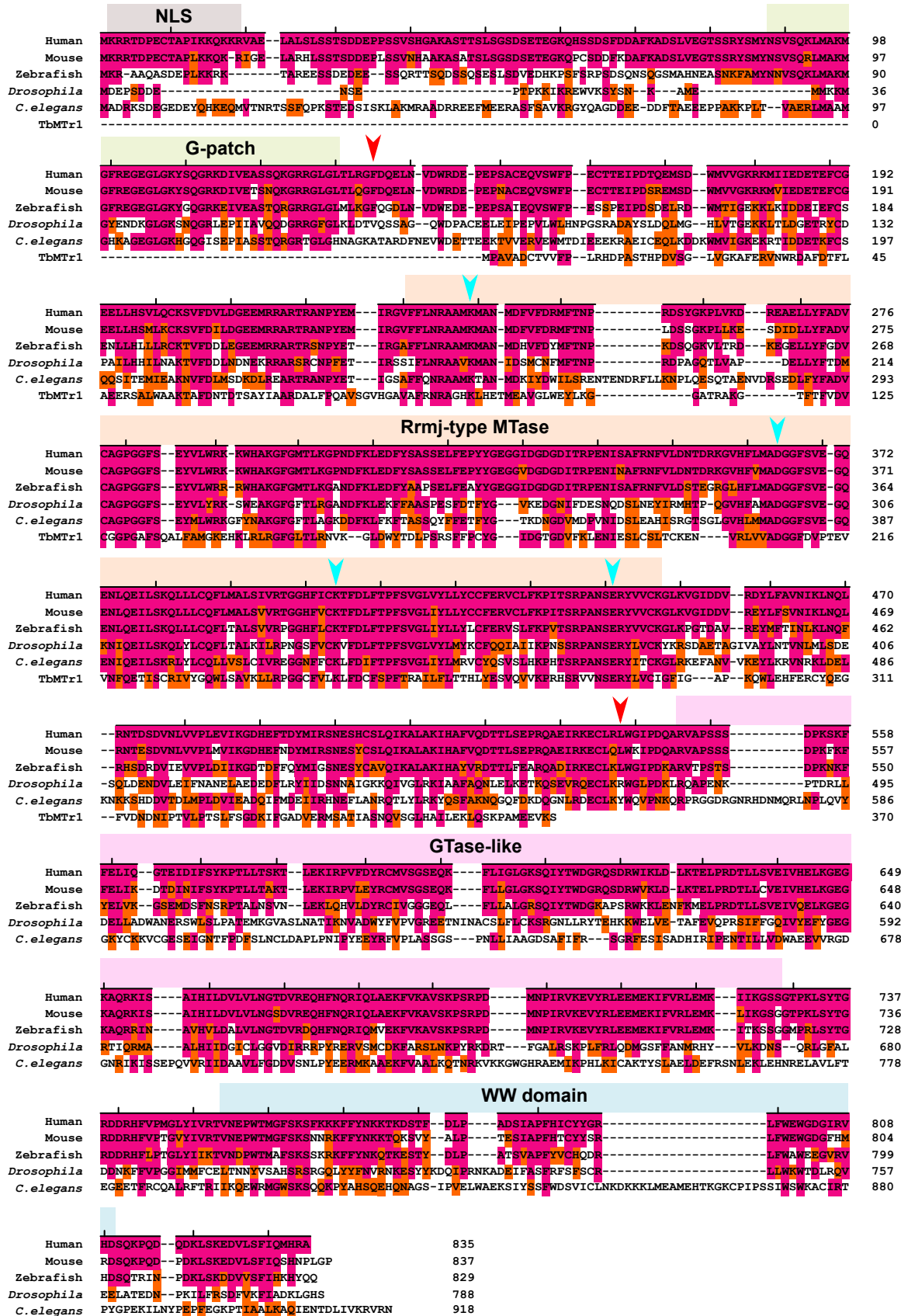
(C) Substrate and co-factor recognition of vaccinia VP39 (red), modelled trypanosome TbMTr2 (grey) and TbMTr3 (blue). Amino acids contacting the cap binding are shown on top and amino acids contacting the co-factor SAH (yellow) are shown at the bottom. The catalytic centre is indicated by a dashed circle 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.

Supplemental Figure S7. mRNA cap 2'-O-ribose methylation is not used for temperature adaptation in *C. elegans* and *Drosophila*.

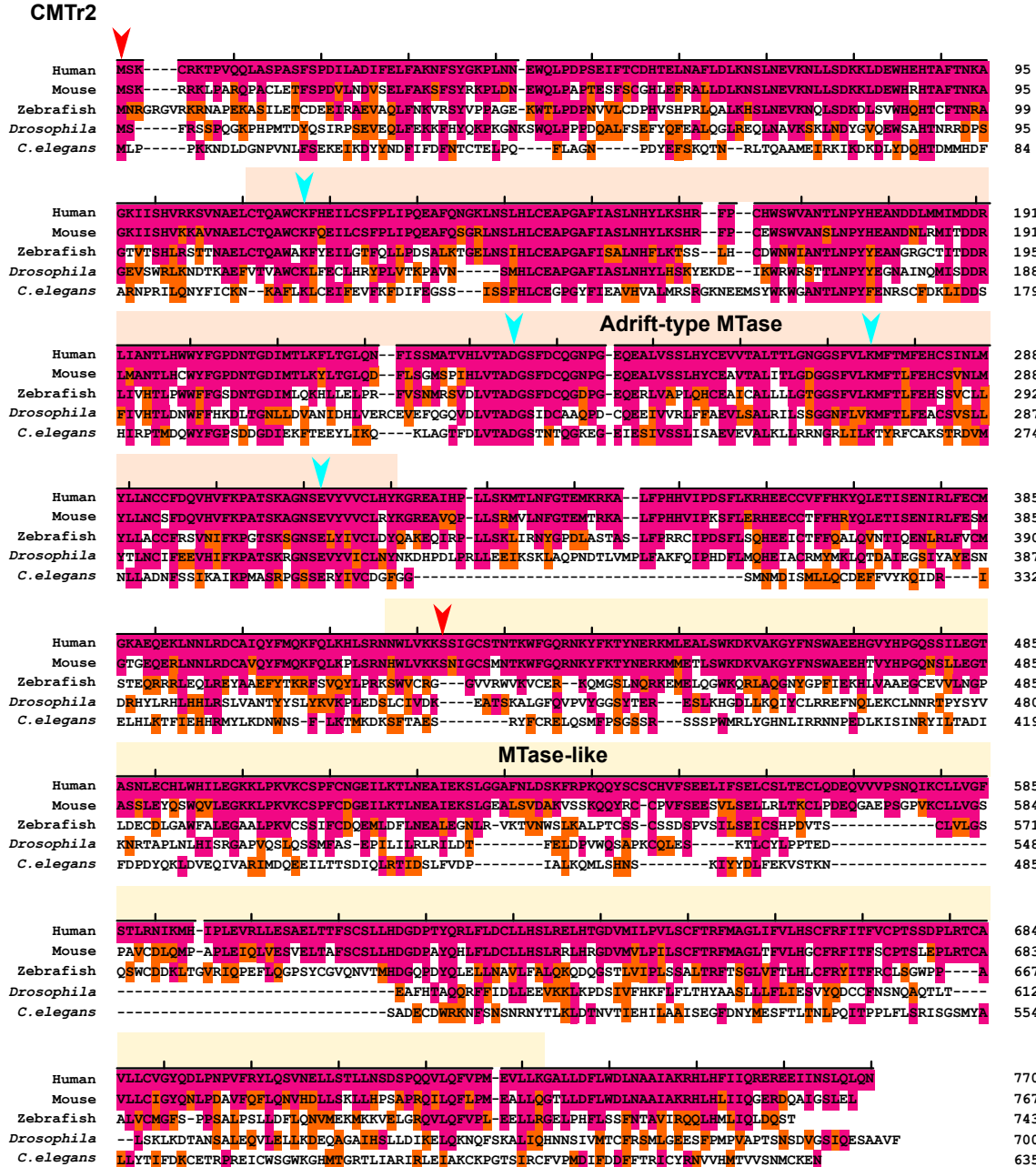
Recapping of mRNA with ^{32}P -alphaGTP from *C. elegans* and *Drosophila* reared at low and high temperature (15°C and 25°C for *C.elegans* and 18°C and 29°C for *Drosophila*). Markers – M1: RNase I digested ^{32}P -alphaGTP capped *in vitro* transcript starting with AGU. M2: RNase I digested ^{32}P -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 are shown on the right. L: Alkaline hydrolysis of a 5' ^{32}P -labeled RNA oligonucleotide with the nucleotide number indicated in white.

Supplemental Figure S1

CMTr1



Supplemental Figure S2



Supplemental Figure S3

PCIF1

Human	-----MANENHC-----SPREEASLLSHSPGT-----SN	24
Mouse	-----MANENHC-----SPRECASLLSHSPGT-----SS	24
Zebrafish	-----MTSENHT-----TIKADSAIVMSPTGS-----TS	24
Drosophila	MAANNNSIINHLAGTEGLSSVIASTGTGPMGSSSPVPSRASSSSASSAAQSAWDQLTASASNMSPTAGAPTGTGPIILPDNLGVAGSSPVASCSS	100
Honey Bee	MNSKFCCTFSTNWKRYHYTQQSVHSTMNEVSGGKQDISNTSTWETLSGGPASSSLTMHHHP--TLQQDT--TPTVAQVIVPTPKVL--QTPLLS--ST	92
T. brucei	-----MVRVTKREGVSNHFC-----SAYVNCEIISAFIKQ-----K	31

WW domain

Human	OSQPCSPKPIRLVQ-----DLPEELVHAGWEKWCWSRRNRPYYFNRTNQSLWEMPVLG--QHDVISDPLGLNAT--PLPQDSSLV	101
Mouse	OSQPCSPKPIRLVQ-----DLPEELVHAGWEKWCWSRRERFPYYFNRTNQSLWEMPVLG--QHDVLSLPLGLNAT--PLPQDSSLV	101
Zebrafish	QAAPFSPSTSKFIQ-----ELFDELIQAGWSKWCWSKRENRPYYFNRTNQSLWEMPVLG--QHDVISDPLGLNAA--PASGEANAD	101
Drosophila	PSTPTKTHAPSPLESMGHTPQGPPTLGPGGYG--EELTAEVLNQGWKRFWSKRENRPYYFNKVTGESLWEMPGTRP--FDPLTDLPLGICHAGGPTDIPNMH	198
Honey Bee	QSLPDHCSQLGHIQSSGLMTQGTGTP--PGTFPESELSPQLQGGWKKFWSKRENRPYFNKLTGESLWVLPPLKQPTPTDPLGICG--VP--PASG	183
T. brucei	REENSISGLCPSAH-----TAVENSTIRAGSRTAFLLHVPISCGPSSPVGKECGAPENNDNSVENEDTGTYLKLVIGDEVVSSIMS	111

Human	ETPPAENKPKRQLSE--EQPSGNGVKKPKIEIPVPTTGQSVSPSSIPGPTPLKMWG--TSPEDKQ-----AALLRPTVEVYWDLDIQTNVAVIKHRGPS	192
Mouse	ETPPVENKSRKRQLSE--EQPSGNGVKKPKIEIPVPTTQSVPSSIPGPTPLKMWG--SSTEDKQ-----AALLRPTVEVYWDLDIQTNVAVIKHRGPS	191
Zebrafish	AG--LNGGQRKRHPSEDASQAGFNSFKRPKVEIBATPTTPTVEIPSPSTPG--VKPWNTTDEKQGOASTPAPAPYRFSVVYWDLDIQTNVAVIREAPA	196
Drosophila	QQQHQHHHLKRRPSD-----DMHIEHNOHHMGGGPPM--KK-----FVLAGP--WDLEVCITNAVIVERPPT	256
Honey Bee	NGTIPSGGTLKRRASE-----DSVV-----PAT--KK-----FVLAGP--WDLEIPTNTVIYERAPS	231
T. brucei	ADDELRSDALRLSC--K-----LCEVFRIEVETITGDNLSHG-----NCLWSEE--GSCSVLTPELRELSR	171

Helical domain

Human	EVLPPHPVEVLLRSQLI--LKLQHYRELCCQREGIE--PPRESFNRMWLERKVVVDKGSDFLLPSNCEPVVSPSMFREIMNDIPRLSRIRKFREEAKRLLF	290
Mouse	EVLPPHPDVELLRSLI--LKLQHYRELCCQREGIE--PPRESFNRMWLERKVVVDKCDPLPSNCEPVVSPSMFREIMNDIPRLSRIRKFREEAKRLLF	289
Zebrafish	DHLPPHPEIEIQRQLT--TKLRQHYELCSCQREGIE--PPRESFNRLWLERKVVVDKGLDPLPSCEDPVISPSMFREIMNDIPRLSRIRKKEEAKRLLF	294
Drosophila	LPLQHPHVEALRAAFT--MKLLKTYEDLCMRRENK--APRDSFNRLWLERKVIDTCCDPLPSSCMPEIISMSYREIMSDIPKIVKPKFTGDARKQLSR	354
Honey Bee	NLPHVHPEAEALRCGLI--AKLRQCYQELCHTRESID--APKDSFNRLWLERKVIDCGSDPLLPSCQCFEISMSYREIMNDIPKIVKPKFTGDARKQLSR	329
T. brucei	WIMQSLNEATVASSHRNGVTLDCGATVCLKSEVLNDEPKRELSSEFVRKRNKYPENAFQYRVCDVADMANGQY--WVEVEPLITGPESTGSVARRRFD	268

Human	YAEAAARLLIESRSASPDSSRKVVKNVEDTFSWLKRDHSASKEDYMDRLEHLRQCQGPVHSAAAKDSVEGICSKIIYHISLEYVKRIRKHLAILKENNISE	390
Mouse	YAEAAARLLIESRSASPDSSRKVVKNVEDTFSWLKRDHSASKEDYMDRLEHLRQCQGPVHSAAAKDSVEGICSKIIYHISLEYVKRIRKHLAVLKENNISE	389
Zebrafish	YAEAAAKMIDSRNATPPSRKVVKNVEDTMNWLRRDHSASKEDYMDRLEHLRQCQGPVHSAVAKDSVEGICSKIIYHISAEYVRRIRQAHLTLKECNISV	394
Drosophila	YAEAAKQIIIESRSAPAESRKVVKNVEDTFQWLRRTVGASIEDQDLRAHLKRCQCPHLIVETVKS SVETLCVKIYHLSADHARKIRERHSQLLKEHGPIE	454
Honey Bee	YAEAAKMIIESRAASSESARKVVKNNAEDTFQWLRRTVGATFDDQDLRAHLKRCQCPHLITETVKSASVEGICLKIYHLSLEYAKKVRDKNNQILLKDNGLGN	429
T. brucei	RDQLLLSPPOFVYFDELLPVDEINNSRGTSACLFLSAVLAVGRIRSERAAIGDLSMWDNYPKNVTLMQYASRDIRRVHVELCAARDALHVEGGTEGTD	368

Human	E---VEAEVEEPLRVICYF--VRLAVSAPPMPSPVEMHMENNVCIRYKGMVK-----VSRNYFSKLWLLYRSCIDDS--AFERFLPRVWCLLRRY	475
Mouse	E---VEASELEPLRVICYF--VRLAVSAPPMPSPVEMHMENSVVCIRYKGMVK-----VSRSYFSKLWLLYRSCVDDS--AFERFLPRVWCLLRRY	474
Zebrafish	DG---TESAEVQDLRVICYF--VRLSIPAPPQTRVELHFENDACLRFKGMVK-----VSRGHFNKLELLYRSCIDDP--RFEKFLSRVWCILKRY	480
Drosophila	PTP--PP--PPPHLKVVICYF--IQFAVPSPRMPTIEYLQDRDHMIKYPATINQPDQAQYINLTYLQKLEQLYRHNCFFDK--KFDLFIGRVWCILKRY	546
Honey Bee	VIP--LGGFASAQRKVCYF--VQFSLPTERLPQVDYLPBEREQTMLRFHGTV-----CINNIHLAKLEHLRYRNCFFDK--KFEMFLPRVWCILKRY	516
T. brucei	VKKEELRLRLTDYVTLTRYKGTIEVLQRTVERLSLLWDARLAVRRVVGARR-----AHELNLDVAVETQAASDFSEKIHVFSTVEPFEMRLFTLLRY	461

m⁶Am MTase domain

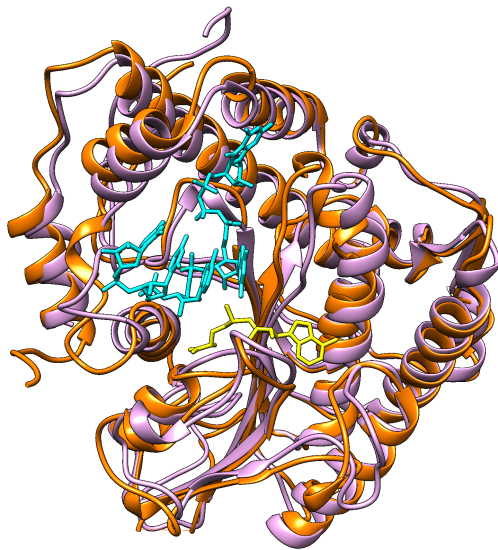
Human	QMMFGVGLY--EGTGLQGSILPVHVFALHRLFGVSFEFCASPLNLCYFROYCSAFPDTDGYFGSRGPCLDFAPLSGSFEANPPFCEELMDAMVSHFERLL	572
Mouse	QMMFGVGLY--EGTGLQGSILPVHVFETLHRLFGVSFEFCASPLNLCYFROYCSAFPDTDGYFGSRGPCLDFTPLSGSFEANPPFCEELMDAMVSHFEKLL	571
Zebrafish	QVMFGSSVN--EGSGLQGSILPVVFEEALNKQFGVTFECFASPLNLCYFQFCASAFPDIDGFFGSRGFFLSFSPASGSFEANPPFCEELMDAMVTHFEDLL	577
Drosophila	QTLFLGAINSSQAEALTAALPVVFECLEHROFGVSFEFCASPLNLCYFROYCSAFADTDAYFGSRGFFLDKFPVSGSFQVNPFPCEELIEASLLIDKLL	646
Honey Bee	QTYLG--IN--EQQATQMALPVTVFECLQRSFGVTFECFASPLNLCYFROYCSAFADTDSYFGSRGAFLDREPVSGSFQANPPYCEELMEAMVNHFERLL	611
T. brucei	SSLFG--DLG--YNGQPHAAVPPALMQLCQVDFLQCAFASPLNAQLPLFCSLFPDIDYFFGSLGSFFDIALTAGHFEVNPPEVTAVALQREKLKLLNDT	557

Human	E-----SSPEPLSFIVFPEWREPTPALTRMEQ--SRFKRHQLILPAFEHEYRSQSHICKKEEMHYKAVHNTAVLFLQNDPGFAKWAPTPERLQELSA	666
Mouse	E-----SSAEPLSFIVFPEWREPTPALTRMEQ--SRFKRHQLVLPAFEHEYRSQSHICKKEEMHYKAVHNTAVLFLQNGPFAKWGPTPERLQELTAA	665
Zebrafish	G-----RSSEPLSFIVFPEWRDPPTPALTRMEA--SRFRRHQMTVPAFEHEYRSQSHICKREEIYKAIHGTAVIFLQNNAGFAKWEPTTERIQELLAA	671
Drosophila	T-----DTMEPLSFIVFLPEWKS-----ISKLDD--SMYKRRSMVVLGMAHEYRHGYOHLQKSDVLICKMGQTOVVLQNSAGYARWGNEIRVEALREA	735
Honey Bee	A-----DSTEPLSFVFLPEWRDPAPNALIKLES--SHFKRKQVVVPAMEHEYRHGFOHILPK-----ISFFI-----IEVKSILEQ	681
T. brucei	LPINDSDGAASMLFVVVLSHDLDETERSDGLGGAPVAGNTTLRQNNNCDRKRGRFENDGRDKGGGNVSTDRALRESPLYCLAHVLCSAHSSVYVDGHQ	657

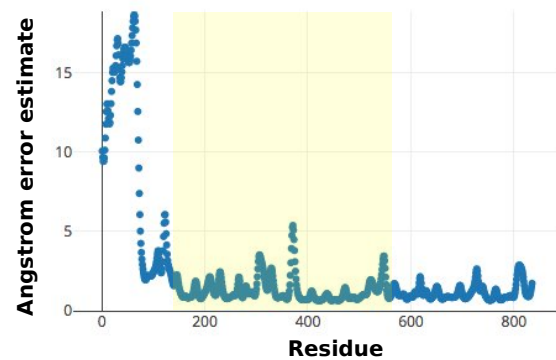
Human	YKQS-----GRSH--SSGSSSSSSSEAKDRDSGREQGPSREPHPT	704
Mouse	YKQS-----GRSHGSSSSSSSSSSSEAKDRDSGREQGPSREPHPT	706
Zebrafish	YKVS-----GRSLP--SPGPSSTNTGKDSKPAPERTAPSQDNSSPVDKTAQDTTNT	721
Drosophila	FRPQDRERERAAAAAAGVTSSQCSPTPPSAASTPTTNSCTISGVNSNSSATTSTSSPSLTSTSSMSPLSHTTPTGNPPQPFMTISNTSS	835
Honey Bee	HMEH-----	685
T. brucei	HMLR-----APLFCISPTRLVLVGNRTARLRYSDAATRLSVREAWRGYTLENRVDS	711

Supplemental Figure S4

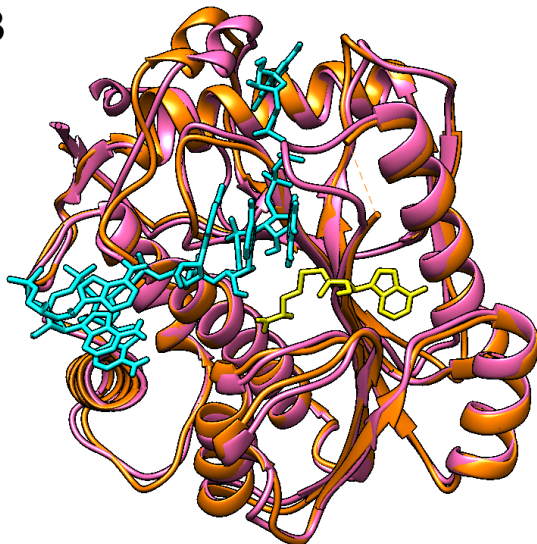
A



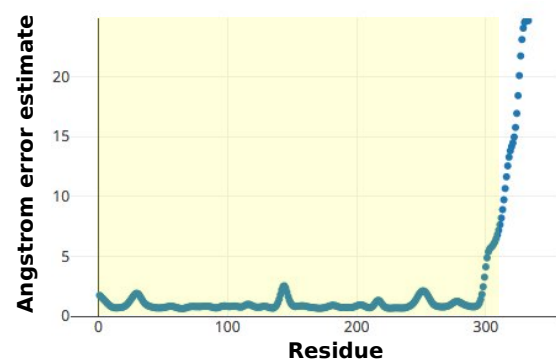
Human CMTr1 (PDB 4n48)
Human CMTr1 (RoseTTA)



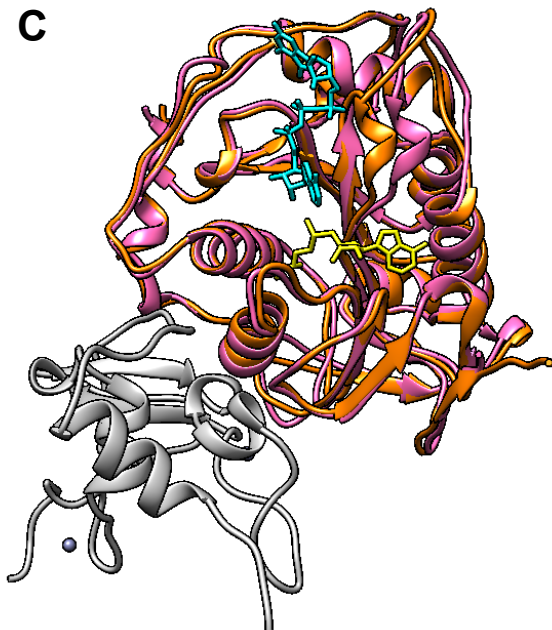
B



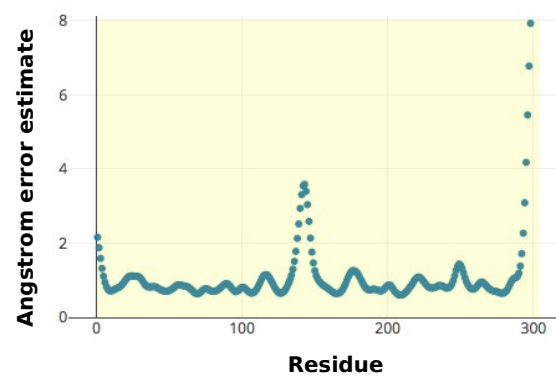
Vaccinia virus VP39 (PDB 1av6)
Vaccinia VP39 (RoseTTA)



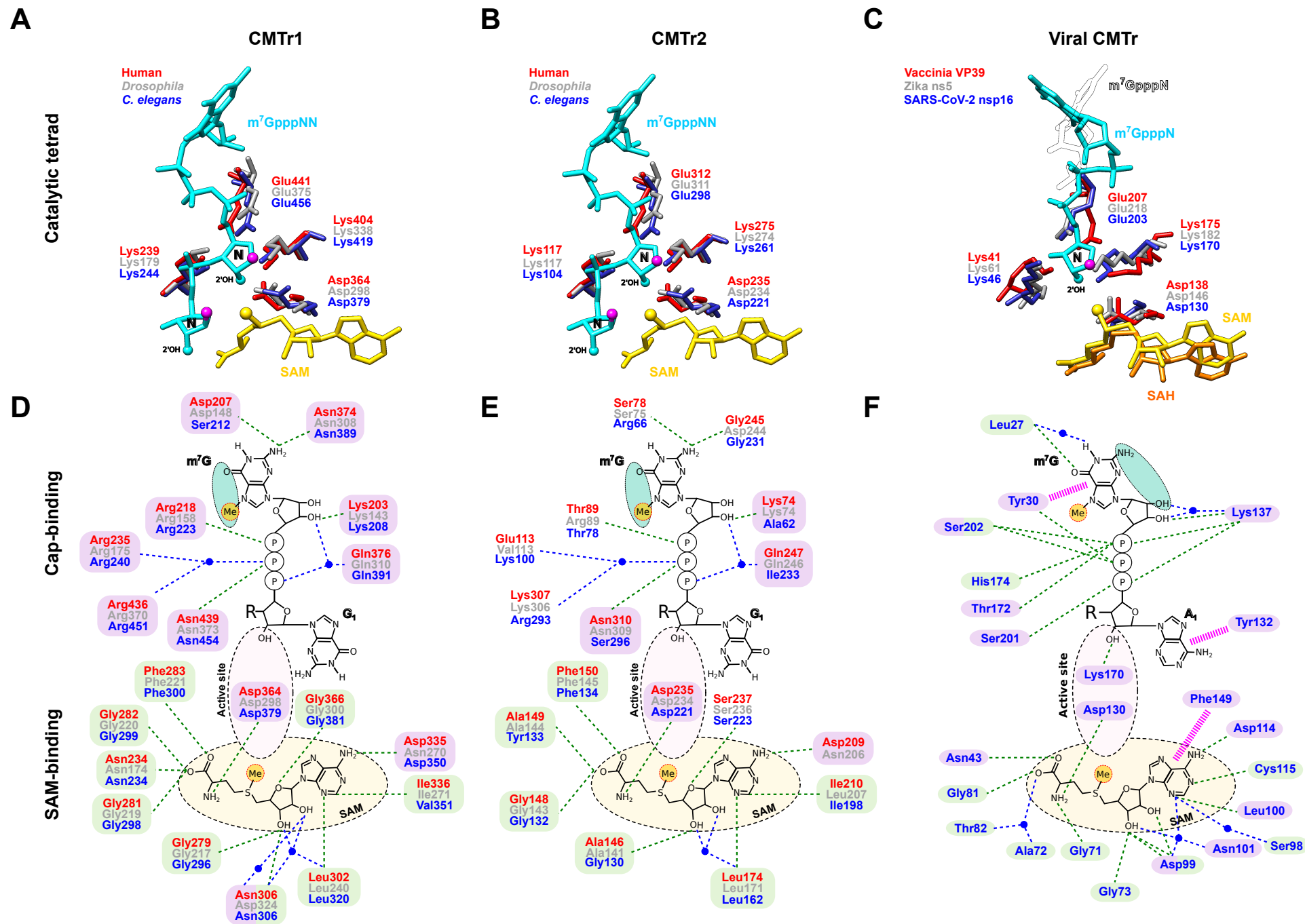
C



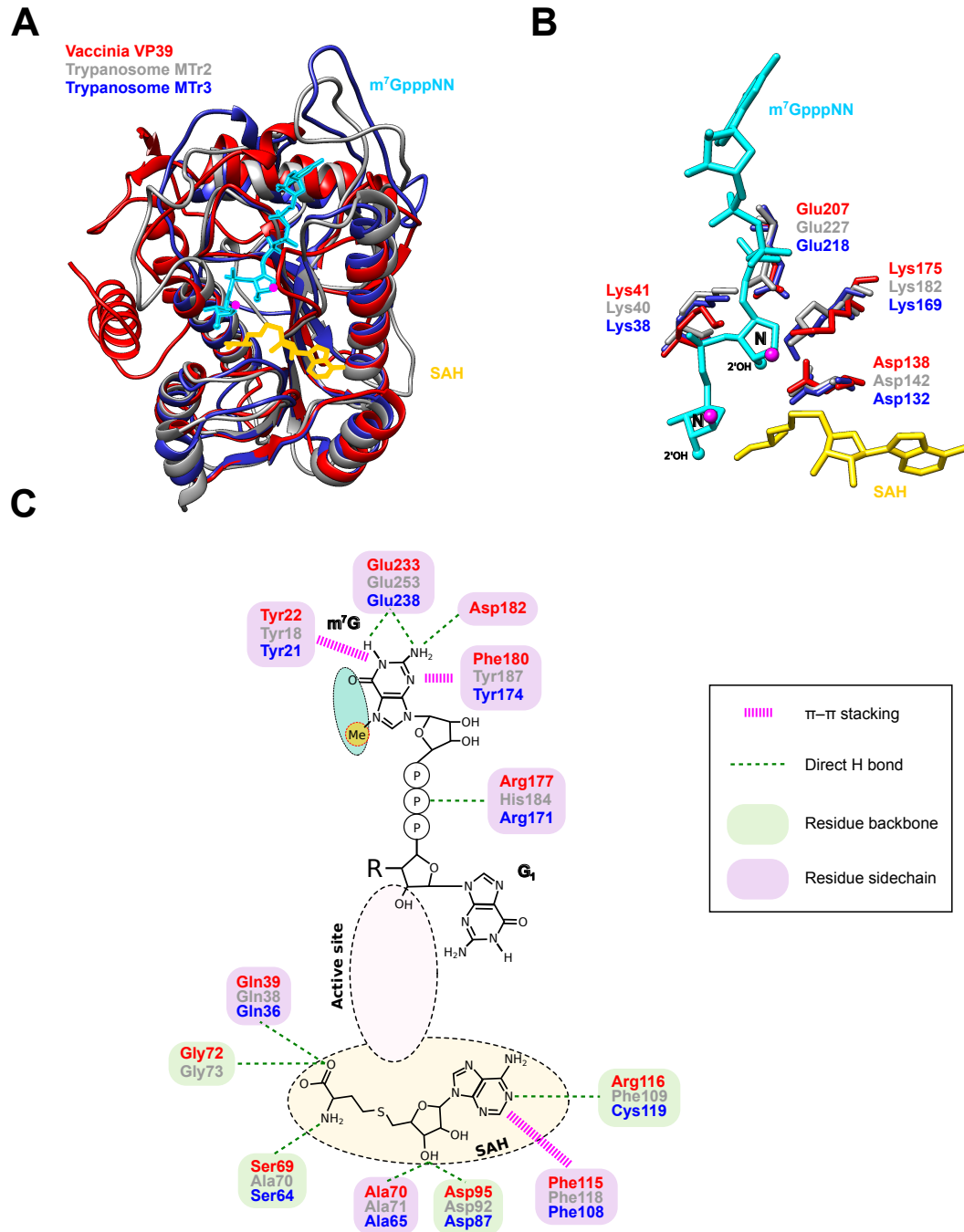
SARS-CoV-2 nsp16/nsp10 (PDB 6wks)
SARS-CoV-2 nsp16 (RoseTTA)



Supplemental Figure S5



Supplemental Figure S6



Supplemental Figure S7

