The C-Terminal Domain of the Sudan Ebolavirus L Protein Is Essential for RNA Binding and Methylation

Coralie Valle, Baptiste Martin, Francoise Debart, Jean-Jacques Vasseur, Isabelle Imbert, Bruno Canard, Bruno Coutard, Etienne Decroly

The C-Terminal Domain of the Sudan Ebolavirus L Protein Is Essential for RNA Binding and Methylation

Coralie Valle[1#], Baptiste Martin[1#], Françoise Debart[2], Jean-Jacques Vasseur[2], Isabelle Imbert[1], Bruno Canard[1], Bruno Coutard[3] & Etienne Decroly[1*]

[1]AFMB, CNRS, Aix-Marseille University, UMR 7257, Case 925, 163 Avenue de Luminy, 13288 Marseille Cedex 09, France

[2] IBMM, UMR 5247 CNRS, Montpellier University, ENSCM, Montpellier, France

[3] Unité des Virus Emergents (UVE: Aix-Marseille Univ-IRD 190-Inserm, 1207-IHU Méditerranée Infection), Marseille, France

# These authors contribute equally to this work.* Corresponding author (etienne.decroly@afmb.univ-mrs.fr)

ABSTRACT

The large protein (L) of mononegaviruses is a key player of virus infection as its N-terminal region bears the RNA-dependent RNA polymerase activity, which is essential for replication/transcription, and its C-terminus contains a cap assembly line composed of a capping domain, and a methyltransferase domain followed by a C-terminal domain (CTD) of unknown function. For viruses belonging to the *Mononegavirales* order, the capping domain, which harbours a polyribonucleotidyltransferase activity, has been shown to transfer a GTP molecule to the 5' end of nascent viral RNA forming the cap structure. The MTase activity of L protein next sequentially methylates the 2'O and the N7 position of the cap structure. In addition, the MTase of *Sudan ebolavirus* (SUDV) induces also internal adenosine 2'O-methylations, which might prevent viral RNA sensing by innate immunity. In this work, we addressed the regulating role of the CTD of SUDV L protein on the MTase activity. We demonstrated that the CTD, which is enriched in basic amino acids, plays a key role in the RNA recognition. Consequently, we showed that the MTase domain is inactive when expressed in absence of CTD (MTase-ΔCTD) whereas the MTase+CTD is enzymatically active. The role of the CTD in the RNA recognition and in the regulation of the MTase activity was further confirmed by mutagenesis. We demonstrated that single mutations in the CTD domain can uncouple or abolish the different MTase activities. Altogether, our results elucidate the role of the SUDV L protein CTD domain in the RNA capping pathway, which is essential for RNA translation into viral proteins and virus escape from detection by innate immunity.

INTRODUCTION

Ebola virus is an emerging virus causing severe epidemics such as the unprecedented devastating outbreak of 2015 in West Africa, which killed more than 11,000 people [1]. Although Ebola faded from public attention, outbreaks continue to occur especially in Democratic Republic of the Congo (DRC) where the current epidemic, which has been declared in August 2018, has already killed more than 1,000 citizens [2]. This virus infects human and non-human primates to provoke a massive viremia leading to haemorrhagic fever, which is fatale in most cases. Despite a vaccine is currently in development [3][4] and some therapeutic antibodies have shown antiviral activities [5], we urgently need to develop new therapeutic compounds to treat infected people and contact patients.

Ebolavirus genus contains five strains: *Zaire ebolavirus* (Ebola virus, EBOV), *Sudan ebolavirus* (Sudan virus, SUDV), *Tai Forest ebolavirus* (Tai Forest virus, TAFV), *Bundibugyo ebolavirus* (Bundibugyo virus, BDBV), and *Reston ebolavirus* (Reston virus, RESTV). These filamentous viruses together with *Marburg marburgvirus* (Marburg virus, MARV), also responsible of acute haemorrhagic fever, and *Lloviu cuevavirus* form the *Filoviridae* family [6]. They are enveloped virus with non-segmented negative strand RNA genome (NNS) belonging to the *Mononegavirales* order. These mononegaviruses, which contain several important human pathogens such as measles, RSV and rabies viruses, display similar genetic organization and share common replication strategies.

The genome of filoviruses is about 19 kb long and codes for seven major proteins: the nucleoprotein (NP), viral proteins VP35 and VP40, the glycoprotein (GP), viral proteins VP30 and VP24 and the "large" protein L [7][8]. The viral life cycle is initiated by the interaction of the GP envelope protein with cell surface determinants allowing virus entry into the cell by macropinocytosis. The NP-encapsidated viral genome is next released in the host cell cytoplasm where the virus replication cycle is initiated. The large protein L, associated with NP, VP35 and VP30 first transcribes viral mRNAs by using a discontinuous transcription mechanism that generates seven monocistronic capped and poly-adenylated RNAs [9][10]. The polymerase complex initiates the mRNA synthesis at a conserved gene-start (GS) sequence and transcribes the genes to the gene-stop sequence (GE) [11]. The poly-A tail is then added by a slippage mechanism of the polymerase occurring in a conserved poly$(U)_6$ tract in the GE sequence [12]. The RNA-dependent RNA polymerase (RdRp) then scans the intergenic sequences and reinitiates the RNA transcription at the next GS sequence. The viral mRNAs of mononegaviruses are co-transcriptionally capped by an unusual sequence of capping reactions. This process was elucidated for vesicular stomatitis virus (VSV) and takes place when the nascent RNA chain has reached 31 nucleotides length [13]. The cap synthesis starts with the formation of a covalent adduct between a conserved histidine residue of the polyribonucleotidyltransferase (PRNTase) embedded into the Cap domain and the 5'-phosphate of viral mRNA. In presence of GDP, the PRNTase transfer a GMP moiety to the 5′-diphosphate of the covalently bound RNA, yielding to the formation of the cap structure (GpppN). The 5' cap structure is subsequently methylated at the 2'O position of the first nucleotide (N1) and at the N7 position of the guanosine (^{m}GpppN$_{m}$ or cap-1) [14]. This cap-1 structure is undistinguishable from cellular caps and plays a critical role for virus replication as it protects viral

mRNAs from cellular 5' exonucleases and it allows the eIF4e-dependent initiation of mRNA translation. Additionally, the cap-2'O methylation is a self-marker that limits the detection of infecting viral RNAs by the innate immunity sensors of the retinoic acid-inducible gene-I (RIG-I) like family. Thus, viruses expressing RNAs lacking cap 2'O methylation (^{m}GpppN-RNA or cap-0) are early detected by RIG-I receptors in infected cells leading to the secretion type I interferon ($INF_{\alpha/\beta}$). Furthermore, INFs stimulate the expression of INF-stimulated genes (ISG) such as INF Induced Protein with Tetratricopeptide Repeats (IFIT) that sequesters miss-capped RNAs (ie: cap-0) leading to restriction of virus propagation (for review see [15]). Capping enzymes such as cap-MTases are thus considered as potent antiviral targets as the inhibition of N7 MTase is supposed to limit RNA translation into viral proteins and 2'O MTase inhibitors should unmask viral RNA to the intracellular self-sensors.

The structure of VSV L protein, recently solved by cryo-electromicroscopy (cryo-EM) at 3.8 Å, reveals its organization in 5 main topological domains [16]. The RdRp, that contains three conserved regions (CRI, CRII, and CRIII), is intimately associated with the PRNTase domain (CRIV) forming a "donut-like" structure observed by negative staining [17]. The RdRp is organised as a right-hand "fingers-palm-thumb" structure, typical of polymerases. The Cap domain (PRNTase), which harbours an original fold, projects a loop near the catalytic site of the polymerase. This structural interplay suggests that the PRNTase domain participates to the initiation of polymerisation similarly to polymerases having a priming loop. In addition, the cryo-EM structure corroborates biochemical analysis showing the interplay between RdRp and capping activities [13]. The EM studies shows that the "donuts-like" structure (RdRp and PRNTase domains) is followed by 3 globular structures corresponding to the connector domain (CD), the MTase domain and a small C-terminal domain (CTD), respectively. The CD forms a globular domain, separating the PRNTase domain from the MTase domain, which is composed of a bundle of eight helices. It is likely that the CD domain plays an organizational role as it interacts with the viral P protein. Conversely, the MTase domain contains a K-D-K-E catalytic tetrad, characteristic of 2'O MTases. The VSV cryo-EM structure reveals its organization in a typical Rossman fold of S-adenosyl methionine (SAM) dependent methytransferases. The 2.2 Å X-ray structure of human metapneumovirus (hMPV) MTase+CTD domain confirms the methyltransferase fold with a SAM-binding site positioning the released methyl group face to the catalytic tetrad (Paesen et al.). In contrast, the MTase+CTD lacks a canonical cap-binding site like observed in flavivirus MTases [18][19][20]. Thus, it is likely that the RNA substrate accommodates an unusually narrow RNA-binding groove formed by the MTase domain overlayed by the CTD enriched in basic amino acids.

The MTase activity of VSV L protein was demonstrated *in vitro* for VSV L protein that sequentially methylates the RNA cap structure at the ribose 2'O and the guanosine N7 positions [21]. The catalytic activity of the MTase domain has been confirmed *in vitro* using MTase+CTD constructs of hMPV and filovirus (SUDV) [22][23]. The MTase+CTD domains of these viruses have also been reported to

methylate the 2'O position of the N1 of capped and triphosphate RNAs. The subsequent methylation of the N7 guanosine was evidenced using longer RNA substrates using SUDV MTase+CTD. Besides this, we also demonstrated recently that the SUDV MTase and hMPV MTase (in a lesser extend) harbour a cap-independent methyltransferase activity inducing internal adenosine 2'O methylations. The role of such epitranscriptomic RNA modification is not yet fully understood, but similar internal 2'O methylations observed inside the HIV RNA genome have been shown to protect viral RNAs from detection by the RIG-like receptor Melanoma Differentiation-Associated protein 5 (MDA5) [24]**.** Altogether, these results suggest that the filovirus MTase evolved towards a dual activity with distinct substrate specificities. The cap-N7 MTase activity that promotes viral protein translation, the cap-dependent and independent 2'O MTase activities inducing cap 2'O methylation and internal 2'O methylation limiting the viral RNA sensing by RIG-I and MDA5 respectively [25][26][27][28].

In this work, we addressed the functional role of the SUDV CTD domain by comparing the MTase activity of MTase+CTD and MTase-ΔCTD constructs. After expression and purification of various SUDV MTase constructs, we demonstrated the critical role of the CTD domain for RNA substrate recognition and thereby controlling the MTase activity. The importance of CTD basic residues in RNA recognition and MTase activity was confirmed by directed mutagenesis and we identified single mutations uncoupling the different activities carried by SUDV MTase. These results highlight the key regulation role of the CTD domain in the RNA methylation process and raise the possibility to identify compounds inhibiting specifically these viral MTases by targeting the CTD.

RESULTS

The CTD domain of SUDV L protein is enriched in basic amino acids and stabilizes the MTase domain.

The C-terminal part of SUDV L protein contains a conserved MTase domain including the conserved region VI (CRVI) followed by a CTD domain of unknown function (**Fig. 1A**). The MTase harbours a K-D-K-E catalytic tetrad, canonic of 2'O MTases, and a SAM binding GxGxG motif highlighted in the multi-alignment of filovirus C-terminal part of L proteins (**Fig. S1**). The MTase domain (residues 1693 to 2036) is followed by the CTD (residues 2037 to 2210), which greatly varies in length within the *Mononegavirales* order (from ~120 residues in *Pneumoviridae* to ~240 residues in *Rhabdoviridae)* (**Fig. S2).** Despite this weak conservation, CTDs are enriched in basic residues in most of *Mononegavirales.* Calculated isoelectric points of each CTD vary between 8.33 and 9.94 except for rubelavirus and bornavirus (**Fig. S2**). The alignment of filovirus CTDs also reveals the presence of 3 strictly conserved basic amino acids (arginine and lysine, indicated by *) and 2 other conserved positions with either arginine or lysine residues (indicated by x) (**Fig. S1**). Two conserved aromatic residues (histidine and phenylalanine, indicated by *), which might participate to RNA recruitment by interacting with

nucleobases by staking, are also present in the CTD. We also analysed the conservation of residues in the protein alignment of the MTase+CTD domains of *Mononegavirales*. Results are represented onto the hMPV structure **(Fig. 1B&C)**. The conservation model indicates that the most conserved residues are localized around the catalytic pocket, the SAM binding site and in a groove enriched in basic amino acids previously proposed to accommodate the RNA substrate [22]. These observations together with the structural analysis of hMPV MTase+CTD showing that MTase domain is overlayed by the CTD to form a narrow RNA binding groove suggest that the CTD plays a key role in RNA recognition.

The CTD domain L protein is essential for the SUDV MTase activity.

We expressed in bacteria the MTase adjoining the CTD (MTase+CTD) and the SUDV MTase domain (MTase-ΔCTD). Both proteins were purified by affinity chromatography on NTA column combined with exclusion chromatography. The purity of both proteins was assessed by SDS-PAGE analysis upon Coomassie blue staining. **Fig. 2A** shows that MTase+CTD and SUDV MTase-ΔCTD migrate at their expected apparent molecular weight of 58 and 39 kDa, respectively. Their identity was confirmed by MALDI-TOF (not shown). Both proteins were also analysed by thermal shift assay (TSA) in order to assess their folding and compare their stability (**Figure 2B**). The proteins revealed typical denaturation curve of folded proteins with a T_m value of 54.3°C and 46.0°C for the MTase+CTD and the MTase-ΔCTD, respectively. These data suggest that the CTD domain increases the stability of the MTase domain. We next demonstrated the involvement of the CTD in the MTase activity by comparing the enzymatic activity of both proteins on a capped RNA (GpppN-RNA) mimicking the conserved 5' sequence of SUDV 5' transcripts. The 13-nucleotides length-capped RNA substrate was incubated with both enzymes in presence of radiolabeled methyl donor (3H-SAM). The reaction products were filtered through a DEAE membrane, and the radioactivity transferred to the RNA substrate was determined (filter binding assay). **Fig. 2C** shows that the MTase-ΔCTD domain was not active on the short capped-RNA substrate whereas the MTase+CTD domain induced the methylation of the capped-RNA substrate.

We also compared RNA binding properties of MTase+CTD and MTase-ΔCTD proteins by fluorescence polarization assay (FP). For this purpose, the capped RNA substrate was labelled at its 3'end by ligation of a pCp-Cy5 residue (GpppN-RNA-pCp-Cy5) and was incubated with increasing concentration of MTase+CTD or MTase-ΔCTD. The interaction of RNA with both proteins was next determined by FP measurement. **Fig. 2D** demonstrates that MTase-ΔCTD does not interact with the capped RNA substrate. Conversely, the MTase+CTD induces an increased FP signal confirming its interaction with the synthetic capped RNA substrate with a K_m value of 740 nM. Altogether, these results demonstrated that the CTD domain of SUDV plays a key role in RNA recognition and, in turn, promotes the MTase activity.

Effect of a single mutation of conserved residues in the SUDV CTD on the MTase activity.

We previously demonstrated that the SUDV MTase+CTD induces 2'O methylation of internal adenosines in addition to N7 and 2'O methylations of the cap [23]. To further characterize the role of the CTD in the RNA recruitment and unravel the involvement of CTD in the regulation of the different MTase activities carried by MTase+CTD, we mutated conserved residues of filoviruses (**Fig.S1**) into alanine. The mutant proteins were expressed in bacteria and purified by affinity chromatography on NTA column. **Figure 3A** shows that the mutated proteins migrate at a molecular weight similar to the WT MTase+CTD. The mutated proteins were also analysed by TSA and show typical denaturation curve of folded proteins (**Fig.S3**) with T_m value ranging from 50.6 °C to 55.6 °C (**Table 1**). **Table 1** indicates that most mutants display T_m values similar to the wild-type MTase+CTD protein. In contrast, we observed that the mutant R2068A has a slightly decreased melting temperature (ΔTm = -3,7°C) suggesting that this mutation affects the protein stability. We next tested the effect of mutations in the CTD on the RNA binding properties by FP assay. We performed FP measurements using the GpppG-RNA-Cy5 and mGpppGm-RNA-Cy5 SUDV-RNA (**Table 1** and **Fig. S4**). Mutations K2043A and K2189A lead to an alteration of interactions between the substrate (GpppG-$SUDV_{12}$) and the protein and a loss of recognition for the product $^{m}GpppG_m$-$SUDV_{12}$ (Kd > 9µM). We next tested the effect of mutations on cap-dependent and cap-independent MTase activities using different RNA substrates. By incubating the WT and mutated MTase+CTD proteins with either $GpppG_m(A_m)$-$SUDV_{12}$, $^{m}GpppA(A_m)$-$SUDV_{12}$ or $^{m}GpppG_m$-$SUDV_{12}$ in presence of radiolabelled methyl donor (3H-SAM), we determined the effect of CTD mutations on cap-N7 MTase, cap-2'O MTase and internal 2'O-A MTase activities, respectively (**Fig. 3B**). We identified three categories of mutations. (i) Mutations leading to the reduction (H2112A and F2113A) or the lost (R2068A, K2118A and K2189A) of all MTase activities. Interestingly, the RNA recognition feature of R2068A, K2118A and K2189A is also strongly impaired and R2068A mutation is also associated to a reduced protein stability compared to the WT (ΔTm - 3.7°C). (ii) Mutations decreasing the 2'O MTase activities (cap-2'O MTase and internal 2'O-A MTase), but not the cap-N7 MTase activity (K2043A). This result could be correlated with the alteration of recognition observed in FP, where the interaction with the RNA substrate $^{m}GpppG_m$-$SUDV_{12}$ is more impaired in comparison to the interaction between mutated protein and substrate GpppG-$SUDV_{12}$. (iii) Mutation reducing more specifically the internal 2'O-A MTase activity (H2067A) than the cap-MTase activities. Altogether, these results demonstrate that mutations in the SUDV CTD of L protein regulate the RNA recognition and the MTase activities. We highlighted mutations uncoupling cap methylations (cap-2'O MTase and cap-N7 MTase) from internal 2'O-A methylation or N7 methylation from 2'O methylations (cap-2'O MTase and internal 2'O-A MTase) demonstrating that the CTD plays a key role in the fine tuning of the N7 and 2'O MTase activities of the L protein.

DISCUSSION

The mononegavirus L protein plays a key role in replication/transcription and RNA capping. Cryo-EM and biochemical studies have evidenced its organization in 5 main topological domains [16] embedding the RdRp, the PRNTase and the MTase activities. Nevertheless, the role of the Connector Domain (CD) and the CTD, framing the MTase domain, is still elusive. In this work, we focused on the regulation role of the CTD on the SUDV MTase activities.

The CTD domain, the less conserved domain of mononegavirus L protein, is enriched in basic amino acid suggesting their involvement in RNA recruitment. This hypothesis is also supported by the X-ray structure of hMPV MTase+CTD showing that the CTD together with the MTase form a kind of conserved RNA binding groove at immediate proximity of the MTase catalytic site (**Fig. 1B &C**). In this work, we experimentally demonstrated that the CTD domain is essential for the RNA recognition as the recognition of capped RNA is strongly impaired in absence of the CTD (**Fig.2D**). The key role of the CTD in RNA recognition was further confirmed by alanine mutagenesis of the CTD. A single mutation of conserved basic and hydrophobic residues was engineered and different mutated proteins were produced and purified. We next demonstrated by TSA that these mutations barely affect the protein folding and stability (**Table 1**). In addition, all these single mutations decrease the recognition of capped RNAs (**Table 1**), but less than observed with the MTase-ΔCTD protein. Our results indicate that some mutations strongly impair the interaction with $^{m}GpppG_{m}$-$SUDV_{12}$, more than with GpppG-$SUDV_{12}$ (K2043A and K2189A) suggesting that these mutations might regulate in a different way the MTase activity targeting the cap structure or adenosine residues within RNA sequence. Thus, all these results demonstrate the crucial role of CTD conserved residues in MTase substrate recognition. These data agree with the X-ray structure of hMPV and the cryo-EM structure of VSV full length protein indicating that the canonical cap binding domain is virtually absent in these proteins. Conversely, it was suggested that the RNA substrate might accommodate a groove enriched in basic amino acids at the interface between the MTase and the CTD domain.

We next evaluated the effect of CTD deletion and single mutations on the MTase activities. The cap-N7 MTase and the cap-2'O MTase activities were followed by incubating the different mutants with specific RNA substrates previously described [23]. In addition, we followed the methylation of adenosines within RNA sequence using a cap-1 RNA substrate ($^{m}GpppG_{m}$-$SUDV_{12}$). We observed that MTase activities are completely abrogated by the deletion of the CTD that hampers the RNA recognition. In contrast, single mutations in the CTD induce more contrasted effects on the SUDV MTase activities. We identified mutations decreasing or abrogating the overall MTase activities (F2113A, H2112A, R2068A, K2118A and K2189A), as well as one mutation uncoupling the N7 MTase activity from the 2'O MTase activities (K2043A), and one mutation affecting internal methylation rather than cap methylations (H2067A). These results are consistent with the RNA binding assay as mutations decreasing efficiently RNA recognition (MTase-ΔCTD > K2118A > R2068A) also strongly impair MTase activities. Conversely, we identified one mutant (K2189A), which still binds partially to GpppG-$SUDV_{12}$, but is barely active, suggesting that such mutation affects the RNA position in the

catalytic site. Finally, we also observed that mutants keeping partially the capacity to recognize RNA (K2043A, H2112A, F2113A, H2067A) also keep cap-N7 and/or cap-2'O MTase activities.

Altogether, these results elucidate one function of the CTD domain of SUDV, which regulates RNA substrate recognition and in turns, participate to the fine-tuning of the MTase activity. Interestingly, we recently demonstrated that the SUDV MTase induces *in vitro* internal adenosine 2'O methylation in addition to the N7 and 2'O methylation of the cap structure [23]. On the other hand, other viruses belonging to the *Mononegavirales* order such as hMPV seem to induce preferentially cap-N7 and cap-2'O methylations beside internal methylation. The different specificities between these enzymes might depend on their capacity to recognize and position the RNA substrate into the MTase catalytic site linked to the CTD recognition properties, which is barely conserved among the mononegavirus families. We do not know yet whether SUDV induces the methylation of adenosines inside its own genomic or messenger RNAs in infected cells. In addition, the role of such methylations in viral RNA sensing, mRNA translation into viral protein, or RNA encapsidation is still elusive. In contrast to other viral systems such as HIV infection, it was recently demonstrated that internal 2'O methylations of RNAs induced by a cellular MTase recruited by the virus during infection (FTSJ3) limits RNA detection by the MDA5 pathway yielding to reduction of interferon secretion [24]. The identification of mutations in the CTD of SUDV uncoupling the different MTase activities of the L protein opens the way to elucidate the function of internal methylations in filovirus infection and would participate to develop new antiviral strategies leading to an activation of the innate immunity system, especially through RIG-I/MDA5 pathway.

Supplementary Information is available in the online version of the paper.

Acknowledgements

The research was made possible by funding from the European Union Seventh Framework Programme (FP7/2007-2013) under SILVER grant agreement no. 260644 and from an MRC grant (MR/L017709/1). The authors thank Diamond Light Source for beamtime (proposal mx8423), and the staff of the MX beamlines for assistance. This work was supported by a Wellcome Trust administrative support grant 075491/Z/04.

Author Contributions. B.M., C.V., B.Co., and E.D. conceived and designed the study. B.M. C.V, B.Co., and E.D. performed experiments, using material prepared by F.D. and J-J.V. B.M., C.V, B.Co and E.D analysed data and wrote the paper. I.I. and B.Ca. participated in the review of the paper.

Methods

Cloning and expression

Codon-optimized CRVI-CTD synthetic gene (Biomers) was designed for expression in c2566 pRARE2. N-terminal truncated construct of CRVI-CTD was synthetized by cloning the sequence of interest from full-length CRVI-CTD into a pDEST14 plasmid with Gateway system for expression in c2566 bacteria. After transformation, bacteria were cultured at 30°C until a OD of 0.6, and induced overnight with 500 mM IPTG (Sigma) at 17°C. Finally, cultures were centrifuged at 8,000 xg for 10 min at 4°C using a Sorval Lynx 6000 centrifuge (Thermo), and pellets were conserved at -80°C until purification.

Purification

Pellets were thawed on ice, and lysed in a 10 times final culture OD volume of optimized lysis buffer (50 mM Tris pH 8, 150 mM NaCl, 5% glycerol, 30 mM imidazole, 1 mM PMSF, 100 µg/mL lysozyme, 1 µg/mL DNase, 0.1% Triton X100) supplemented with BugBuster (Merck Millipore) for MTase+CTD. After clarification (18,000 xg, 30 min, 4 °C), the lysate was incubated with CoNTA resin (Thermo; 0.5 mL/L culture) for 1 h at 4 °C, with gentle shaking. The beads were transferred to a 25-mL column and washed with 2 x 20 mL of buffer W1 (50 mM Tris pH 8, 1 M NaCl, 5% glycerol, 30 mM imidazole) and 2 x 10 mL of buffer W2 (50 mM Tris pH 8, 150 mM NaCl, 5% glycerol). Protein was eluted in buffer E (50 mM Tris pH 8, 150 mM NaCl, 5% glycerol) supplemented with 150 mM imidazole in the case of MTase-ΔCTD and 1 M arginine for CRVI+CTD. Imidazole was removed by gel filtration using a Superdex S75 16/60 (GE Healthcare) for MTase-ΔCTD. Finally, proteins were concentrated on Amicon Ultra (EMD Millipore), and conserved in 50% glycerol at -20°C.

Synthesis of RNA substrates

RNA sequences were chemically synthesized on solid support using an ABI 394 oligonucleotidessynthesizer. After RNA elongation with 2'-*O*-pivaloyloxymethyl phosphoramidite ribonucleotides,[10-11] and 2'-*O*-methyl phosphoramidite ribonucleotides (Chemgenes, USA), the 5'-hydroxyl group was phosphorylated and the resulting *H*-phosphonate derivative[12] was oxidized and activated into a phosphoroimidazolidate derivative to react with guanosine diphosphate (**G**pppRNA)[13-14]. After deprotection and release from the solid support, GpppRNAswere purified by IEX-HPLC and validated to be >95 % pure by MALDI-TOF spectrometry. *N*7-methylation of the purified **G**pppRNA was performed enzymatically using *N*7-hMTase[13-14].

MTase activity assays

To evaluate methyltransferase activities, a radioactive test was set up by mixing 4 µM of protein with 1 µM of purified and validated synthetic RNAs (see all RNAs in **Table S1**), 10 µM of SAM and 0.5

µM of ^{3}H-SAM (Perkin Elmer) in an optimized MTase assay buffer (50 mM Tris-HCl variable pH, 10 mM arginine). Reactions were stopped by a 10-fold dilution in water after 3 h at 30°C. Samples were transferred to DEAE filtermats (Perkin Elmer) using a Filtermat Harvester (Packard Instruments). Methyl transfer was then evaluated as described before [29]. Briefly, the RNA-retaining mats were washed twice with 10 mM ammonium formate pH 8, twice with water and once with ethanol. They were then soaked with liquid scintillation fluid, allowing the measurement of ^{3}H-methyl transfer to the RNA substrates using a Wallac MicroBeta TriLux Liquid Scintillation Counter13.

Fluorescence polarization (PF)

Using T4 RNA ligase 1 (20 units; New England Biolabs), cyanine 5-cytidine-5-phosphate-3-(6-aminohexyl)phosphate (12.5 µM, Jena Bioscience) was ligated to the 3'ends of the RNA substrates (10 µM) in T4 RNA ligase 1 buffer (NEB), 1 mM ATP (16 °C, overnight). Ligase was removed by RNA precipitation in 3 M sodium acetate supplemented with 1 µg/µL of glycogen (Thermo Scientific). The fluorescent RNA was incubated (5 min at room temperature) with increasing concentrations of CRVI+CTD, in 20 mM Tris pH 8 or pH 8,5 150 mM NaCl, 5% glycerol. Fluorescence polarization (FP) measurements were performed in a microplate reader (PHERAstar FS; BMG Labtech) with an optical module equipped with polarizers and using excitation and emission wavelengths of 590 and 675 nm, respectively. Dissociation constants (K_d) were determined using Hill slope curve fitting (Prism).

Thermal shift assay (TSA)

A mix of 4 µM protein, buffered by resuspension buffer (50 mM Tris pH 8, 150 mM NaCl, 5% glycerol) with 0,02% (v/v) SYPRO Orange dye (Thermo Scientific) was made up to a total volume of 20 µL. Samples were placed in a semi skirted 96 well PCR plate (BioRad), sealed and heated in an Mx3005p qPCR machine (BioRad) from 25 to 95°C at a rate of 1 °C.min^{-1}. Fluorescence changes were monitored with excitation and emission wavelengths at 492 and 610 nm respectively.

References

[1] J. M. Shultz, Z. Espinel, M. Espinola, and A. Rechkemmer, "Distinguishing epidemiological features of the 2013-2016 West Africa Ebola virus disease outbreak.," *Disaster Heal.*, vol. 3, no. 3, pp. 78–88, 2016.

[2] "Ebola | Ebola situation reports: Democratic Republic of the Congo," *WHO*, 2019.

[3] A. M. Henao-Restrepo *et al.*, "Efficacy and effectiveness of an rVSV-vectored vaccine expressing Ebola surface glycoprotein: interim results from the Guinea ring vaccination cluster-randomised trial," *Lancet*, vol. 386, no. 9996, pp. 857–866, Aug. 2015.

[4] C. R. Wells *et al.*, "Ebola vaccination in the Democratic Republic of the Congo," *Proc. Natl. Acad. Sci.*, vol. 116, no. 20, p. 201817329, Apr. 2019.

[5] E. O. Saphire, S. L. Schendel, B. M. Gunn, J. C. Milligan, and G. Alter, "Antibody-mediated protection against Ebola virus.," *Nat. Immunol.*, vol. 19, no. 11, pp. 1169–1178, Nov. 2018.

[6] J. H. Kuhn *et al.*, "Proposal for a revised taxonomy of the family Filoviridae: classification, names of taxa and viruses, and virus abbreviations.," *Arch. Virol.*, vol. 155, no. 12, pp. 2083–103, Dec. 2010.

[7] A. Sanchez, M. P. Kiley, B. P. Holloway, and D. D. Auperin, "Sequence analysis of the Ebola virus genome: organization, genetic elements, and comparison with the genome of Marburg virus," *Virus Res.*, vol. 29, no. 3, pp. 215–240, Sep. 1993.

[8] L. H. Elliott, A. Sanchez, B. P. Holloway, M. P. Kiley, and J. B. McCormick, "Ebola protein analyses for the determination of genetic organization.," *Arch. Virol.*, vol. 133, no. 3–4, pp. 423–36, 1993.

[9] E. Mühlberger, S. Trommer, C. Funke, V. Volchkov, H. D. Klenk, and S. Becker, "Termini of all mRNA species of Marburg virus: sequence and secondary structure.," *Virology*, vol. 223, no. 2, pp. 376–80, Sep. 1996.

[10] M. Weik, J. Modrof, H.-D. Klenk, S. Becker, and E. Mühlberger, "Ebola virus VP30-mediated transcription is regulated by RNA secondary structure formation.," *J. Virol.*, vol. 76, no. 17, pp. 8532–9, Sep. 2002.

[11] S. P. J. Whelan, J. N. Barr, and G. W. Wertz, "Transcription and replication of nonsegmented negative-strand RNA viruses.," *Curr. Top. Microbiol. Immunol.*, vol. 283, pp. 61–119, 2004.

[12] P. C. Jordan *et al.*, "Initiation, extension, and termination of RNA synthesis by a paramyxovirus polymerase," *PLOS Pathog.*, vol. 14, no. 2, p. e1006889, Feb. 2018.

[13] G. Tekes, A. A. Rahmeh, and S. P. J. Whelan, "A freeze frame view of vesicular stomatitis virus transcription defines a minimal length of RNA for 5' processing.," *PLoS Pathog.*, vol. 7, no. 6, p. e1002073, Jun. 2011.

[14] A. J. Shatkin, "Capping of eucaryotic mRNAs.," *Cell*, vol. 9, no. 4 PT 2, pp. 645–53, Dec. 1976.

[15] E. Decroly, F. Ferron, J. Lescar, and B. Canard, "Conventional and unconventional mechanisms for capping viral mRNA.," *Nat. Rev. Microbiol.*, vol. 10, no. 1, pp. 51–65, Dec. 2011.

[16] B. Liang *et al.*, "Structure of the L Protein of Vesicular Stomatitis Virus from Electron Cryomicroscopy," *Cell*, vol. 162, no. 2, pp. 314–327, Jul. 2015.

[17] A. A. Rahmeh *et al.*, "Molecular architecture of the vesicular stomatitis virus RNA polymerase.," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 107, no. 46, pp. 20075–80, Nov. 2010.

[18] M.-P. Egloff, D. Benarroch, B. Selisko, J.-L. Romette, and B. Canard, "An RNA cap (nucleoside-2'-O-)-methyltransferase in the flavivirus RNA polymerase NS5: crystal structure and functional characterization.," *EMBO J.*, vol. 21, no. 11, pp. 2757–68, Jun. 2002.

[19] D. Ray *et al.*, "West Nile virus 5'-cap structure is formed by sequential guanine N-7 and ribose 2'-O methylations by nonstructural protein 5.," *J. Virol.*, vol. 80, no. 17, pp. 8362–70, Sep. 2006.

[20] Y. Zhou *et al.*, "Structure and Function of Flavivirus NS5 Methyltransferase," *J. Virol.*, vol. 81, no. 8, pp. 3891–3903, Apr. 2007.

[21] A. A. Rahmeh, J. Li, P. J. Kranzusch, and S. P. J. Whelan, "Ribose 2'-O Methylation of the Vesicular Stomatitis Virus mRNA Cap Precedes and Facilitates Subsequent Guanine-N-7 Methylation by the Large Polymerase Protein," *J. Virol.*, vol. 83, no. 21, pp. 11043–11050,

Nov. 2009.
[22] G. C. Paesen *et al.*, “X-ray structure and activities of an essential Mononegavirales L-protein domain.,” *Nat. Commun.*, vol. 6, p. 8749, Nov. 2015.
[23] B. Martin *et al.*, “The methyltransferase domain of the Sudan ebolavirus L protein specifically targets internal adenosines of RNA substrates, in addition to the cap structure,” *Nucleic Acids Res.*, vol. 46, no. 15, pp. 7902–7912, Sep. 2018.
[24] M. Ringeard, V. Marchand, E. Decroly, Y. Motorin, and Y. Bennasser, “FTSJ3 is an RNA 2′-O-methyltransferase recruited by HIV to avoid innate immune sensing,” *Nature*, vol. 565, no. 7740, pp. 500–504, Jan. 2019.
[25] R. Züst *et al.*, “Ribose 2’-O-methylation provides a molecular signature for the distinction of self and non-self mRNA dependent on the RNA sensor Mda5.,” *Nat. Immunol.*, vol. 12, no. 2, pp. 137–43, Feb. 2011.
[26] K. Karikó, M. Buckstein, H. Ni, and D. Weissman, “Suppression of RNA recognition by Toll-like receptors: the impact of nucleoside modification and the evolutionary origin of RNA.,” *Immunity*, vol. 23, no. 2, pp. 165–75, Aug. 2005.
[27] C. Schuberth-Wagner *et al.*, “A Conserved Histidine in the RNA Sensor RIG-I Controls Immune Tolerance to N1-2’O-Methylated Self RNA.,” *Immunity*, vol. 43, no. 1, pp. 41–51, Jul. 2015.
[28] S. C. Devarkar *et al.*, “Structural basis for m7G recognition and 2′-O-methyl discrimination in capped RNAs by the innate immune receptor RIG-I,” *Proc. Natl. Acad. Sci.*, vol. 113, no. 3, pp. 596–601, Jan. 2016.
[29] M. Bollati *et al.*, “Recognition of RNA cap in the Wesselsbron virus NS5 methyltransferase domain: implications for RNA-capping mechanisms in Flavivirus.,” *J. Mol. Biol.*, vol. 385, no. 1, pp. 140–52, Jan. 2009.

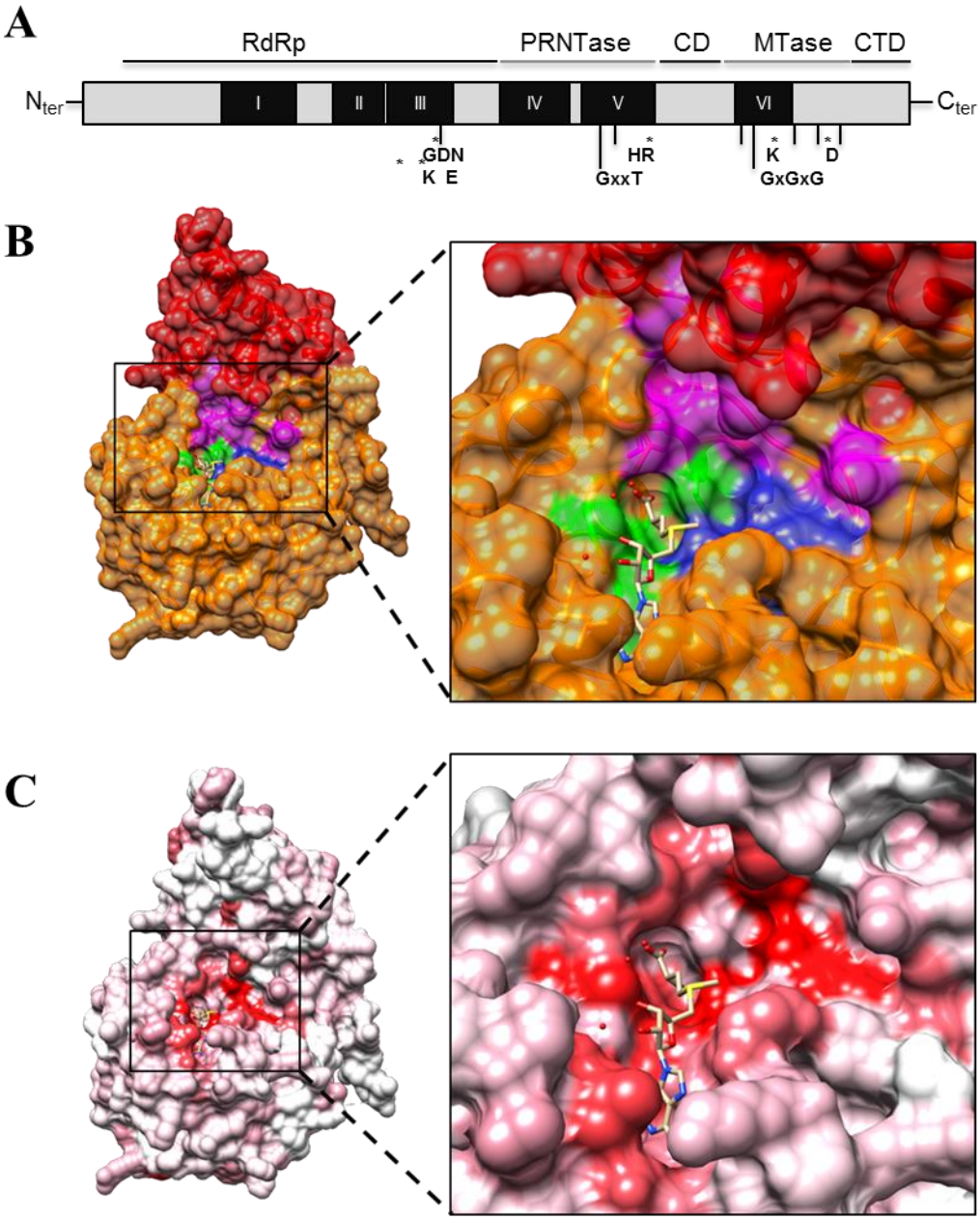


Figure 1 – Bioinformatic analysis of the C-terminal domain (MTase+CTD) of mononegavirus L protein based on hMPV structure.
(A) Sequence organization of the mononegavirus L protein revealed six conserved regions (CRI to CRVI, black boxes) that contain motifs responsible for the different activities of the L (motifs mapped with asterisks). The SUDV MTase domain was defined as the fragment encompassing amino acids 1713–2046 using the alignments with the VSV L protein. The SAM-binding site motifs (GxGxG) and the 2'*O* catalytic tetrad K-D-K-E have also been identified (asterisks). The CTD of SUDV L follows the MTase domain (amino acids 2047–2211). **(B)** The X-ray structure of hMPV methyltransferase (MTase, orange) and C-terminal (CTD, red) domains at a resolution of 2.2 Å is presented (Paesen *et al.*, 2015). The residues involved in the catalytic tetrad, SAM-binding site and RNA-binding groove are mapped respectively in blue, green and purple. (**C**) Conservation model of hMPV MTase+CTD domains of mononegavirus L based on proteins alignment presented in **Fig. S1**. Colors indicate the level of residue conservation (white: not conserved, pink: barely conserved, red: highly conserved).

1

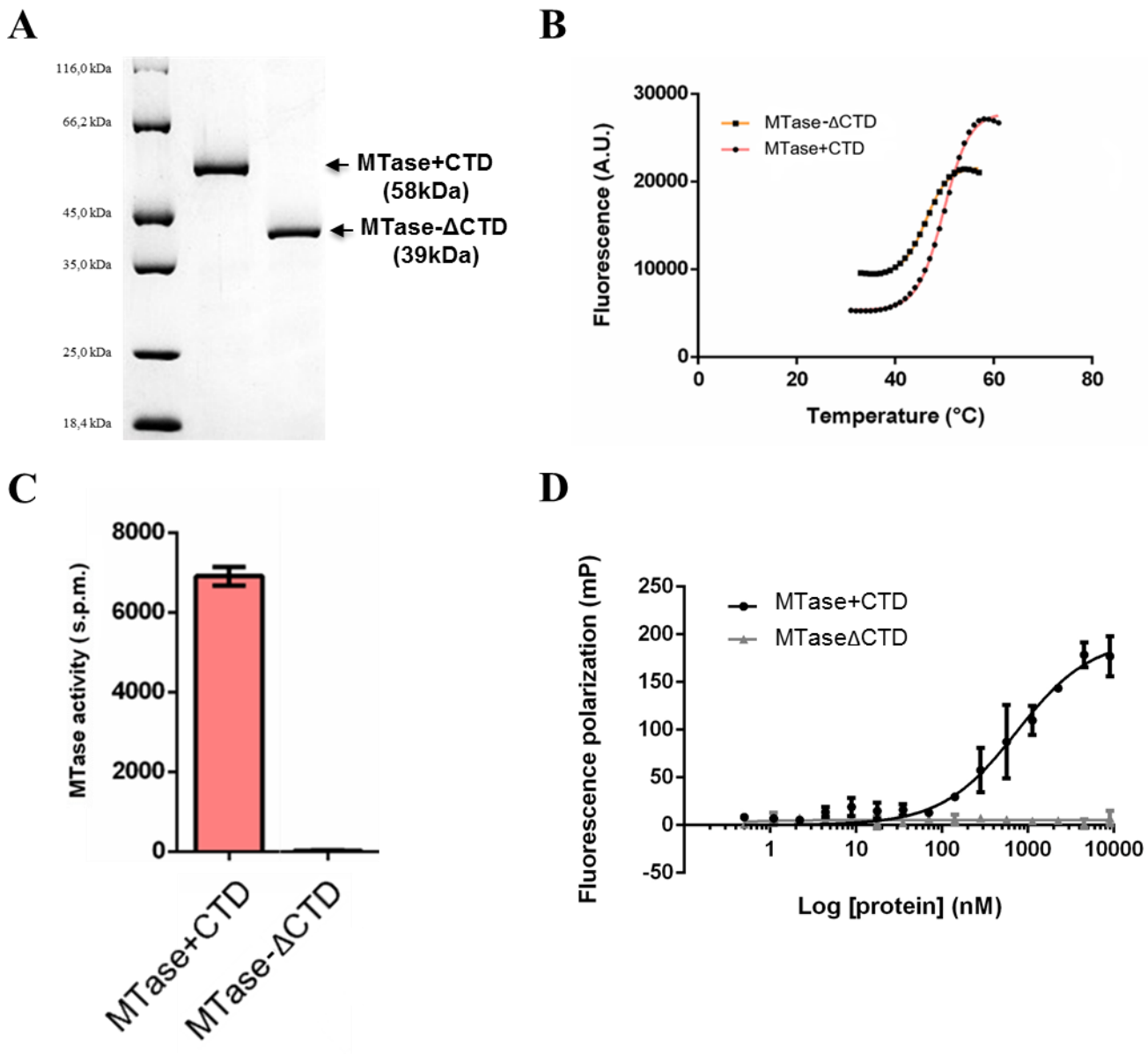
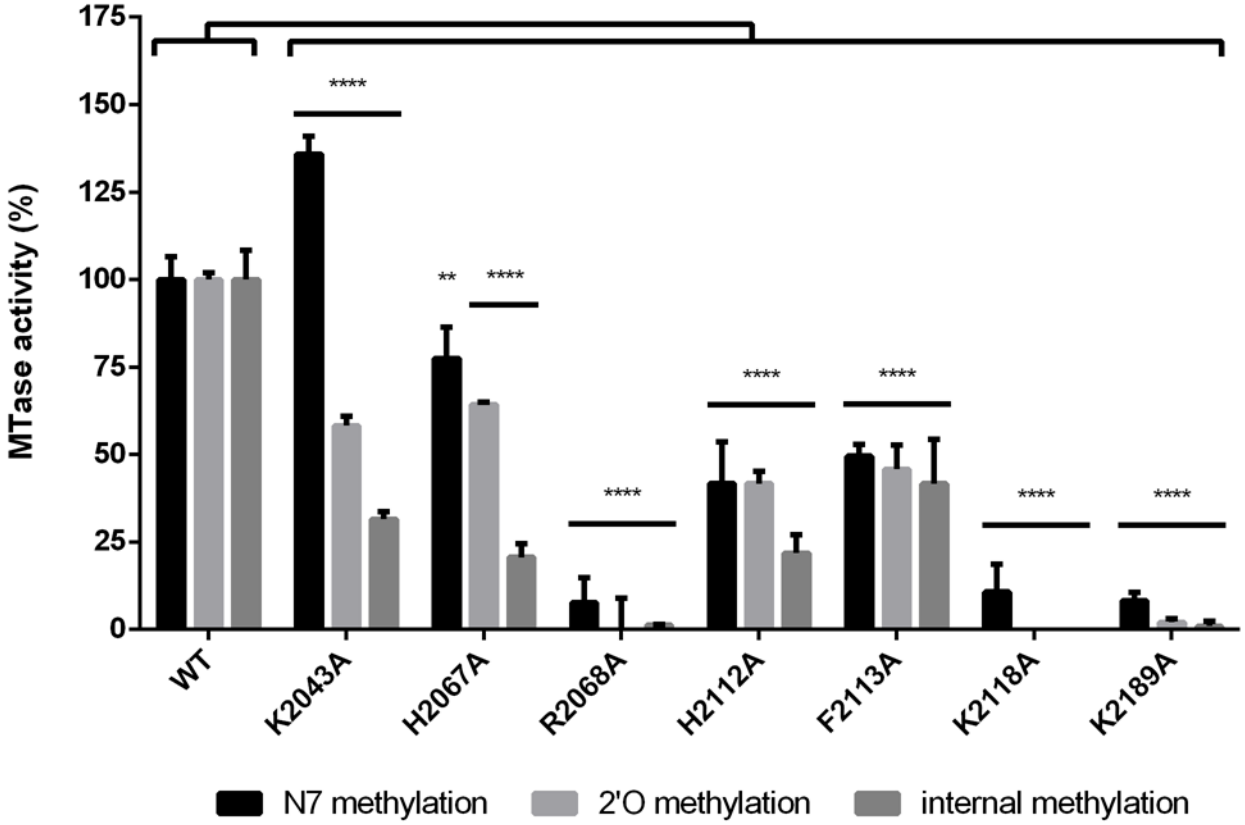


Figure 2 – **Production and purification of SUDV MTase+CTD and MTase-ΔCTD and effect of CTD on MTase activity and RNA binding.**

(**A**) MTase+CTD (58 kDa) and MTase-ΔCTD (39 kDa) of *Sudan ebolavirus* (SUDV), purified by two-steps chromatography, were separated by SDS-PAGE analysis before Coomassie blue staining. (**B**) Thermostability assay of SUDV MTase+CTD and MTase-ΔCTD constructs (n=1). Melting temperature indexes (T_m) have been calculated following a Boltzmann sigmoidal regression and evaluated at both 54,6°C and 46°C, respectively. Data represent raw data. **(C)** MTase activity evaluation of *Sudan ebolavirus* (SUDV) MTase-ΔCTD (orange) and MTase+CTD (red) incubated with a 13-mers capped RNA substrate mimicking the capped 5' extremity of viral mRNA (GpppG-$SUDV_{12}$). MTase activity was determined by filter binding assay (n=6). Data represent mean ± standard deviation. **(D)** RNA-binding assay by fluorescence polarization of a 13-mers capped SUDV RNA substrate labelled in 3' by pCp-Cy5 on SUDV methyltransferase domain (MTase-ΔCTD, grey) and SUDV MTase associated with the C-terminal domain (MTase+CTD, black) (n=3). A dissociation coefficient (K_d) has been calculated by one site specific binding regression with Hill slope. The MTase+CTD domain affinity for GpppG-$SUDV_{12}$ is estimated to be 740 ± 1,1 nM. Data represent mean ± standard deviation.

1



<u>Figure 3</u>- Production and purification of mutated MTase+CTD and effect of a single mutation of conserved residues in CTD on MTase activity and RNA binding.

(A) Mtase+CTD and mutated MTase+CTD (58 kDa) of *Sudan ebolavirus* (SUDV), purified by chromatography, were separated by SDS-PAGE analysis before Coomassie blue staining. (B) MTase activity evaluation of *Sudan ebolavirus* (SUDV) wild-type (WT) and mutated SUDV MTase+CTD incubated with either GpppGm(Am)-$SUDV_{12}$, $^{m}GpppA(A_m)$-$SUDV_{12}$ or $^{m}GpppG_m$-$SUDV_{12}$ in presence of radiolabelled methyl donor (3H-SAM), to get access to cap-N7 MTase, cap-2'O MTase and internal 2'O-A MTase activities, respectively . MTase activity was determined by filter binding assay.

1

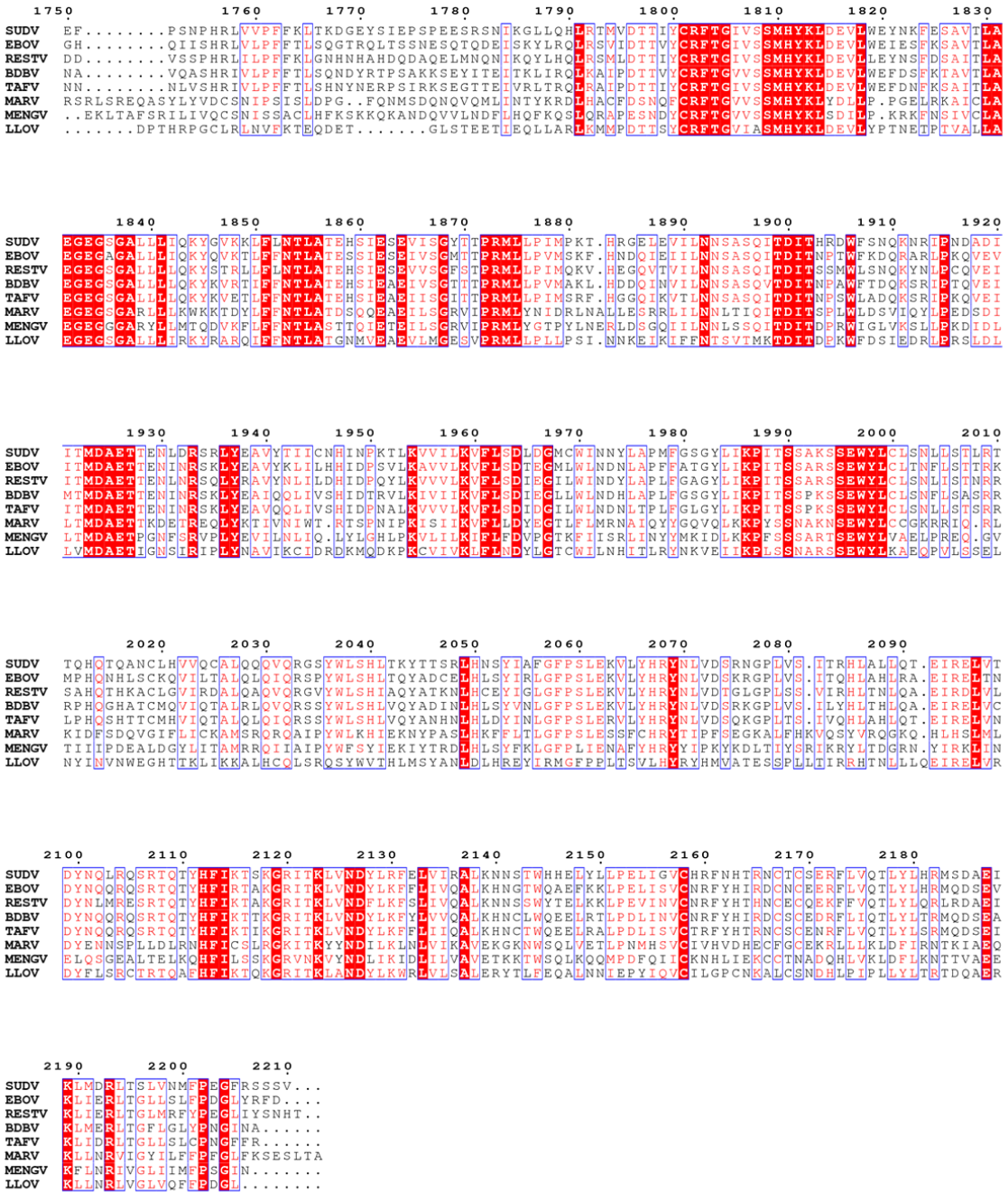
	Thermal Shift Assay			Fluorescence polarization	
				GpppG-SUDV$_{12}$	mGpppGm-SUDV$_{12}$
	Mean	SD	ΔTm	Kd (µM)	Kd (µM)
WT	54.3	0.13	-	0.740	0.570
K2043A	55.7	0.03	1.4	2	> 9
H2067A	54.7	0.58	0.4	5	4.4
R2068A	50.6	0.01	-3.7	> 9	> 9
H2112A	54.0	0.39	-0.3	4.5	5.4
F2113A	53.5	0.33	-0.8	6.6	3.1
K2118A	53.7	0.24	-0.6	> 9	> 9
K2189A	55.0	0.83	0.7	1.9	8.6

Table 1- Effect of a single mutation of conserved residues in CTD on MTase activity and RNA binding.

Thermostability assay (TSA) of SUDV MTase+CTD and SUDV MTase+CTD mutated proteins. Melting temperature indexes (T_m) have been calculated following a Boltzmann sigmoidal regression.

RNA-binding assay by fluorescence polarization of a 13-mers SUDV-specific capped RNA (GpppG-SUDV$_{12}$ and ^{m}GpppG$_m$-SUDV$_{12}$) labelled at the 3'end by pCp-Cy5 on SUDV MTase+CTD and mutated MTase+CTD proteins. A dissociation coefficient (K_d) has been calculated by one site specific binding regression.

1



Supplementary figure 1 - Alignment of MTase+CTD domain of filovirus L protein

Sequences of filovirus methyltransferase (MTase) and C-terminal (CTD) domains have been identified from L protein sequences picked up on NCBI Protein data bank: EBOV (Zaire ebolavirus, AAG40171.1), SUDV (Sudan ebolavirus, YP_138527.1), TAFV (Tai Forest ebolavirus, ALT19766.1), BDBV (Bundibugyo ebolavirus, AKB09568.1), RESTV (Reston ebolavirus, APA16576.1), MARV (Marburg virus, CAA82542.1), LLOV (Lloviu cuevavirus, YP_004928143.1) and MENGV (Mengla dianlovirus, AZL87829.1). Alignment has been generated with Seaview, then treated with ESPript. Orange and red lines model MTase and CTD domains, respectively. The 2'O MTase-specific catalytic tetrad K-D-K-E (Ferron *et al.*, 2002) and SAM-binding motif GxGxG have been indicated by blue and green respectively. Strictly conserved amino acids are indicated by *, and conserved positions with either arginine or lysine residues are indicated by X.

1

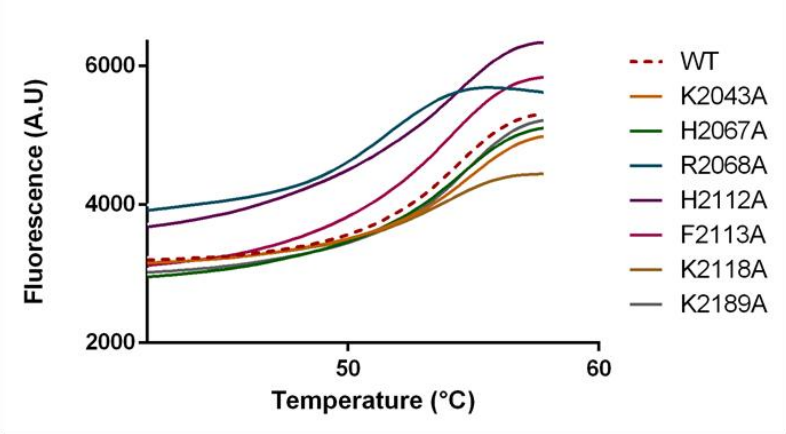
2

Order	Family	Genus	Species	Domain	Total residue number	Basic residue number	% basic residues	pI	ΔpI
Mononegavirales	*Filoviridae*	*Ebolavirus*	EBOV (AAG40171.1)	MTase	363	37	10	7,73	1,46
				CTD	171	24	14	9,19	
				MTase-CTD	534	61	11	8,81	
			SUDV (YP_138527.1)	MTase	346	35	10	6,49	3,22
				CTD	172	23	13	9,71	
				MTase-CTD	518	58	11	8,68	
			TAFV (ALT19766.1)	MTase	354	35	10	8,50	0,66
				CTD	169	20	12	9,16	
				MTase-CTD	523	55	11	8,93	
			BDBV (AKB09568.1)	MTase	354	39	11	7,66	1,12
				CTD	169	20	12	8,78	
				MTase-CTD	523	59	11	8,53	
			RESTV (APA16576.1)	MTase	351	33	9	7,20	1,94
				CTD	173	22	13	9,14	
				MTase-CTD	524	55	10	8,65	
		Marburgvirus	MARV (CAA82542.1)	MTase	445	53	12	8,48	0,88
				CTD	176	24	14	9,36	
				MTase-CTD	621	77	12	8,95	
		Cuevavirus	LLOV (YP_004928143.1)	MTase	339	43	13	8,55	0,86
				CTD	167	21	13	9,41	
				MTase-CTD	506	64	13	8,99	
	Paramyxoviridae	*Respirovirus*	SeV (AAB06283.1)	MTase	343	37	11	5,69	3,85
				CTD	205	31	15	9,54	
				MTase-CTD	548	68	12	7,26	
		Morbillivirus	MeV (BAB60955.1)	MTase	338	34	10	7,78	2,16
				CTD	165	27	16	9,94	
				MTase-CTD	503	61	12	9,32	
		Rubulavirus	MuV (BAA01432.1)	MTase	340	19	6	4,95	0,49
				CTD	227	26	11	5,44	
				MTase-CTD	567	45	8	5,13	
		Henipavirus	HeV (O89344.3)	MTase	342	34	10	6,02	3,61
				CTD	165	27	16	9,63	
				MTase-CTD	503	61	12	8,62	
	Pneumoviridae	*Metapneumovirus*	hMPV (Q91L20)	MTase	285	34	12	8,57	0,92
				CTD	124	19	15	9,49	
				MTase-CTD	409	53	13	8,97	
		Orthopneumovirus	hRSV (AAX23996.1)	MTase	323	36	11	8,94	0,48
				CTD	127	16	13	9,42	
				MTase-CTD	450	52	12	9,14	
	Rhabdoviridae	*Vesiculovirus*	VSV (5A22_A)	MTase	317	30	9	5,81	3,54
				CTD	197	24	12	9,35	
				MTase-CTD	514	54	11	7,58	
		Lyssavirus	RABV (ABZ81226.1)	MTase	285	27	9	6,98	1,36
				CTD	217	24	11	8,34	
				MTase-CTD	502	51	10	8,02	
	Bornaviridae	*Bornavirus*	BDV (P52639.2)	MTase	184	18	10	6,21	0,12
				CTD	88	9	10	6,33	
				MTase-CTD	272	27	10	6,25	

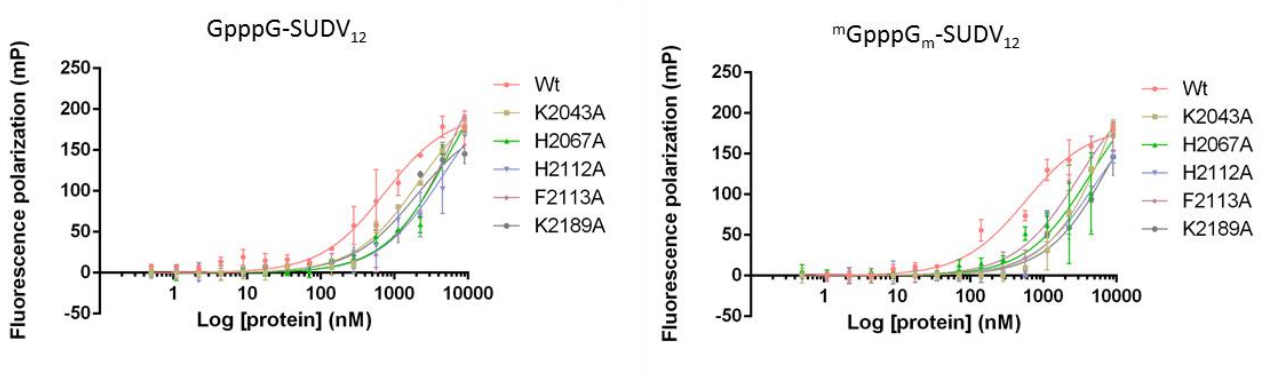
<u>Supplementary figure 2</u>– Analysis of the residue content of mononegavirus MTase and CTD domains

Based on alignment of different Mononegavirales, residue content of the methyltransferase (MTase) and C-terminal (CTD) domains of mononegavirus L proteins has been analysed. For each virus, MTase, CTD and MTase+CTD total number of residues, number of basic residues and the percentage of basic residues have been determined. Isoelectric points have also been estimated using ProtParam. Differences between MTase and CTD pIs have been calculated and modeled by a blue gauge normalized to the greater difference.

1



Supplementary figure 3 - Thermal Shift Assay
Thermostability assay of SUDV MTase+CTD and SUDV MTase+CTD mutated proteins. Melting temperature indexes (T_m) have been calculated following a Boltzmann sigmoidal regression.



Supplementary figure 4 - RNA-binding assay by fluorescence polarization of SUDV-specific capped RNA of 13 mers (GpppG-$SUDV_{12}$ and $^{m}GpppG_{m}$-$SUDV_{12}$) labelled at the 3'end by pCp-Cy5 on SUDV MTase+CTD and mutated MTase+CTD proteins. A dissociation coefficient (K_d) has been calculated by one site specific binding regression.

Table S1–List of synthetic RNAs

RNA	Sequence	Origin
GpppG-SUDV$_{12}$	GpppGAUGAAGAUUAAG	chemical synthesis
GpppG$_m$(A$_m$)-SUDV$_{12}$	GpppGA$_m$UGA$_m$A$_m$GA$_m$UUA$_m$A$_m$G	chemical synthesis
^{m}GpppG$_m$-SUDV$_{12}$	^{m}GpppG$_m$AUGAAGAUUAAG	chemical synthesis
^{m}GpppA(A$_m$)-SUDV$_{12}$	^{m}GpppAUGA$_m$UGA$_m$A$_m$GA$_m$UUA$_m$	chemical synthesis

1