<table>
<thead>
<tr>
<th><strong>Title</strong></th>
<th>Molecular basis for specific viral RNA recognition and 2'-O-ribose methylation by the dengue virus nonstructural protein 5 (NS5)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Author(s)</strong></td>
<td>Zhao, Yongqian; Soh, Tingjin Sherryl; Lim, Siew Pheng; Chung, Ka Yan; Swaminathan, Kunchithapadam; Vasudevan, Subhash G.; Shi, Pei-Yong; Lescar, Julien; Luo, Dahai</td>
</tr>
<tr>
<td><strong>Date</strong></td>
<td>2015</td>
</tr>
<tr>
<td><strong>URL</strong></td>
<td><a href="http://hdl.handle.net/10220/40374">http://hdl.handle.net/10220/40374</a></td>
</tr>
<tr>
<td><strong>Rights</strong></td>
<td>© 2015 The Author(s) (Published by National Academy of Sciences). This is the author created version of a work that has been peer reviewed and accepted for publication by Proceedings of the National Academy of Sciences of the United States of America, The Author(s) (Published by National Academy of Sciences). It incorporates referee’s comments but changes resulting from the publishing process, such as copyediting, structural formatting, may not be reflected in this document. The published version is available at: [<a href="http://dx.doi.org/10.1073/pnas.1514978112">http://dx.doi.org/10.1073/pnas.1514978112</a>].</td>
</tr>
</tbody>
</table>
Molecular basis for specific viral RNA recognition and 2’-O ribose methylation by the
dengue virus NS5 protein

Yongqian Zhao\textsuperscript{1,2,\#}, Tingjin Sherryl Soh\textsuperscript{3,4,\#}, Siew Pheng Lim\textsuperscript{4}, Ka Yan Chung\textsuperscript{3,4},
Kunchithapadam Swaminathan\textsuperscript{2}, Subhash G. Vasudevan\textsuperscript{1,2}, Pei-Yong Shi\textsuperscript{4,*}, Julien Lescar\textsuperscript{3,5,6,*}, & Dahai Luo\textsuperscript{6,*}

\textsuperscript{1}Program in Emerging Infectious Diseases, DUKE-NUS Graduate Medical School, 8 College
Road, Singapore 169857
\textsuperscript{2}NUS Graduate School for Integrative Sciences and Engineering, National University of
Singapore, 28 Medical Drive, Singapore 117456
\textsuperscript{3}School of Biological Sciences, Nanyang Technological University. 60 Nanyang Drive,
Singapore 637551
\textsuperscript{4}Novartis Institute for Tropical Diseases, 10 Biopolis Road, Singapore 138670
\textsuperscript{5}UPMC UMRS CR7 - CNRS ERL 8255-INSERM U1135 Centre d’Immunologie et des
Maladies Infectieuses. Centre Hospitalier Universitaire Pitié-Salpêtrière, Faculté de Médecine
Pierre et Marie Curie, Paris, France.
\textsuperscript{6}NTU Institute of Structural Biology, Nanyang Technological University, EMB 06, 59
Nanyang Drive, Singapore 636921
\textsuperscript{7}Lee Kong Chian School of Medicine, Nanyang Technological University, EMB 03, 59
Nanyang Drive, Singapore 636921

Running title: Structure of the Dengue Virus NS5 protein bound to Cap0 viral RNA

Keywords: Dengue virus, Non-structural protein 5 methyltransferase-polymerase, 2’O ribose
methyl-transferase; cap 0 RNA, innate immunity evasion

\# These two authors contributed equally to the work;

& Current affiliation: Department of Biochemistry and Molecular Biology, University of
Texas Medical Branch, Galveston, TX 77555-1055, United States. Email: peshi@utmb.edu

* To whom correspondence may be addressed. E-mails: pei_yong.shi@novartis.com;
 julien@ntu.edu.sg; luodahai@ntu.edu.sg
Abstract

Dengue virus (DENV) causes several hundred million human infections and more than 20,000 deaths annually. Neither an efficacious vaccine conferring immunity against the four circulating serotypes nor specific drugs are currently available to treat this emerging global disease. Capping of the Dengue virus RNA genome is an essential structural modification that protects the RNA from degradation by 5’ exo-ribonucleases, ensures efficient expression of viral proteins and allows escape from the host innate immune response. The large flavivirus NS5 protein (105 kDa) has RNA methyl-transferase activities at its N terminal region, which is responsible for capping the virus RNA genome. The methyl transfer reactions are thought to occur sequentially using the strictly conserved flavivirus 5’ RNA sequence as substrate (G_{ppp}AG-RNA), leading to the formation of the 5’ RNA cap: G_{0ppp}AG-RNA $\rightarrow$ m$^7$G_{0ppp}AG-RNA (named “cap-0”) $\rightarrow$ m$^7$G_{0ppp}A_{m2'-O}G-RNA (named “cap-1”). To elucidate how viral RNA is specifically recognized and methylated, we determined the crystal structure of a ternary complex between the full-length NS5 protein from dengue virus, an octameric cap-0 viral RNA substrate bearing the authentic DENV genomic sequence (5'-m$^7$G_{0ppp}A$_1$G$_2$U$_3$U$_4$G$_5$U$_6$U$_7$-3’) and S-adenosyl-L-homocysteine (SAH), the by-product of the methylation reaction. The structure provides for the first time a molecular basis for specific adenosine 2’O methylation, rationalizes mutagenesis studies targeting the K61-D146-K180-E216 enzymatic tetrad as well as residues lining the RNA binding groove and offers novel mechanistic and evolutionary insights into cap-1 formation by NS5, which underlies innate immunity evasion by flaviviruses.

Significance Statement

Dengue is the most prevalent mosquito-borne viral disease, endemic in more than 100 tropical and subtropical countries. NS5, the largest viral protein, is a key replication enzyme with both methyltransferase and RNA polymerase activities. We present the first crystal structure of the full-length NS5 protein from DENV bound to the authentic 5’ end viral RNA fragment. This structure captures the viral enzyme in the act of transferring a methyl group to 2’O ribose of the first nucleotide of the viral genome, providing an atomic level understanding of specific 2’O methylation and cap formation by the flavivirus methyltransferase. The structure also suggests an evolutionary origin for the methyltransferase domain of NS5 and strategies for designing novel antiviral inhibitors.
Introduction

Several members of the Flavivirus genus from the Flaviviridae family are major human pathogens, such as the four Dengue virus serotypes (DENV1-4), West Nile virus (WNV), Japanese encephalitis virus (JEV) and Yellow fever virus (YFV). Recent large-scale DENV vaccine trials using a tetravalent formulation and three booster injections have shown only limited cross-protection against the four DENV serotypes and no specific antiviral drug has reached the market so far (1-3). The flavivirus genome consists in a (+) sense single-stranded RNA of ~11 kb with a type 1 cap structure, followed by the strictly conserved dinucleotide sequence “AG”: 5’-m7GpppAmp2’ O-G-3’ (4, 5). Addition of the cap moiety to the 5’ end of the viral genome is crucial for viral replication because this structure ensures efficient production of viral polyproteins by the host translation machinery, protection against degradation by 5’-3’ exo-ribo nucleases and also because it conceals the triphosphate end from recognition by host cell innate immune sensors, preventing activation of a RIG-I-mediated interferon response and production of pro-inflammatory cytokines (6). Following (+) strand RNA synthesis by the C-terminal RNA dependent RNA polymerase (RdRp) domain of NS5, cap formation in flaviviruses results from several sequential enzymatic reactions carried out by (i) the RNA triphosphatase activity of the NS3 protease-helicase that hydrolyzes the γ-phosphate group of the viral 5’ untranslated region (UTR) yielding a diphosphate RNA, (ii) a guanylyl-transferase activity proposed to reside in the MTase domain of NS5, that transfers a GMP molecule to the 5’-diphosphate RNA, (iii) NS5 mediated sequential N7 and 2’O methylations according to the following scheme: GpppAG-RNA→m7GpppAG-RNA (“cap 0”) →m7GpppAm2’ O-G-RNA (“cap 1”) (5, 7-9).

During flavivirus RNA replication, 5’-guanosine N7 methylation was shown to be essential for the translation of virus polyprotein (10) while 2’O methylation on the penultimate A nucleotide conceals the viral genome from host immune sensors (6). More recently, 2’O methylation at internal adenosine (but not at G, C or U positions) by the flavivirus NS5 protein was demonstrated. The functional consequence of methylation at internal adenosines was an attenuation of viral RNA translation and replication (9). In vitro, the MTase domain of NS5 catalyzes these two enzymatic reactions with distinct requirements of RNA substrates and buffers: while 5’-guanosine N7 methyl transfer is optimal on a 211 nt segment of the 5’ UTR at pH 6 and is inhibited by MgCl2, adenosine ribose 2’O methylation only requires a short RNA with “AG” as the first two RNA nucleotides and is maximum at pH 9-10 in the presence of Mg2+ ions. Thus, NS5 plays a crucial role both in virus replication
and evasion of the host innate immune response (6, 10-12) and constitutes an attractive therapeutic target for anti-viral drug and vaccine development (2, 13).

Several crystal structures of flavivirus methyl-transferases (MTases) have been reported either as free enzymes, bound to GTP (11), to the broad antiviral nucleoside analogue ribavirin (14), to short cap analogues (15, 16) and to a capped-RNA octamer (17). Collectively, these structures uncovered a GTP binding site, the S-adenosyl methionine (SAM) methyl-donor binding pocket and a basic cleft at the protein surface that was proposed to accommodate the incoming RNA substrate. However, in the absence of a viral RNA in a catalytically meaningful position, the mechanism accounting for specific viral RNA methylation, including the structural basis for specific adenosine 2’O methylation remains elusive. Moreover, the size of the RNA substrate that can be accommodated by the putative RNA-binding cleft is also unknown, as any requirement for a specific RNA conformation. Determination of the structure of NS5 bound to a viral RNA would give valuable information to guide the design of specific NS5 inhibitors.

To elucidate how the flavivirus RNA is recognized and methylated, we determined the crystal structure of a ternary complex between the full length NS5 protein (DENV serotype 3), an authentic cap-0 viral RNA substrate (5’-m7G0pppA1G2U3U4G5U6U7-3’) and SAH, the by-product of the methylation reaction. Together with mutagenesis data informed by the present structure, this work reveals a unique and specific interaction between the protein and viral RNA and provides the molecular basis for the methyl transfer reaction. Furthermore, despite a low sequence identity between the two proteins, the RNA recognition mode is reminiscent of how the human 2’O ribose methyltransferase CMTr1 binds mRNA for cap formation, suggesting that the viral methyl-transferase might derive from its eukaryotic homologue.

Results and Discussion

Overall structure of the ternary complex of DENV3 NS5, cap 0 vRNA, and SAH.

An overall view of bound RNA in the MTase domain of NS5 is given in Fig. 1a. The electrostatic surface representation in Fig. 1b reveals that the RNA moiety occupies a significant portion of a large basic patch of the NS5 protein. The ordered part of the single strand RNA substrate consists in the 5’-m7G0pppA1G2U3U4 sequence (Fig. 1c and Fig. S1). The crystal structure of the ternary complex between NS5, cap 0 RNA and SAH (Fig. 1) was refined to R=0.2027, Rfree=0.251 at a resolution of 2.70 Å with good stereo-chemical
parameters (Table 1). Bound RNA extends between F25 which is stacked with $m^7G_0$ at the 5’ end, and the methyl donor pocket, spanning an overall distance of ~17 Å across the RNA binding groove. The $m^7G_0$ ring makes hydrogen bonds through its N2 atom with the carbonyl oxygen atom of L20 and N18 as well as via its ribose 2’OH with the N18 side chain (Fig. 1c) (10, 11). The N7-methyl of the base enhances stacking interactions with F25. The negative charges of the triphosphate group that links $m^7G_0$ and A$_1$ are neutralized by a hydrated Mg$^{2+}$ ion and hydrogen bonds with residues S213 and S150 that project from the protein surface (Table S1a). The remainder of the RNA (G$_2$U$_3$U$_4$) adopts an A-helical conformation along an axis that criss-crosses with the $m^7G_{0ppp}$ moiety, giving an overall “L”-shape to the bound cap-0 RNA (Fig. 1c). The observation of four stacked bases A$_1$G$_2$U$_3$U$_4$ in addition to $m^7G_{0ppp}$ (Fig. 1c) is consistent with previous RNase foot-printing assays showing that a total of four nucleotides become protected by the flavivirus MTase during 2’O methylation (10, 18). Only weak electron density is visible for G$_3$U$_6$U$_7$ that extend out of the MTase binding groove. In the present conformation, no constraint on the length of the RNA substrate that can be accommodated by the MTase domain, seems imposed by the NS5 RdRp domain (Fig. 1 and S1). To assess possible conformational changes upon substrate binding, we superimposed the present NS5-capped RNA complex with the free NS5 enzyme (PDB code: 4V0Q). This returned a r.m.s value of 0.27 Å for 739 C$_{\alpha}$ atoms, suggesting an essentially preformed RNA binding groove with no domain reorientation between the MTase and RdRp domains upon binding capped RNA (12). Remarkably, the RNA substrate is positioned such that the 2’O atom of residue A$_1$ lies next to the sulphur atom of SAH and adjacent to the K180 side chain from the “K$_{61}$-D$_{146}$-K$_{180}$-E$_{216}$” enzymatic motif, poised to accept a methyl group from a SAM methyl donor (Fig. 1c and 2a).

**Specificity of viral RNA recognition by NS5 MTase**

The crystal structure explains the specific recognition of the flavivirus RNA 5’ cap by NS5 and in particular the requirement for an adenosine at position 1 and a guanosine at position 2 (Fig.1d, 1e and S2). The first nucleotide of the flavivirus genome, always an adenosine (A$_1$), resides ~ 8 Å away from and nearly perpendicular to $m^7G_0$ to which it is connected via the triphosphate 5’-5’ linkage (Fig. 1c and 1d). Remarkably, the A$_1$ base fits snugly in a pocket shaped by residues I147-G148-E149-S150 from NS5 and SAH (Fig. 1d) with one hydrogen bond connecting the hydroxyl side chain of S150 and the adenine N7 atom. To understand the origin of specific adenosine 2’O methylation by flaviviruses, we modelled X$_1$=G, U or C ($m^7G_{0ppp}$X$_1$-RNA) in place of A$_1$ and found that the N2 amine group
from G₁ would sterically collide with SAH (nearest distance ~ 1 Å) (Fig. S2b). Conversely, pyrimidine bases would leave a large empty space in the pocket, leading to energetically less favourable Van der Waals interactions with the protein (Fig. S2a). G₂, the second nucleotide of the viral genome, stacks with A₁ and makes a hydrogen bond with the carboxylic side-chain of E111 via its N2 atom and with the Mg²⁺ ion (Fig. 1e and S2b). We explore below the importance of this polar contact in the context of dengue virus replication. Modelling X₂=A, U or C (7meG₀pppAX₂-RNA) in place of G₂, showed that bases other than G disrupt favourable contacts and in some cases, introduce unfavourable charge-charge repulsion (Fig. S2b). Thus specific recognition of the 5’ end of the flavivirus genome is seen as the consequence of optimum shape complementarity and specific hydrogen bonding between the NS5 protein and the SAM methyl donor. This observation is consistent with the strict conservation of the 5’ G₀pppA₁G₂ sequence during flavivirus genomes evolution (Fig. S2c) (19). Likewise, residues from the NS5 MTase domain that form the tip of the RNA binding pocket and establish specific polar interactions with the A₁G₂ dinucleotide are also conserved.

**Enzymatic mechanism for viral RNA 2’O methylation**

RNA is bound in the DENV MTase catalytic site, such that the 2’O ribose of residue A₁ is placed next to the SAH moiety. This observation suggests that the present complex represents a snapshot along the 2’O methylation reaction pathway, possibly the reaction product (Fig. 2 and S3a): (i) in the first step, the side chain of K180 activates the 2’OH of the adenosine ribose of the RNA methyl acceptor, either by proton abstraction or destabilization of the hydroxyl oxygen orbital (20), (ii) second, the activated ribose 2’O makes a nucleophilic attack on the methyl group of the positively charged sulphur centre via an in-line S_N2 reaction. At pH=8.5, a value close to the pKa of the lysine ε- amino-group side-chain, where the 2’O MTase reaction is maximal (21, 22), the side-chain of K180 can be easily deprotonated. Moreover, because K180 is sandwiched between the acidic side-chains of D146 and E216, its side chain could alternate between a charged and uncharged species, at various stages of the 2’O methylation cycle. This scheme is consistent with the available mutagenesis data: the D146A single mutant is completely devoid of 2’O methylation activity (Table S1a); its side-chain carboxylic group is at the right distance to stabilize the electrophilic sulphur group of SAM as well as K180, following activation of the 2’-OH adenosine ribose (Fig. 2a and S3) (23). Mutagenesis and thermo-denaturation experiments also indicate that residues F25, R57, and K61 play a major role in the reaction: alanine mutations at these positions essentially abolish the 2’O MTase activity (Table S1a).
Moreover, the pattern of RNA-enhanced thermo-stabilization for the DENV MTase is consistent with the observation that amino acids F25, R57 and R61 all make key contacts with RNA in the present complex. One attractive hypothesis suggested by the present structure, is that the SAM cofactor could act as a “sensor” by probing for the base identity and for the presence of methyl group(s) in the RNA substrate. Depending on its methylation state, the RNA substrate could be either in a favourable position in the binding groove conducive to methyl-transfer or sterically collides with the methyl donor favouring its dissociation from the MTase domain.

The catalytic mechanism proposed for NS5 is shared with the human mRNA cap-specific 2’O-ribose MTase CMTr1 and the Vaccinia virus VP39 proteins (whose structures with capped RNA are known), and other RNA 2’O MTases across all animal kingdoms and viruses whose free enzyme structures have been reported (Fig. S4 and Table S2) (24, 25). However, in contrast to the human CMTr1 or vaccinia virus VP39 that methylates Cap 0 RNA substrates in a sequence-independent manner, the flavivirus NS5 MTase has a demonstrated RNA substrate sequence specificity, which is evident in our structure analysis and functional studies (Fig. 1, 3 and 4). A requirement for Mg$^{2+}$ ions was reported for the COMT MTase as well as for FtsJ 2’O methylation activities (26, 27) and Mg$^{2+}$ has stimulatory effects on the DENV 2’O MTase activity (28). The structure of COMT suggested that Mg$^{2+}$ ions play an important role for RNA stabilization but not for catalysis. Given the close similarity between the active sites of VP39 and FtsJ, Mg$^{2+}$ ions is thought to be dispensable for 2’O methylation (27). Thus, activity enhancement by the presence of Mg$^{2+}$ in DENV NS5 appears to be due to structural and electrostatic stabilization of the protein RNA complex.

**Flavivirus NS5 MTase is closest to the human mRNA cap-specific 2’-O-ribose CMTr1**

To gain insights into the evolution leading to the present NS5 MTase domain, we performed an homology search against the various available MTase structures (29). The flavivirus NS5 protein belongs to a family of SAM-dependent RNA MTases that methylate diverse RNA species including mRNA, rRNA, tRNA and siRNA, either at the bases or at 2’ hydroxyl groups of specific nucleosides. Within the flavivirus genus, both the primary sequence and the 3D structures of MTase domain are highly conserved, with sequence identities ranging from 51 to 100 % (Z scores of 32.2-43.8). Surprisingly, despite a low sequence identity of 14% between the two proteins, the closest NS5 structural homologue with a r.m.s.d. of 2.9 Å for 222 of the 406 CMTr1 α-carbon atoms and a Z score of 15.8, is the human mRNA cap-
specific 2’-O-ribose MTase CMTr1 (PDB code: 4N48). Moreover, the seven next closest homologues are also of non-viral origin (Fig. S4 and Table S2). These structures share the general αβα Rossmann-fold core domain that shapes the SAM binding pocket and the highly conserved K-D-K-E catalysis tetrad, which is consistent with a common catalytic mechanism for methyl transfer. Structural divergence beyond the core domain likely reflects the present variety of RNA substrate preferences (eg 2’-O-ribose of capped RNA vs ribosomal RNA methylation) (30, 31). Thus, the Flavivirus NS5 MTase domain might derive from its eukaryotic homologue. Interestingly, the Flavivirus genus is the only clade in Flaviviridae family that encodes a MTase at the N terminus of NS5. Considering that the RdRp domain is highly conserved in all viruses from Flaviviridae family but absent from the host cells, the present flavivirus NS5 protein might have originated from a domain fusion event through the acquisition of a RNA 2’-O-MTase domain from the host, following the separation from the Hepatitis, Pestivirus and GB genera of the Flaviviridae family (12).

Residue E111 is key for RNA recognition and virus replication

To investigate functional implications of the present ternary complex for virus replication, we performed a mutagenesis study targeting the highly conserved residue E111 of NS5 and the second base (G2) of the flavivirus RNA (Fig. 1d, 1e and Fig. 4). We targeted residue E111 because it establishes the sole direct specific polar contact made between the protein and the G2 base (Fig. 1c). We evaluated the effects of E111 mutants both in vitro using an NS5 MTase enzymatic assay and also in cell culture by examining the growth kinetics of the corresponding virus mutants. Reversing the charge through the E111R mutation abolishes 2’-O MTase activity (Fig. 4a). In contrast, mutants E111A and E111Q retain > 93% activity, which is consistent with a limited impact of these mutations on protein-RNA binding affinity (Fig. 4a and Table S1b). Substituting X2=U or C in place of G2 (7meG0pppAX2-RNA) introduces unfavourable electrostatic contacts between their O2 atom and the carboxylic group of E111 (for X2=A2, between N6 and Mg2+) (Fig. S2b). Mutations of G2 to U, C, or A reduce the 2’-O methylation activity of NS5 down to 19%, 0% and 33% respectively in WNV (10) and to 55% 24% and 84% in DENV4 (Table S1b). Moreover, the consensus sequence G0pppA1G2X (Fig. S2c) binds more tightly than G0pppA1C2X to the DENV MTase (16). Consistently, specific 2’O-methylations have been identified in viral genomic RNA internal adenosines and polyA but not on polyG, polyC or polyU substrates (9). In addition to modulating the NS5 MTase activity, we found that E111 plays a crucial role in
viral replication (Fig. 4e). The E111Q and E111A mutations attenuated virus replication and produced 3-4 times less infectious virus, while the E111R mutation severely impaired viral replication and yielded no virus (Fig. 4e). In summary, the targeted mutagenesis on both NS5 protein and viral RNA proves the direct relevance of the structure to dengue virus replication and infection.

**Concluding Remarks**

The structure presented here and the functional studies also suggest possible routes for the design of inhibitors targeting the NS5 MTase catalytic site. One possible strategy would consist in extending SAM analogues towards the RNA binding groove - for instance by linking the A₁ moiety to nucleoside analogues (eg: ribavirin) that bind at the GTP binding site, next to residue F25. This strategy may lead to larger but more specific and potent flavivirus MTase inhibitors (32). Our study also indicates that attenuated-viruses could be designed for instance by targeting the E111 residue. Lastly, the present complex provides a striking example of the specific interactions between an enzyme and its RNA substrate (34).
Methods

Methods used in this study are briefly summarized below. Full descriptions are given in Supporting Information Appendix.

Protein expression and purification. Protocol for NS5 protein expression and purification was done as previously described (12).

Crystallization and data collection. Crystallization was set-up at 20 °C using the hanging drop vapour diffusion and published conditions (12). Native crystals obtained over 2–5 days by mixing a volume of 1 µl of NS5 (6-895) at 4-6 mg/ml with 1 µl of precipitation solution (0.2 M magnesium acetate, 0.1 M sodium cacodylate, pH=6.4, 10-20% (w/v) PEG 8000) were soaked with 1mM of 7meG0pppA1G2U3U4G5U6U7-3’ overnight.

Structure solution and refinement. The crystals have a solvent content of 50.2% and a Matthews coefficient (Vm) of 2.47 (35). Refinement was initiated with structure 4V0Q using REFMAC5 (36) and PHENIX (37) and interspersed with manual model rebuilding sessions using Coot (38). The refined coordinates were deposited in the PDB under accession code XXXX.

Construction of mutant m7Gppp-DENV4 5'UTR nt-110 RNAs for 2’-O MTase assay. DENV4 5'UTR nt-110 DNA with substitutions at the position of 5’ nucleotide G2 was engineered by PCR using Phusion DNA Polymerase under standard procedures.

DENV4 2’-O MTase assay. DENV4 2’-O MTase assay was performed as described previously(39, 40).

Construction of mutant DENV2 full-length infectious clone. DENV2 full-length cDNA clones with NS5 E111A, E111Q and E111R mutations were constructed with a pACYC-NGC FL and a TA-NGC (shuttle E) vector as previously described (41).
**Author Contributions.** Experiment design: KS, PYS, SGV, JL and DL. Crystallization, data collection, structure determination and analysis: YZ, JL, and DL. Enzymatic and virology experiments: SS, KYC, SPL and PYS. Manuscript writing: JL and DL with inputs from all authors.

**Acknowledgements.** We thank Drs Liew Chong Wai and Chen Ming Wei, scientists from Swiss Light Source beam-line PXIII and National Synchrotron Radiation Research Centre beam-line ID13 for their help with diffraction data collection. We thank Dr George Chandy for critically reading our manuscript. This work was supported by (1) the start-up grant to DL lab from Lee Kong Chian School of Medicine, Nanyang Technological University, (2) National Medical Research Council grant CBRG14May051 to JL, (3) MOE Tier 1 complexity grant to JL lab, and (4) National Medical Research Council grant # NMRC/1315/2011 to SV lab. YZ is supported by a NGS scholarship.

**Conflict of Interest.** The authors declare no competing financial interests.
Reference:

34. Hedstrom L (2010) Enzyme Specificity and Selectivity. eLS.
Table 1: Data collection and refinement statistics

<table>
<thead>
<tr>
<th></th>
<th>NS5-Capped RNA7</th>
</tr>
</thead>
<tbody>
<tr>
<td>Resolution range (Å)</td>
<td>47.33 - 2.70 (2.85 - 2.70)*</td>
</tr>
<tr>
<td>Space group</td>
<td>P 2\text{I}2\text{I}2</td>
</tr>
<tr>
<td>Unit cell parameters a, b, c (Å)</td>
<td>94.66 151.39 69.43</td>
</tr>
<tr>
<td>Measured reflections</td>
<td>114,953 (17,172)</td>
</tr>
<tr>
<td>Unique reflections</td>
<td>28,027 (4,018)</td>
</tr>
<tr>
<td>Multiplicity</td>
<td>4.1 (4.3)</td>
</tr>
<tr>
<td>Completeness (%)</td>
<td>99.7 (99.6)</td>
</tr>
<tr>
<td>Mean I/sigma(I)</td>
<td>9.6 (2.1)</td>
</tr>
<tr>
<td>R&lt;sub&gt;merge&lt;/sub&gt;&lt;sup&gt;a&lt;/sup&gt;</td>
<td>0.126 (0.625)</td>
</tr>
<tr>
<td>R&lt;sub&gt;work&lt;/sub&gt;&lt;sup&gt;b&lt;/sup&gt;</td>
<td>0.2027 (0.2925)</td>
</tr>
<tr>
<td>R&lt;sub&gt;free&lt;/sub&gt;&lt;sup&gt;c&lt;/sup&gt;</td>
<td>0.2510 (0.3496)</td>
</tr>
<tr>
<td>Number of non-hydrogen atoms</td>
<td>7,102</td>
</tr>
<tr>
<td>macromolecules</td>
<td>6,889</td>
</tr>
<tr>
<td>ligands</td>
<td>45</td>
</tr>
<tr>
<td>water</td>
<td>168</td>
</tr>
<tr>
<td>Protein residues</td>
<td>851</td>
</tr>
<tr>
<td>RMS(bonds, Å)</td>
<td>0.004</td>
</tr>
<tr>
<td>RMS(angles, °)</td>
<td>0.85</td>
</tr>
<tr>
<td>Ramachandran plot (%)</td>
<td></td>
</tr>
<tr>
<td>favoured</td>
<td>95.45</td>
</tr>
<tr>
<td>allowed</td>
<td>4.31</td>
</tr>
<tr>
<td>outliers</td>
<td>0.24</td>
</tr>
<tr>
<td>Clash score</td>
<td>6.37</td>
</tr>
<tr>
<td>Average B-factor</td>
<td></td>
</tr>
<tr>
<td>macromolecules</td>
<td>38.90</td>
</tr>
<tr>
<td>ligands</td>
<td>49.00</td>
</tr>
<tr>
<td>solvent</td>
<td>53.10</td>
</tr>
<tr>
<td></td>
<td>28.60</td>
</tr>
</tbody>
</table>

*The numbers in parentheses refer to the last (highest) resolution shell.

<sup>a</sup> $R_{\text{merge}} = \sum |I_j - <I>| / \sum |I_j|$, where $I_j$ is the intensity of an individual reflection, and $<I>$ is the average intensity of that reflection.

<sup>b</sup> $R_{\text{work}} = \sum |F_o| - |F_c| / \sum |F_o|$, where $F_o$ denotes the observed structure factor amplitude, and $F_c$ the structure factor amplitude calculated from the model.

<sup>c</sup> $R_{\text{free}}$ is as for $R_{\text{work}}$ but calculated with 5% (3044) of randomly chosen reflections omitted from the refinement.
Figure Legends

Figure 1. Structure of the ternary complex between DENV3 NS5, capped RNA and SAH. (a) Crystal structure of NS5<sub>6-896</sub> (MTase domain: cyan; RdRp domain green) bound to RNA (5'-<sup>m7</sup>G<sub>0</sub>pppA<sub>1</sub>G<sub>2</sub>U<sub>3</sub>U<sub>4</sub>GUU-3', magenta sticks) and SAH (yellow sticks). The RNA and SAH binding site is boxed. (b) Electrostatic surface representation of NS5 (positive charges in blue, negative charges in red). Capped RNA (magenta) and SAH (yellow) are shown as sticks. (c) Close-up view of the RNA and SAH binding sites as in b. Protein residues binding RNA or SAH are represented as sticks (cyan) and labelled. Capped RNA nucleotides are labelled. Dashed lines indicate polar interactions. Mg<sup>2+</sup> is displayed as a green sphere. (d) Interactions established between adenine A<sub>1</sub> from the bound RNA with NS5, illustrating the tight shape complementarity with adenine only. (e) Hydrogen bonds formed between E111, the capped RNA second nucleotide G<sub>2</sub> and the bound Mg<sup>2+</sup> ion.

Figure 2. Proposed enzymatic mechanism for 2'-O methylation by the flavivirus NS5 MTase. (a) Close-up view of the MTase active site with the K<sub>61</sub>-D<sub>146</sub>-K<sub>180</sub>-E<sub>216</sub> catalytic tetrad shown as yellow sticks and SAH in magenta sticks. Colour code for the bound RNA: G<sub>0</sub> carbon atoms: green, A<sub>1</sub> in cyan, and G<sub>2</sub>U<sub>3</sub>U<sub>4</sub> in grey. The close contacts between the 2'-oxygen atom of adenine A<sub>1</sub>, the amino group of K<sub>180</sub> and the sulphur group of SAH, are indicated by dashed lines and the corresponding distances are given. (b) Schematic view of an active ternary complex comprising cap-0 RNA and the methyl donor SAM, based on the present structure. Proposed enzymatic mechanism for 2'-O methylation: First, the K<sub>180</sub> side chain activates the 2'OH adenosine ribose by proton abstraction or destabilization of the hydroxyl oxygen orbital; the activated ribose 2'-O then makes a nucleophilic attack on the methyl group of the sulphur centre, via an in-line S<sub>N</sub>2 reaction. The 2'-O MTase reaction is favoured at a basic pH (10), close to the K<sub>180</sub> ε- amino-group pK<sub>a</sub> where it can be easily deprotonated. K<sub>180</sub> and K<sub>61</sub> are sandwiched between the acidic side-chains of D<sub>146</sub> and E<sub>216</sub> and could alternate between uncharged and charged species, when progressing through the reaction. The stereochemistry of the reactants (distances and angles) conform to what is expected from an in-line S<sub>N</sub>2 reaction.

Figure 3. Comparison of NS5 MTase with other human and viral 2-O MTases. (a) The DENV3 NS5 MTase (this work) (b), vaccinia virus MTase VP39 (c), human CMTr1 (d), SARS coronavirus MTase nsp16
The comparison highlights the variety of shape complementarity and surface charge distributions between the enzymes, cofactor and RNA substrate. The VP39 G₀ pocket is deeper and more distant from the RNA binding groove compared to NS5 and CMTr1. As a result, G₀ is deeply buried, with a protruding ribose. In contrast, no apparent G₀ binding pocket is found in nsp16. Triphosphates display various conformations: extended in VP39 and CMTr1 and bent in NS5. In all three complexes, four RNA bases occupy the positively charged groove. The RNA binding groove in nsp16 is shallow and less positively charged. The SAM binding pocket is the most conserved feature, except at the adenine binding site: In NS5, a positively charged pocket is located next to N7, while in CMTr1, the corresponding pocket is deeper and larger; in VP39, it is open and negatively charged. In nsp16, no apparent pocket is visible. Positive proteins surfaces are shown in blue, negative surface in red, RNA moieties (yellow carbons) and SAH/SAM (magenta carbons) as sticks. (e) Multiple sequence alignment. The KDKE catalytic tetrad is highlighted in yellow. E₁₁₁ from DENV NS5 is also highlighted.

**Figure 4. Specific recognition of the capped viral RNA is essential for virus replication.**

(a) *In vitro* 2’-O MTase activities of WT NS5 and E₁₁₁ mutants using WT and mutant RNA templates. Data are normalized to the activity of the WT NS5 MTase on WT viral 5’ UTR RNA which is set to 100%. (b) Intracellular viral RNA levels detected by qRT-PCR. (c) Extracellular viral RNA levels in the supernatants detected by qRT-PCR. (d) Virus titres based on the plaque assay shown in (e). (e) Plaque assay for WT and mutants at 25, 72 and 120 hours post infection.
Figure 2

(a) 

(b)