



OPEN

# Identification of the viral RNA promoter stem loop A (SLA)-binding site on Zika virus polymerase NS5

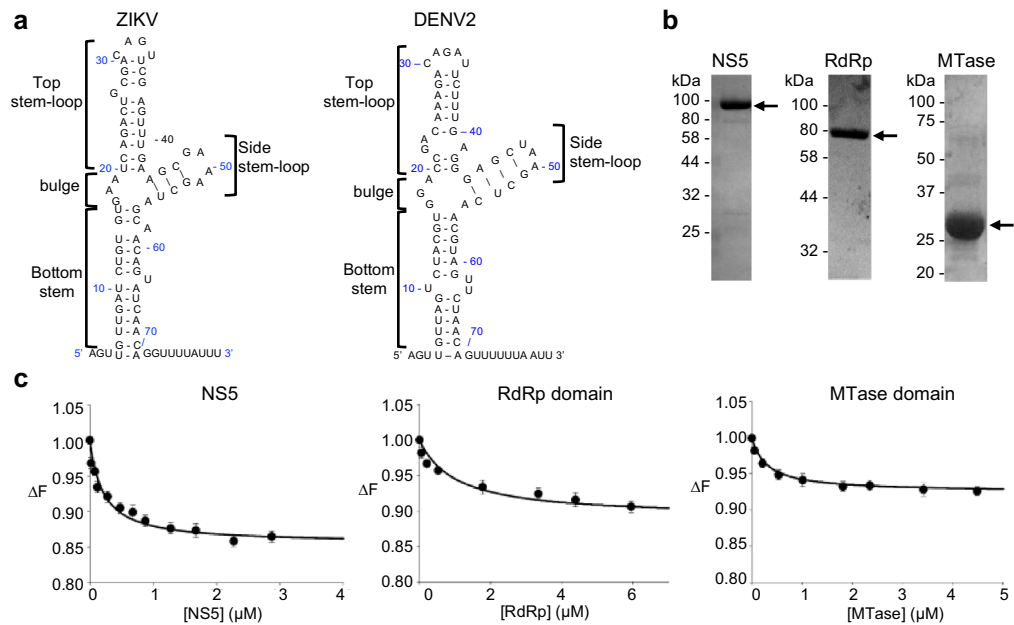
Paul J. Bujalowski, Włodzimierz Bujalowski &amp; Kyung H. Choi✉

Zika virus has recently emerged as an important human pathogen that has spread to more than 60 countries. Infection of a pregnant woman with Zika virus can cause severe brain malformations in the child such as microcephaly and other birth defects. Despite the medical importance of Zika virus infection, the mechanism of viral replication, a process commonly targeted by antiviral therapeutics, is not well understood. Stem-loop A (SLA), located in the 5' untranslated region of the viral genome, acts as a promoter for viral replication and thus is critical for recognition of the viral genome by the viral polymerase NS5. However, how NS5 engages SLA is not clear. We have quantitatively examined the intrinsic affinities between Zika virus SLA and NS5, and identified the SLA-binding site on NS5. Amino acid substitutions in the thumb subdomain of the RNA-dependent RNA polymerase (RdRp) and the methyltransferase (MTase) domain reduced SLA-binding affinity, indicating that they each are part of the SLA-binding site. Furthermore, stopped-flow kinetic analysis of Zika NS5-, RdRp- and MTase-SLA interactions identified distinct intermediates during NS5 and SLA complex formation. These data suggest a model for SLA recognition and the initiation of flaviviral replication by NS5.

Zika virus (ZIKV) is a member of the family *Flaviviridae*, and since 2007 has emerged as a major threat to global health, causing a series of epidemics in Micronesia, the South Pacific and both Americas<sup>1</sup>. Currently, the virus has been reported in 65 countries<sup>2,3</sup>. ZIKV is closely related to human pathogens such as tick-borne encephalitis virus, dengue virus (DENV), West Nile virus, yellow fever virus, and Japanese encephalitis virus<sup>4,5</sup>. ZIKV can be transmitted sexually and through blood transfusions, and spread from a pregnant woman to her fetus. ZIKV infection of a pregnant woman can cause severe brain malformations in her child including microcephaly and other birth defects<sup>6,7</sup>. Moreover, in adults, ZIKV has been linked to Guillain-Barré syndrome, a condition wherein a person's immune system attacks their peripheral nerves<sup>8,9</sup>. In severe cases, this can lead to near-total paralysis and death.

The structure of the ZIKV genome is similar to other flaviviral genomes. The 11 kb long positive sense RNA contains an open reading frame (ORF) as well as 5' and 3' untranslated regions (UTRs) that regulate viral replication<sup>10-12</sup>. The ORF is translated into a polyprotein, C-prM-E-NS1-NS2A-NS2B-NS3-NS4A-NS4B-NS5, which is subsequently cleaved into three structural (C, prM, and E) and seven non-structural (NS) proteins. The viral NS proteins, along with viral RNA and unidentified host proteins, self-assemble on the cytoplasmic side of the ER membrane to form a viral replication complex that carries out genome replication. Although all NS proteins are required to form a viral replication complex, only NS3 and NS5 enzyme activities are involved in viral genome replication. NS3 consists of an N-terminal serine protease domain, which requires NS2B as a cofactor, and a C-terminal helicase domain. The helicase domain also has nucleoside-triphosphatase (NTPase) activity and 5'-RNA triphosphatase activity, which hydrolyzes the  $\gamma$ -phosphate of RNA for RNA capping<sup>13</sup>. NS5 is the viral replicative polymerase responsible for genome synthesis<sup>4,14,15</sup>. Full-length NS5 encompasses approximately 900 amino acids and is composed of two functionally different domains<sup>16-21</sup>. The smaller N-terminal domain functions as a methyltransferase (MTase), whose activity is necessary for the 5'-RNA cap formation and methylation that are important for viral RNA recognition by the host cell translational apparatus<sup>22,23</sup>. The

Department of Biochemistry and Molecular Biology, Sealy Center for Structural Biology and Molecular Biophysics, The University of Texas Medical Branch, Galveston, TX 77555, USA. ✉email: kychoi@utmb.edu



**Figure 1.** Full-length Zika NS5, the RdRp domain, and the MTase domain interact with SLA. **(a)** The predicted secondary structures of the Zika and dengue virus SLA. The secondary structure of the first 80 nt of ZIKV (KU527068.1) used in this study was predicted by mFold<sup>42</sup>. The predicted secondary structure of DENV2 SLA (NC\_001474) is shown on the right. Both SLAs consists of the top stem-loop, side stem-loop, and bottom stem with a bulge between the top and bottom stem. Sequence identity between ZIKV and DENV2 SLA is 64%. **(b)** SDS-PAGE analysis of full-length NS5, RdRp and MTase. The positions of protein molecular weight markers are shown on the left. The expected sizes of the NS5, RdRp, and MTase are 100, 72, and 30 kDa, respectively. **(c)** Fluorescence titrations of fluorescein-labeled ZIKV SLA with NS5, the RdRp domain, and the MTase domain. The binding reactions were carried out with 15 nM fluorescein-labeled SLA in buffer B1 (50 mM Tris, pH 8.0, 150 mM NaCl, 1 mM MgCl<sub>2</sub>, 2 mM  $\beta$ -mercaptoethanol and 10% glycerol) at 20 °C with  $\lambda_{ex}$  = 480 nm,  $\lambda_{em}$  = 520 nm. The solid lines are nonlinear least squares fits of the titration curves (Eq. 11) with  $K_{NS5} = 3.5 \times 10^6 M^{-1}$  ( $K_d = 0.28 \mu M$ ,  $R^2 = 0.98$ ),  $K_{RdRp} = 1.9 \times 10^6 M^{-1}$  ( $K_d = 0.53 \mu M$ ,  $R^2 = 0.97$ ), and  $K_{MTase} = 1.7 \times 10^6 M^{-1}$  ( $K_d = 0.59 \mu M$ ,  $R^2 = 0.99$ ). Each experiment was carried out in triplicate. Error bars represent the standard deviations.

larger C-terminal domain possesses RNA-dependent RNA polymerase (RdRp) activity, containing the active site for RNA synthesis<sup>24</sup>.

In flaviviruses, the 5'-UTR of the viral genome includes a structured stem-loop, called stem-loop A (SLA, nt 1-70) (Fig. 1a), which functions as a promoter and recruits the viral polymerase NS5 to initiate RNA synthesis at the 3'-end of the genome<sup>10-12,25,26</sup>. Although NS5 can synthesize RNA from short RNA templates, NS5 cannot synthesize RNA from a sub-genomic RNA template that lacks SLA<sup>11</sup>. A predicted 'Y'-shaped secondary structure of SLA is conserved in all dengue virus serotypes (DENV1-4) as well as in other flaviviruses such as ZIKV and West Nile virus<sup>12,27,28</sup>, suggesting that flaviviruses share a common promoter structure and initiation mechanism. Additionally, SLA is required for 5' cap methylation at the N7 position, indicating that the MTase domain of NS5 recognizes structural features of the SLA to carry out the methylation reaction<sup>29</sup>. Despite the importance of SLA in initiating both of these critical aspects of viral replication (RNA synthesis and RNA capping), the exact roles each of the NS5 domains play in recognizing and binding SLA are not well understood. In particular, the SLA-binding site on NS5 has not been identified. Thus, it is not clear whether both NS5 domains are involved in SLA binding or whether one of the two domains regulates promoter binding.

To elucidate the mechanism of the first step of flavivirus replication, i.e., the NS5-RNA genome recognition process, we have previously examined DENV NS5 interactions with SLA, ssRNA, and dsRNA using fluorescence titration techniques<sup>30,31</sup>. DENV NS5 binds SLA specifically with 83 nM affinity and a 1:1 stoichiometry<sup>30</sup>. Here, we have examined the binding affinities of ZIKV SLA with the full-length NS5 and its two individual RdRp and MTase domains, and identified the location of the SLA-binding site on the surface of NS5. Our data show that SLA interacts with specific positively charged surface located on both the RdRp and MTase domains of NS5. We further determined the kinetics of Zika NS5-, RdRp- and MTase-SLA complex formation, and observed the presence of intermediates in the NS5-SLA complex reaction.

## Results

**Full-length Zika NS5 and the individual RdRp and MTase domains interact with Zika SLA.** Flavivirus replication is initiated when the viral polymerase NS5 identifies the viral genome via specific recognition of the SLA<sup>10-12</sup>. Accordingly, flavivirus NS5 has been shown to bind SLA with high affinity<sup>10,30,32</sup>. We have previ-

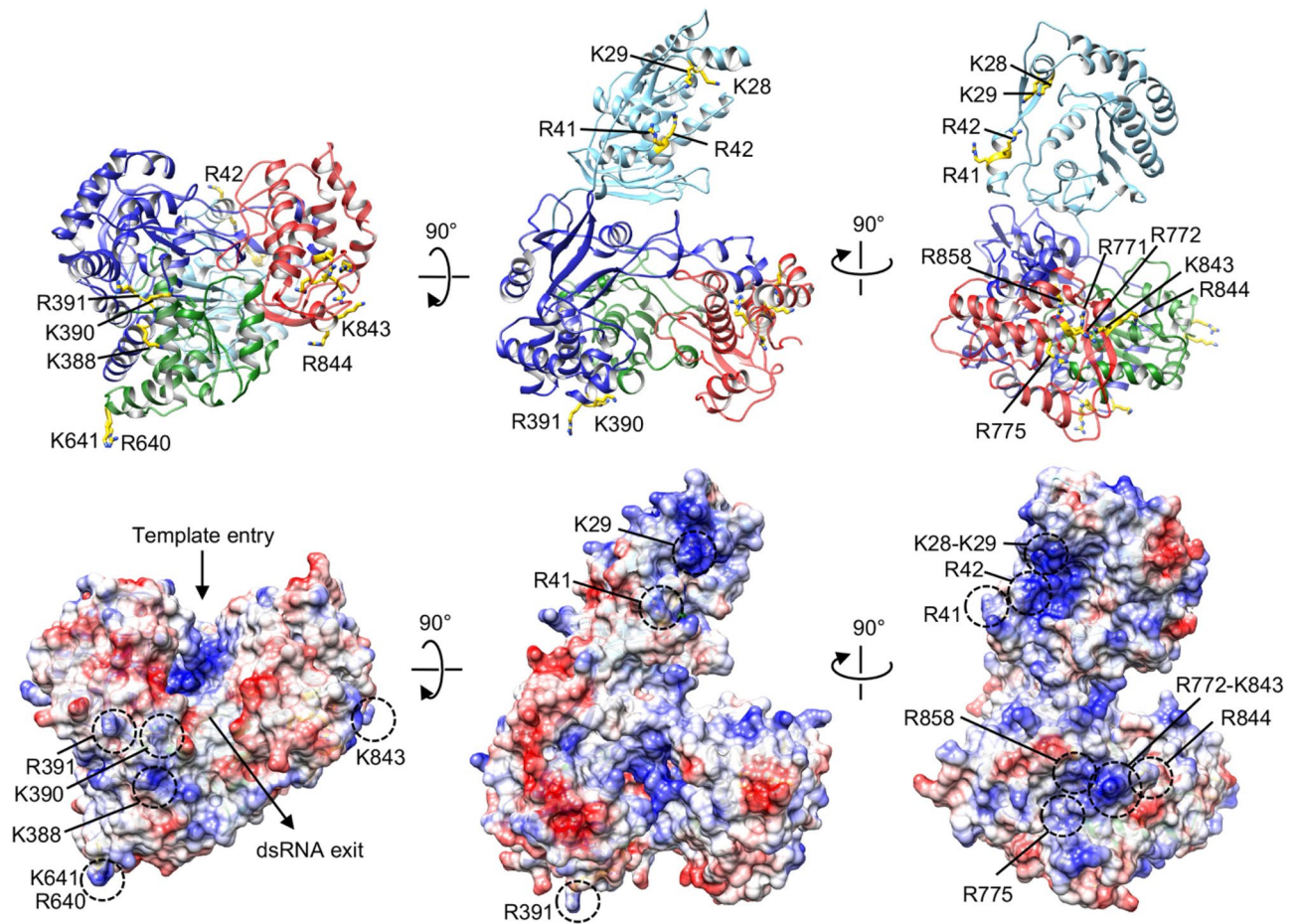
Protein	Location of mutations	Binding constant, K ( $10^6 \text{ M}^{-1}$ ) <sup>a</sup>	Dissociation constant, $K_d$ ( $\mu\text{M}$ )	Fold change
WT NS5		$3.5 \pm 0.6$	0.28	–
MTase domain		$1.7 \pm 0.3$	0.59	2.1
RdRp domain		$1.9 \pm 0.3$	0.53	1.8
NS5 mutant				
K388A-K390A-R391A	Palm	$3.5 \pm 0.6$	0.28	1.0
R640A-R641A	Fingers	$3.5 \pm 0.6$	0.28	1.0
R775A	Thumb	$3.5 \pm 0.6$	0.28	1.0
R771A-R772A-R775A	Thumb	$2.0 \pm 0.3$	0.50	1.8
R771A-R772A-R775A-K843A-R844A	Thumb	$1.4 \pm 0.2$	0.71	2.5
K843A-R844A	Thumb	$3.5 \pm 0.6$	0.28	1.0
R858A	Thumb	$3.5 \pm 0.6$	0.28	1.0
R891A	Thumb/C-terminus	$3.5 \pm 0.6$	0.28	1.0

**Table 1.** Effects of Zika NS5 mutations on SLA-binding affinity. <sup>a</sup>All binding constants were measured in Buffer B1 (50 mM Tris–HCl, pH 8.0, 150 mM NaCl, 1 mM MgCl<sub>2</sub>, 2 mM  $\beta$ -mercaptoethanol, and 10% glycerol).

ously determined that DENV NS5 and SLA interact with 83 nM affinity using fluorescence measurements<sup>30</sup>. However, the roles each of the NS5 domains play in recognizing and binding SLA are not well understood. Here, we used ZIKV NS5 and the individual RdRp and MTase domains, and examined their binding affinities using the ZIKV SLA (80 nt, Fig. 1a) fluorescein-labeled at the 3' end (SLA-F). Sequence identities between DENV and ZIKV NS5 and SLA are 67% and 64%, respectively. Fluorescence titrations of SLA-F with ZIKV NS5, RdRp domain and MTase domain (Fig. 1b) were carried out in buffer B1 (50 mM Tris adjusted to pH 8.0 with HCl at 20 °C, 150 mM NaCl, 1 mM MgCl<sub>2</sub>, 2 mM  $\beta$ -mercaptoethanol, and 10% glycerol). The formation of the complex between the ZIKV NS5 proteins and SLA-F causes a decrease in nucleic acid fluorescence (Fig. 1c). The maximum observed value of fluorescence quenching ranged from 15% for the full-length NS5, 9% for the RdRp domain, to 7% for the MTase domain. The solid lines in Fig. 1c are the nonlinear least-squares fits of Eq. (11) to the experimental data (see “Methods” section). The binding constants are listed in Table 1. The full-length ZIKV NS5 binds SLA with a binding constant of  $K_{\text{NS5}} = 3.5 \times 10^6 \text{ M}^{-1}$  ( $K_d = 0.28 \mu\text{M}$ ). The binding constants for the RdRp domain and MTase domain are  $K_{\text{RdRp}} = 1.9 \times 10^6 \text{ M}^{-1}$  ( $K_d = 0.53 \mu\text{M}$ ) and  $K_{\text{MTase}} = 1.7 \times 10^6 \text{ M}^{-1}$  ( $K_d = 0.59 \mu\text{M}$ ), respectively. Thus, both MTase and RdRp domains individually interact with SLA-F with approximately two-fold lower affinities than the full-length NS5.

**The thumb subdomain of the RdRp domain is responsible for binding to SLA.** Because both RdRp and MTase domains are capable of independently binding to SLA with similar affinities, the SLA-binding site on NS5 likely involves both the RdRp and MTase domains. To identify the SLA-binding site in NS5, we calculated the electrostatic surface of the ZIKV NS5 structure<sup>19</sup> to look for positively charged patches that would be complementary to the negatively charged SLA (Fig. 2). The RdRp domain is composed of three subdomains, fingers, palm and thumb subdomains, and positively charged residues were selected from all three subdomains for substitutions in NS5. Figure 2 shows the locations of targeted residues within NS5. The selected NS5 mutants were generated by site-directed mutagenesis and purified similar to the wild-type protein (Fig. 3a). The binding constants for NS5 mutants were then measured using SLA-F in buffer B1 (Fig. 3b–c). The binding constants are listed in Table 1. The binding assay showed that mutations in the fingers and palm subdomains did not affect SLA binding (Fig. 3b). For example, a double mutant within the fingers subdomain (R640A-R641A) and a triple mutant in the palm subdomain (K388A-K390A-R391A) bind to SLA with indistinguishable affinities compared to the wild-type NS5. Thus, these regions in the fingers and palm subdomains are not directly involved in interactions with SLA. The residues K388, K390, and R391 in the palm subdomain are located near the product exit site of the RdRp domain likely involved in the dsRNA product binding (Fig. 2). Mutations of K387A-K388A or K388A-K389A in DENV4 NS5 (corresponding to K388, H389 and K390 in ZIKV NS5) have shown to reduce virus titers by 1.4 – 2.4 log<sub>10</sub> PFU/ml<sup>33</sup>. Our result thus suggests that the dsRNA product binding site does not overlap with the SLA-binding site.

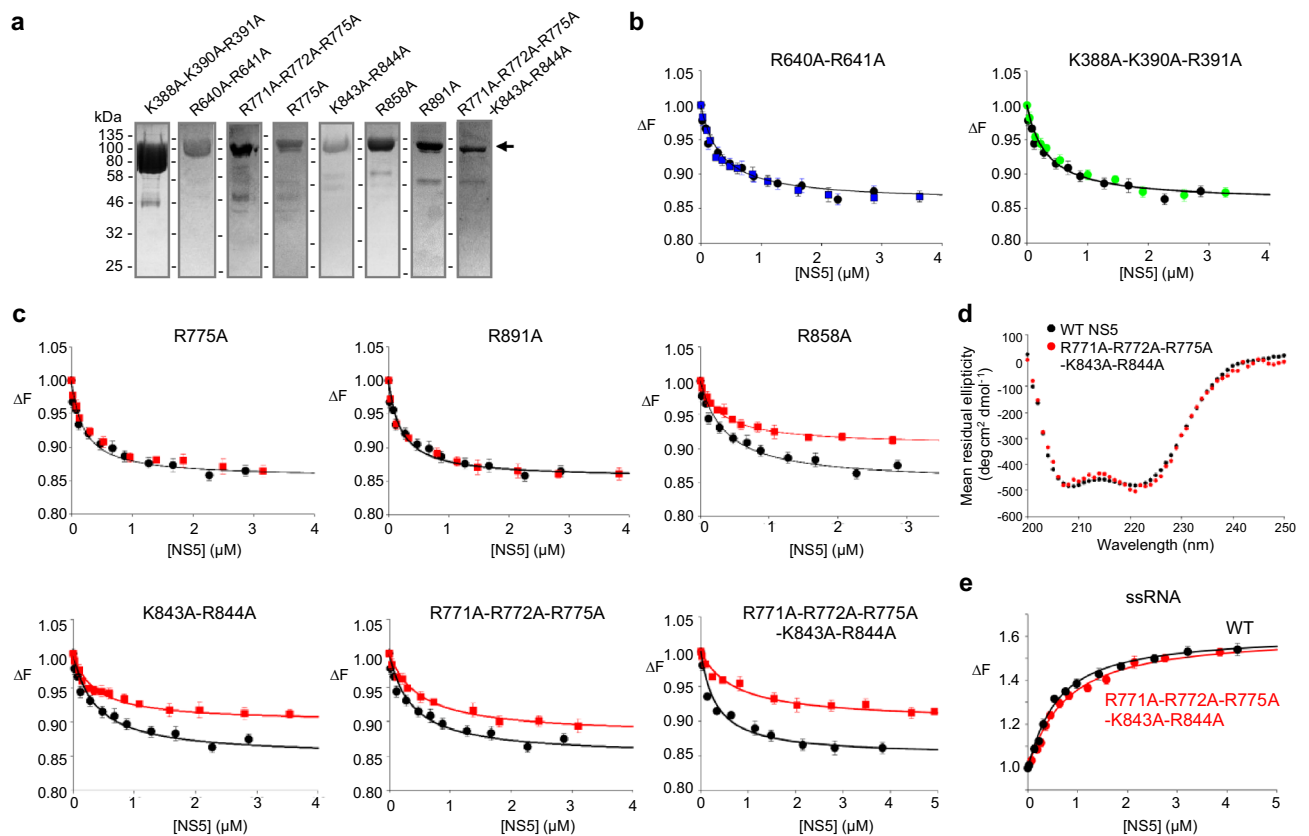
In contrast, several mutants in the thumb subdomain showed reduced binding affinities and/or altered the magnitude of the fluorescence signal (Fig. 3c). The thumb subdomain contains a large area of positively charged surface that include R771, R772, R775, K843, and R844 (Fig. 2). Thus, we generated six mutations in the area (Fig. 3a,c). The triple mutant, R771A-R772A-R775A showed the binding affinity of  $2.0 \times 10^6 \text{ M}^{-1}$  ( $K_d = 0.50 \mu\text{M}$ ), a 1.8-fold reduction from the wild-type affinity (Fig. 3c and Table 1). Since K843 and R844 are located in the positively charged patch near the mutated R771, R772, and R775, we tested if additional mutations within the patch reduce the binding to SLA even further. NS5 mutant containing R771A, R772A, R775A, K843A, and R844A indeed displayed a weaker affinity to SLA with  $K = 1.4 \times 10^6 \text{ M}^{-1}$  ( $K_d = 0.71 \mu\text{M}$ ) (Fig. 3c and Table 1). To determine if the reduced affinity of the R771A-R772A-R775A-K843A-R844A mutant is caused by the structural changes of the mutant, we compared the CD spectra for wild-type NS5 and the mutant (Fig. 3d). The CD spectra were similar, indicating that the secondary structure of NS5 has not been affected by mutations. Thus, the reduced



**Figure 2.** Mapping of Zika NS5 residues potentially involved in the recognition of SLA. Ribbon diagrams (top) and electrostatic surfaces (bottom) of Zika NS5 (PDB accession code 5U0B) are shown in three orientations with color-coded subdomains of the RdRp domain (thumb, red; palm, green; fingers, blue) and the MTase domain (cyan). Residues selected for mutagenesis are labeled and colored in yellow in ribbon diagram.

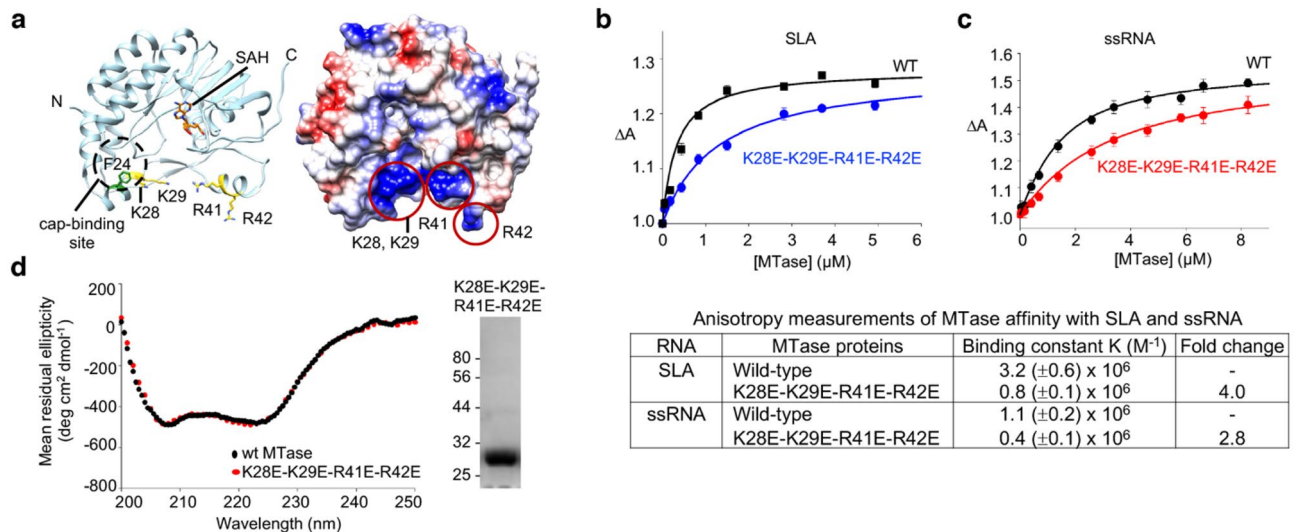
affinity of the R771A-R772A-R775A-K843A-R844A mutant is caused by the loss of its direct interaction with SLA. We also generated single and double mutations in the region, but they did not significantly affect SLA binding, suggesting that other positively charged residues within the patch may compensate for the loss of their positively charged side chains (Fig. 2). For example, R775A shows the same binding affinity as the wild-type NS5 (Fig. 3c and Table 1). A single mutant of R858A showed decreased magnitude of the fluorescence signal, although the binding affinity remained the same as for wild-type NS5 (Fig. 3c and Table 1). This indicates that Ala substitution in R858A affects the conformation of the bound SLA. A double mutant K843A-R844A also showed a decreased magnitude of the fluorescence signal suggesting that K843 and R844 also induce changes in the nucleic acid structure, although the mutations alone were not able to significantly reduce SLA binding (Fig. 3c and Table 1). We also tested if the C-terminal region of NS5 is involved in interactions with SLA. A single mutation of R888 to Ala in DENV NS5 was shown to produce a completely non-viable virus<sup>34</sup>. In the crystal structure of DENV NS5, R888 in the C-terminal helix is located near R772 of the positively charged patch<sup>16</sup>, suggesting that R888 could form a part of the SLA-binding surface. The corresponding residue R891 in ZIKV NS5 is disordered in the crystal structure<sup>19</sup>. However, the R891A mutant in ZIKV NS5 binds SLA with the wild-type affinity, suggesting that R891 at the C-terminal end of NS5 is not involved in the interactions with SLA (Fig. 3c and Table 1).

To determine if the SLA-binding site on the thumb subdomain is specific for SLA interaction, we next measured NS5 protein interactions with a nonspecific single-stranded RNA (ssRNA). Because the identified SLA-binding site does not overlap with the template-binding site (where ssRNA binds, Fig. 2), the SLA-binding site mutant should be able to bind nonspecific ssRNAs. We thus determined the binding affinity of NS5 proteins (wild-type and the R771A-R772A-R775A-K843A-R844A mutant) with etheno-labeled A(pA)<sub>19</sub>, εA(peA)<sub>19</sub> (Fig. 3e). Wild-type and mutant NS5 bind εA(peA)<sub>19</sub> with a similar binding affinity of  $K_{wt} = 2.0 \times 10^6 \text{ M}^{-1}$  ( $K_d = 0.50 \text{ } \mu\text{M}$ ) and  $K_{mut} = 1.5 \times 10^6 \text{ M}^{-1}$  ( $K_d = 0.67 \text{ } \mu\text{M}$ ). Thus, the mutations that affect SLA binding in NS5 do not significantly change nonspecific ssRNA interaction, indicating that the positively charged patch in the thumb subdomain (R771, R772, R775, K843, and R844) specifically interacts with SLA.



**Figure 3.** Mutations in the thumb subdomain decrease NS5 and SLA interactions. **(a)** SDS-PAGE of purified ZIKV NS5 mutants. Positions of molecular weight markers are shown on the left. **(b)** Fluorescence titrations of the fingers and palm subdomain mutants with SLA. NS5 mutants R640A-R641A (fingers subdomain, blue) and K388A-K390A-R391A (palm subdomain, green) interactions with fluorescein-labeled SLA in buffer B1 were measured by the fluorescence change at 20 °C. Titration of NS5 wild-type is shown in black. The solid lines are nonlinear least squares fits of the titration curves to Eq. (11) for the wild-type and mutants. Both R640A-R641A and K388A-K390A-R391A bind SLA with wild-type binding affinity of  $K = 3.5 \times 10^6 \text{ M}^{-1}$  ( $K_d = 0.28 \text{ } \mu\text{M}$ ,  $R^2 = 0.98$  and  $R^2 = 0.99$  respectively). Each experiment was carried out in triplicate. Error bars represent the standard deviations. **(c)** Fluorescence titrations of the thumb subdomain mutants with SLA. The titration curves were fit to Eq. (11) using nonlinear least squares fits. The R775A and R891A mutants show identical SLA binding curves to wild-type NS5 ( $R^2 = 0.98$  and  $R^2 = 0.99$ , respectively). The R858A and K843A-R844A mutants bind SLA with wild-type affinity, but show a reduced maximal value of the fluorescence signal ( $R^2 = 0.97$  and  $R^2 = 0.98$ , respectively). The R771A-R772A-R775A and R771A-R772A-R775A-K843A-R844A bind Zika SLA with reduced affinities of  $K = 2.0 \times 10^6$  ( $K_d = 0.50 \text{ } \mu\text{M}$ ,  $R^2 = 0.98$ ) and  $1.4 \times 10^6 \text{ M}^{-1}$  ( $K_d = 0.71 \text{ } \mu\text{M}$ ,  $R^2 = 0.98$ ), respectively. Each experiment was carried out in triplicate. Error bars represent the standard deviations. **(d)** CD spectra of wild-type NS5 and mutant R771A-R772A-R775A-K843A-R844A. The CD data indicate that there is no structural change in the NS5 mutant. **(e)** Interactions of Zika NS5 proteins with 20-mer ssRNA. Fluorescence changes of the wild type (black) and mutant NS5 (R771A-R772A-R775A-K843A-R844A, red) upon binding  $\epsilon\text{A}(\text{pEA})_{19}$  ( $\lambda_{\text{ex}} = 325 \text{ nm}$ ;  $\lambda_{\text{em}} = 410 \text{ nm}$ ) were measured in buffer B1. The solid lines are nonlinear least squares fits of the titration curve (Eq. 11) with  $K_{\text{wt}} = 2.0 \times 10^6$  ( $K_d = 0.50 \text{ } \mu\text{M}$ ,  $R^2 = 0.98$ ) and  $K_{\text{mut}} = 1.5 \times 10^6 \text{ M}^{-1}$  ( $K_d = 0.67 \text{ } \mu\text{M}$ ,  $R^2 = 0.98$ ). Each experiment was carried out in triplicate. Error bars represent the standard deviations.

**Mapping the SLA-binding region within the MTase domain.** Our binding data show that the MTase domain alone interacts with SLA (Fig. 1c). Crystal structures of flavivirus MTases co-crystallized with GTP, cap analog GpppA, cofactor S-adenosyl-L-methionine (SAM) or its product S-adenosyl-L-homocysteine (SAH) identified the catalytic site and cofactor binding sites<sup>23,35</sup>. The structures also suggest a putative RNA-binding site, consisting of K28, K29, R37, R41, R42, R57, and R213 (Fig. 4a). Several of these residues (K28, K29, R57, and R213) in DENV MTase have previously been mutated to Ala and their interactions with a short RNA examined<sup>36</sup>. The reduction in RNA binding varied by one to five-fold, with greatest effect observed for the K29A mutation. This putative RNA-binding site with a positively charged patch is located on the same surface as the positively charged patch in the RdRp domain (Fig. 2, right). To determine whether these MTase residues play a role in the NS5-SLA complex formation, we introduced a four-residue mutation, K28E-K29E-R41E-R42E, in the MTase domain construct and measured its interaction with SLA-F by fluorescence. However, the change in fluorescence signal for the MTase mutant was too small to accurately determine the binding constant. To increase the resolution of the analysis, we utilized the anisotropy signal to examine MTase and SLA interactions (see

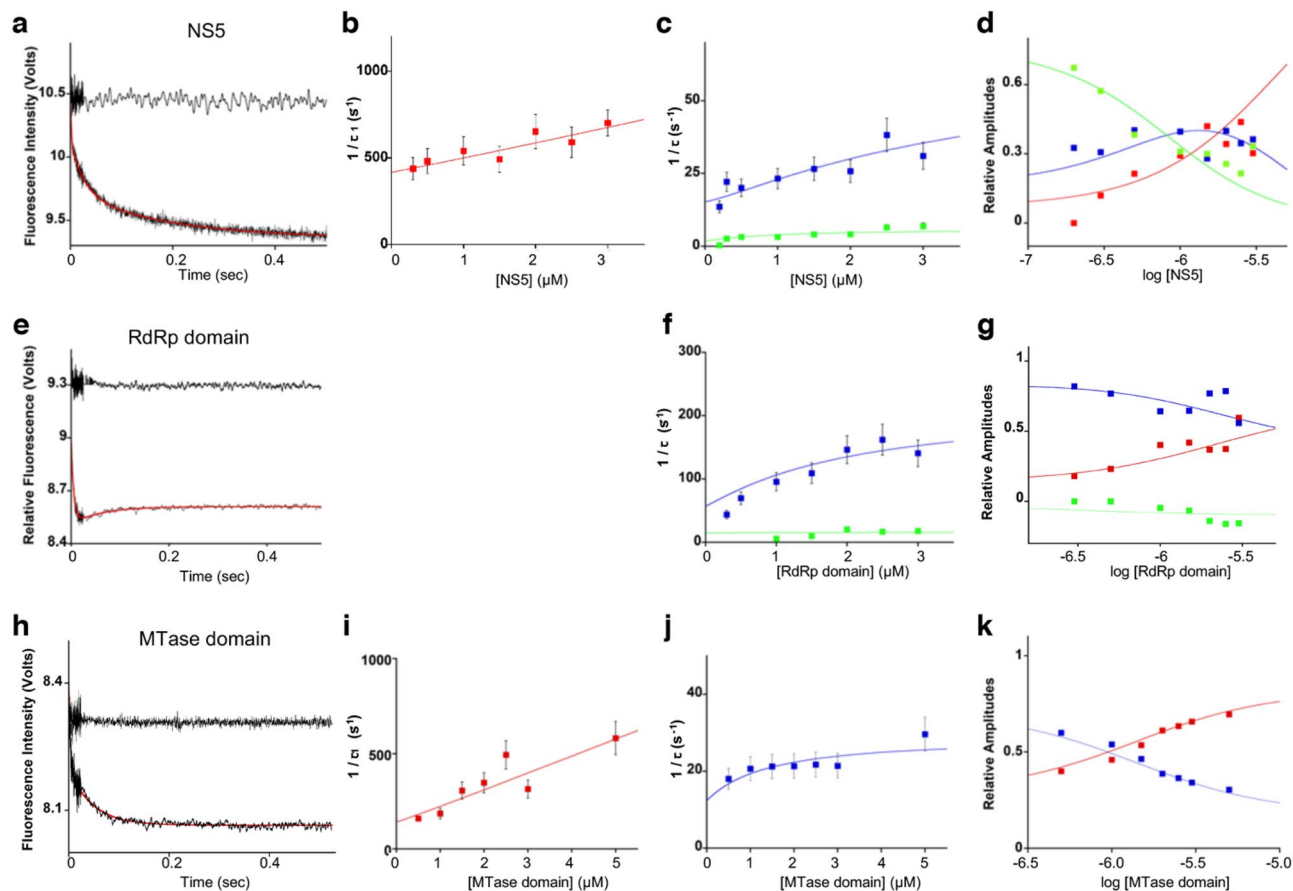


**Figure 4.** Mutations of the MTase domain affect SLA interactions. **(a)** ZIKV MTase structure. The locations of the mutated residues in the ZIKV MTase domain (PDB accession code 5U0B) are shown in a ribbon diagram (left) and an electrostatic surface (right). The coproduct S-adenosyl homocysteine (SAH) and the cap-binding site F24 are also indicated in the ribbon diagram. **(b)** Interactions of Zika SLA with MTase proteins. MTase wild-type (WT, black) and K28E-K29E-R41E-R42E (blue) interactions with fluorescein-labeled SLA in buffer B2 were analyzed by measuring the relative increase of fluorescence anisotropy of the labeled nucleic acid. The solid lines are nonlinear least squares fits of the titration curves with  $K_{\text{wt}} = 3.2 \times 10^6$  ( $K_d = 0.31 \mu\text{M}$ ,  $R^2 = 0.97$ ) and  $K_{\text{mut}} = 0.8 \times 10^6 \text{ M}^{-1}$  ( $K_d = 1.25 \mu\text{M}$ ,  $R^2 = 0.98$ ). The binding constants and relative binding affinities of MTase proteins are listed below. **(c)** Interactions of 20-mer ssRNA with Zika MTase proteins. MTase WT (black) and K28E-K29E-R41E-R42E (red) interactions with  $\epsilon\text{A}(\text{pEA})_{19}$  were measured by fluorescence anisotropy ( $\lambda_{\text{ex}} = 325 \text{ nm}$ ,  $\lambda_{\text{em}} = 410 \text{ nm}$ ). The solid lines are nonlinear least squares fits of the titration curves with  $K_{\text{wt}} = 1.1 \times 10^6$  ( $K_d = 0.91 \mu\text{M}$ ,  $R^2 = 0.98$ ) and  $K_{\text{mut}} = 0.4 \times 10^6 \text{ M}^{-1}$  ( $K_d = 2.5 \mu\text{M}$ ,  $R^2 = 0.98$ ). The binding constants and relative binding affinities of MTase proteins are listed below. Each experiment **(b,c)** was measured in triplicate. Error bars represent the standard deviations. **(d)** Comparison of CD spectra between the wild-type MTase and the mutant K28E-K29E-R41E-R42E. The CD spectrum of the mutant MTase is identical to that of the wild-type MTase. The purity of the mutant MTase is analyzed by SDS-PAGE (right).

“Methods” section). Anisotropy titrations of SLA-F with wild-type and the mutant (K28E-K29E-R41E-R42E) MTase domain in buffer B2 (50 mM Tris at pH 8.0, 100 mM NaCl, 1 mM  $\text{MgCl}_2$ , 2 mM  $\beta$ -mercaptoethanol, and 10% glycerol) at 20 °C are shown in Fig. 4b. The wild-type MTase domain binds SLA with a binding affinity  $K_{\text{wt}} = 3.2 \times 10^6 \text{ M}^{-1}$  ( $K_d = 0.31 \mu\text{M}$ ). This value measured by anisotropy is similar to the binding constant determined by the fluorescence titration in the same buffer B2,  $2.5 \times 10^6 \text{ M}^{-1}$  ( $K_d = 0.40 \mu\text{M}$ ) (Fig. S1). The affinity of the mutant MTase K28E-K29E-R41E-R42E measured by anisotropy was  $K_{\text{mut}} = 0.8 \times 10^6 \text{ M}^{-1}$  ( $K_d = 1.25 \mu\text{M}$ ), i.e., it decreased fourfold compared to the wild-type MTase (Fig. 4b). The reduced SLA-binding affinity of the mutant was not caused by structural changes induced by the mutations, since the comparison of the CD spectra between the wild-type and mutant MTase show identical curves (Fig. 4d). These results indicate that the mutated residues are most likely involved in NS5–SLA interaction.

To determine if the decreased affinity is specific for SLA compared to non-specific ssRNA, the associations of the wild-type and mutant MTase were measured by fluorescence anisotropy with the fluorescently labeled ssRNA,  $\epsilon\text{A}(\text{pEA})_{19}$  in buffer B2 (Fig. 4c). The wild-type MTase binds the ssRNA with  $K_{\text{wt}} = 1.1 \times 10^6 \text{ M}^{-1}$  ( $K_d = 0.91 \mu\text{M}$ ). The binding constant of the mutant MTase K28E-K29E-R41E-R42E for the ssRNA is  $K_{\text{mut}} = 0.4 \times 10^6 \text{ M}^{-1}$  ( $K_d = 2.5 \mu\text{M}$ ), a 2.8-fold reduction compared to the wild-type. The SLA-binding site identified in the MTase (K28, K29, R41, R42) partially overlaps with the capped RNA-binding site (K14, L17, N18, L20, F25, K29) determined by the crystal structures of MTase in complex with capped RNA<sup>35,37</sup>. Thus, it is likely that the SLA-binding site mutant in the MTase domain also reduces the ssRNA interaction. Nevertheless, the mutant MTase (K28E-K29E-R41E-R42E) reduced the SLA binding affinity more significantly than the ssRNA binding affinity (four-fold vs. 2.8-fold), indicating that the positively charged patch containing K28, K29, R41, and R42 is involved in SLA binding.

**Stopped-flow kinetics of ZIKV NS5, RdRp, and MTase binding to SLA.** Next, we examined the kinetic mechanism of ZIKV SLA binding to full-length NS5 and the isolated RdRp and MTase domains by fluorescence stopped-flow methods. All stopped-flow experiments were performed under pseudo-first-order conditions with respect to the protein concentration by mixing SLA-F with a large excess of ZIKV NS5, the RdRp domain or the MTase domain. An example of the stopped-flow kinetic trace of SLA-F fluorescence after mixing with ZIKV NS5 is shown in Fig. 5a. The best fit for the NS5–SLA kinetic traces was obtained using a triple-exponential function, indicating that the simplest mechanism that accounts for the observed change of relaxation times is a minimum three-step binding process. The SLA-F alone showed no fluorescence changes when mixed



**Figure 5.** Dynamics of Zika NS5, RdRp and MTase upon SLA binding. (**a–d**) The kinetic measurement for ZIKV NS5 and SLA interaction. Fluorescence stopped-flow traces after mixing the fluorescein-labeled ZIKV SLA (SLA-F) with NS5 or buffer (top trace) are shown in (**a**). The solid red line is the nonlinear least-squares fit of the experimental curve using Eq. (12) with a three-step relaxation process. The dependence of the reciprocal relaxation times  $1/\tau_1$  (red),  $1/\tau_2$  (blue) and  $1/\tau_3$  (green) upon the total NS5 concentration are shown in (**b,c**). The dependence of the normalized, individual amplitudes of the observed relaxation processes,  $A_1$  (red),  $A_2$  (blue) and  $A_3$  (green) upon the logarithm of the total NS5 concentration are shown in (**d**). The solid lines in (**b–d**) are the global fits of the experimental curves with a three-step relaxation process. (**e–g**) The kinetic measurement for ZIKV RdRp domain and SLA interaction. A fluorescence stopped-flow traces after mixing Zika SLA-F with the RdRp domain or buffer (top trace) are shown in (**e**). The solid red line is the nonlinear least-squares fit of the experimental curve with a two-step relaxation process (Eq. (12)). A fast step beyond the resolution of the stopped-flow instrument precedes the two observed relaxation steps (see text). The dependence of the reciprocal relaxation times  $1/\tau_2$  (blue) and  $1/\tau_3$  (green) upon the total RdRp concentrations are shown in (**f**). The dependence of the normalized relaxation amplitudes  $A_1$  (red),  $A_2$  (blue) and  $A_3$  (green) upon the logarithm of the total RdRp concentrations are shown in (**g**).  $A_1$  was obtained by subtracting  $A_2$  and  $A_3$  from the total amplitude. The solid lines in (**f,g**) are global fits of the curves to the three-step sequential mechanism. (**h–k**) The kinetic measurement for ZIKV MTase domain and SLA interaction. A fluorescence stopped-flow traces after mixing the Zika SLA-F with the MTase domain or buffer (top trace) are shown in (**h**). The solid red line is the nonlinear least-squares fit of the experimental curve with a two-step relaxation process (Eq. (12)). The dependence of reciprocal relaxation times  $1/\tau_1$  (red) and  $1/\tau_2$  (blue) upon the MTase concentration are shown in (**i,j**), respectively. The dependence of amplitudes  $A_1$  (red) and  $A_2$  (blue) upon the logarithm of the MTase concentrations are shown in (**k**). The solid lines in (**i–k**) are the global fits of the curves with a two relaxation-step process.

with the buffer solution without NS5 protein (Fig. 5a, the trace on top). To determine the kinetic mechanism, the stopped-flow experiments were carried out at various NS5 concentrations (0.3–3  $\mu\text{M}$ ) with 15 nM of SLA-F. The kinetic traces were then fit using a triple-exponential function to determine the reciprocal relaxation times ( $\lambda_i$  or  $1/\tau_i$ ,  $i = 1, 2, 3$ ) and the fractional individual amplitudes ( $A_i$ ,  $i = 1, 2, 3$ ) of the NS5–SLA complex formation with nonlinear least-squares fits (Eq. (12) in “Methods” section). The dependencies of the reciprocal relaxation times,  $1/\tau_1$ ,  $1/\tau_2$ , and  $1/\tau_3$  plotted as a function of the NS5 protein concentrations are shown in Fig. 5b and c. The plot of the reciprocal relaxation time of the first exponential phase,  $1/\tau_1$  versus NS5 concentrations showed a linear dependence, indicating a bimolecular association (Fig. 5b). The reciprocal relaxation times  $1/\tau_2$  and  $1/\tau_3$  increased in a hyperbolic manner with NS5 concentrations, suggesting intramolecular transitions of the NS5–SLA complex (i.e., conformational changes) in these steps (Fig. 5c). The dependencies of the individual

Parameter <sup>a</sup>	NS5	RdRp domain	MTase domain
$K_{\text{overall}} (M^{-1})$	$3.5 (\pm 0.6) \times 10^6$	$1.9 (\pm 0.3) \times 10^6$	$1.7 (\pm 0.3) \times 10^6$
$\Delta F_{\text{max}}$	0.15	0.09	0.07
$K_1 (M^{-1})$	$2.64 (\pm 0.40) \times 10^5$	$5.0 (\pm 1.0) \times 10^5$	$7.2 (\pm 1.4) \times 10^5$
$k_1 (M^{-1} s^{-1})$	$9.5 (\pm 1.8) \times 10^7$	$1.0 (\pm 0.2) \times 10^9$	$9.0 (\pm 1.8) \times 10^7$
$k_{-1} (s^{-1})$	$360 \pm 60$	$2000 \pm 300$	$125 \pm 25$
$K_2$	$4.6 \pm 0.5$	$2.0 \pm 0.5$	$1.07 \pm 0.25$
$k_2 (s^{-1})$	$55 \pm 10$	$140 \pm 50$	$15 \pm 3$
$k_{-2} (s^{-1})$	$12.0 \pm 1.8$	$70 \pm 9$	$14.0 \pm 2.1$
$K_3$	$1.6 \pm 0.4$	$0.07 \pm 0.02$	
$k_3 (s^{-1})$	$4.0 \pm 0.6$	$1.0 \pm 0.1$	
$k_{-3} (s^{-1})$	$2.5 \pm 0.4$	$15.0 \pm 2.2$	

**Table 2.** Thermodynamic and kinetic parameters characterizing Zika NS5, the RdRp domain and the MTase domain interactions with Zika SLA. <sup>a</sup>The  $K_{\text{overall}}$  and  $\Delta F_{\text{max}}$  values were determined from the independent fluorescence binding experiments.

amplitudes,  $A_1$ ,  $A_2$ , and  $A_3$ , on the NS5 concentration are shown in Fig. 5d. The individual amplitudes are normalized and expressed as fractions of the total amplitude,  $A_i/\Sigma A_i$ . At low NS5 concentrations, the total observed amplitude,  $A_T$ , is dominated by  $A_3$ . With increasing protein concentration, contributions of  $A_1$  and  $A_2$  increase, with  $A_1$  becoming dominant at higher protein concentrations.

To globally fit the kinetic data at various NS5 concentrations and obtain rate constants, the stopped-flow kinetic data were also analyzed using the matrix projection operator technique<sup>38,39</sup>. The three-step binding process is described by Eq. (1):

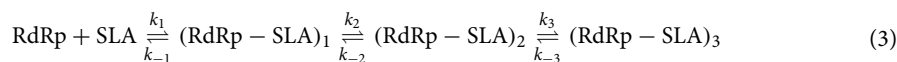


The partial equilibrium constants for each step of the above reaction are related to the  $K_{\text{overall}}$  as,

$$K_{\text{overall}} = K_1(1 + K_2 + K_2K_3) \quad (2)$$

where  $K_1 = k_1/k_{-1}$ ,  $K_2 = k_2/k_{-2}$ , and  $K_3 = k_3/k_{-3}$ . The six kinetic rate constants for the three kinetic steps, as well as fluorescence changes characterizing individual intermediates, are related through the overall binding constants and the maximum fluorescence change  $\Delta F_{\text{max}}$ . Using the overall binding constants ( $K_{\text{overall}} = 3.5 \times 10^6 M^{-1}$ ) and the maximum fluorescence change ( $\Delta F_{\text{max}} = 0.15$ ) that are independently determined (Fig. 1c), partial equilibrium constants and rate constants were determined (Table 2). The rate constants indicate that the formation of the NS5-SLA complex proceeds first by the fast bimolecular association of NS5 and SLA ( $k_1 = 9.5 \times 10^7 M^{-1} s^{-1}$  and  $k_{-1} = 360 s^{-1}$ ) followed by two structural rearrangement steps ( $k_2 = 55 s^{-1}$ ,  $k_{-2} = 12 s^{-1}$  and  $k_3 = 4 s^{-1}$ ,  $k_{-3} = 2.5 s^{-1}$ ).

The stopped-flow kinetic traces of SLA-F fluorescence after mixing with ZIKV RdRp were analogously analyzed. The best fit for the RdRp-SLA kinetic traces were obtained using a double-exponential function (Fig. 5e). However, the binding process is likely more complex because the sum of the amplitudes of the fitted relaxation steps is significantly smaller than the observed total amplitude, i.e., the fluorescence of free SLA-F is significantly higher than the fluorescence of the first observed data point (Fig. 5e). Thus, there is a fast step which is beyond the resolution of the stopped-flow instrument, which precedes the two observed relaxation steps. Therefore, in the case of the RdRp-SLA interaction, the simplest minimum mechanism is a three-step binding process, where initial recognition occurs through very fast bimolecular association of SLA and RdRp.



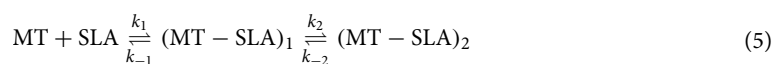
Even though we cannot determine the value of the reciprocal relaxation time for the first fast relaxation process,  $1/\tau_1$ , we can determine its amplitude,  $A_1$  by Eq. (4), where  $A_T$  is the experimentally observed total amplitude of the kinetic trace at a given total concentration of the protein<sup>38</sup>.

$$A_1 = A_T - A_2 - A_3 \quad (4)$$

The dependences of the reciprocal relaxation times ( $1/\tau_2$  and  $1/\tau_3$ ) and normalized amplitudes,  $A_1$ ,  $A_2$ , and  $A_3$  of three observed relaxation processes on the concentration of the ZIKV RdRp domain are shown in Fig. 5f and g. At low RdRp domain concentrations, the total observed amplitude,  $A_T$ , is dominated by  $A_2$ . With increasing protein concentrations, the contribution of  $A_1$  to  $A_T$  increases while  $A_3$  remains negative throughout the process. The partial equilibrium constants and kinetic rate constants characterizing the intermediates for the three steps are related to the overall binding constant  $K_{\text{overall}}$  and the maximum fluorescence change  $\Delta F_{\text{max}}$ , as described above (Table 2). In the RdRp-SLA complex formation, the initial recognition occurs through a very fast bimolecular association of SLA and RdRp ( $k_1 = 1.0 \times 10^9 M^{-1} s^{-1}$  and  $k_{-1} = 2000 s^{-1}$ ), which is followed by two structural rearrangement steps ( $k_2 = 140 s^{-1}$ ,  $k_{-2} = 70 s^{-1}$  and  $k_3 = 1.0 s^{-1}$ ,  $k_{-3} = 15.0 s^{-1}$ ). Note the initial bimolecular association and dissociation ( $k_1$  and  $k_{-1}$ ) of RdRp and SLA are 10 times faster than the initial association of NS5 and SLA.



Lastly, a stopped-flow kinetic trace for the MTase domain and SLA interaction is shown in Fig. 5h. The solid red line is a nonlinear least-squares fit of the data using a double-exponential function, indicating that the simplest mechanism that can account for the data is a two-step binding process (Eq. 5).



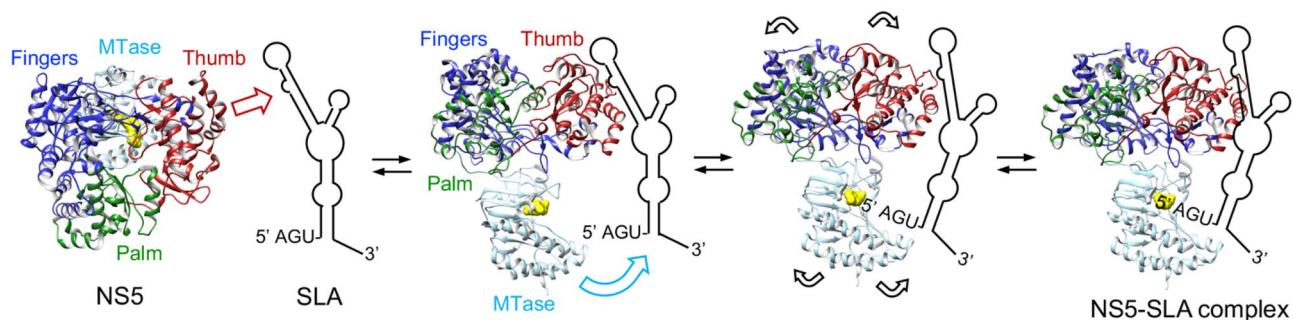
The dependencies of the reciprocal relaxation time  $1/\tau_1$  and  $1/\tau_2$ , characterizing the observed kinetic steps as a function of the total MTase domain concentration are shown in Fig. 5i and j, respectively. The plot of the reciprocal relaxation time of the first exponential phase ( $1/\tau_1$ ) versus MTase concentrations showed a linear dependence, indicating a MTase–SLA complex is formed after the bimolecular event (Fig. 5i). The dependence of the normalized amplitudes,  $A_1$  and  $A_2$  upon the total concentration of the ZIKV MTase domain is shown in Fig. 5k. The global fit of the kinetic data at various MTase concentrations provides rate constants for the two steps (Table 2). The kinetic constants indicate that a fast bimolecular association of SLA and the MTase domain ( $k_1 = 9.0 \times 10^7 \text{ M}^{-1} \text{ s}^{-1}$  and  $k_{-1} = 125 \text{ s}^{-1}$ ) is followed by only one structural rearrangement step ( $k_2 = 15 \text{ s}^{-1}$  and  $k_{-2} = 14 \text{ s}^{-1}$ ) (Table 2).

## Discussion

**The SLA-binding site on NS5 is located in the thumb and MTase domains.** Replication of the viral genome is a complex process that involves the viral RNA and polymerase, as well as other viral and cellular proteins. In the case of flaviviruses, the SLA at the 5' end of viral RNA plays a critical role by functioning as a promoter of the replication process<sup>10–12</sup>. In order to understand how ZIKV NS5 recognizes the SLA promoter for initiation of RNA synthesis, we examined the roles that each of the NS5 domains play in binding to SLA. Both RdRp and MTase domains bind to SLA. Moreover, the SLA-binding affinities for both domains were similar to each other and were approximately two-fold lower than the affinity of the full-length NS5 for SLA. This indicates that ZIKV NS5 recognizes SLA using both RdRp and MTase domains. We have thus identified the SLA-binding sites within the RdRp and MTase domains. In the RdRp domain, only the thumb subdomain mutants show reduced SLA binding. Alanine substitutions of R771, R772 and R775, and additional K843 and R844 substitutions in ZIKV NS5 significantly reduced the SLA binding affinity from 280 nM (wild-type NS5) to 500 and 710 nM, respectively (Fig. 3c and Table 1). This suggests that a large positively charged patch consisting of R771, R772, R775, K843 and R844 in the thumb subdomain engages SLA and mediates the formation of the NS5–SLA complex. Interestingly, the single NS5 mutations (R775A and R891A) or a double mutation (K843A–R844A) did not significantly affect the ZIKV NS5–SLA interactions (Table 1), suggesting that the SLA-binding surface on NS5 encompasses a cluster of residues, and other positively charged residues within the patch can compensate for the loss of one or two positively charged residues. The SLA-binding site identified here does not overlap with the template-binding site or the dsRNA exit site in the RdRp domain, suggesting that NS5 has two RNA binding sites, one for ssRNA template, and the other for SLA. Consistent with this idea, wild-type and the SLA-binding site mutant NS5 (R771A–R772A–R775A–K843A–R844A) bind ssRNA with similar affinities (Fig. 3e). In the MTase domain, the mutations of K28E–K29E–R41E–R42E reduced SLA binding affinity from 310 nM (wild-type MTase) to 1.25  $\mu\text{M}$ , suggesting that this four-residue mutation within the MTase domain eliminates specificity for SLA. An NS5 construct containing mutations in both RdRp and MTase patches was expressed, but the protein was insoluble upon *E. coli* expression preventing further investigations. The identified SLA-binding residues in MTase and RdRp domains are highly conserved in flavivirus NS5 (Fig. S2), and residues in the thumb subdomain have previously been shown to be important for viral replication. Individual mutations of R770A and R773A, and a double mutation of K840A–R841A in a DENV2 replicon (corresponding to R772, R775, K843 and R844 in ZIKV NS5) delayed or abolished viral replication<sup>32</sup>. These residues are not directly involved in the catalysis of RNA synthesis, and thus these mutants may abolish viral replication by diminishing interactions with SLA.

The two positively charged patches in the thumb subdomain of RdRp and MTase domains are located on the same surface of NS5, strongly suggesting that the SLA-binding site encompasses both domains. During the viral replication cycle, NS5 interacts with SLA at least twice. First, NS5 must recognize the SLA at the 5' end of the genome to initiate negative-strand RNA synthesis at the 3' end (i.e. to carry out the polymerase function of NS5)<sup>10–12</sup>. Second, NS5 must again bind SLA to methylate the cap at the 5' end of the nascent positive-strand RNA (i.e., to carry out the MTase function of NS5)<sup>29</sup>. These two functionally different NS5 and SLA interactions would engage the same binding site, since it seems unlikely that NS5 possesses two separate SLA-binding sites. In such a case, both MTase and RdRp domains are likely involved in SLA binding. The positively charged patches in RdRp and MTase are separated by  $\sim 55$ – $60$  Å. Based on the secondary structure prediction (Fig. 1a), ZIKV SLA is a long molecule consisting of at least 23 base pairs ( $\sim 60$  Å) without counting the bulge between the top and bottom stems or the top loop. Thus, SLA could simultaneously bind both the RdRp and MTase patches.

**Zika NS5 initially recognizes SLA via the RdRp domain.** We have examined the dynamics of the NS5–SLA interaction using stopped-flow techniques. The NS5 recognition of SLA occurs through a three-step mechanism. The first step is a fast, bimolecular association of NS5 and SLA, which is followed by two steps in which structural rearrangements of the NS5–SLA complex occur (Fig. 6). Interestingly, the RdRp domain interactions with SLA also occur through a three-step process, but the first step is significantly faster than for the full-length NS5 ( $k_{1-\text{RdRp}} = 1.0 \times 10^9 \text{ M}^{-1} \text{ s}^{-1}$  vs  $k_{1-\text{NS5}} = 9.5 \times 10^7 \text{ M}^{-1} \text{ s}^{-1}$ ). On the other hand, the RdRp has a large reverse rate constant ( $k_{-1-\text{RdRp}} = 2000 \text{ s}^{-1}$ ), indicating that the RdRp–SLA complex dissociates more easily than the NS5–SLA complex where the reverse rate constant is much lower ( $k_{-1-\text{NS5}} = 360 \text{ s}^{-1}$ ). This indicates that the MTase domain in the intact NS5 slows the recognition step but increases the stability of the formed complex.



**Figure 6.** Model of Zika NS5–SLA recognition and complex formation. Zika NS5 recognizes SLA via the thumb subdomain of the RdRp domain. Following the fast binding of the RdRp domain and SLA, two additional conformational changes that involve both RdRp and MTase domains occur. Kinetic measurements suggest that the MTase domain is not involved in the initial SLA recognition but helps to stabilize the NS5–SLA complex. The RdRp and MTase domains are colored as in Fig. 2. The active site in MTase domain is indicated by the bound S-adenosyl homo cysteine (yellow), the byproduct of the MTase reaction. The SLA molecule is oriented based on the requirement of the 5' cap for MTase activity and previous binding data. The 5' terminal AGU is required to bind the MTase domain<sup>35</sup>, and thus positioned near the MTase active site. The top stem-loop of SLA is oriented toward the RdRp domain because the top stem-loop is suggested to bind the RdRp domain in the RNase protection assay<sup>10</sup>.

Accordingly, binding of the MTase domain to SLA is characterized by only a two-step mechanism and the initial recognition step is much slower than the RdRp domain ( $k_{1-MTase} = 9.0 \times 10^7 \text{ M}^{-1} \text{ s}^{-1}$  vs.  $k_{1-RdRp} = 1.0 \times 10^9 \text{ M}^{-1} \text{ s}^{-1}$ ). Therefore, NS5 likely recognizes SLA first through the RdRp domain. Taken together, we propose an NS5–SLA interaction model wherein the first fast bimolecular association between NS5 and SLA is carried out via the RdRp domain involving the thumb subdomain (Fig. 6). This fast recognition step is then followed by two steps corresponding to two conformational changes, during which NS5 locks itself onto the SLA through the RdRp and MTase domain to form a stable NS5–SLA complex. Crystal structures of flavivirus NS5s show two different relative arrangements of the RdRp and MTase domains, suggesting that interaction between the RdRp and MTase domains is not fixed spatially<sup>16–20,40</sup>. Thus, it is likely that NS5 adopts several conformations by rearrangement of RdRp and MTase domains to accommodate different RNA molecules such as structured SLA, ssRNA, and dsRNA during viral replication.

## Methods

**Reagents and buffers.** A Zika SLA construct containing the first 80 nucleotides with 5' triphosphate and a 3' fluorescein tag was obtained from Midland Certified Reagents (Midland, TX). SLA concentration was determined using the extinction coefficient of SLA  $\epsilon_{260} = 682,500 \text{ cm}^{-1} \text{ M}^{-1}$ . The etheno-modified A(pA)<sub>19</sub>,  $\epsilon\text{A}(\text{p}\epsilon\text{A})_{19}$  was obtained by modification of the oligomer with chloroacetaldehyde as described previously<sup>31</sup>. The concentration of  $\epsilon\text{A}(\text{p}\epsilon\text{A})_{19}$  was determined using the extinction coefficient,  $\epsilon_{257} = 3,700 \text{ cm}^{-1} \text{ M}^{-1}$  (nucleotide)<sup>31</sup>.

**Expression and purification of Zika NS5.** Full-length ZIKV NS5 (GenBank: KY559015.1) and its RdRp domain constructs were obtained from Dr. Cheng Kao (Indiana University)<sup>19</sup>. Full-length ZIKV NS5 and the RdRp domain were expressed with an N-terminal hexahistidine tag and a Small Ubiquitin-like Modifier (SUMO) tag. The gene encoding the MTase domain (GenBank: KU955593.1) was synthesized with an N-terminal hexahistidine tag. All three proteins, the full-length NS5, the RdRp and the MTase domain, were expressed in Rosetta (DE3) *Escherichia coli* cells. Briefly, the cells were grown at 37 °C in Luria Broth (LB) containing 100 µg/ml ampicillin to an OD<sub>600</sub> of 0.6–0.7. Protein expression was induced by the addition of 1 mM isopropyl β-D-1-thiogalactopyranoside and cell growth continued overnight at 18 °C. For protein purification, the cells were resuspended in lysis buffer containing 50 mM Tris (pH 8.0), 500 mM NaCl, 2 mM β-mercaptoethanol, and 1 tablet of EDTA-free protease inhibitor (Roche Applied Science, Penzberg, Germany), and lysed by sonication. After centrifugation to clear the insoluble debris, the protein in the soluble fraction was purified by Talon metal affinity chromatography (Clontech, Mountain View, CA). NS5 was eluted with a 10–200 mM gradient of imidazole in 50 mM Tris (pH 7.0) buffer containing 500 mM NaCl, 2 mM β-mercaptoethanol and 10% glycerol. The fractions containing NS5 protein were collected and concentrated using an Ultrafree centrifugal filter device (Millipore, Billerica, MA). For the full-length NS5 and RdRp domain, the SUMO and hexahistidine tag was cleaved from the protein by incubation with SUMO protease (1:100 w/w ratio). Subsequently, the cleaved fractions were loaded on a Superdex 200 size exclusion column (GE Healthcare Life Sciences, Marlborough, MA), equilibrated with buffer containing 20 mM Tris–HCl (pH 7.0), 500 mM NaCl, 2 mM β-mercaptoethanol and 20% glycerol. The MTase domain was purified similarly using Talon metal affinity and size-exclusion chromatography.

Mutations within the ZIKV NS5 and the MTase domain were introduced by site-directed mutagenesis using the PfuUltra II Fusion HS DNA Polymerase kit (Agilent Technology, CA). The primers used for mutagenesis are listed in Table S1. Mutations were confirmed by DNA sequencing at the molecular genomics facility at UTMB. Mutant proteins were purified similarly to wild-type proteins. Purified wild-type and mutant proteins were analyzed by SDS-PAGE (Figs. 1b, 3a, 4d). Uncropped images of SDS-PAGE gels are shown in Fig. S3.

**Fluorescence measurements.** Steady-state fluorescence titrations were performed using an ISS PC1 spectrofluorometer (ISS, Urbana, IL). Polarizers were placed in excitation and emission channels and set at 90° and 55° (magic angle), respectively, in order to avoid possible artifacts due to fluorescence anisotropy of the sample. Experiments were carried out by adding Zika NS5 proteins or the RdRp/MTase domain to a solution containing fluorescein-labeled Zika SLA. A standard buffer B1, containing 50 mM Tris adjusted to pH 8.0 with HCl at 20 °C, 150 mM NaCl, 1 mM MgCl<sub>2</sub>, 2 mM β-mercaptoethanol, and 10% glycerol, was used in fluorescence experiments. Each experiment was carried out in triplicate. The binding of ZIKV NS5 to SLA was monitored by measuring changes in the signal originating from the fluorescein attached to SLA ( $\lambda_{\text{ex}} = 480$  nm,  $\lambda_{\text{em}} = 520$  nm). Titration curves were fit using Kaleida Graph software (Synergy Software, PA). In these experiments, the relative fluorescence change ( $\Delta F_{\text{obs}}$ ) is defined as  $(F_i - F_0)/F_0$ , where  $F_i$  is the fluorescence of the RNA at a given titration point and  $F_0$  is the initial value of the fluorescence of the sample. Both  $F_0$  and  $F_i$  were corrected for background fluorescence at the applied excitation wavelength. Interactions between the NS5 proteins (wild-type and R771A-R772A-R775A-K843A-R844A) and ssRNA were measured with etheno-labeled  $\epsilon\text{A}(\text{p}\epsilon\text{A})_{19}$  in buffer B1 (50 mM Tris adjusted to pH 8.0 with HCl at 20 °C, 150 mM NaCl, 1 mM MgCl<sub>2</sub>, 2 mM β-mercaptoethanol, and 10% glycerol). The  $\epsilon\text{A}(\text{p}\epsilon\text{A})_{19}$  samples were excited at  $\lambda_{\text{ex}} = 325$  nm and emission was recorded at  $\lambda_{\text{em}} = 410$  nm<sup>31</sup>.

Interactions between the MTase domain proteins (wild-type and K28E-K29E-R41E-R42E) and fluorescein-labeled SLA were analyzed using fluorescence anisotropy ( $r$ ), because fluorescence intensity changes were too small to accurately determine the binding constant for the mutant. Buffer B2, containing 50 mM Tris adjusted to pH 8.0 with HCl at 20 °C, 100 mM NaCl, 1 mM MgCl<sub>2</sub>, 2 mM β-mercaptoethanol, and 10% glycerol, was used for anisotropy measurement. Fluorescence anisotropy ( $r$ ) is defined as:

$$r = \frac{I_{VV} - GI_{VH}}{I_{VV} + 2GI_{VH}} \quad (6)$$

where  $I_{VV}$  and  $I_{VH}$  are fluorescence intensities where the first and second subscripts refer to vertical (V) polarization of the excitation and vertical (V) or horizontal (H) polarization of the emitted light. The factor,  $G = I_{HV}/I_{HH}$ , corrects for the different sensitivity of the emission monochromator for vertically and horizontally polarized light. The samples in the fluorescence anisotropy experiments were excited at 480 nm and the emission recorded at 520 nm (fluorescein peaks for excitation and emission, respectively). Each experiment was carried out in triplicate. Interactions between the MTase domain proteins (wild-type and K28E-K29E-R41E-R42E) and etheno-labeled  $\epsilon\text{A}(\text{p}\epsilon\text{A})_{19}$  were measured in buffer B2 using fluorescence anisotropy. The fluorescence signal of  $\epsilon\text{A}(\text{p}\epsilon\text{A})_{19}$  were monitored with  $\lambda_{\text{ex}} = 325$  nm and  $\lambda_{\text{em}} = 410$  nm<sup>31</sup>. Each experiment was carried out in triplicate.

**Fluorescence signal analysis.** The binding constant,  $K_1$ , characterizing the SLA association with NS5, is defined as

$$K_1 = \frac{[C_1]_F}{[SLA]_F [NS5]_F} \quad (7)$$

where  $[C_1]_F$  is the concentration of the formed complex.

The observed fluorescence of the sample at any point of the titration is defined as

$$F_{\text{obs}} = F_F [SLA]_F + F_C [C_1]_F \quad (8)$$

where  $F_F$  and  $F_C$  are the molar fluorescence intensities of the free SLA and the formed complex  $C_1$ , respectively. Thus,

$$F_{\text{obs}} = F_F [SLA]_F + F_C K_1 [SLA]_F [NS5]_F \quad (9)$$

The mass conservation equation for the total SLA concentration,  $[SLA]_T$ , in the sample is

$$[SLA]_T = [SLA]_F + [C_1]_F = [SLA]_F (1 + K_1 [NS5]_F) \quad (10)$$

Using Eqs. (9) and (10), the relative observed change of the SLA fluorescence,  $\Delta F_{\text{obs}}$ , is then

$$\Delta F_{\text{obs}} = \frac{F_{\text{obs}}}{F_F [SLA]_T} = \frac{1}{1 + K_1 [NS5]_F} + \Delta F_{\text{max}} \left[ \frac{K_1 [NS5]_F}{1 + K_1 [NS5]_F} \right] \quad (11)$$

where  $\Delta F_{\text{max}} = F_C/F_F$  is the maximum value of the observed relative fluorescence quenching. Fluorescence titration curves were fitted using Eq. (11) to determine binding constants. The goodness of the fits were determined using standard chi-square test, which provides the confidence level for the selected binding model of  $\geq 0.98$ .

**Stopped-flow kinetics.** All fluorescence stopped-flow kinetic experiments were performed using SX.MV20 stopped-flow instrument (Applied Photophysics Ltd. Leatherhead, UK). Fluorescein-labeled SLA (15 nM) was mixed with 7–8 concentrations of NS5 (300 nM–3 μM), RdRp (300 nM–3 μM) or MTase (500 nM–5 μM) under pseudo-first-order conditions ( $[\text{protein}] \gg [\text{RNA}]$ ), and the fluorescence signal upon complex formation was monitored as a function of time with  $\lambda_{\text{ex}} = 480$  nm and  $\lambda_{\text{em}} = 520$  nm. Generally, 10–12 traces were collected and averaged for each concentration of proteins. The kinetic curves were fitted to extract relaxation times and corresponding amplitudes using nonlinear least-squares fitting software provided by the manufacturer, with the exponential function defined as

$$F(t) = F(\infty) + \sum_{i=1}^n A_i \exp(-\lambda_i t) \quad (12)$$

where  $F(t)$  is the fluorescence intensity at time  $t$ ,  $F(\infty)$  is the fluorescence intensity at  $t = \infty$ ,  $A_i$  is the amplitude corresponding to  $i$ th relaxation process,  $\lambda_i$  is the time constant (reciprocal relaxation time) characterizing the relaxation process, and  $n$  is the number of relaxation processes. Global analyses of the stopped-flow kinetic data were performed using the matrix projection operator technique<sup>38,39</sup>. All analyses of the data were performed using Mathematica (Wolfram, Urbana, IL) and Kaleida Graph (Synergy Software, PA) as previously described<sup>38,39</sup>.

**Circular dichroism (CD) spectrometry.** The far-UV circular-dichroism measurements for wild-type and mutant proteins were performed on a model No. J-815 spectrometer (JASCO, Oklahoma City, OK). The protein concentration was ~0.1 mg/ml in 50 mM Tris buffer (pH 7.4) containing 150 mM NaCl, 10% glycerol and 1 mM  $MgCl_2$ . CD spectra were collected at 20 °C in a 0.2-cm path-length cuvette from 195 to 260 nm. Three scans were averaged in the range of 200–250 nm. Data were converted from millidegrees to mean residue ellipticity (MRE) using the formula:

$$[\theta] = (\theta \times 10^6)/(c \times l \times n)$$

where  $[\theta]$  is the molar residue ellipticity,  $\theta$  is the CD signal in millidegrees,  $c$  is the protein concentration in  $\mu$ M,  $l$  is the path length in mm, and  $n$  is the number of amino acid residues. The CD data analysis program BeStSel was used to analyze the data and estimate the secondary structure<sup>41</sup>.

Received: 16 November 2019; Accepted: 3 July 2020

Published online: 06 August 2020

## References

- Weaver, S. C. *et al.* Zika virus: history, emergence, biology, and prospects for control. *Antiv. Res.* **130**, 69–80. <https://doi.org/10.1016/j.antiviral.2016.03.010> (2016).
- Petersen, E. *et al.* Rapid spread of Zika virus in the Americas-implications for public health preparedness for mass gatherings at the 2016 Brazil Olympic games. *Int J Infect Dis* **44**, 11–15. <https://doi.org/10.1016/j.ijid.2016.02.001> (2016).
- Petersen, L. R., Jamieson, D. J. & Honein, M. A. Zika virus. *N Engl J Med* **375**, 294–295. <https://doi.org/10.1056/NEJMc1606769> (2016).
- Chambers, T. J., Hahn, C. S., Galler, R. & Rice, C. M. Flavivirus genome organization, expression, and replication. *Annu. Rev. Microbiol.* **44**, 649–688. <https://doi.org/10.1146/annurev.mi.44.100190.003245> (1990).
- Westaway, E. G. Flavivirus replication strategy. *Adv. Virus Res.* **33**, 45–90 (1987).
- Malakar, J. *et al.* Zika virus associated with microcephaly. *N. Engl. J. Med.* **374**, 951–958. <https://doi.org/10.1056/NEJMoa1600651> (2016).
- Driggers, R. W. *et al.* Zika virus infection with prolonged maternal viremia and fetal brain abnormalities. *N. Engl. J. Med.* **374**, 2142–2151. <https://doi.org/10.1056/NEJMoa1601824> (2016).
- Musso, D., Nilles, E. J. & Cao-Lormeau, V. M. Rapid spread of emerging Zika virus in the Pacific area. *Clin. Microbiol. Infect.* **20**, O595–596. <https://doi.org/10.1111/1469-0691.12707> (2014).
- Oehler, E. *et al.* Zika virus infection complicated by Guillain–Barre syndrome—case report, French Polynesia, December 2013. *Eur. Commun. Dis. Bull.* **19**, 20720 (2014).
- Filomatori, C. V., Iglesias, N. G., Villordo, S. M., Alvarez, D. E. & Gamarnik, A. V. RNA sequences and structures required for the recruitment and activity of the dengue virus polymerase. *J. Biol. Chem.* **286**, 6929–6939. <https://doi.org/10.1074/jbc.M110.162289> (2011).
- Filomatori, C. V. *et al.* A 5' RNA element promotes dengue virus RNA synthesis on a circular genome. *Genes Dev.* **20**, 2238–2249. <https://doi.org/10.1101/gad.1444206> (2006).
- Gebhard, L. G., Filomatori, C. V. & Gamarnik, A. V. Functional RNA elements in the dengue virus genome. *Viruses* **3**, 1739–1756. <https://doi.org/10.3390/v3091739> (2011).
- Bartelma, G. & Padmanabhan, R. Expression, purification, and characterization of the RNA 5'-triphosphatase activity of dengue virus type 2 nonstructural protein 3. *Virology* **299**, 122–132 (2002).
- Ferrer-Orta, C., Arias, A., Escarmis, C. & Verdaguer, N. A comparison of viral RNA-dependent RNA polymerases. *Curr. Opin. Struct. Biol.* **16**, 27–34. <https://doi.org/10.1016/j.sbi.2005.12.002> (2006).
- Choi, K. H. & Rossmann, M. G. RNA-dependent RNA polymerases from flaviviridae. *Curr. Opin. Struct. Biol.* **19**, 746–751. <https://doi.org/10.1016/j.sbi.2009.10.015> (2009).
- Klema, V. J. *et al.* Dengue virus nonstructural protein 5 (NS5) assembles into a dimer with a unique methyltransferase and polymerase interface. *PLoS Pathog.* **12**, e1005451. <https://doi.org/10.1371/journal.ppat.1005451> (2016).
- Lu, G. & Gong, P. Crystal structure of the full-length Japanese encephalitis virus NS5 reveals a conserved methyltransferase-polymerase interface. *PLoS Pathog.* **9**, e1003549. <https://doi.org/10.1371/journal.ppat.1003549> (2013).
- Zhao, Y. *et al.* A crystal structure of the dengue virus NS5 protein reveals a novel inter-domain interface essential for protein flexibility and virus replication. *PLoS Pathog.* **11**, e1004682. <https://doi.org/10.1371/journal.ppat.1004682> (2015).
- Zhao, B. *et al.* Structure and function of the Zika virus full-length NS5 protein. *Nat. Commun.* **8**, 14762. <https://doi.org/10.1038/ncomms14762> (2017).
- Upadhyay, A. K. *et al.* Crystal structure of full-length Zika virus NS5 protein reveals a conformation similar to Japanese encephalitis virus NS5. *Acta Crystallogr. Sect. F Struct. Biol. Commun.* **73**, 116–122. <https://doi.org/10.1107/S2053230X17001601> (2017).
- Klema, V. J., Padmanabhan, R. & Choi, K. H. Flaviviral replication complex: coordination between RNA synthesis and 5'-RNA capping. *Viruses* **7**, 4640–4656. <https://doi.org/10.3390/v7082837> (2015).
- Issur, M., Despins, S., Bougie, I. & Bisaillon, M. Nucleotide analogs and molecular modeling studies reveal key interactions involved in substrate recognition by the yeast RNA triphosphatase. *Nucleic Acids Res.* **37**, 3714–3722. <https://doi.org/10.1093/nar/gkp227> (2009).
- Egloff, M. P., Benarroch, D., Selisko, B., Romette, J. L. & Canard, B. An RNA cap (nucleoside-2'-O-)-methyltransferase in the flavivirus RNA polymerase NS5: crystal structure and functional characterization. *EMBO J.* **21**, 2757–2768. <https://doi.org/10.1093/emboj/21.11.2757> (2002).

24. Brinton, M. A. Replication cycle and molecular biology of the West Nile virus. *Viruses* **6**, 13–53. <https://doi.org/10.3390/v6010013> (2014).
25. Li, X. F. *et al.* RNA elements within the 5' untranslated region of the West Nile virus genome are critical for RNA synthesis and virus replication. *J. Gen. Virol.* **91**, 1218–1223. <https://doi.org/10.1099/vir.0.013854-0> (2010).
26. Dong, H., Zhang, B. & Shi, P. Y. Terminal structures of West Nile virus genomic RNA and their interactions with viral NS5 protein. *Virology* **381**, 123–135. <https://doi.org/10.1016/j.virol.2008.07.040> (2008).
27. Brinton, M. A. & Dispoto, J. H. Sequence and secondary structure analysis of the 5'-terminal region of flavivirus genome RNA. *Virology* **162**, 290–299. [https://doi.org/10.1016/0042-6822\(88\)90468-0](https://doi.org/10.1016/0042-6822(88)90468-0) (1988).
28. Liu, Z. Y. *et al.* Characterization of cis-acting RNA elements of Zika virus by using a self-splicing ribozyme-dependent infectious clone. *J. Virol.* **91**, 15–20. <https://doi.org/10.1128/JVI.00484-17> (2017).
29. Dong, H. *et al.* Distinct RNA elements confer specificity to flavivirus RNA cap methylation events. *J. Virol.* **81**, 4412–4421. <https://doi.org/10.1128/JVI.02455-06> (2007).
30. Bujalowski, P. J., Bujalowski, W. & Choi, K. H. Interactions between the dengue virus polymerase NS5 and stem-loop A. *J. Virol.* **91**, e00047-00017. <https://doi.org/10.1128/JVI.00047-17> (2017).
31. Szymanski, M. R. *et al.* Full-length dengue virus RNA-dependent RNA polymerase-RNA/DNA complexes: stoichiometries, intrinsic affinities, cooperativities, base, and conformational specificities. *J. Biol. Chem.* **286**, 33095–33108. <https://doi.org/10.1074/jbc.M111.255034> (2011).
32. Iglesias, N. G., Filomatori, C. V. & Gamarnik, A. V. The F1 motif of dengue virus polymerase NS5 is involved in promoter-dependent RNA synthesis. *J. Virol.* **85**, 5745–5756. <https://doi.org/10.1128/JVI.02343-10> (2011).
33. Hanley, K. A., Lee, J. J., Blaney, J. E. Jr., Murphy, B. R. & Whitehead, S. S. Paired charge-to-alanine mutagenesis of dengue virus type 4 NS5 generates mutants with temperature-sensitive, host range, and mouse attenuation phenotypes. *J. Virol.* **76**, 525–531 (2002).
34. Tay, M. Y. *et al.* The C-terminal 50 amino acid residues of dengue NS3 protein are important for NS3-NS5 interaction and viral replication. *J. Biol. Chem.* **290**, 2379–2394. <https://doi.org/10.1074/jbc.M114.607341> (2015).
35. Egloff, M. P. *et al.* Structural and functional analysis of methylation and 5'-RNA sequence requirements of short capped RNAs by the methyltransferase domain of dengue virus NS5. *J. Mol. Biol.* **372**, 723–736. <https://doi.org/10.1016/j.jmb.2007.07.005> (2007).
36. Henderson, B. R., Saeedi, B. J., Campagnola, G. & Geiss, B. J. Analysis of RNA binding by the dengue virus NS5 RNA capping enzyme. *PLoS ONE* **6**, e25795. <https://doi.org/10.1371/journal.pone.0025795> (2011).
37. Yap, L. J. *et al.* Crystal structure of the dengue virus methyltransferase bound to a 5'-capped octameric RNA. *PLoS ONE* **5**, 15–20. <https://doi.org/10.1371/journal.pone.0012836> (2010).
38. Bujalowski, W. & Jezewska, M. J. Kinetic mechanism of the single-stranded DNA recognition by *Escherichia coli* replicative helicase DnaB protein. Application of the matrix projection operator technique to analyze stopped-flow kinetics. *J. Mol. Biol.* **295**, 831–852. <https://doi.org/10.1006/jmbi.1999.3378> (2000).
39. Jezewska, M. J., Szymanski, M. R. & Bujalowski, W. Kinetic mechanism of the ssDNA recognition by the polymerase  $\times$  from African swine fever virus: dynamics and energetics of intermediate formations. *Biophys. Chem.* **158**, 9–20. <https://doi.org/10.1016/j.bpc.2011.04.010> (2011).
40. Bussetta, C. & Choi, K. H. Dengue virus nonstructural protein 5 adopts multiple conformations in solution. *Biochemistry* **51**, 5921–5931. <https://doi.org/10.1021/bi300406n> (2012).
41. Micsonai, A. *et al.* Accurate secondary structure prediction and fold recognition for circular dichroism spectroscopy. *Proc. Natl. Acad. Sci. U.S.A.* **112**, E3095–3103. <https://doi.org/10.1073/pnas.1500851112> (2015).
42. Zuker, M. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res.* **31**, 3406–3415 (2003).

## Acknowledgements

We thank Dr. Cheng Kao at Indiana University for the ZIKV NS5 clones. This work was supported by National Institutes of Health [AI 087856, AI 137627]; and University of Texas Medical Branch [John Sealy Memorial Endowment fund to KHC, Jeane B. Kempner postdoctoral fellowship to P.J.B.].

## Author contributions

P.J.B. and K.H.C. designed the overall experiments. P.J.B. performed the experiments. P.J.B, W.B., and K.H.C analyzed data and wrote the paper.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary information** is available for this paper at <https://doi.org/10.1038/s41598-020-70094-y>.

**Correspondence** and requests for materials should be addressed to K.H.C.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2020