RESUBMITTED MANUSCRIPT

Chikungunya virus nsP4 RNA-dependent RNA polymerase core domain displays detergent-sensitive primer extension and terminal adenylyltransferase activities

Running title: Characterization of Chikungunya virus nsP4 polymerase

Ming Wei CHEN[a*], Yaw Bia TAN[a], Jie ZHENG[b^], Yongqian ZHAO[a], Bee Ting LIM[b,c], Tobias CORNVIK[b,c], Julien LESCAR[b], Lisa Fong Poh NG[d], Dahai LUO[a#].

[a]Lee Kong Chian School of Medicine, Nanyang Technological University, Singapore; [b]School of Biological Sciences, Nanyang Technological University, Singapore; [c]Protein Production Platform, Nanyang Technological University, Singapore; [d]Singapore Immunology Network (SIgN), Agency for Science, Technology and Research (A*STAR), Singapore.

#Correspondence can be addressed to Dahai Luo. Email: luodahai@ntu.edu.sg

Phone: (65) 65923986

Address: 59 Nanyang Drive, Experimental Medicine Building, 03-07-01, Singapore 636921.

*Present address: Protein Production Platform, Nanyang Technological University, Singapore.

^Present address: Department of Molecular Therapeutics, The Scripps Research Institute, Jupiter, Florida, USA.

Keywords: Chikungunya, RdRP, HDX-MS, fluorescence anisotropy, homology modeling

ABSTRACT

Chikungunya virus (CHIKV) is an important arboviral infectious agent in tropical and subtropical regions, often causing persistent and debilitating disease. The viral enzyme non-structural protein 4 (nsP4), as RNA-dependent RNA polymerase (RdRP), catalyzes the formation of negative-sense, genomic and subgenomic viral RNAs. Here we report a truncated nsP4 construct that is soluble, stable and purified recombinantly from *Escherichia coli*. Sequence analyses and homology modelling indicate that all necessary RdRP elements are included. Hydrogen/deuterium exchange with mass spectrometry was used to analyze solvent accessibility and flexibility of subdomains. Fluorophore-conjugated RNA ligands were designed and screened by using fluorescence anisotropy to select a suitable substrate for RdRP assays. Assay trials revealed that nsP4 core domain is conditionally active upon choice of detergent species, and carries out both primed extension and terminal adenylyltransferase activities. The polymerization assay can be further developed to screen for antiviral compounds *in vitro*.

1. INTRODUCTION

Chikungunya virus is one of the most important emerging RNA viruses today (Burt et al., 2012). CHIKV belongs to the Alphavirus genus under the family Togaviridae (Schwartz and Albert, 2010; Strauss and Strauss, 1994); the mosquito species *Aedes aegypti* and *Aedes albopictus* serve as its vector while human is the ultimate host. CHIKV infection results in a febrile disease known as chikungunya fever (Borgherini et al., 2008; Mavalankar et al., 2007). Though rarely fatal in healthy adults, the infection often causes long-lasting joint pain and arthritis, possibly because of latent infection (Borgherini et al., 2008; Hoarau et al., 2010; Labadie et al., 2010; Levine et al., 1994). The increased prevalence of disease, the current lack of specific treatment and vaccine, and the severity of associated long-term sequelae make CHIKV a threat to global healthcare, with high socioeconomic costs (Couturier et al., 2012; Mavalankar et al., 2007; Oon and Ng, 2014; Queyriaux et al., 2008). Co-circulation of dengue, Zika and Chikungunya viruses in endemic areas resulting in co-infection (Villamil-Gómez et al.) and adaptation of CHIKV to new mosquito hosts (Ledermann et al., 2014) may further elevate the crisis.

CHIKV contains one copy of single-chain RNA genome, just below 12 kb (Rupp et al., 2015). Upon entry into host cell, the RNA genome is released into the cytoplasm and acts as the messenger RNA for the non-structural polyprotein P1234, which is the precursor to the CHIKV replication complex (RC). The RC matures as the polyprotein is cleaved sequentially, eventually becoming four non-structural proteins (nsP1, nsP2, nsP3 and nsP4). Within the RC, nsP4 encodes the RNA-dependent RNA polymerase (RdRP), which is responsible for replicating viral RNA (Tomar et al., 2006). An intriguing feature of alphaviral nsP4 proteins is the presence of an

N-terminal region (approximately 100 residues) before the core RdRP domain; the N-terminal "domain" is predicted to be mostly disordered and has little homology to known sequences. The absolutely conserved Tyr-1 residue was found to be important for the synthesis of minus-sense RNA (Rupp et al., 2011), and can only be mutated to other aromatic residues for activity to be preserved (Shirako and Strauss, 1998). Studies on the Sindbis virus nsP4 indicate that it is able to synthesize the negative sense anti-genome *de novo* and has terminal adenylyl transferase (TATase) activity (Rubach et al., 2009). However, the related full-length Sindbis virus (SINV) nsP4 has been shown to be insoluble when expressed in *E. coli* without additional solubilizing tags and optimized lysis buffers (Rubach et al., 2009). Therefore, it remains of interest to determine whether the C-terminal RdRP domain of CHIKV nsP4 may function on its own as the minimal catalytic core.

As viral RdRP proteins are the central component to RNA virus replication, they are an attractive target for antiviral development (Velkov et al., 2014; Waheed et al., 2013). Nucleoside and non-nucleoside inhibitors which specifically target the enzymes would be highly desirable as an infection cannot be established without viral RNA synthesis. In this work, we report a truncated nsP4 RdRP domain construct (nsP4-Δ118) from CHIKV that can be solubly expressed in *Escherichia coli* and purified to high homogeneity. Homology modelling suggests that all RdRP catalytic elements are present in the construct. The protein performs primed extension and TATase activity on short RNA oligonucleotides in the presence of a specific detergent N,N-dimethyldodecylamine N-oxide (LDAO). To our knowledge, this is the first reported truncated and active CHIKV nsP4 protein. The protein construct can be further utilized for structural studies and development of RdRP-inhibitor assays.

2. RESULTS

2.1. CHIKV nsP4 RdRP domain is a soluble protein. Full-length CHIKV nsP4 (nsP4-FL) expressed recombinantly in *E. coli* could not be solubilized. We then tried to design core RdRP domain constructs that are devoid of the extraneous N-terminal region. Protein BLAST (Boratyn et al., 2012) suggested that residues 151 – 599 constitute the core RdRP domain, placing nsP4 in protein family pfam00978 that includes RdRPs from bromoviruses, tobamoviruses and togaviruses (Figure 1A). However, the N-terminal region does not have significant homology to known protein families. Next, GlobPlot (Linding et al., 2003b) was used to identify subdomains and potentially disordered regions within nsP4. Residues 12 – 95 in the N-terminus were predicted to be globular in solution, but large portions of it may be intrinsically disordered as predicted by DisEMBL (Linding et al., 2003a). Residues 115 – 611 were predicted to form a globular domain and cover the entire core RdRP region. We then cloned, expressed and attempted to purify multiple constructs that are based on the suggested domain boundary. Other constructs were also made based on secondary structure prediction; a list of insoluble constructs is given in Table S1. Only one construct with the first 118 residues removed (nsP4-Δ118) (Figure 1B) was soluble upon cell lysis. The tagged protein could be purified by using a three-step protocol to >95% purity by SDS-PAGE analysis (Figure 1C).

2.2. Homology modelling reveals an RdRP fold similar to picornaviral polymerases. Top hits that surfaced on BLAST queries against the PDB belonged to RdRPs of the Picornaviridae family, which are also positive-sense ssRNA viruses. The hits included RdRPs of Rhinovirus B14 (PDB ID: 1xr5)(Love et al., 2004), Enterovirus 71 (EV71; PDB ID: 4ika)(Chen et al., 2013), Coxsackie virus (PDB ID: 3ddk)(Campagnola et al., 2008), Encephalomyocarditis virus (EMCV;

PDB ID: 4nz0)(Vives-Adrian et al., 2014), and Norwalk virus (PDB ID: 2b43)(Högbom et al., 2009) of the Caliciviridae. The RdRP amino acid sequences aligned with nsP4 sequence only within the predicted core RdRP region, resulting in sequence identity below 20% (Figure 2A). nsP4 RdRP is more similar to the abovementioned "small" RdRPs as opposed to "large" flaviviral RdRPs that have C-terminal extensions containing extra features such as priming loops (Caillet-Saguy et al., 2014; Yap et al., 2007) (Figure S1). In the absence of a crystal structure of nsP4, a structural model of CHIKV RdRP domain was generated by homology modelling using I-TASSER server (Yang et al., 2015). The C-score for the best-ranked model was -1.20 falling within the range for correct global topology.

The calculated nsP4-Δ118 model showed a classical, basic RdRP architecture with well-defined fingers, thumb and palm domains (Figure 2B). On top of the palm domain and between the fingers and thumb domains was a tunnel across the structure, corresponding to the RNA-binding channel and RdRP active site. To compare the model to the crystal structures mentioned above, each structure was separately superimposed to the calculated nsP4-Δ118 model by using the secondary-structure matching algorithm in WinCoot (Emsley et al., 2010), with root-mean-squared deviations of 1.6 – 1.9 Å. The fingers, thumb and palm domains were well-aligned as seen in the overlay of backbone carbon atom traces (Figure 2C). In particular, the catalytic GDD motif (motif C) and the "B-loop" (motif B) which has been shown to contact template strand RNA directly were in good agreement with the reference structures (Figure 2D). Comparison with a structure of poliovirus 3DPol in complex with both template and product RNA strands (PDB ID: 4k4s) also confirmed the correctness of the modelled RNA-binding channel. Overall, the nsP4-Δ118 model is an accurate structural prediction where the active site and RNA-binding

elements are concerned. However, the N-terminal portion of nsP4-Δ118 which precedes the predicted RdRP core had no counterpart within the experimental structures.

2.3. nsP4-Δ118 is flexible in the fingers domain. Hydrogen/deuterium exchange coupled to mass spectrometry (HDX-MS) was used to analyse solvent accessibility and structural dynamics (Konermann et al., 2011; Zhao et al., 2015; Zheng et al., 2015) of nsP4-Δ118 protein. In brief, the proteins were exposed to deuterated water to incur HDX, denatured, and digested with pepsin. The peptides were then subjected to liquid chromatography mass spectrometry, resulting in peptide mass spectra which indicated increasing HDX over time. HDX-MS results for nsP4-Δ118 shown as a graphical heat map (Figure 3A) revealed regions of the protein that underwent different degrees of HDX. The results confirmed the identity of our nsP4-Δ118 preparations by virtue of the peptide masses, and suggested good conformational homogeneity based on the highly reproducible HDX percentages (Figure 3A). Interestingly, the fingers, thumb and palm domains displayed high, medium and low levels of HDX, respectively. The heat map of HDX was subsequently projected onto the calculated nsP4-Δ118 model (Figure 3B, 3C). A long stretch of the N-terminal sequence (A243 – Y322), encompassing the fingers domain, displayed HDX above 70%. A peptide with especially high HDX (95 – 98%; Y256 – L268) could be mapped to two short helices in the fingers domain. For the interest of future studies, these highly mobile regions may be involved in protein protein interactions within the RC.

2.4. nsP4-Δ118 prefers binding to RNA duplex with 5'-overhangs. In order to assess the RNA-binding properties of nsP4-Δ118 and thereby select a suitable substrate for RNA polymerase assays, a series of FAM-conjugated short RNA molecules were used to conduct fluorescence

anisotropy (FA) based RNA binding kinetics study (Table S2). We included single-stranded RNA (r06), RNA hairpin with a blunt end (oDL518), RNA hairpin with 3'-overhang (r11), and RNA hairpin with 5'-overhang (r10) to investigate the RNA preference of nsP4. All RNAs except oDL518 showed good binding kinetics and had submicromolar Kd values (Figure 4A), indicating that single-stranded regions are required for binding to nsP4-Δ118. Next, r10 was further developed into a series carrying different 5'overhangs (r10a, r10b and r10c carried $(U)_4$, $(U)_8$ and $(U)_4A(U)_3$, respectively). FA with these RNAs showed that nsP4-Δ118 had the highest affinity for r10b, followed by r10c, r10a and r10 (Figure 4B). Shortening the 5'-overhang from 8-nt (r10b, r10c) to 4-nt (r10a) almost doubled the Kd value. Taken together, the FA results show that nsP4-Δ118 prefers single-stranded RNA with a free 5'-overhang that is longer than 4-nt.

2.5. nsP4-Δ118 displays detergent-sensitive primed elongation on hairpin substrates and terminal adenylyltransferase (TATase) activity. To examine if nsP4-Δ118 was a minimal functional RdRP, we set up polymerase activity assays using the r10-like short RNA as substrates and monitored the reactions as the fluorophore-conjugated RNAs increase in length on a denaturing PAGE gel. A list of RNAs involved in the RdRP assays is provided in Table S3. Initially no elongation was detected on r10 or r10b with up to 10 μM nsP4-Δ118 under commonly prescribed assay conditions (30 °C for two hours). Polymerase activity became evident when incubation was carried out at lower temperature (room temperature; 22 – 25 °C) overnight, in the presence of both magnesium and manganese(II) ions. Reduced r10b band intensity and accumulation of a diffuse high molecular weight (HMW) smear would be observed (Figure 5A). This is in contrast to a control reaction by a dengue virus serotype 4 RdRP (D4R), which gave a distinct end-product band as major product.

Further investigation revealed that nsP4-Δ118 was strongly affected by detergents used in the assay buffer. For example, the non-ionic Triton X-100 caused incomplete reaction (r10b band faded but not completely) that favoured HMW smear formation (presumably due to primed extension followed by TATase activity). Ionic detergents such as sodium dodecyl sulfate and sodium deoxycholate completely inactivated the enzyme. However, LDAO resulted in the accumulation of a major product band, with most of the starting r10b substrate utilized. 3 mM LDAO appeared to be the minimum amount required for complete reaction with majority of product in a single gel band (Figure 5B), a concentration that is about 1.5 times the critical micellar concentration (c.m.c.). The lack of any detectable polymerase activity in the GNN mutant (Figure 5C), in which the two catalytic aspartates in the GDD motif are mutated to asparagines, confirmed that the results discussed above were not due to trace contaminant proteins.

We examined nsP4-Δ118's detergent preference further by screening several related detergents with different head groups and acyl-chain lengths. LDAO-like amine oxide detergents with successively shorter acyl chains (UDAO and DDAO) at 1.5 c.m.c. showed significant lower activity as LDAO (Figure 5D). Detergents with the same 12-carbon acyl chain as LDAO but different head groups also gave negative results. These included inactivation (sarkosyl), HMW smear (dodecyltrimethylammonium chloride and lauryl sucrose), and incomplete extension (dodecyldimethylglycine and LysoFos choline 12). Amongst the detergents tested only LDAO allowed the formation of a major product. In comparison, D4R activity is not fully abolished even in the presence of strong detergents (Figure S2).

Kinetics of nsP4-Δ118 with r10b substrate was assessed with a time-course experiment following buffer optimization. A shift of band intensity from the original r10b to a HMW end product is visible when the reaction was sampled at different time points (Figure 6A). A fully extended, blunt-ended variation of r10b ("Ref1") is provided for reference. Elongated products can be detected at the 3-h time point and continued to accumulate during overnight incubation at room temperature. Replacing ATP with 2'-deoxy-ATP ("dA") and 3'-deoxy-ATP ("3'dA") established that nsP4-Δ118 cannot incorporate these ATP analogues into product strands. On the other hand, D4R was able to reach reaction endpoint by the 0.5-h time point. Interestingly, the final products of D4R RdRP activity in the presence of both magnesium(II) and manganese(II) migrate higher than the reference, and this behaviour is not observed when only magnesium is used as cofactor metal (Figure S3).

We then tested nsP4 RdRP activity on a true duplex RNA, which is assembled from a primer strand (5'-FAM-labelled r16) and a template strand (non-labelled r15). An $(A)_8$ tract on r15 serves as template for the elongation of r16. Free r16, in the absence of r15, was elongated by nsP4-Δ118 via TATase activity (Figure 6B). The products were visible as a ladder, with each successive elongated product showing weaker intensity. No product was observed when non-ATP nucleotides were used. In contrast, r16 annealed to r15 displayed a finite series of elongation products with no gradual decrease in band intensity, though the reaction was not complete when compared to a reference RNA of fully elongated length ("Ref2"). Reactions using 2'- and 3'-deoxy-ATP did not yield products. In conclusion, nsP4-Δ118 performed primed extension and template-independent polyadenylation. The polymerase activitiy is sensitive to specific detergents.

3. DISCUSSION

CHIKV nsP4 required extensive construct boundary design in order to be solubly expressed in *E. coli* in good yield suitable for structural and biochemical analyses. Based on our bioinformatics analysis, we designed several nsP4 RdRP constructs and successfully obtain a soluble protein construct (nsP4-Δ118) which encompasses all known RdRP motifs. This construct is significantly shorter than a previously reported SINV truncated nsP4, which was generated by deleting 97 N-terminal residues (SINΔ97nsP4) (Tomar et al., 2006). CHIKV nsP4 constructs designed around the equivalent N-terminus of SINΔ97nsP4 did not yield soluble protein. Despite the overall similarity between the two nsP4 proteins (73% identity), residues 90 – 120 form one of the most divergent tracts within alphaviral nsP4 sequences (Figure S4), which may explain the difference in the construct boundaries.

HDX-MS is a powerful tool to study protein structural dynamics in solution. We have previously used HDX-MS to study the protein dynamics of dengue NS5 polymerase (Zhao et al., 2015) which is a methyltransferase-RdRP tandem protein. HDX-MS not only revealed interdomain contacts in solution but also provided an NS5 flexibility profile that agreed well with local crystallographic B-factors. HDX-MS performed on nsP4-Δ118 revealed subdomain-based differences in solvent accessibility and possibly flexibility, suggesting that nsP4-Δ118 protein is a well-folded RdRP. By projecting the HDX heat map onto the I-TASSER model, we observed that HDX is highest for the fingers domain, followed by the thumb domain and lastly the palm domain. This may be due to the role of the fingers domain in nucleic acid translocation (Najmudin et al., 2000). Useful structural information such as local flexibility and solvent accessibility has been derived from HDX-MS experiments, which will facilitate structure determination by crystallography in future.

We then developed non-radioactive, fluorophore-based RNA-binding and polymerase assays to study this well-folded RdRP domain protein. Our FA studies indicate that a single-stranded 5'-overhang region, preferably 8-nt or longer, facilitates binding to nsP4-Δ118. The r10 series of labelled hairpin RNAs have a duplex 10-bp in length, which suits the RdRP's footprint as seen in the elongation complex crystal structure of poliovirus 3DPol (PDB ID: 4k4s) (Gong et al., 2013). We employed r10b as the preferred self-primed substrate as it binds to nsP4-Δ118 with high affinity (Figure 4). The r16/r15 duplex, which is similar to r10b but without the hairpin loop element, also works well in our end-point RdRP assay and can help distinguish primed extension and TATase activities (Figure 6B). Deoxy-ATP analogues cannot be incorporated into elongation products, further supporting that the truncated enzyme performs genuine RdRP activity, albeit a weak one.

The polymerase activity of nsP4-Δ118 is also extremely sensitive to the influence of detergents (Figure 5), highlighting the need for carefully designed RdRP assay conditions that use truncated nsP4 constructs. Given the membrane-bound nature of alphaviral replication complexes, it is possible that nsP4 is also involved in membrane association. RNA synthesis of replication complex solubilized in detergent is lowered (Gomatos et al., 1980), while recombinantly produced nsP4 is only fully active when membrane isolates containing all other non-structural proteins are supplied (Lemm et al., 1998). RNA replication may also depend on membrane lipid composition as seen in another RNA virus (Xu and Nagy, 2015). These previous findings may explain the conditional catalytic activity and detergent preferences of nsP4-Δ118. A major impetus for acquiring a highly soluble, stable and functional nsP4 RdRP protein is the development of high-throughput compound screening assays targeting RNA polymerase activity. Fluorophore-labelled hairpin RNAs, serving as self-primed RdRP substrates, are useful for this

purpose as changes in RNA length can be detected with minimal amounts of RNA. The tedious and potentially hazardous use of radioactive nucleotides can also be circumvented. Our polymerase activity assay can be further optimized towards attaining a simple, high-throughput format that does not rely on cell extracts, full-length genomic or replicon RNA, or radioactive nucleotides.

4. MATERIALS AND METHODS

4.1. Cloning, expression and purification of nsP4-Δ118. Constructs were designed based on bioinformatics analysis by using protein BLAST (Altschul et al., 1997), GlobPlot (Linding et al., 2003b) and DisEMBL (Linding et al., 2003a), and also secondary structure prediction with JPred (Cole et al., 2008). Synthetic CHIKV strain LR2006 (GenBank accession ID: **DQ443544.2**) non-structural polyprotein gene was purchased (Bio Basic, Ontario) to serve as template. Ligation-independent cloning (LIC) of constructs into SGC vectors (pNIC28-Bsa4, pNIC-CH2 or pNIC-GST) (Savitsky et al., 2010) was performed with PCR amplicons flanked by appropriate LIC adaptor sequences. Resultant clones would carry an N-terminal hexahistidine tag and a TEV protease site. LIC products were transformed into TOP10 cells; clones were isolated, purified by Miniprep (Qiagen, Germany), and sequenced (1^{st} BASE, Singapore). GDD-to-GNN mutations were introduced by using site-directed mutagenesis with forward primer (5'-CGCGTGCGCGGCCTTCATCGGCAATAATAACATAATACATGGAGTCGTC-3') and reverse primer (5'-GACGACTCCATGTATTATGTTATTATTGCCGATGAAGGCCGCGCACGCG-3'). Correct clones were transformed into Rosetta 2 (DE3) cells and subjected to solubility testing. Briefly, each clone was grown in a modified autoinductive terrific broth supplemented with 100 μg/mL kanamycin, 64 μg/mL chloramphenicol, 0.8% v/v glycerol, 2 mM $MgCl_2$, 0.5% w/v lactose and 0.015% w/v glucose. Cultures were grown to OD600 of 2 and then cooled down to 20 °C for overnight expression. Cells were harvested by centrifugation at 4000 *g* and the pellets stored at -20 °C. Lysis buffer (50 mM Tris-Cl, 500 mM NaCl, 10% v/v glycerol, 5 mM 2-mercaptoethanol, pH 7.5) was added to cell pellets approximately 10 mL to 1 g pellet. Resuspended cells were lysed by sonication, clarified by centrifugation, and added with 100 μL Ni-NTA resin (BioRad).

Resin was washed with buffer containing up to 20 mM imidazole and the bound protein(s) eluted by using buffer containing 300 mM imidazole. Insoluble and elution fractions were analysed by SDS-PAGE. In our laboratory only nsP4-Δ118 was found to be soluble. Expression of nsP4-Δ118 was scaled up to 4.8 L using the same media composition and expression protocol. Cells were lysed with 5 mL lysis buffer per g pellet; the lysate was clarified by centrifugation at 45000 *g* and subjected to Ni-NTA affinity purification. The eluate was loaded on a 5 mL HiTrap Q column (GE Healthcare) and a linear gradient of increasing NaCl concentration was used to elute protein fractions. Target protein normally eluted at approx. 300 mM NaCl. Clean fractions were pooled and loaded on an S200 16/600 column (GE Healthcare) equilibrated in gel filtration buffer (20 mM HEPES, 300 mM NaCl, 5% v/v glycerol, 2 mM DTT, pH 7.5). Protein fractions corresponding to the monomer were concentrated to 11.8 mg/mL, aliquoted, flash-frozen and stored at -80 °C. Final yield was approximately 5 mg and the purity >95% as by SDS-PAGE.

4.2. Homology modelling of nsP4-Δ118. The sequence of nsP4-Δ118 including the peptide tag was submitted to the I-TASSER webserver (Yang et al., 2015) without imposing user-defined parameters. Close homologs in the Protein Data Bank (PDB) were identified by querying nsP4-Δ118 sequence against the PDB in protein BLAST. The experimental structures were then structurally aligned to the calculated nsP4-Δ118 model by using the secondary-structure matching alignment function in WinCoot (Emsley et al., 2010). Root-mean-squared deviation values of the alignments were generated and the aligned structures were visualized on PyMOL (DeLano Scientific).

4.3. Hydrogen/Deuterium Exchange of nsP4-Δ118. The buffer for HDX on-exchange was the same composition except that H_2O was replaced with D_2O (99.99%) and glycerol was removed. 4 μL of nsP4-Δ118 was mixed with 16 μL D_2O buffer (final D_2O concentration was 80%) or H_2O buffer for 0s samples, and incubated at 4 °C at 600s time interval for on-exchange. Ice-cold 20 μL quench solution consisting of 1 M guanidine hydrochloride and 1.5 % (v/v) formic acid was added to each on-exchanged samples after specific time intervals prior to rapid freezing by liquid nitrogen and stored frozen.

4.4. LC-MS Under Quench Condition. For capillary-flow LC, buffer A was H_2O containing 0.3% (v/v) formic acid. Buffer B was acetonitrile containing 0.3% (v/v) formic acid. Protein samples were digested online by passing them through an immobilized pepsin-coupled column (2.1 mm internal diameter x 30 mm, Invitrogen), and were de-salted for 3 min on a house-packed C4 trap (0.75 mm i.d. x 10 mm, C4 beads purchased from Michrom). The mobile phase for on-line pepsin digestion was buffer A and the flow rate was 150 μL min^{-1} driven by an LC loading pump (Dionex 3000 RSLC). A 20 min gradient on a house-packed C4 column (0.3 mm i.d. x 50 mm, C4 beads purchased from Michrom) was used to separate and elute peptic peptides prior to MS analysis. All parts were connected by a 1/16" outer diameter x 50 μM i.d. PEEK tubing and the flow rate was 15 μL min^{-1} driven by LC NC pumps. The gradient started from 5 % buffer B and increased to 35 % buffer B within 20 min, followed by washing with 90 % buffer B for 3 min and equilibration with 1 % buffer B for 5 min. MS raw files were acquired in the range of m/z 300-2000 for 30 min in positive mode on a LTQ-Orbitrap mass spectrometer (Thermo Fisher Scientific) equipped with an ESI source (capillary temperature 275 ℃ and spray voltage of 5 kV). All the HDX systems were strictly performed at 0 °C (fully buried on ice & water) and on-line

pepsin digestion was carried out at 16 °C. Blank injections were made between every two samples to remove carryover peptides. Data for each time point was repeated three times. All HDX data were normalized to 100 % D_2O content, corrected for an estimated average deuterium recovery of 70 %, and analyzed by the software HD Desktop (Pascal et al., 2009). Initial peptic peptide identifications were performed with the same HDX set up as described above. Four microliters of protein sample (20 μM) was injected into the HDX MS system. Product ion (MS/MS) spectra were acquired in linear ion trap LTQ with eight most abundant ions selected in the precursor (MS) scan with a 7.5 sec exclusion time. MS and tandem MS files were extracted and searched by using Integrated Proteomics Pipeline (http://integratedproteomics.com/products/ip2/) for high-confidence peptide identification (searching parameters: false positive rate 0.05, precursor delta mass cutoff 20, best peptide FP threshold 0.01, best peptide delta mass threshold 10).

4.5. Mapping of HDX-MS results onto nsP4-Δ118 homology model. Heat map data were divided into five color-coded tiers: blue, cyan, green, yellow and red in increasing order of HDX. The calculated model of nsP4-Δ118 was mapped with the colors; regions without HDX-MS peptide coverage were colored grey.

4.6. Fluorescence anisotropy (FA) studies. RNA molecules were designed based on the sequence of CHIKV subgenomic promoter and purchased from IDT (Supplementary Table 1). A FAM fluorophore is introduced into each RNA via a dT base. HPLC-purified RNA oligonucleotides were resuspended in ME50 buffer (10 mM MOPS, 50 mM NaCl, 1 mM EDTA, pH 7.4) to 100 μM; 1 μM working stock solutions were made in water. To perform FA, 30-μL

reactions containing 25 mM HEPES pH 7.5, 0.01% Triton X-100, 0.5 mM TCEP, 50 mM NaCl, 5 mM $MgCl_2$, 5 nM RNA and nsP4-Δ118 serial dilutions up to 10 μM were prepared on a 384-well opaque-bottom black plate. Triplicates of 12-point protein concentration series were made. The mixtures were incubated at room temperature for 60 min before the plate was measured on a Synergy H1 plate reader (BioTek, Vermont) fitted with a 485 nm/510 nm fluorescence polarization optical module. $I_{\parallel}$ (intensity of fluorescent light parallel to plane of excitation) and $I_{\perp}$ (intensity of fluorescent light perpendicular to plane of excitation) were measured, and FA was calculated as given:

$$FA = \frac{I\parallel - G * I\perp}{I\parallel + 2G * I\perp}$$

where G (= 0.87) is a correction factor specific to the instrument. FA data plotted against protein concentration by using GraphPad Prism 6 and fitted to a one-site total binding equation, in the form of a hyperbole with y-intercept

$$y = y0 + \frac{ymax * X}{Kd + x}$$

where x = protein concentration, y0 = FA at zero protein concentration, ymax = FA at saturating protein concentration, and Kd = dissociation constant. The RNA species were then compared by using the derived Kd values.

4.7. RNA polymerase assays. Ribonucleotides, 2'-deoxy-ATP and 3'-deoxy-ATP were obtained from Promega, Axil Scientific (Singapore) and Sigma-Aldrich, respectively. The RNA polymerase activity of nsP4-Δ118, specifically its elongation activity on a primed substrate, was assayed by using fluorescent RNAs. For initial tests with RNA r10, 50 nM RNA was incubated with up to 10 μM nsP4-Δ118 for 60 min at 30 °C protected from light, in reaction buffer

consisting of 25 mM HEPES pH 7.5, 0.01% Triton X-100, 2.5 mM TCEP, 2.5 mM $MgCl_2$, 2.5 mM $MnCl_2$, 1 mM of each NTP and 0.4 unit/μL RiboLock (Promega). For detergent screen experiments, detergent stock solutions standardized at 10 times critical micellar concentration (c.m.c.) were taken from the Detergent Screen kit (Hampton Research). The finalized assays with RNAs r10b and r14 were carried out at room temperature overnight, in 25 mM HEPES pH 7.5, 3 mM LDAO, 20 mM KCl, 2.5 mM TCEP, 5 mM $MgCl_2$, 2.5 mM $MnCl_2$, 2 mM ATP and 0.4 unit/μL RiboLock. For r16/r15 duplex experiments, 1.0 μM r16 and 1.1 μM r15 were mixed in 20 mM KCl, heated to 95 °C and gradually cooled down in order to anneal the RNAs. 50 nM of duplex RNA was used in the assay. Reactions were stopped by adding equal volumes of RNA loading dye (95% v/v formamide, 1% w/v SDS, 0.02% w/v bromophenol blue, 0.01% w/v xylene cyanol, 20 mM EDTA pH 8) and heating the mixtures at 65 °C for 5 min. Samples were then resolved on a 18% or 20% urea PAGE gel and imaged by using Typhoon imager (GE Healthcare). All experiments were accompanied by a dengue serotype 4 RdRP construct ("D4R") as control enzyme, cloned and prepared according to our previous work (Yap et al., 2007).

ACKNOWLEDGMENTS

We thank NTU Protein Production Platform for the extensive small-scale expression and purification trials. We also thank Wint Wint Phoo for providing purified D4R, and Stephan Reich and Simon Lattmann for suggestions regarding RdRP assays. This work was supported by Lee Kong Chian School of Medicine, Nanyang Technological University, and the Singapore Ministry of Education [grant number 2T1-06/14].

Conflict of Interest. The authors declare no competing financial interests.

Figure Legends

Figure 1. Construct design and purification of a soluble, truncated CHIKV nsP4 protein. The numbering of amino acid residues in this paper shall be based on post-processing CHIKV nsP4 alone, i.e. starting from Y1 and ending with K611. **(A)** The 611-residue sequence was analysed to elucidate component domains within nsP4. Protein BLAST (Altschul et al., 1997) aligned residues 151 – 599 to the RdRP_2 group in the Conserved Domains Database, while the N-terminal region has no clear domain assignment. GlobPlot (Linding et al., 2003b) suggested the presence of two globular domains in nsP4; Domain 2 (residues 115 – 611) is longer than the RdRP core region and formed the basis of our construct design. DisEMBL (Linding et al., 2003a) suggested five major disordered regions based on the prediction of "hot loops", indicating that a large part of the N-terminal Domain 1, and the subsequent linker region, to be intrinsically disordered. **(B)** Construct nsP4-Δ118, with the first 118 residues removed and fused to an N-terminal hexahistidine tag and TEV protease cleavage site, could be solubly expressed in *E. coli*. **(C)** Analytical size-exclusion chromatography of purified nsP4-Δ118. Purified protein was injected onto a Superdex 200 10/300 column and eluted as a single symmetrical peak at an expected elution volume. Left inset: Calibration of the analytical column. Ovalbumin (43 kD), conalbumin (75 kD), aldolase (158 kD) and ferritin (440 kD) were used as reference. The protein size is estimated to be 60 kD (calculated: 57.2 kD) indicating its monomeric nature. Right inset: SDS-PAGE analysis of the purified protein showing >95% purity at an expected molecular size.

Figure 2. RdRP sequence alignment and homology modelling of nsP4-Δ118. **(A)** Sequence alignment of nsP4-Δ118 against picornaviral RdRPs with existing crystal structures in the PDB. PDB IDs 3ddk, 4k4s, 4ika, 4nz0 and 1xr5 refer to RdRP structures of coxsackie virus, poliovirus,

Enterovirus 71, encephalomyocarditis virus and rhinovirus B14, respectively. Secondary structure annotations shown are based on the coxsackie virus RdRP structure 3ddk while RdRP sequence motifs are also indicated. Structure-based sequence alignment was carried out by using the RaptorX Structure Alignment Server (Wang et al., 2013; Wang et al., 2011) and then output graphically by using ESPript (Gouet et al., 2003). **(B)** The structural model of nsP4-Δ118 was calculated by using I-TASSER (Yang et al., 2015), and showed expected features of an RdRP such as the palm, fingers and thumb domains, and an RNA-binding substrate channel. The model is coloured blue to red from the N-terminus to the C-terminus, and was used in the abovementioned structure-based sequence alignment. **(C)** Backbone-carbon traces of superimposed RdRPs in the same view. Crystal structures of select picornaviral RdRPs were superimposed against the model of nsP4-Δ118 as reference, by using SSM superimposition in Wincoot (Emsley et al., 2010). Black: model of nsP4-Δ118; green: 3ddk, r.m.s.d. of 1.57 Å; orange: 4ika, 1.67 Å; blue: 4k4s, 1.68 Å; yellow: 1xr5, 1.72 Å; red: 4nz0, 1.91 Å. General vicinity of the catalytic site is indicated by a red box. **(D)** Close-up view of catalytic site motif B (shown as loop-helix) and motif C (strand-loop-strand) of the superimposed structures. The RNA-binding channel and surrounding peptide backbones are indicated. Good agreement is seen in major structural elements including the catalytic GDD loop in motif C.

Figure 3. HDX-MS analysis of nsP4-Δ118. The purified protein was exposed to deuterated water, denatured, digested by pepsin, and subjected to mass spectrometry. **(A)** Heat map of nsP4-Δ118 peptides. Peptide populations detected by MS are represented by individual bars, indicated with percentage of HDX, standard deviation, and the charge state of peptide. Peptides are coloured by degree of HDX (low HDX: blue, high HDX: red). The protein sequence is numbered

according to full-length nsP4 residues; the N-terminal purification tag sequence is shown in green. **(B)** Colour-mapping of HDX-MS results of nsP4-Δ118 onto the homology model. The fingers domain shows the highest mobility and flexibility followed by the thumb and the palm domains.

Figure 4. RNA-binding analysis of nsP4-Δ118 by using fluorescence anisotropy. Structural representations of the RNAs are indicated. FAM is shown as pointed star; for hairpin RNAs, 5' strand is shown on top and the 3' strand at the bottom. Data points were fitted to hyperbola to derive Kd values. **(A)** Comparison of r06 (ssRNA), r10 (hairpin with 5'-overhang), r11 (hairpin with 3'-overhang) and oDL518 (blunt hairpin duplex). Kd for oDL518 was not derived due to poor isotherm fitting. **(B)** Further investigation of r10-like hairpin RNAs. The 5'-overhang of r10 was modified to $(U)_4$, $(U)_8$ and $(U)_4A(U)_3$ in r10a, r10b and r10c, respectively.

Figure 5. Gel-based RdRP assays using FAM-labelled RNAs as substrate. Incubation products were resolved on 18% or 20% denaturing polyacrylamide gels; increase in RNA molecular size is shown as retarded migration from sample wells at the top. **(A)** Elongation of r10b with distinct detergent species. RNA: without enzyme; (-): nsP4-Δ118 without any detergent in assay buffer; D4R: control enzyme (dengue serotype 4 RdRP) without detergent; 1: Triton X-100; 2: sodium deoxycholate; 3: sodium dodecyl sulfate; 4: sulfobetaine 3-10; 5: LDAO. **(B)** Optimization of LDAO concentration in assay buffer. 1 – 4 mM LDAO, corresponding to about 0.5 – 2.0 times LDAO c.m.c., was tested. By 3 mM all original substrate r10b is utilized and a prominent major product band is acquired. **(C)** Effect of mutating catalytic GDD motif. wt: wildtype nsP4-Δ118 protein. GNN: aspartate-to-asparagine mutant, which shows no visible activity. **(D)** Testing of

detergents with other amine oxide detergents (same head group as LDAO but different acyl tail lengths) and detergents with 12-carbon tails and various head groups. 5: LDAO; 6: UDAO; 7: DDAO; 8: dodecyltrimethylammonium chloride; 9: sarkosyl; 10: lauroyl sucrose; 11: dodecyldimethylglycine; 12: dodecylphosphocholine; 13: LysoFos choline 12.

Figure 6. Assessment of reaction kinetics, nucleotide requirements and activity on single- and double-stranded RNA substrates. **(A)** Hairpin substrate r10b could not be elongated in the presence of 3'-deoxy-ATP (3'dA) and 2'-deoxy-ATP (dA) despite overnight incubation. Elongation products formed by nsP4-Δ118 in the presence of ATP were detectable at the 3-h time point and were more prominent after overnight incubation. Ref1, a reference blunt-end hairpin RNA which represents the fully elongated product, was included for reference. Ref1 sample was a synthesized RNA and not a reaction product. D4R completed the reaction by the 0.5-h time point; however the end products are larger than Ref1. See also Figure S3. **(B)** Purified nsP4-Δ118 elongated r16, a single-stranded 5'-FAM-RNA in the presence of ATP but not the other NTPs ("UCG"). D4R did not show the same TATase activity with only ATP added. When duplexed to r15 template, r16 was elongated by both D4R and nsP4-Δ118, though the nsP4-Δ118 reaction did not complete. Ref2, a synthesized 5'-FAM-RNA with the same length as fully elongated r16 with r15 as template, was included as reference. No product formation was observed when U/C/GTP or the deoxy-ATP analogues were used.

REFERENCES

Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W., Lipman, D.J., 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic acids research 25, 3389-3402.

Boratyn, G.M., Schaffer, A.A., Agarwala, R., Altschul, S.F., Lipman, D.J., Madden, T.L., 2012. Domain enhanced lookup time accelerated BLAST. Biology direct 7, 12.

Borgherini, G., Poubeau, P., Jossaume, A., Gouix, A., Cotte, L., Michault, A., Arvin-Berod, C., Paganin, F., 2008. Persistent arthralgia associated with chikungunya virus: a study of 88 adult patients on reunion island. Clinical infectious diseases : an official publication of the Infectious Diseases Society of America 47, 469-475.

Burt, F.J., Rolph, M.S., Rulli, N.E., Mahalingam, S., Heise, M.T., 2012. Chikungunya: a re-emerging virus. Lancet 379, 662-671.

Caillet-Saguy, C., Lim, S.P., Shi, P.-Y., Lescar, J., Bressanelli, S., 2014. Polymerases of hepatitis C viruses and flaviviruses: Structural and mechanistic insights and drug development. Antiviral Research 105, 8-16.

Campagnola, G., Weygandt, M., Scoggin, K., Peersen, O., 2008. Crystal Structure of Coxsackievirus B3 3Dpol Highlights the Functional Importance of Residue 5 in Picornavirus Polymerases. Journal of Virology 82, 9458-9464.

Chen, C., Wang, Y., Shan, C., Sun, Y., Xu, P., Zhou, H., Yang, C., Shi, P.-Y., Rao, Z., Zhang, B., Lou, Z., 2013. Crystal Structure of Enterovirus 71 RNA-Dependent RNA Polymerase Complexed with Its Protein Primer VPg: Implication for a trans Mechanism of VPg Uridylylation. Journal of Virology 87, 5755-5768.

Cole, C., Barber, J.D., Barton, G.J., 2008. The Jpred 3 secondary structure prediction server. Nucleic acids research 36, W197-201.

Couturier, E., Guillemin, F., Mura, M., Leon, L., Virion, J.M., Letort, M.J., De Valk, H., Simon, F., Vaillant, V., 2012. Impaired quality of life after chikungunya virus infection: a 2-year follow-up study. Rheumatology 51, 1315-1322.

Emsley, P., Lohkamp, B., Scott, W.G., Cowtan, K., 2010. Features and development of Coot. Acta crystallographica. Section D, Biological crystallography 66, 486-501.

Gomatos, P.J., Kääriäinen, L., Keränen, S., Ranki, M., Sawicki, D.L., 1980. Semliki Forest Virus Replication Complex Capable of Synthesizing 42S and 26S Nascent RNA Chains. Journal of General Virology 49, 61-69.

Gong, P., Kortus, M.G., Nix, J.C., Davis, R.E., Peersen, O.B., 2013. Structures of Coxsackievirus, Rhinovirus, and Poliovirus Polymerase Elongation Complexes Solved by Engineering RNA Mediated Crystal Contacts. PLoS ONE 8, e60272.

Gouet, P., Robert, X., Courcelle, E., 2003. ESPript/ENDscript: Extracting and rendering sequence and 3D information from atomic structures of proteins. Nucleic acids research 31, 3320-3323.

Hoarau, J.J., Jaffar Bandjee, M.C., Krejbich Trotot, P., Das, T., Li-Pat-Yuen, G., Dassa, B., Denizot, M., Guichard, E., Ribera, A., Henni, T., Tallet, F., Moiton, M.P., Gauzere, B.A., Bruniquet, S., Jaffar Bandjee, Z., Morbidelli, P., Martigny, G., Jolivet, M., Gay, F., Grandadam, M., Tolou, H., Vieillard, V., Debre, P., Autran, B., Gasque, P., 2010. Persistent chronic inflammation and infection by Chikungunya arthritogenic alphavirus in spite of a robust host immune response. J Immunol 184, 5914-5927.

Högbom, M., Jäger, K., Robel, I., Unge, T., Rohayem, J., 2009. The active form of the norovirus RNA-dependent RNA polymerase is a homodimer with cooperative activity. Journal of General Virology 90, 281-291.
Konermann, L., Pan, J., Liu, Y.-H., 2011. Hydrogen exchange mass spectrometry for studying protein structure and dynamics. Chemical Society Reviews 40, 1224-1234.
Labadie, K., Larcher, T., Joubert, C., Mannioui, A., Delache, B., Brochard, P., Guigand, L., Dubreil, L., Lebon, P., Verrier, B., de Lamballerie, X., Suhrbier, A., Cherel, Y., Le Grand, R., Roques, P., 2010. Chikungunya disease in nonhuman primates involves long-term viral persistence in macrophages. The Journal of clinical investigation 120, 894-906.
Ledermann, J.P., Guillaumot, L., Yug, L., Saweyog, S.C., Tided, M., Machieng, P., Pretrick, M., Marfel, M., Griggs, A., Bel, M., Duffy, M.R., Hancock, W.T., Ho-Chen, T., Powers, A.M., 2014. Aedes hensilli as a potential vector of Chikungunya and Zika viruses. PLoS neglected tropical diseases 8, e3188.
Lemm, J.A., Bergqvist, A., Read, C.M., Rice, C.M., 1998. Template-dependent initiation of Sindbis virus RNA replication in vitro. Journal of virology 72, 6546-6553.
Levine, B., Hardwick, J.M., Griffin, D.E., 1994. Persistence of alphaviruses in vertebrate hosts. Trends in microbiology 2, 25-28.
Linding, R., Jensen, L.J., Diella, F., Bork, P., Gibson, T.J., Russell, R.B., 2003a. Protein disorder prediction: implications for structural proteomics. Structure 11, 1453-1459.
Linding, R., Russell, R.B., Neduva, V., Gibson, T.J., 2003b. GlobPlot: Exploring protein sequences for globularity and disorder. Nucleic acids research 31, 3701-3708.
Love, R.A., Maegley, K.A., Yu, X., Ferre, R.A., Lingardo, L.K., Diehl, W., Parge, H.E., Dragovich, P.S., Fuhrman, S.A., 2004. The Crystal Structure of the RNA-Dependent RNA Polymerase from Human Rhinovirus. Structure 12, 1533-1544.
Mavalankar, D., Shastri, P., Raman, P., 2007. Chikungunya epidemic in India: a major public-health disaster. The Lancet infectious diseases 7, 306-307.
Najmudin, S., Coté, M.L., Sun, D., Yohannan, S., Montano, S.P., Gu, J., Georgiadis, M.M., 2000. Crystal structures of an N-terminal fragment from moloney murine leukemia virus reverse transcriptase complexed with nucleic acid: functional implications for template-primer binding to the fingers domain1. Journal of Molecular Biology 296, 613-632.
Oon, L.L., Ng, L.C., 2014. Chikungunya in Singapore - the Battle Continues. Annals of the Academy of Medicine, Singapore 43, 325-327.
Pascal, B.D., Chalmers, M.J., Busby, S.A., Griffin, P.R., 2009. HD desktop: an integrated platform for the analysis and visualization of H/D exchange data. Journal of the American Society for Mass Spectrometry 20, 601-610.
Queyriaux, B., Simon, F., Grandadam, M., Michel, R., Tolou, H., Boutin, J.P., 2008. Clinical burden of chikungunya virus infection. The Lancet infectious diseases 8, 2-3.
Rubach, J.K., Wasik, B.R., Rupp, J.C., Kuhn, R.J., Hardy, R.W., Smith, J.L., 2009. Characterization of purified Sindbis virus nsP4 RNA-dependent RNA polymerase activity in vitro. Virology 384, 201-208.
Rupp, J.C., Jundt, N., Hardy, R.W., 2011. Requirement for the amino-terminal domain of sindbis virus nsP4 during virus infection. Journal of virology 85, 3449-3460.

Rupp, J.C., Sokoloski, K.J., Gebhart, N.N., Hardy, R.W., 2015. Alphavirus RNA synthesis and non-structural protein functions. The Journal of general virology 96, 2483-2500.
Savitsky, P., Bray, J., Cooper, C.D., Marsden, B.D., Mahajan, P., Burgess-Brown, N.A., Gileadi, O., 2010. High-throughput production of human proteins for crystallization: the SGC experience. Journal of structural biology 172, 3-13.
Schwartz, O., Albert, M.L., 2010. Biology and pathogenesis of chikungunya virus. Nature reviews. Microbiology 8, 491-500.
Shirako, Y., Strauss, J.H., 1998. Requirement for an aromatic amino acid or histidine at the N terminus of Sindbis virus RNA polymerase. Journal of virology 72, 2310-2315.
Strauss, J.H., Strauss, E.G., 1994. The alphaviruses: gene expression, replication, and evolution. Microbiological reviews 58, 491-562.
Tomar, S., Hardy, R.W., Smith, J.L., Kuhn, R.J., 2006. Catalytic core of alphavirus nonstructural protein nsP4 possesses terminal adenylyltransferase activity. Journal of virology 80, 9962-9969.
Velkov, T., Carbone, V., Akter, J., Sivanesan, S., Li, J., Beddoe, T., Marsh, G.A., 2014. The RNA-dependent-RNA polymerase, an emerging antiviral drug target for the Hendra virus. Current drug targets 15, 103-113.
Villamil-Gómez, W.E., Rodríguez-Morales, A.J., Uribe-García, A.M., González-Arismendy, E., Castellanos, J.E., Calvo, E.P., Álvarez-Mon, M., Musso, D., Zika, dengue, and chikungunya co-infection in a pregnant woman from Colombia. International Journal of Infectious Diseases.
Vives-Adrian, L., Lujan, C., Oliva, B., van der Linden, L., Selisko, B., Coutard, B., Canard, B., van Kuppeveld, F.J.M., Ferrer-Orta, C., Verdaguer, N., 2014. The Crystal Structure of a Cardiovirus RNA-Dependent RNA Polymerase Reveals an Unusual Conformation of the Polymerase Active Site. Journal of Virology 88, 5595-5607.
Waheed, Y., Bhatti, A., Ashraf, M., 2013. RNA dependent RNA polymerase of HCV: a potential target for the development of antiviral drugs. Infection, genetics and evolution : journal of molecular epidemiology and evolutionary genetics in infectious diseases 14, 247-257.
Wang, S., Ma, J., Peng, J., Xu, J., 2013. Protein structure alignment beyond spatial proximity. Scientific reports 3, 1448.
Wang, S., Peng, J., Xu, J., 2011. Alignment of distantly related protein structures: algorithm, bound and implications to homology modeling. Bioinformatics (Oxford, England) 27, 2537-2545.
Xu, K., Nagy, P.D., 2015. RNA virus replication depends on enrichment of phosphatidylethanolamine at replication sites in subcellular membranes. Proceedings of the National Academy of Sciences of the United States of America 112, E1782-1791.
Yang, J., Yan, R., Roy, A., Xu, D., Poisson, J., Zhang, Y., 2015. The I-TASSER Suite: protein structure and function prediction. Nature methods 12, 7-8.
Yap, T.L., Xu, T., Chen, Y.L., Malet, H., Egloff, M.P., Canard, B., Vasudevan, S.G., Lescar, J., 2007. Crystal structure of the dengue virus RNA-dependent RNA polymerase catalytic domain at 1.85-angstrom resolution. Journal of virology 81, 4753-4765.
Zhao, Y., Soh, T.S., Zheng, J., Chan, K.W., Phoo, W.W., Lee, C.C., Tay, M.Y., Swaminathan, K., Cornvik, T.C., Lim, S.P., Shi, P.Y., Lescar, J., Vasudevan, S.G., Luo,

D., 2015. A crystal structure of the Dengue virus NS5 protein reveals a novel inter-domain interface essential for protein flexibility and virus replication. PLoS pathogens 11, e1004682.
Zheng, J., Yong, H.Y., Panutdaporn, N., Liu, C., Tang, K., Luo, D., 2015. High-resolution HDX-MS reveals distinct mechanisms of RNA recognition and activation by RIG-I and MDA5. Nucleic acids research 43, 1216-1230.

FIGURES

Figure 1

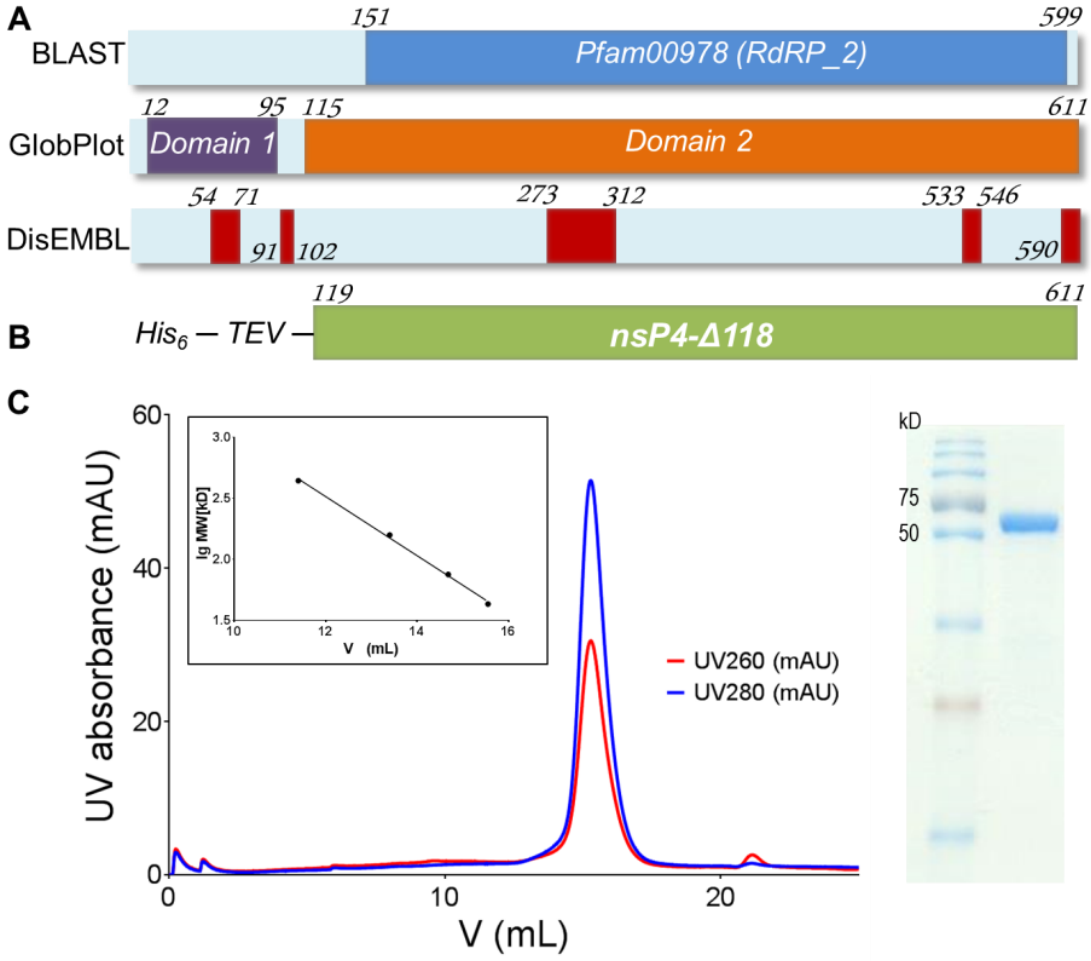


Figure 2

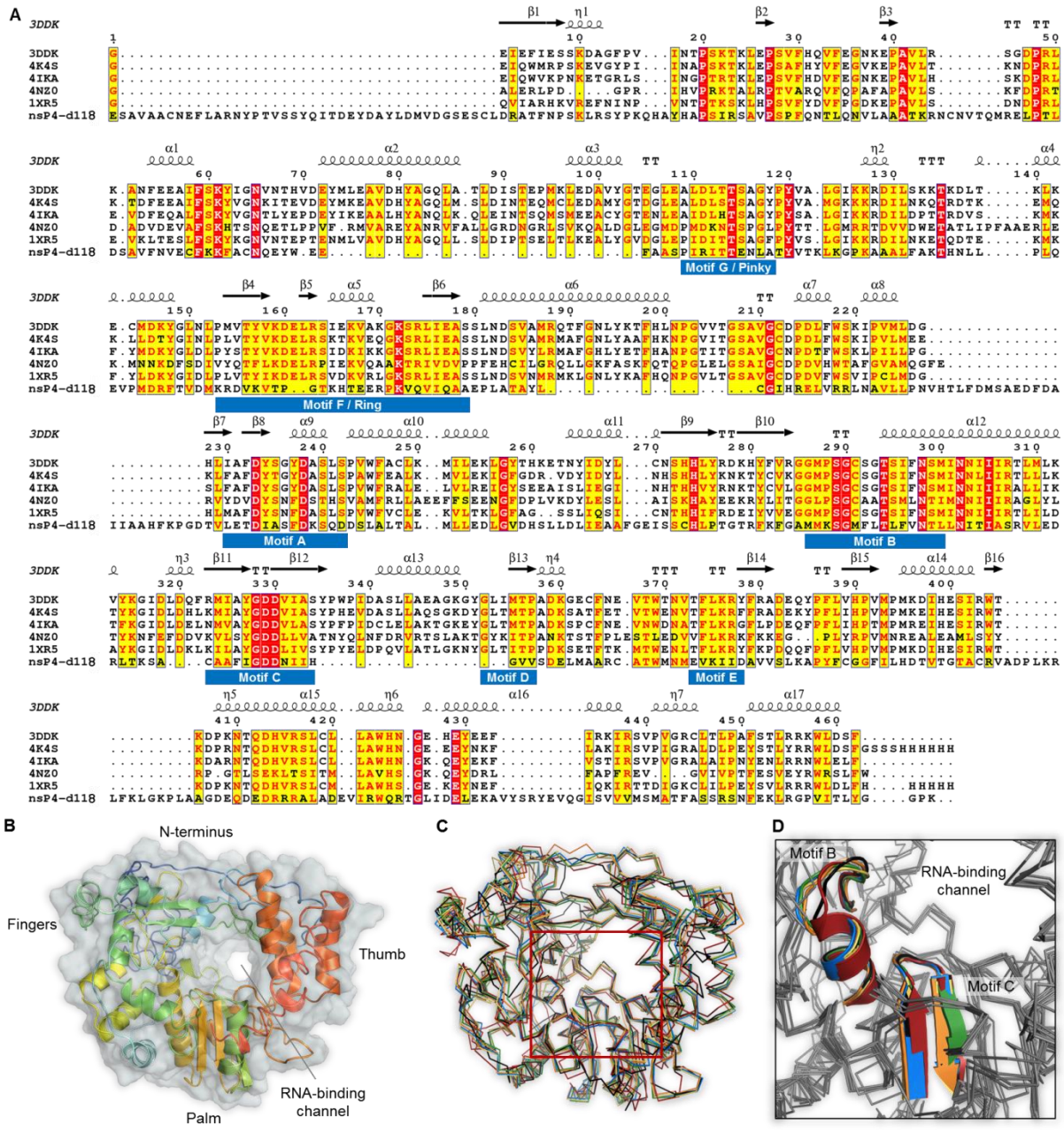

A
3DDK
4K4S
4IKA
4NZ0
1XR5
nsP4-d118
Motif G / Pinky
Motif F / Ring
Motif A
Motif B
Motif C
Motif D
Motif E
B
N-terminus
Fingers
Thumb
RNA-binding channel
Palm
C
D
Motif B
RNA-binding channel
Motif C

Figure 3

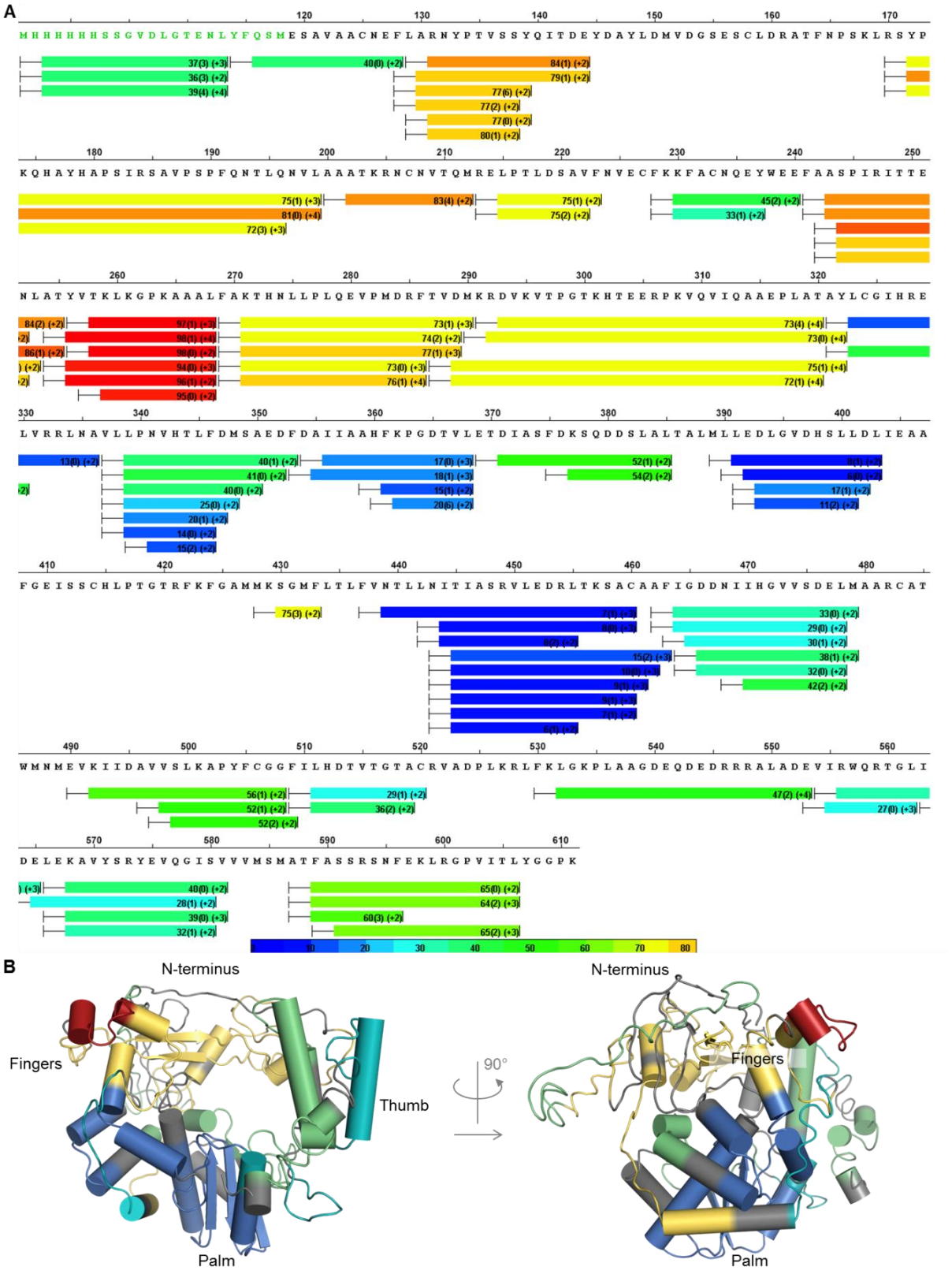


Figure 4

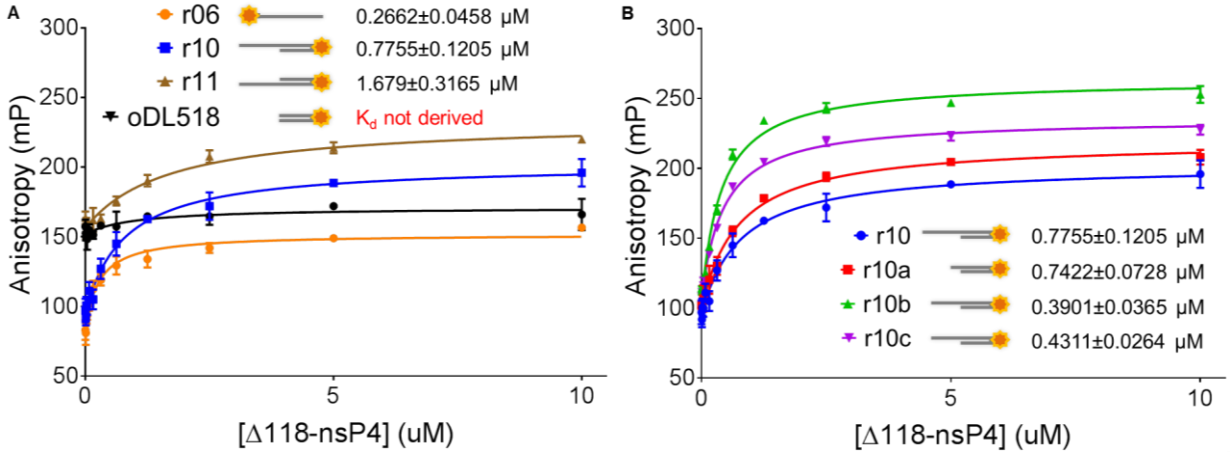

A
r06
r10
r11
oDL518
0.2662±0.0458 μM
0.7755±0.1205 μM
1.679±0.3165 μM
Kd not derived
Anisotropy (mP)
[Δ118-nsP4] (uM)
B
r10
r10a
r10b
r10c
0.7755±0.1205 μM
0.7422±0.0728 μM
0.3901±0.0365 μM
0.4311±0.0264 μM
Anisotropy (mP)
[Δ118-nsP4] (uM)

Figure 5

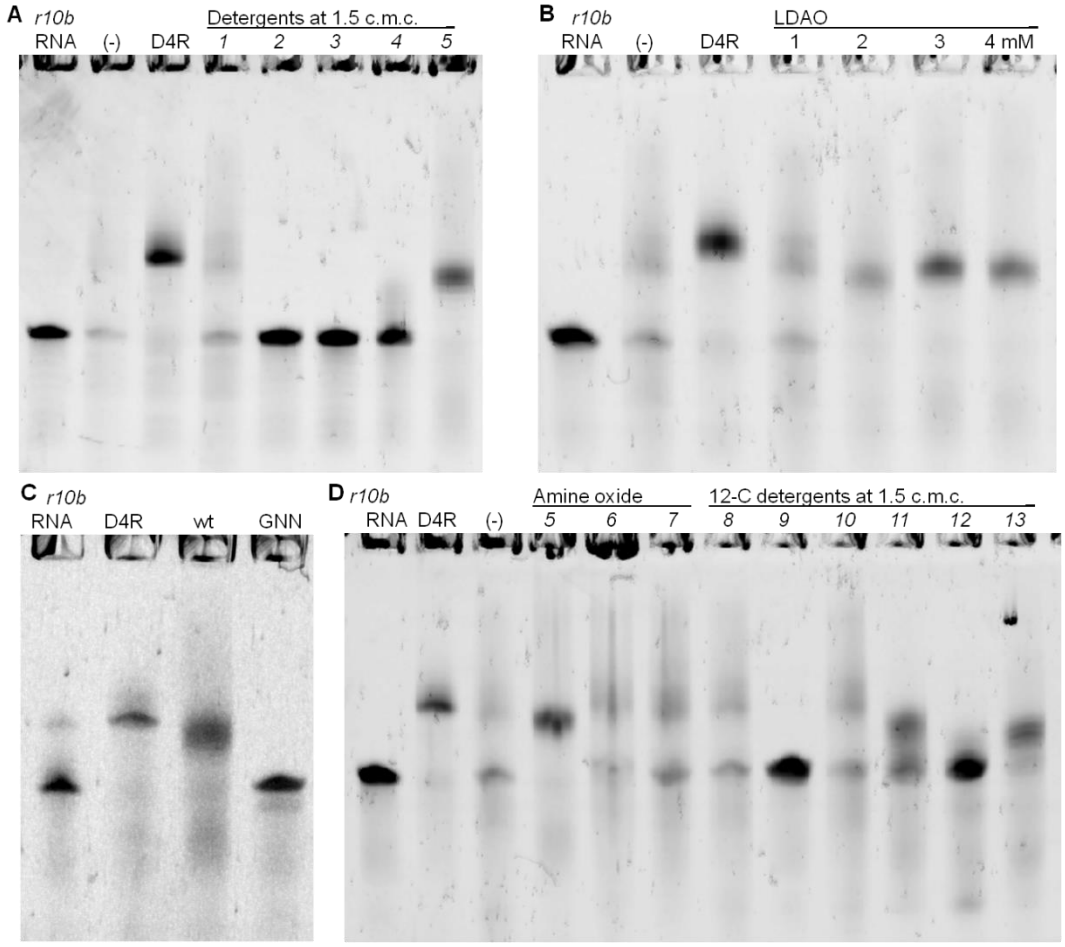
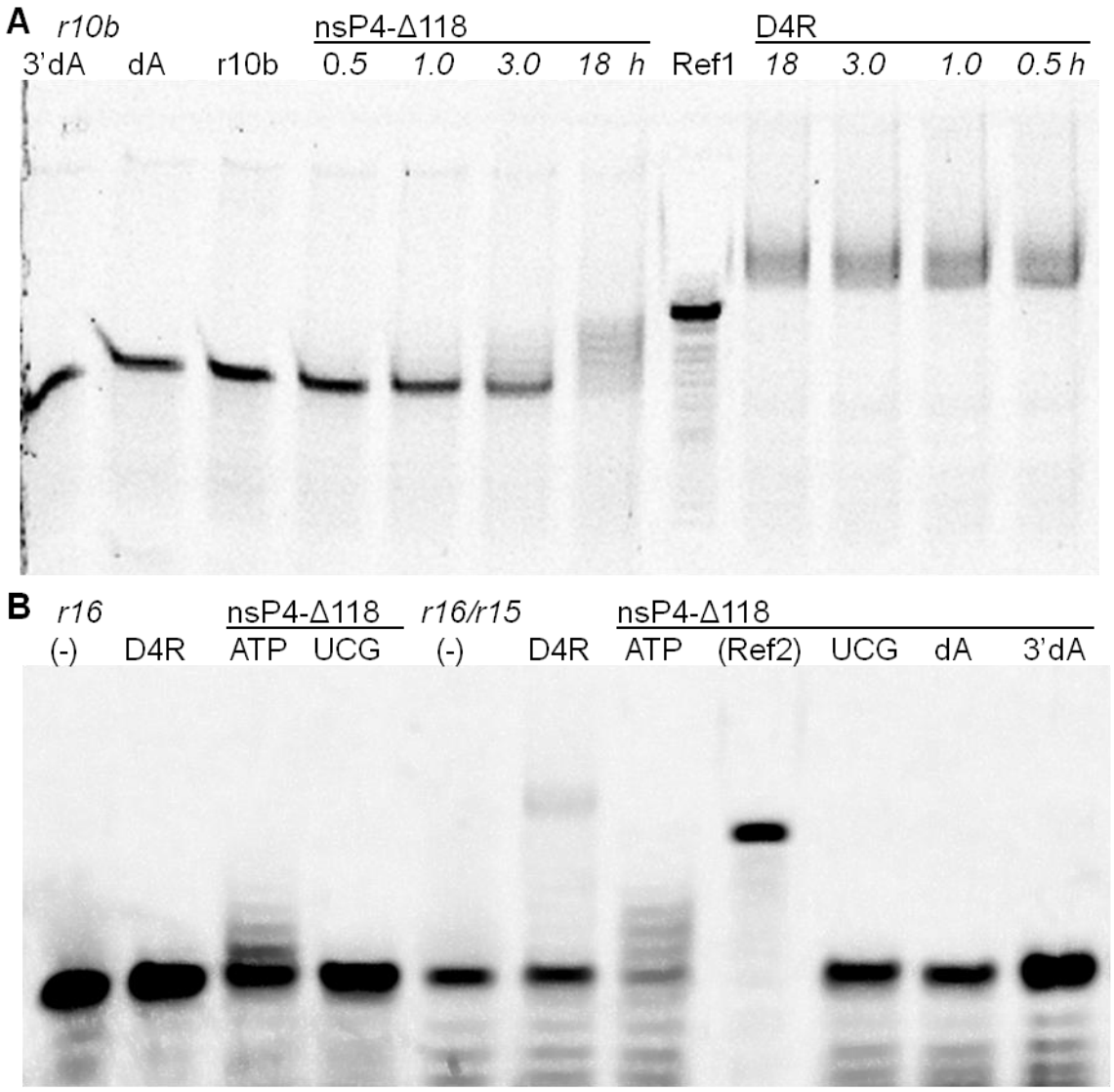

A r10b
RNA (-) D4R 1 2 3 4 5
Detergents at 1.5 c.m.c.
B r10b
RNA (-) D4R 1 2 3 4 mM
LDAO
C r10b
RNA D4R wt GNN
D r10b
RNA D4R (-) 5 6 7 8 9 10 11 12 13
Amine oxide
12-C detergents at 1.5 c.m.c.

Figure 6


A
r10b
nsP4-Δ118
D4R
3'dA
dA
r10b
0.5
1.0
3.0
18 h
Ref1
18
3.0
1.0
0.5 h
B
r16
nsP4-Δ118
r16/r15
nsP4-Δ118
(-)
D4R
ATP
UCG
(-)
D4R
ATP
(Ref2)
UCG
dA
3'dA