A crystal structure of alphavirus nonstructural protein 4 (nsP4) reveals an intrinsically dynamic RNA-dependent RNA polymerase

Yaw Bia Tan[1,2], Laura Sandra Lello[3], Xin Liu[4], Yee-Song Law[1,2], Congbao Kang[6], Julien Lescar[2,5], Jie Zheng[4], Andres Merits[3*], Dahai Luo[1,2,5*]

[1] Lee Kong Chian School of Medicine, Nanyang Technological University, 59 Nanyang Drive, Singapore 636921

[2] NTU Institute of Structural Biology, Nanyang Technological University, 59 Nanyang Drive, Singapore 636921

[3] University of Tartu, Institute of Technology, Nooruse 1, 50411 Tartu, Estonia

[4] Shanghai Institute of Materia Medica, China Academy of Sciences, 555 Zu Chong Zhi Road, Zhang Jiang Hi-Tech Park, Pudong, Shanghai, China

[5]School of Biological Sciences, Nanyang Technological University, 60 Nanyang Drive, Singapore

[6] Experimental Drug Development Centre, Agency for Science, Technology and Research (A*STAR), 10 Biopolis Rd, #05-01/06 Chromos, Singapore 138670

* Correspondence to:

Andres Merits, Email: andres.merits@ut.ee

Dahai Luo, Email: luodahai@ntu.edu.sg

Keywords: Antiviral, nonstructural protein 4, Positive-sense RNA virus, RNA-dependent RNA polymerase.

Running title: Crystal structure of alphavirus RNA polymerase

Abstract

Alphaviruses such as Ross River virus (RRV), chikungunya virus and Venezuelan equine encephalitis virus are mosquito-borne pathogens that can cause arthritis or encephalitis diseases. Nonstructural protein 4 (nsP4) of alphaviruses possesses RNA-dependent RNA polymerase (RdRp) activity essential for viral RNA replication. No 3D structure has been available for nsP4 of any alphaviruses despite its importance for understanding alphaviral RNA replication and for the design of antiviral drugs. Here, we report a crystal structure of the RdRp domain of nsP4 from RRV determined at a resolution of 2.6 Å. The structure of the alphavirus RdRp domain appears most closely related to RdRps from pestiviruses, noroviruses and picornaviruses. Using hydrogen-deuterium exchange mass spectrometry (HDX-MS) and nuclear magnetic resonance (NMR) methods, we showed that in-solution nsP4 is highly dynamic with an intrinsically disordered N-terminal domain. Both full-length nsP4 and the RdRp domain were able to catalyze *in vitro* RNA polymerization. Structure-guided mutagenesis using a *trans*-replicase system identified nsP4 regions critical for viral RNA replication.

Significance Statement

Alphaviruses are a group of viruses that comprise several important pathogens that provoke encephalitis or arthritis in humans. Through its RdRp activity, nsP4 of alphaviruses plays a crucial role in viral RNA replication. We determined the crystal structure of nsP4, disclosing a dynamic structure resembling the core catalytic domain of RNA polymerases from pestiviruses, picornaviruses, and noroviruses. These findings expand our knowledge of the molecular basis for alphavirus RNA replication and should accelerate the design of specific antiviral compounds targeting RNA polymerase activity.

Main Text

Introduction

Alphaviruses (family *Togaviridae*) and flaviviruses (family *Flaviviridae*) comprise numerous important arthropod-borne viruses (arboviruses) that cause diseases in humans. Fast urbanization, massive international travel, and global warming have triggered the re-emergence of several arboviral diseases, such as dengue, Zika and chikungunya fevers, affecting millions worldwide (1-3). Many arboviruses infect humans through *Aedes* mosquitoes which are abundantly present in tropical/subtropical regions and are expanding into temperate climate territories (4). Arbovirus members of the genus *Alphavirus* have been divided into Old World and New World alphaviruses, and cause clinically distinct symptoms (5, 6). New World alphaviruses such as eastern equine encephalitis virus (EEEV) and Venezuelan equine encephalitis virus (VEEV) are primarily associated with encephalitis in horses and humans that causes death or long-term neurological impairment. Despite the generally nonfatal nature of infections, Old World alphaviruses, such as Chikungunya virus (CHIKV), cause a debilitating illness with persisting painful arthritis (7). Between March 2005 and April 2006, a major outbreak in the Reunion resulted in approximately 255,000 cases of chikungunya fever, approximately one-third of the total population of this French island. Human diseases caused by Ross River virus (RRV) and Sindbis virus (SINV) are generally milder, although they are also arthritogenic. Without any approved treatment, re-emerging outbreaks of alphaviruses highlight the unmet needs for vaccines and specific antiviral drugs. Available treatments for alphaviral infection are mostly symptom relief to minimize fever, joint pain, and associated inflammation. Several compounds, such as β-d-N 4-hydroxycytidine (NHC), favipiravir and Compound-A, were claimed to be effective in inhibiting VEEV, CHIKV and SINV RNA synthesis but have thus far not progressed towards clinical trials (8).

Alphaviruses have a positive-sense (+) RNA genome of approximately 11.8 kilobases in length with a 5’ type-0 cap and a 3’ polyadenylated tail. The two open reading frames (ORFs) in alphavirus RNA are flanked by *cis*-acting sequence elements: 5’ and 3’ untranslated regions (UTRs) and subgenomic (SG) promoters (9). RRV and CHIKV invade host cells by binding to the receptor Mxra8 (10), while SINV recognizes NRAMP as the main receptor (11). Receptor binding is followed by clathrin-assisted endocytosis, entry by fusion between the host cell and the viral membrane, and the release of the nucleocapsid containing the viral RNA genome into the cytoplasm. Upon its release, the genome functions as mRNA for polyproteins P123 and P1234 translated from its first ORF, and the expression of P1234 occurs through the readthrough of an inframe opal codon. These polyproteins are precursors for four nonstructural proteins (nsP1-4) that are formed via tightly regulated autoproteolysis and assemble into a viral RNA replication complex (RC). Alphavirus RCs are vesicular organelles called spherules and represent sites for viral RNA synthesis that are protected from the host immune response. The following are the

three types of synthesized viral RNAs: the full-length negative-strand intermediate, the full-length positive-strand genomic RNA and the shorter positive-strand SG RNA. The synthesis of SG RNA is driven by the SG promoter, and the RNA is used to express structural proteins encoded by the second ORF of the alphavirus genome. The newly made RNA genome and structural proteins assemble into progeny virions that are released by budding from the plasma membrane and are thereafter capable of further infection.

Alphavirus P1234 and its processing products contain all virus-encoded components required for RC assembly and possess all enzymatic activities needed for the synthesis and modification of viral RNAs. nsP4 contains RNA-dependent RNA polymerase (RdRp) activity that plays a key role in *de novo* RNA synthesis and terminal adenylyltransferase (TATase) activity (12-14). In addition to nsP4, all other nsPs are also important for effective genome replication (13, 14). Many molecular and structural details of the alphaviral nsP1-3 proteins have already been reported except for the C-terminal hypervariable region of nsP3, which is intrinsically disordered. In contrast, structural information for nsP4 has been lacking, in large part due to its intrinsic flexibility, which is important for its biological functions (15-23). Out of a total of ~610 amino acid residues, the 109 N-terminal residues of nsP4 comprise a functionally poorly defined N-terminal domain (NTD) unique to alphaviruses. Residues 110 to the end of the protein comprise the alphaviral RdRp domain. In general, active nsP4 from alphaviruses shows insoluble recombinant protein expression. Substantial differences exist, however, between nsP4 proteins of different alphaviruses: recombinant nsP4 of SINV possesses significant RNA polymerase activity (24), while the recombinant nsP4 RdRp domain of CHIKV displays only low template elongation activity (12). The weak activities of recombinant nsP4 proteins are due to their poor solubility and tendency to aggregate. As a result, nsP4 has resisted structural determination for several decades. Consequently, it has not been possible to use rational design for the development of compounds to target alphavirus nsP4 activities. The availability of experimentally determined structures of nsP4 from any alphavirus would therefore be regarded as an important step to accelerate the high-throughput screening of inhibitors *in silico,* while the development of robust enzymatic RdRp assays would facilitate testing of these compounds. Furthermore, structural information on nsP4 would also be crucial for a better understanding of the structure and function of RC as a whole, and thus present a strong foundation for drug discovery efforts targeting alphavirus RC.

Here, we report the first crystal structure of an alphavirus RdRp from RRV determined at 2.6 Å resolution. Having a dynamic structure, the protein contains several disordered segments and has some unassigned electron density in the RNA binding tunnel. Complementary structural analysis performed using hydrogen-deuterium exchange mass spectrometry (HDX-MS) and nuclear magnetic resonance (NMR) helped better understand nsP4 structural dynamics and the relative conformations adopted by its NTD and RdRp domains in solution. We also discovered

that both full-length nsP4 (nsP4FL) and its RdRp domain can catalyze RNA polymerization in test-tube reactions. Finally, structure-guided mutagenesis experiments were used to gather functional insights into nsP4 in alphaviral RNA replication.

Results

Determination of the structure of nsP4 from RRV reveals a highly dynamic RdRp

We obtained soluble recombinant nsP4FL and RdRp proteins of RRV (**Figure 1a**) using the *E. coli* expression system. The latter recombinant protein, lacking the N-terminal 109 amino acids of nsP4, was purified as a soluble active enzyme that crystallized. Crystals of the RdRp protein initially diffracted to a maximum resolution of 3.8-4.0 Å. To improve diffraction, we substituted hydrophilic glutamine residues (Q192 and Q196) into the RdRp that are presumably exposed at the surface of the folded protein with alanines. The obtained surface-entropy double mutant $RdRp^{SER3}$ formed crystals with increased size and allowed native X-ray diffraction intensities to be collected to a maximum resolution of 2.60 Å (**Table 1**). Because no solution was found using molecular replacement (MR), L-selenomethionine (SeM)-labeling of $RdRp^{SER3}$ was used for single-wavelength anomalous diffraction (SAD) phasing of data collected at the Se K edge at 2.83 Å resolution using a crystal isomorphous to native $RdRp^{SER3}$ crystals (**Table 1**). A total of 401 out of 502 residues could be built, including twenty alpha-helices, four 3_{π} helices and ten β-strands (**Figures 1b-c, Table 1**). Of note, no bound ligands, such as RNA, ATP-Mg^{2+} or trimethylamine N-oxide (TMAO), were identified in the $RdRp^{SER3}$ crystal structure despite their presence under crystallization conditions (see Methods). An anomalous Fourier difference map using data collected at the Se K edge allowed unambiguous mapping of eleven Se-Met residues (out of a total of thirteen) onto the structure, giving confidence of the correctness of the atomic model (**Figure S2**). In the present crystal form, $RdRp^{SER3}$ adopts the right-hand structure characteristic of RdRps with fingers-palm-thumb subdomains sequentially arranged from the amino-to-carboxylic ends. The fingertips (subdivided into the index, pinky, ring and middle finger subdomains) extend from the fingers towards the thumb to form the RdRp-distinctive encircled ring conformation. To fold into this topology, the nsP4 index finger extends its N-terminal helix horizontally towards the thumb to which it docks via a hydrophobic interface (**Figures 1b-c**). While the core catalytic domain including the palm region is well defined with continuous electron density (**Figures 1b-c**), the 101 missing residues from the current model (**Figure 2a**) belong to the N-terminal fingertips and C-terminal thumb subdomains: index finger (residues 135-184; denoted as “flex” region hereafter), pinky and ring fingers (residues 210-213 and **Figure S1a** 291-307, respectively), and C-terminus of the protein (residues 537-542, 575-579, 588-593 and 600-611) (**Figure 2a; Figure S1b**). The flex region is suspected to insert itself into the central RNA binding tunnel, where some connected pieces of electron densities resembling a polypeptide chain (but not RNA bases) could be observed directly above GDD motif C and motif E (**Figures 1 and 3;**

Figure S1c). The relatively large number of missing residues of the RRV RdRp structure suggests a highly dynamic structure adaptable to many conformational changes that occur during the RNA replication cycle, which are not visible in the state captured in the present crystal form.

The structural dynamics of nsP4FL and its domains in solution

To gain insights into the NTD:RdRp domain interaction interface and the interaction of the protein with its ligands, we explored the molecular motions of recombinant nsP4 in solution. HDX-MS is a proteomics approach allowing deuterated water molecules (D_2O) to exchange deuterium atoms with hydrogens located in residues decorated on the solvent-exposed surface of nsP4. An increased percentage of deuteration (D%) reflects the flexibility of the RRV RdRp structure in solution that can be correlated with the temperature B-factor obtained in the crystalline form. Due to the incompleteness of the $RdRp^{SER3}$ crystallographic structure, we generated structure models of nsP4FL, RdRp and the NTD using template-based homology models (predicted structures are denoted as “model(s)” hereafter) using Robetta ternary-structure prediction.

Overall, the HDX profiles of both nsP4FL and RdRp were almost identical (**Figure 2a**). To illustrate the structural dynamics in 3D, we plotted the HDX rates to the models for comparisons (**Figures 2b-d**). In the nsP4FL structural dynamics model, the palm and thumb subdomains as well as the middle and ring fingers are largely protected from the exchange and therefore constitute a buried structural core. In contrast, the flexible N-terminal pinky fingertip harboring the motif G and flex region close to the thumb C-terminus (residues 560-585) became highly deuterated in the HDX experiment (**Figures 2a, 2b**). Thus, these are unprotected regions, a finding that is in agreement with their high mobility found in the crystal structure (**Figure 1**). The HDX protection profile of the middle and ring fingers and the last 20 amino acids at the protein C-terminus is in contrast with the RRV RdRp crystal structure, where these regions had weak or missing electron densities. In the nsP4FL protein, the linker to NTD appeared to have higher D% values, indicating increased dynamics of the NTD and polymerase domains in solution, possibly indicating the presence of various relative conformations in solution. Based on the high D% HDX profile observed, the NTD in solution appears quite flexible, except at residues 53-74, where a predicted helix-loop-helix (HLH) secondary structure motif is located (**Figure 2b)**. Overall, the similar HDX profiles of the NTD protein by itself and within the nsP4FL indicate that there is no intramolecular interaction between the NTD and the RdRp region of the nsP4FL.

Given its low molecular mass of 12.6 kilodalton, the recombinant NTD is amenable to structural characterization via NMR in solution. The NTD contains 107 amino acids with an expected number of 99 cross-peaks (excluding Pro residues) in the ^{1}H-^{15}N-HSQC spectrum. From these, approximately 60 cross-peaks could be observed, making backbone resonance assignments challenging (**Figure S3**). The narrow dispersion of the cross-peaks in the spectrum

suggests that the NTD contains flexible and helical regions that support the NTD homology model and are in agreement with the HDX profile (**Figure 2b**).

Comparison of the RRV RdRp structure with that of other viral RdRps

The nsP4 polymerase domain of RRV adopts a conserved tertiary structure harboring features common to viral RdRps. A structural homology search of the Protein Data Bank (PDB) database (25) revealed that the closest structural homologs of RRV RdRp are RdRps from the human Norwalk virus from the *Caliciviridae* family (NV; PDB 5TSN) (26), enterovirus-71 (EV71, PDB 6KWQ) from the *Piconaviridae* family (27) and classical swine fever virus (CSFV, PDB 5Y6R) as well as bovine viral diarrhea virus (BVDV) from the *Flaviviridae* family (28) (**Figure 3**). In RRV nsP4, the conserved polymerase motifs forming a fully encircled right-hand fold are arranged in the primary structure as G-F-A-B-C-D-E, showing evolutionary conversation with the *Picornaviridae* (but not to the birnavirus) RNA polymerase (29) (**Figures 1a, 3; Figure 3**). The core polymerase domain at the palm of viral RdRp contains structurally well-conserved motifs A, B, C and E and is the most highly conserved region in terms of both structure and sequence (**Figure 3**). The Phe residue on motif E of nsP4 points towards the GDD residues in the β-turn of motif C, as in other viral RdRps (**Figure 3a**; yellow motif E and red motif C). In contrast, motif D is likely to constitute an alphavirus distinctive feature that separates it from other viral RdRps. Residues from motif D of nsP4 adopt an elongated beta-strand conformation next to motifs A-C, which form a Greek-key-like beta-sheet motif. The least conserved motifs lie in the pinky finger (motif G) and ring finger (motif F) subdomains. The fingertips are typically folded in beta sheets extending towards the thumb like a lid over the palm. The thumb subdomain of nsP4 shows poor structural homology with RdRp homologs, and its C-terminal region is folded in a different orientation without beta-sheet formation. These structural differences make RRV RdRp an alpha-helix-rich structure compared to its closest viral RdRp homologs.

RRV RdRp displayed polymerase activity and RNA species binding preferences

A previously described fluorescent RNA-based assay (12) was optimized to increase robustness and sensitivity and used for analysis of the RNA polymerase activities of the obtained recombinant proteins. In this assay, RNA intermediates (RIs) result from the incorporation of 1-7-A residues complementary to the polyuridine overhang on the T1 RNA template, while the RNA product (RP) is the fully extended RNA obtained by incorporation of 8-A residues (**Figure 4a**). Reactions were run for 24 hours (h) with time-point sampling at 0, 2, 4, 6, 8, 10 and 24 h.

The recombinant nsP4FL, RdRp and $RdRp^{SER3}$ were enzymatically active and capable of extending the T1 template. The RIs were observed from 2 h onwards for all of these enzymes. Then, the polymerase reaction remained largely stagnant for over 10 h and subsequently accelerated to form a mixture of RPs and RIs at 24 h (**Figure 4b**). After 24 h, nsP4FL ranks as

the most efficient for the synthesis of RPs and RIs with 6- or 7-A incorporations, followed by RdRp^{SER3} and RdRp. The conclusion that the activity of the surface-engineered mutant RdRp^{SER3} exceeded that of its wild-type (wt) counterpart was based on the observation that for the mutant enzyme, no RIs of less than 4-A added were detected at 24 h, while for its wild-type RdRp, a faint ladder consisting of the template and RIs with 1-3-A added was still visible even at this ending time point (**Figure 4b**). In contrast to the active enzymes, GNN formed much smaller amounts of RIs, and the level of RPs remained below the detection limit even at the 24 h time point.

No primer extension was observed for the recombinant protein corresponding to the nsP4 NTD (**Figure 4c**). This allowed probing of RdRp and RdRp^{SER3} for possible cooperative interactions with the NTD. In the corresponding polymerase assay, these enzymes were mixed with NTD using 1:1, 1:2 and 1:5 molar ratios. Under these conditions, no or very few RIs were detected at the 2 h time point, and it was concluded that the presence of an individual NTD delayed the polymerization reaction. No increase in polymerase activity of RdRp^{SER3} or RdRp was observed using various molar ratios of NTD at any time point (**Figure 4c**). Rather, an increasing molar amount of NTD had a slight negative impact on the polymerase activity. The inhibitory effect was more evident for RdRp, in which case a significant amount of T1 template remained unused even at the 24 h time-point at 1:1 and 1:2 RdRp:NTD molar ratios (**Figure 4c**). Thus, the increased activity observed for nsP4FL could not be reproduced by adding NTD to RdRp *in trans*.

Next, we used fluorescence polarization (FP) to assess binding between the nsP4FL, RdRp or NTD of RRV and RNA. Binding was measured with various species of fluorescently labeled RNAs (**Figure 4d**); the impact of ATP added to the reaction mixture on complex formation was also analyzed. nsP4FL preferentially binds to single-stranded RNA (ssRNA (U)) with a K_d of 125.7 ± 6.8 nM (**Figure 4d, left panel**). Binding to a hairpin RNA with a 5' overhang (hpRNA (R)) was less efficient, with K_d=332.3 ± 11.6 nM (**Figure 4d, right panel**). The addition of ATP only caused subtle changes to the measured K_d values for both RNA species. The recombinant protein corresponding to the isolated NTD did not bind to either RNA species. Unexpectedly, the binding kinetics of RdRp could not be determined due to poor curve fitting. However, when compared with nsP4FL, a prominent 10- to 100-fold decrease in binding affinity was observed for RdRp for both RNA species (**Figure 4d**), suggesting an indirect contribution of NTD to RNA binding, possibly via *cis* stabilization of the RdRp catalytic domain in nsP4FL.

Activities of RRV replicase are affected by structure-guided mutations in nsP4

Having determined the crystal structure of RRV RdRp, we next analyzed the functional importance of several of its key structural features. As the introduction of unfavorable mutations into an infectious clone tends to result in no data (lethal phenotype) or in the rapid appearance of revertants or compensating mutations (30, 31), we instead used a novel two-component *trans*-replication system (**Figure 5a**), which has extremely high sensitivity (Lello et al, unpublished

manuscript). Briefly, the system is based on the ability of separately expressed P123 and nsP4 of alphaviruses to form a functional RNA replicase. Coexpression of P123 and nsP4 with replication-competent template RNA also triggers the formation of RCs (spherules) similar to those formed in virus-infected cells (32), i.e., represents a relevant tool to study alphavirus RNA replication. Plasmids for the expression of RRV P123 and nsP4 as well as the corresponding template RNA have been described (Lello et al, unpublished manuscript). The template RNA contains sequences coding for the following two reporters: firefly luciferase (Fluc) under the genomic promoter and *Gaussia* luciferases (Gluc) under the SG promoter (**Figure 5a**). For simplicity, the synthesis of full-length RNA serving as the template for Fluc expression is hereafter termed "replication", and the synthesis of RNA serving as the template for Gluc expression is termed "transcription". The efficiency of replication and transcription were estimated by the variations in corresponding reporter expression (33).

Mutations were introduced into nsP4 of RRV (**Figure S6**), and the activities of the replicases harboring mutant nsP4 proteins were compared with those of the wt replicase. Substitution of Arg63 and Arg67 with Ala residues did not result in significant changes in RNA replication, while transcription was slightly but significantly ($p<0.001$) reduced (**Figure 5b, Figure S6**). In contrast, only minimal transcription activity was detected for replicase harboring the Strep-FLAG tag after residue 92. The same tag inserted after residue 109 had a somewhat smaller impact; however, it still reduced the transcription level ~230-fold when compared with wt replicase ($p<0.0001$) (**Figure 5b**). Thus, replacement of charged residues with Ala residues in the NTD of RRV nsP4 was well-tolerated compared to insertion of peptide tags.

Replacement of Cys125 with the Val residue had no significant effect on the replication and transcription activities of the RRV RNA replicase. However, replacement of the same residue with Asp resulted in an ~60-fold decrease in replication and an ~14-fold decrease in transcription ($p<0.0001$ in both cases). Analysis of the nsP4 structure revealed that residues 136-185 of RRV nsP4 form a flexible loop. Replacement of this region with a Strep-FLAG tag resulted in the complete loss of RNA replicase activity, indicating a crucial function played by this region. This conclusion was corroborated by neutralizing the charges in this loop. Replacement of Asp143 and Asp150 with Ala resulted in a large ~43-fold decrease in replication and an even larger ~310-fold decrease in transcription ($p<0.0001$ in both cases). Interestingly, a similar replacement of Asp146 and Asp153 residues had a smaller impact: replication was only reduced ~5-fold ($p<0.0001$) (Figure 5b). Likewise, substituting charged amino acid residues with Ala in the thumb subdomain resulted in a decrease in RNA replicase activity: replacement of Asp543 and Asp545 caused an ~25-fold decrease in replication and an ~35-fold decrease in transcription ($p<0.0001$). The effect of substitution of the Arg546 and Arg548 residues caused a less prominent ~11-fold decrease in replication ($p<0.0001$), while the impact on transcription was more pronounced (~120-fold, $p<0.0001$) (**Figure 5b**). Finally, the introduction of a stop codon after residue 601, resulting in the

loss of the ten C-terminal residues of nsP4 that could not be resolved in the present crystal structure, resulted in almost complete loss of RNA replicase activity. The insertion of a Strep-FLAG tag after residue 601 had a similar effect (**Figure 5b**). Taken together, these data indicate that the presence of the ten C-terminal residues and their location relative to the rest of nsP4 are crucial for RRV nsP4 activity.

Discussion

The RRV RdRp crystal structure and its evolutionary connection with other known viral RdRps

The structure of alphavirus RdRp features an overall conserved fold in its palm subdomain but rather surprisingly not in the fingers and thumb subdomains. This observation could be derived from crystal lattice packing forces that lead to an alteration of either an inherently flexible structure or some partial misfolding of the recombinant protein. However, an alternative is that the crystal structure of RRV RdRp captured here could result from divergent evolution leading to the peculiar folds of its fingers and thumb at odds with the nonalphaviral RdRps. These tertiary-structural differences could indicate a new viral mechanism for functional adaptations in mammalian and insect hosts.

One of the most conspicuous changes visible from the crystal structure is the elongated beta-strand adopted by motif D next to the catalytic GDD motif C located in the palm subdomain. This structure contrasts with the canonical beta-strand-loop motif found in other viral RdRps and whose high flexibility is important for correct NTP selectivity (29). This observation may also explain the absence of bound ATP or dATP in the central catalytic cleft above the palm despite their presence in the crystallization buffer. Moreover, both the flexible fingers and the C-terminus of the thumb adopt noncanonical directions compared to the well-conserved palm. One obvious difference lies in the ring-finger-carrying motif F extending towards the solvent rather than inserting below the middle finger. Assignment of residues in the N-terminal helix of the index finger that make hydrophobic interactions with the thumb remains ambiguous due to residues in the flex region that are missing in the resolved structure. The characteristic and structurally conserved regions among the viral RdRps in the palm subdomain and the center-facing fingertips are essential for RNA priming and synthesis (29, 34).

Solution structural dynamics revealed the relationship between the NTD and RdRp domains of nsP4.

The combination of HDX-MS, NMR and Robetta homology modeling provided insights into the structural dynamics of RRV nsP4 in solution. The highly dynamic regions were identified mostly at the fingers and thumb of the RdRp and in the NTD. Dynamics are potentially required by the alphavirus RdRp before converting from its apoenzyme state to the stabilized elongation

complex with RNA, such as in EV71 RdRp from *Piconaviridae* (27). The homology model of NTD features a mixture of disordered regions and a predicted helical secondary structure consisting of an HLH motif (aa 53-74) (**Figure 2b**). Despite the challenging assignments of the cross-peaks to the NTD residues in the incomplete NMR spectra, these missing cross-peaks provided clues that the NTD may have active conformational exchanges. The HDX-protected region at the HLH motif and the NMR-detected cross-peaks both suggest secondary structure elements in the NTD, which agree well with the NTD homology model. Furthermore, the HDX-MS profiles indicate that the NTD does not interact with the catalytic RdRp domain of nsP4 but possibly with other proteins, such as viral nsP1-3 and/or host factors. The NTD of the viral RdRp is an adaptive domain that often undergoes divergent structural evolution to perform novel functions in various virus families or genera to target different hosts (34). Thus, other unknown mechanisms unique to the alphavirus NTD might be at play.

The importance of NTD for RNA synthesis

Analyses of RNA polymerase activity and RNA binding kinetics have provided several clues about the functions of the nsP4 NTD. Based on HDX-MS data, the NTD does not directly interact with the RdRp domain. It is dispensable for RNA polymerase activity yet appears to enhance *de novo* RNA synthesis (**Figure 4b**). The highest accumulation of fully extended RPs was observed when the NTD and RdRp domains were present *in cis* (i.e., in the context of the RRV nsP4FL). Thus, the NTD may serve as a cofactor to the RdRp domain to effectively dock onto the RNA template, likely in a cooperative manner. Alternatively, the NTD could be required for proper folding of the polymerase domain. The NTD could also constitute an additional layer to regulate alphavirus RNA polymerase activity in a way reminiscent of how the fidelity of BVDV RdRp is modulated during RNA synthesis (34). Intriguingly, in the absence of NTD, recombinant $RdRp^{SER3}$ had higher RNA polymerase activity than its wt counterpart, possibly due to its lowered surface entropy preventing precipitation. Moreover, the addition of detergents such as lauryldimethylamine oxide (LDAO) and Triton X-100 in polymerase assay (PA) buffer may also provide suitable conformational flexibility to RdRp to prevent aggregation and to maintain its active state.

Mutagenesis mapped functional regions of nsP4 crucial for virus RNA replication

nsP4 is key to viral RNA synthesis and essential for virus replication. To characterize the function of selected regions of the alphavirus RdRp for RNA replication and transcription in the cellular environment, a biologically relevant *trans*-replicase system was used. The importance of

the NTD for viral RNA replication was reflected by the strong negative impact caused by the Strep-tag inserted at various sites within the first 110 residues of nsP4. The flex region and the index fingertip are equally critical for viral RNA replication and transcription. We also found that several charged residues in the flex region (Asp143 and Asp150), in the fingers and at the helix tip of the thumb (Arg543 and Arg545) are critical, presumably by providing electrostatic interactions either within the RdRp or/and with an RNA template. Finally, the last ten thumb residues protected from solvent exchange in the HDX experiment were shown to be indispensable for RNA replicase activity. These observations are consistent with the presence of an intramolecular interface within alphavirus nsP4 critical for viral RNA replication. In this respect, the C-terminus of the alphavirus RdRp could constitute a regulatory motif similar to that found in the flavivirus RdRp (29).

Although nsP4 is the most conserved nsP, several known properties of recombinant nsP4 of SINV (13, 24) indicate higher activity than what was observed for the RRV counterpart. We have also found that nsP4 of SINV and RRV are not fully functionally interchangeable: RRV P123 can form active RNA replicase with nsP4 either from RRV or SINV, while SINV P123 forms active replicase only with its cognate nsP4 (Lello et al, unpublished manuscript). Based on sequence similarities, alphaviruses are divided into different complexes and clades (35). RRV is a member of the Semliki Forest virus complex, while SINV, the type member of the genus, belongs to the western equine encephalitis complex, making nsP4 of SINV a good candidate for validating to what extent findings made for nsP4 of RRV are applicable to nsP4 proteins of alphaviruses in general. Matching experiments performed using SINV *trans*-replicase provided results very similar to that of RRV (**Figure 5c**). The exceptions were mutations introduced in the NTD, which is the least conserved region of nsP4. The negative impact of the Strep-tag insertion was much smaller in the case of SINV nsP4: if the tag was introduced after residue 108, this insertion resulted in a minor ~7-fold reduction in replication and a ~5-fold reduction in transcription. From a practical point of view, inserting an affinity tag at this position could be used for affinity pull-down of a functional SINV nsP4 and complexes formed by this protein. These findings are also consistent with our data, indicating that the NTD of nsP4 is flexible and not directly involved in RNA synthesis. In contrast, all mutations introduced into the RdRp domain of SINV nsP4 caused effects very similar to those observed for RRV. The largest difference between the two viruses was observed for the substitution of the Asp145 and Asp152 residues with Ala residues, which resulted in prominent reductions in both replication and transcription for SINV (by factors of ~38-fold and ~15-fold, respectively); these effects were significantly larger than those observed for corresponding mutations in RRV nsP4 ($p<0.001$ for replication and $p<0.01$ for transcription). Nevertheless, our data clearly indicate that there is high functional conservation in the RdRp

domain even for distantly related alphaviruses, likely resulting from the very similar structures of these proteins. Therefore, it can be assumed that structural information for RRV RdRp is highly applicable for other alphaviruses and can be used for the rational design of drugs targeting many, or possibly all, medically important alphaviruses.

Materials and Methods

Molecular cloning

The sequences encoding nsP4FL (aa 1-611) and its RdRp domain (aa 110-611) were obtained using PCR on an RRV-T48 cDNA plasmid template. The PCR fragments were subcloned into the pSUMO-LIC vector using a ligation-independent cloning method in which the expressed recombinant protein had an N-terminal hexahistidine-linked small ubiquitin-like modifier protein (His_6-SUMO) fusion tag which can act as solubility enhancer. The RdRp surface entropy reduction construct ($RdRp^{SER3}$) harboring two glutamine-to-alanine substitutions (Q192A and Q196A), nsP4FL polymerase-weakend mutant (GDD residues mutated to GNN; named as GNN) and construct for expression of NTD (aa 1-109) were obtained using site-directed mutagenesis with the Toyobo KOD-FX Neo cloning kit. These plasmids were propagated using the *E. coli* TOP10 strain.

Human RNA polymerase I (HSPolI) promoter-based plasmids for the production of replication-competent RNA templates of RRV (HSPolI-FG-RRV) and SINV (HSPolI-FG-SINV) have been previously described (36). Plasmids encoding P123 of RRV (CMV-P123-RRV) and SINV (CMV-P123-SINV) as well as plasmids expressing the corresponding nsP4 fused to ubiquitin at their N-termini (CMV-ubi-nsP4-RRV and CMV-ubi-nsP4-SINV) will be described elsewhere (Lello et al, unpublished manuscript). Variants of CMV-ubi-nsP4-RRV and CMV-ubi-nsP4-SINV encoding the polymerase negative variant of nsP4 with GDD to GAA mutation in the polymerase active site were obtained using site-directed mutagenesis and designated CMV-ubi-$nsP4^{GAA}$-RRV and CMV-ubi-$nsP4^{GAA}$-SINV, respectively. Variants of CMV-ubi-nsP4-RRV and CMV-ubi-nsP4-SINV containing mutations listed in Table 2 were constructed using synthetic DNA fragments (Genscript). Sequences of all plasmids were verified using Sanger sequencing and are available from the authors upon request.

Recombinant protein expression and purification

The recombinant proteins were expressed using the phage-resistant *E. coli* Rosetta II T1R strain developed in-house. Bacteria were grown in 2-4 L Luria-Bertani (LB) Millar Broth (Axil Scientific)

supplemented with 50 µg/mL kanamycin (Gold Bio) and 34 µg/mL chloramphenicol (Gold Biotechnology) in a 37°C shaking incubator (INFORS) until reaching an optical density (OD_{600}) measurement between 0.8 and 1.0. The shaker's temperature was then adjusted to 18°C, expression was induced by the addition of isopropyl β-D-1-thiogalactopyranoside (IPTG) to a final concentration of 0.5 mM, and expression was carried out for 16-24 h. For selenomethionine labeling, an adapted protocol from Molecular Dimension was used for extended expression at 18°C for 16 h (37). All the expression cultures were harvested by centrifugation for 10 min at 5000 x g, and bacterial pellets were frozen and stored at -80°C until recombinant protein purification. For $^{13}C/^{15}N$-labeled NTDs, expression and induction were performed as described above except for the use of M9 minimal medium supplemented with 2 g/L ^{13}C-glucose and/or 1 g/L $^{15}NH_4Cl$.

The purifications of the recombinant nsP4 proteins (labeled or unlabeled) were conducted at 4°C using gentle treatments to ensure the structural integrity of the obtained samples. The bacterial pellet was suspended in IMAC A buffer (2X PBS, 500 mM NaCl, 5 mM β-mercaptoethanol and 10% glycerol) supplemented with 1X protease inhibitor cocktail (DMSO free, MedChemExpress), 0.1-0.25% Tween-20 (Bio Basic) and 5 mM imidazole (Sigma Aldrich). The suspension was briefly sonicated before homogenization through 1000-bar pressurized PANDAPLUS2000 (GEA). The crude lysate was clarified via centrifugation in an ultracentrifuge XPN-100 (Beckman Coulter) at 235,000 x g. The clarified lysate was incubated with Ni-NTA FF beads (Bio Basic). Beads with bound recombinant protein were transferred to a column and subsequently washed with 15 column volumes (CVs) of IMAC A2 buffer (2X PBS, 500 mM NaCl, 5 mM β-mercaptoethanol, 10% glycerol and 10 mM imidazole). The recombinant protein was eluted in 5 CV of IMAC B buffer (2 X PBS, 500 mM NaCl, 5 mM β-mercaptoethanol, 10% glycerol and 300 mM imidazole) and treated with Ulp1-SUMO-protease (1:25 w/w ratio of protease: protein) for 16-20 h at 4°C in the presence of 250 mM NaCl to remove the N-terminal His_6-SUMO tag. The cleaved recombinant protein was loaded onto a Heparin 5 HP column (Cytiva) at an adjusted concentration of 100 mM NaCl and subsequently eluted using 900 mM NaCl. The eluate was further separated through size-exclusion chromatography (SEC) in a HiLoad S75 or S200 16/600 Superdex® column (Cytiva) on an AKTA Pure 25 M fast protein liquid chromatography (FPLC) system (Cytiva).

Products obtained at all purification steps were examined for their purity by SDS-PAGE. The fractions corresponding to the elution peak were selected and concentrated to 2-50 mg/mL through the use of Amicon® Ultra-15 Centrifugal Unit (Merck), frozen and stored at -80°C for downstream experiments.

X-ray crystallography

The purified recombinant nsP4 proteins were screened with Hampton Research commercial screening kits (PEG-ION HT, PEG-RX HT, INDEX HT, SALT-RX, CRYSTAL-SCREEN HT, etc.) and tools at 121 μM with and without ligands, aided by a Mosquito robotic liquid handler (TTP Lab) and periodically imaged in the RockMaker (Formulatrix®) maintained at 20°C.

Under the optimal conditions, 121 μM protein was incubated with 2-5 mM $MgCl_2$, 2-5 mM deoxyadenosine triphosphate or adenosine triphosphate (dATP/ATP) and 135 μM synthesized self-annealing duplex-RNA R2 (dsRNA: 5' GCAUGGGCGCCC 3') for 1-2 h. The protein-RNA mixture developed into a crystal size of 0.2-0.5 mm under crystallization conditions of 24% (w/v) PEG 3350, 0.1 M Na HEPES/Tris HCl pH 8.0, 0.2 M ammonium acetate, and 3% trimethylamine N-oxide (TMAO) in 5-7 days. The crystals were dehydrated under the same conditions but supplemented with 30% (w/v) PEG 3350 and 10% glycerol as cryoprotectants before snap freezing on a cryoloop in liquid nitrogen.

The X-ray diffraction datasets were collected at Taiwan Photon Source (TPS) Beamline TPS-05A, Japan Spring-8 Beamline BL32XU (38, 39) and Australian Synchrotron Beamline MX2 (40, 41), all with a nitrogen stream of 100 K. The data were collected for at least 200 degrees with an oscillation step of 0.1-0.5 degrees with a 70-90% attenuated beam source and 260-300 mm detector distance for the native method (wavelength: ~1 Å) and selenium single-wavelength anomalous diffraction (Se-SAD) method at the selenium K-edge (wavelength: ~0.979 Å). The data collected were preprocessed on-the-fly by an XDS-based program (42) and further processed in the *CCP4 7.1* (43) and *Phenix* 1.19 software suites (44) for SAD/MR phasing and automated model building. The crystal structures were refined to the highest quality in *Coot* and *Phenix* (44, 45) to generate Table 1 summarizing the crystallographic parameters.

Hydrogen-deuterium exchange detected by mass spectrometry

HDX-MS analysis: 5 μM recombinant nsP4FL or RdRp proteins in a buffer containing 50 mM Na HEPES, pH 7.4, 150 mM NaCl, 5% glycerol, 5 mM MgCl2, and 2 mM DTT were used to carry out HDX reactions at 4°C. Four microliters of protein/protein complex with ligand/peptide were diluted in 16 μL D_2O in exchange buffer (50 mM Na HEPES, pH 7.4, 50 mM NaCl, and 2 mM DTT), incubated for various HDX time points (e.g., 0, 10, 60, 300, 900 s) at 4°C and quenched by mixing with 20 μL of ice-cold 1 M tris(2-carboxyethyl)phosphine (TCEP) 100 mM NaH_2PO_4. Each quenched sample was immediately injected into the LEAP Pal 3.0 HDX platform. Upon injection, samples were passed through an immobilized pepsin column (2 mm × 2 cm) at 120 μL $min^{-1,}$ and the digested peptides were captured on a C18 PepMap300 trap column (Thermo Fisher) and

desalted. Peptides were separated across a 2.1 mm × 5 cm C18 separating column (1.9 mm Hypersil Gold, Thermo Fisher) with a linear gradient of 4-40% CH3CN and 0.3% formic acid over 6 min. Sample handling, protein digestion and peptide separation were conducted at 4°C. Mass spectrometric data were acquired using a Fusion Orbitrap mass spectrometer (Thermo Fisher) with a measured resolving power of 65,000 at m/z 400. HDX analyses were performed in duplicate or triplicate with single preparations of each protein state. The intensity-weighted mean m/z centroid value of each peptide envelope was calculated and subsequently converted into a percentage of deuterium incorporation. Statistical significance for the differential HDX data was determined by an unpaired Student's t-test for each time point, a procedure that is integrated into the HDX Workbench software (46). Corrections for back exchange were made based on an estimated 70% deuterium recovery and accounting for the known 80% deuterium content of the deuterium exchange buffer.

Peptide Identification: Product ion spectra were acquired in data-dependent mode with the top eight most abundant ions selected for the production analysis per scan event. The MS/MS data files were submitted to Proteome Discover 2.4 (Thermo Fisher) for high-confidence peptide identification.

Data Rendering: The HDX data from all overlapping peptides were consolidated to individual amino acid values using a residue averaging approach. Briefly, for each residue, the deuterium incorporation values and peptide lengths from all overlapping peptides were assembled. A weighting function was applied in which shorter peptides were weighted more heavily and longer peptides were weighted less. Each of the weighted deuterium incorporation values was then averaged to produce a single value for each amino acid. The initial two residues of each peptide, as well as prolines, were omitted from the calculations.

NMR data collection and analysis

The NMR spectrum of the NTD was collected at 298 K on a Bruker Avance II magnet with a proton frequency of 600 MHz equipped with a cryoprobe. A^{13}C/^{15}N-labeled (0.3 mM) NTD in a buffer that contained 20 mM Na-PO_4, pH 6.5, 150 mM NaCl, and 1 mM DTT was used in data acquisition. Experiments including 3D-HNCACB, HNCOCACB, HNCA, HNCOCA and HNCO were collected for resonance assignment. It is challenging to obtain the assignment due to the absence of signals from quite a few residues. The data were processed using NMRPipe (47) and visualized using NMRView (48).

Structure analysis

The 3D structure of RRV RdRp was queried against the full PDB database through the Dali server (Dali; http://ekhidna2.biocenter.helsinki.fi/dali/) for matching structural homologs. The amino acid sequences of the selected Dali top hits were aligned through the NCBI COBALT Multiple Alignment Tool (https://www.ncbi.nlm.nih.gov/tools/cobalt/cobalt.cgi) or MUSCLE (https://www.ebi.ac.uk/Tools/msa/muscle/) to generate a multiple sequence alignment (MSA) file. The sequence-conservation map overlay with the RdRp secondary structure was generated through ESPRIPT (https://espript.ibcp.fr) (49) with the input of the MSA file and $RdRp^{SER3}$ PDB. The nsP4FL, RdRp and NTD homology models of RRV nsP4 were predicted based on comparative modeling (CM) mode with the template of the best-refined $RdRp^{SER3}$ structure by sampling for 5 models through the Robetta server (https://robetta.bakerlab.org/). The homology models were superimposed on other experimental crystal structures for comparison and analysis. These homology models were colored and annotated for the HDX-MS experiment. All structural displays were created in PyMOL 2.4.0 (Schrodinger LLC.) (50).

RNA polymerase assay

The RNA polymerase assay was conducted using 1 µM recombinant nsP4 proteins and 50 nM internally labeled fluorescent T1 hairpin-RNA template (5' FAM-UUUUUUUUUAGGACCGCAUUGCACUCCGCGGUCCUAA 3') in PA buffer containing 25 mM HEPES pH 8.0, 2.5 mM $MnCl_2$, 5.0 mM $MgCl_2$, 20 mM KCl, 5 mM TCEP, 3 mM LDAO, 0.01% Triton X-100, and 2 U/µL murine RNase inhibitor (NEB) in nuclease-free water (HyClone). The reactions were incubated for 0 to 24 h at 25°C in the dark. At selected incubation time points, the reactions were quenched using 2X RNA loading dye (95% formamide, 0.02% SDS, 0.01% xylene cyanol, 0.02% bromophenol blue, and 1 mM EDTA). The samples were loaded onto 17.5% urea-PAGE and run at 200 V for 1.5-2 h. The urea-PAGE was imaged on a Bio-Rad Gel-Doc XR+ Imager using the Fluorescence Blot ALEXA488 protocol to determine the RNA polymerase activity.

RNA binding measurement using fluorescence polarization

Twelve serial dilutions (dilution factor 0.6) were prepared from 12.5 µM stocks of the recombinant nsP4 proteins. The obtained dilutions were incubated for binding with 10 nM hairpin-RNA R10b with a 5' overhang (5' GUGCGUACCUAUUUAGGACCGCU-FAM-CGGCGGUCCUAA 3') or 10 nM ssRNA R11 (5' FAM-U_{14} 3') with and without ATP in fluorescence polarization (FP) buffer consisting of 25 mM Na HEPES pH 7.5, 0.01% Triton X-100, 5 mM TCEP, 5 mM $MgCl_2$, and 5%

glycerol. The reaction was performed for 30 min at 25°C on a Greiner 384 FLUOTRAC™ plate in the dark before measuring its end-point kinetics in a BioTek Synergy H1 plate reader equipped with an FP-capable optical cube with an excitation wavelength of 485 nm and emission wavelength of 528 nm corrected at 50 gain (G). The anisotropy data were computed from the fluorescence intensity measurements of $I_{//}$ (the fluorescent plane parallel to its excitation plan) and $I_{\perp}$ (the fluorescent plane perpendicular to its excitation source) using formula (1). The fluorescent anisotropy was plotted against recombinant protein concentrations fitting through a quadratic model depicted in formula (2) using GraphPad Prism 9 software (www.graphpad.com) for analysis.

Equation (1): $FA = \frac{I_{\parallel} - I_{\perp}}{I_{\parallel} + 2I_{\perp}}$

Equation (2): $y = y_0 + \frac{y_{max} * x}{K_d + x}$,

where y = anisotropy measures (from zero, y_0 to maximum protein concentration, y_{max}), x = protein concentration, and K_d = dissociation constant.

Trans-replication assay

U2OS human bone osteosarcoma cells (ATCC HTB-96) were maintained in Iscove's modified Dulbecco's m*edium* (Gibco) containing 10% FBS, 2 mM L-glutamine, 100 U/mL penicillin and 0.1 mg/mL streptomycin at 37°C in a 5% CO_2 atmosphere. Cells grown on 12-well plates were cotransfected with the following combination of plasmids (each used at 1 µg per transfection): HSPolI-FG-RRV, CMV-P123-RRV, CMV-ubi-nsP4-RRV or its mutant versions (Table 2) for RRV *trans*-replicase and HSPolI-FG-SINV, CMV-P123-SINV, CMV-ubi-nsP4-SINV or its mutant versions (Table 2) for SINV *trans*-replicase. Transfections were performed using Lipofectamine LTX with PLUS reagent (Thermo Fisher Scientific) according to the manufacturer's instructions. Transfected cells were incubated at 37°C for 18 h. All transfections were performed in triplicate, and experiments were repeated at least twice. After incubation, cells were lysed, and the activities of Fluc and Gluc were measured using the Dual-Luciferase-Reporter assay (Promega). Fluc and Gluc activities measured for cells transfected using plasmids expressing active replicases or their mutants under investigation (Table 2) were normalized to those obtained for the corresponding control cells transfected with HSPolI-FG-RRV, CMV-P123-RRV and CMV-ubi-nsP4^{GAA}-RRV or HSPolI-FG-SINV, CMV-P123-SINV and CMV-ubi-nsP4^{GAA}-SINV.

Acknowledgments

We acknowledge the following grants: MOE2016T22097 from the Ministry of Education of Singapore, NMRC OFIRG17nov084 from Ministry of Health of Singapore, and PRG1154 from Estonian Research Council. We would like to thank the Protein Production Platform from NTU (PPP, NTU), NTU Institute of Structural Biology (NISB, NTU), Lee Kong Chian School of Medicine for providing the facilities and the general support from members of the D.L and A.M laboratories to complete our research. We thank the technical services provided by the "Synchrotron Radiation Protein Crystallography Facility of the National Core Facility Program for Biotechnology, Ministry of Science and Technology" and the "National Synchrotron Radiation Research Center", a national user facility supported by the Ministry of Science and Technology of Taiwan, ROC – for data collection at Beamline NSRRC TPS 05A. We would also like to acknowledge the technical support from Japan SPring-8 Center RIKEN Targeted Proteins Beamline BL32XU and Australian Synchrotron, which is part of ANSTO, for the access to MX1/MX2 Beamlines (MX2 for the utilization of Australian Cancer Research Foundation detector) for data collection.

Author Contributions

A. M. and D. L. designed research; Y. B. T., L. S. L., X. L., Y.-S. L., C. K., and J. Z. performed research; Y. B. T., L. S. L., X. L., Y.-S. L., C. K., J. L., J. Z., A. M., and D. L. processed and analyzed data in contributions for figure compilations; and Y. B. T., L. S. L., A. M., and D. L. prepared the manuscript with advice from all authors.

Declaration of interests

All authors declare no competing interests

Data deposition

Crystal structure data have been deposited in the wwPDB OneDep System under PDB ID code XXX.

References

1. A. Wilder-Smith *et al.*, Epidemic arboviral diseases: priorities for research and public health. *The Lancet Infectious Diseases* **17**, e101-e106 (2017).
2. E. O. Nsoesie *et al.*, Global distribution and environmental suitability for chikungunya virus, 1952 to 2015. *Eurosurveillance* **21**, 30234 (2016).
3. J. Lessler *et al.*, Assessing the global threat from Zika virus. *Science* **353** (2016).
4. W. Chikungunya, https://www.who.int/denguecontrol/arbo-viral/other_arboviral_chikungunya/en/.
5. J. Jain, K. Mathur, J. Shrinet, R. K. Bhatnagar, S. Sunil, Analysis of coevolution in nonstructural proteins of chikungunya virus. *Virol J* **13**, 86 (2016).
6. N. L. Forrester *et al.*, Genome-Scale Phylogeny of the Alphavirus Genus Suggests a Marine Origin. *Journal of Virology* **86**, 2729-2738 (2012).
7. D. C. de Andrade, S. Jean, P. Clavelou, R. Dallel, D. Bouhassira, Chronic pain associated with the Chikungunya Fever: long lasting burden of an acute illness. *BMC Infect Dis* **10**, 31 (2010).
8. Y. Wada *et al.*, Discovery of a novel antiviral agent targeting the nonstructural protein 4 (nsP4) of chikungunya virus. *Virology* **505**, 102-112 (2017).
9. J. H. Strauss, E. G. Strauss, The alphaviruses: gene expression, replication, and evolution. *Microbiological Reviews* **58**, 491-562 (1994).
10. R. Zhang *et al.*, Mxra8 is a receptor for multiple arthritogenic alphaviruses. *Nature* **557**, 570-574 (2018).
11. P. P. Rose *et al.*, Natural resistance-associated macrophage protein is a cellular receptor for sindbis virus in both insect and mammalian hosts. *Cell Host Microbe* **10**, 97-104 (2011).
12. M. W. Chen *et al.*, Chikungunya virus nsP4 RNA-dependent RNA polymerase core domain displays detergent-sensitive primer extension and terminal adenylyltransferase activities. *Antiviral Research* **143**, 38-47 (2017).
13. S. Tomar, R. W. Hardy, J. L. Smith, R. J. Kuhn, Catalytic core of alphavirus nonstructural protein nsP4 possesses terminal adenylyltransferase activity. *J Virol* **80**, 9962-9969 (2006).
14. J. K. Rubach *et al.*, Characterization of purified Sindbis virus nsP4 RNA-dependent RNA polymerase activity in vitro. *Virology* **384**, 201-208 (2009).
15. B. Götte, L. Liu, G. M. McInerney, The Enigmatic Alphavirus Non-Structural Protein 3 (nsP3) Revealing Its Secrets at Last. *Viruses* **10** (2018).
16. N. J. Foy *et al.*, Hypervariable domains of nsP3 proteins of New World and Old World alphaviruses mediate formation of distinct, virus-specific protein complexes. *Journal of virology* **87**, 1997-2010 (2013).
17. Y. S. Law *et al.*, Inter-domain Flexibility of Chikungunya Virus nsP2 Helicase-Protease Differentially Influences Viral RNA Replication and Infectivity. *Journal of virology* 10.1128/JVI.01470-20 (2020).
18. Y. S. Law *et al.*, Structural insights into RNA recognition by the Chikungunya virus nsP2 helicase. *Proceedings of the National Academy of Sciences of the United States of America* **116**, 9558-9567 (2019).
19. Y. A. Karpe, P. P. Aher, K. S. Lole, NTPase and 5′-RNA Triphosphatase Activities of Chikungunya Virus nsP2 Protein. *PloS one* **6**, e22336 (2011).
20. C. Li *et al.*, mRNA Capping by Venezuelan Equine Encephalitis Virus nsP1: Functional Characterization and Implications for Antiviral Research. *Journal of virology* **89**, 8292-8303 (2015).
21. R. Jones, G. Bragagnolo, R. Arranz, J. Reguera, Capping pores of alphavirus nsP1 gate membranous viral replication factories. *Nature* **589**, 615-619 (2021).

22. K. Gottipati, M. Woodson, K. H. Choi, Membrane binding and rearrangement by chikungunya virus capping enzyme nsP1. *Virology* **544**, 31-41 (2020).
23. K. Zhang *et al.*, Structural insights into viral RNA capping and plasma membrane targeting by Chikungunya virus nonstructural protein 1. *Cell host & microbe* 10.1016/j.chom.2021.02.018 (2021).
24. J. K. Rubach *et al.*, Characterization of purified Sindbis virus nsP4 RNA-dependent RNA polymerase activity in vitro. *Virology* **384**, 201-208 (2009).
25. L. Holm, Benchmarking fold detection by DaliLite v. 5. *Bioinformatics* **35**, 5326-5327 (2019).
26. M. M. Shaik, N. Bhattacharjee, M. Feliks, K. K. Ng, M. J. Field, Norovirus RNA-dependent RNA polymerase: A computational study of metal-binding preferences. *Proteins* **85**, 1435-1445 (2017).
27. W. Shi *et al.*, A nucleobase-binding pocket in a viral RNA-dependent RNA polymerase contributes to elongation complex stability. *Nucleic Acids Res* **48**, 1392-1405 (2020).
28. K. H. Choi *et al.*, The structure of the RNA-dependent RNA polymerase from bovine viral diarrhea virus establishes the role of GTP in de novo initiation. *Proc Natl Acad Sci U S A* **101**, 4425-4430 (2004).
29. S. Venkataraman, B. Prasad, R. Selvarajan, RNA Dependent RNA Polymerases: Insights from Structure, Function and Evolution. *Viruses* **10** (2018).
30. Y. S. Law *et al.*, Structural Insights into RNA Recognition by the Chikungunya Virus nsP2 Helicase. *Proceedings of the National Academy of Sciences* **(In Press)** (2019).
31. M. Teppor, E. Žusinaite, A. Merits, Phosphorylation Sites in the Hypervariable Domain in Chikungunya Virus nsP3 Are Crucial for Viral Replication. *J Virol* **95** (2021).
32. K. Hellström *et al.*, Partially Uncleaved Alphavirus Replicase Forms Spherule Structures in the Presence and Absence of RNA Template. *J Virol* **91** (2017).
33. L. S. Lello *et al.*, Cross-utilisation of template RNAs by alphavirus replicases. *PLoS Pathog* **16**, e1008825 (2020).
34. H. Jia, P. Gong, A Structure-Function Diversity Survey of the RNA-Dependent RNA Polymerases From the Positive-Strand RNA Viruses. *Front Microbiol* **10**, 1945 (2019).
35. R. Chen *et al.*, ICTV Virus Taxonomy Profile: Togaviridae. *J Gen Virol* **99**, 761-762 (2018).
36. L. S. Lello *et al.*, Cross-utilisation of template RNAs by alphavirus replicases. *PLoS Pathog* **16**, e1008825 (2020).
37. V. Ramakrishnan, J. Finch, V. Graziano, P. Lee, R. Sweet, Crystal structure of globular domain of histone H5 and its implications for nucleosome binding. *Nature* **362**, 219-223 (1993).
38. K. Hirata *et al.*, ZOO: an automatic data-collection system for high-throughput structure analysis in protein microcrystallography. *Acta Crystallographica Section D: Structural Biology* **75**, 138-150 (2019).
39. K. Yamashita, K. Hirata, M. Yamamoto, KAMO: towards automated data processing for microcrystals. *Acta Crystallographica Section D: Structural Biology* **74**, 441-449 (2018).
40. D. Aragao *et al.*, MX2: a high-flux undulator microfocus beamline serving both the chemical and macromolecular crystallography communities at the Australian Synchrotron. *Journal of Synchrotron Radiation* **25**, 885-891 (2018).
41. N. P. Cowieson *et al.*, MX1: a bending-magnet crystallography beamline serving both chemical and macromolecular crystallography communities at the Australian Synchrotron. *Journal of Synchrotron Radiation* **22**, 187-190 (2015).
42. W. Kabsch, XDS. *Acta Crystallographica Section D* **66**, 125-132 (2010).
43. M. D. Winn *et al.*, Overview of the CCP4 suite and current developments. *Acta Crystallographica Section D: Biological Crystallography* **67**, 235-242 (2011).

44. D. Liebschner *et al.*, Macromolecular structure determination using X-rays, neutrons and electrons: recent developments in Phenix. *Acta Crystallographica Section D: Structural Biology* **75**, 861-877 (2019).
45. P. Emsley, B. Lohkamp, W. G. Scott, K. Cowtan, Features and development of Coot. *Acta Crystallographica Section D: Biological Crystallography* **66**, 486-501 (2010).
46. B. D. Pascal *et al.*, HDX workbench: software for the analysis of H/D exchange MS data. *Journal of the American Society for Mass Spectrometry* **23**, 1512-1521 (2012).
47. F. Delaglio *et al.*, NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J Biomol NMR* **6**, 277-293 (1995).
48. B. A. Johnson, Using NMRView to visualize and analyze the NMR spectra of macromolecules. *Methods Mol Biol* **278**, 313-352 (2004).
49. X. Robert, P. Gouet, Deciphering key features in protein structures with the new ENDscript server. *Nucleic acids research* **42**, W320-W324 (2014).
50. L. Schrödinger, The PyMOL Molecular Graphics System, Version 2.4, open-source.

Figures and Tables

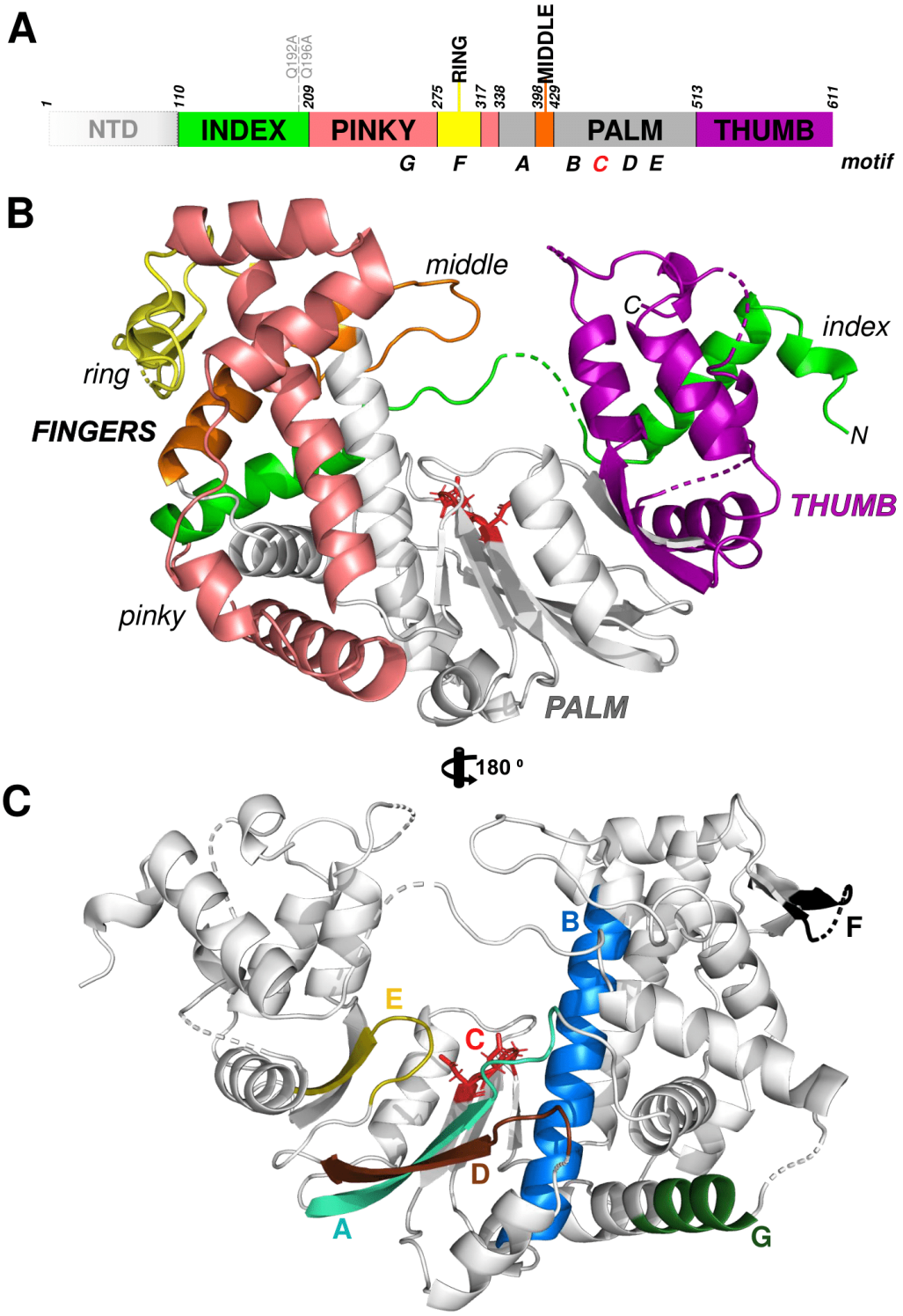
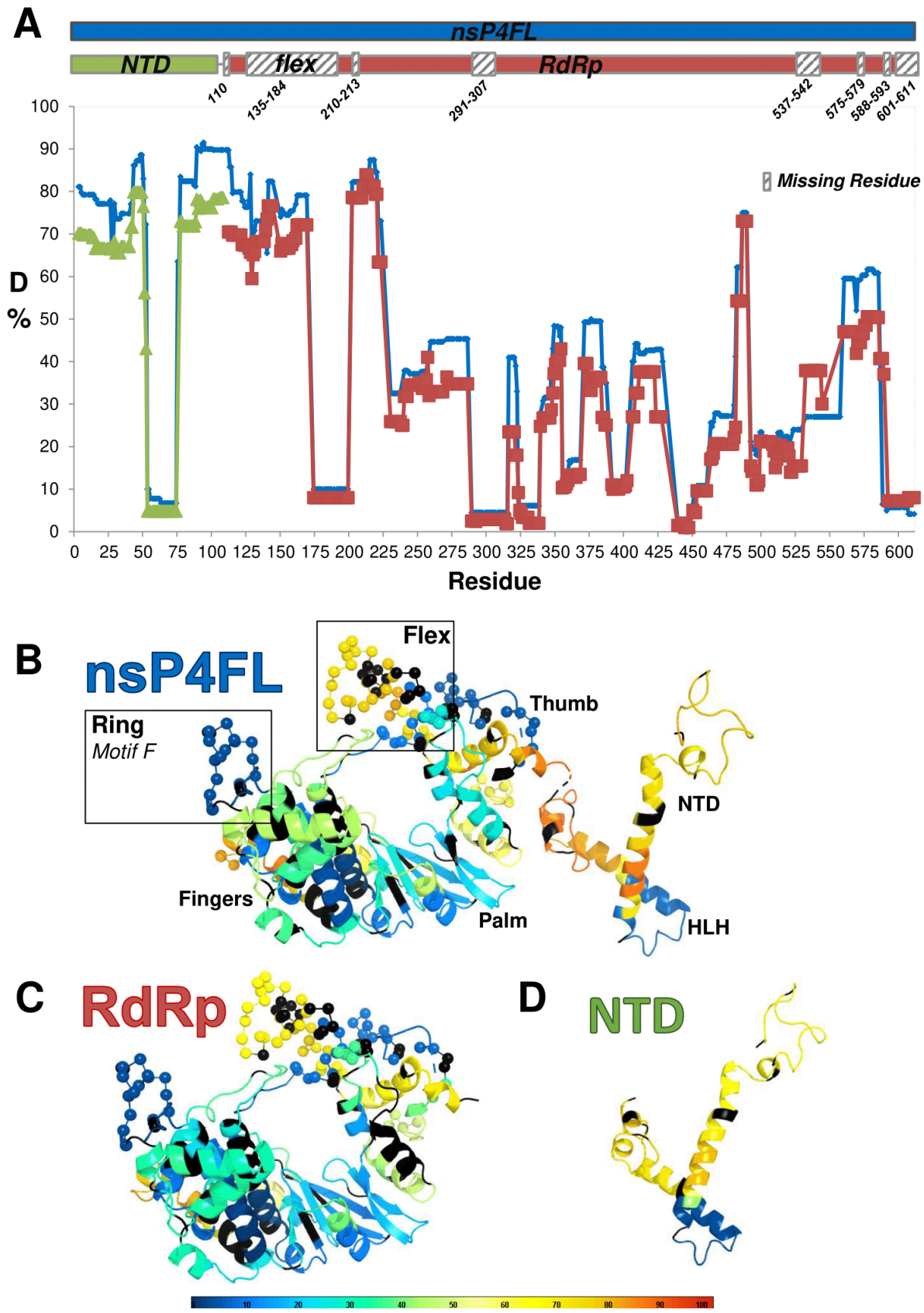


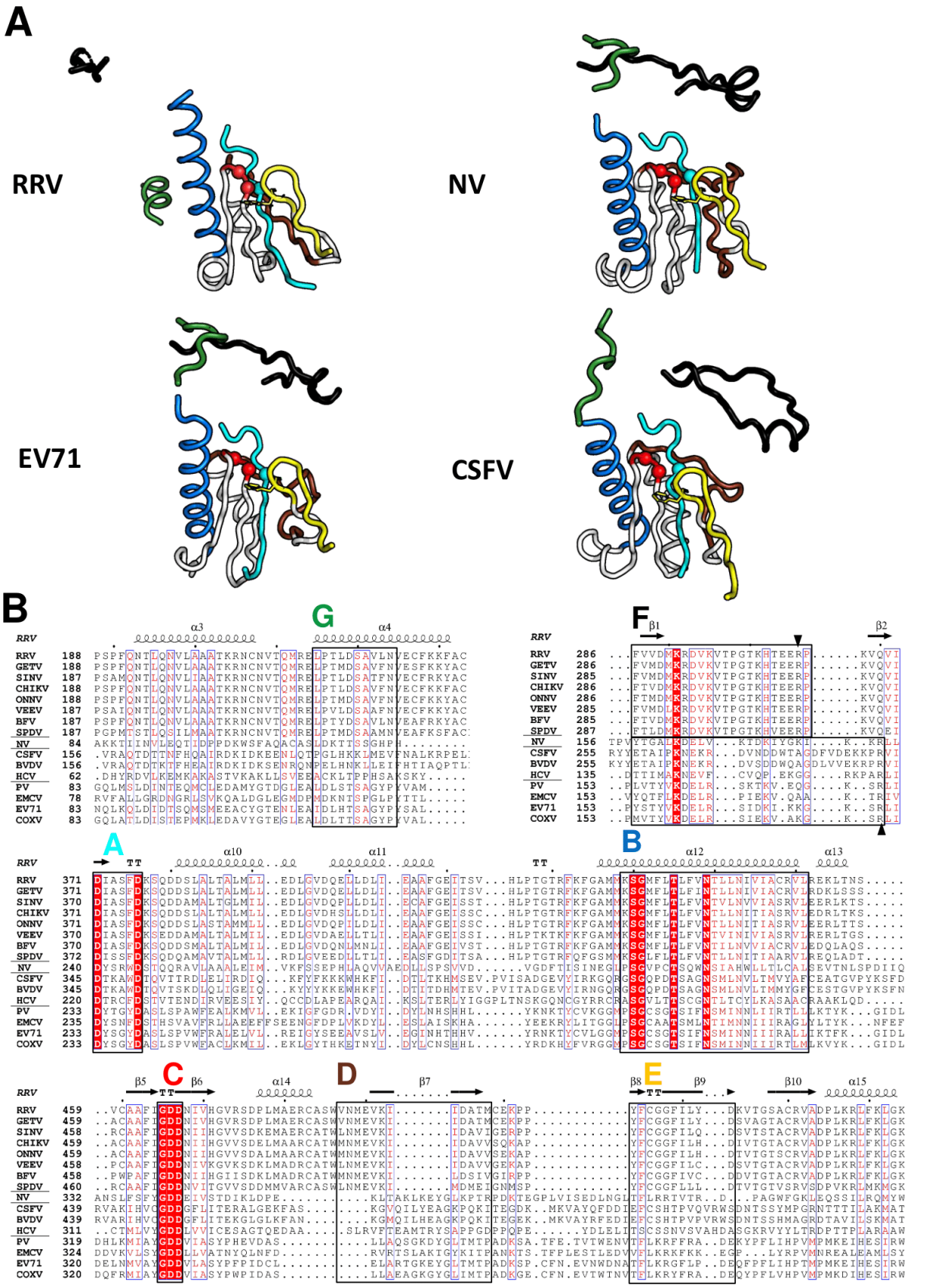
Figure 1. Crystal structure of RRV nsP4 RdRp domain. (A) Schematic presentation of RRV nsP4. The surface-entropy mutations (Q192A, Q196A) and (sub)domain distribution are shown

as follows: fingers with colored fingertips (index in green, middle in orange, ring in yellow and pinky in pink), palm in gray and thumb in purple. The locations of the A-G motifs are shown, and domain boundaries are annotated with residue numbers. (B-C) The annotated crystal structure of RRV RdRp ($RdRp^{SER3}$) colored for subdomain spatial distribution in (B) from the front view (right-hand view) and for the motifs in the 180° rear view in (C). Motif C is highlighted in red, and its conserved GDD active-site residues are displayed as sticks.


A
nsP4FL
NTD
flex
RdRp
110
135-184
210-213
291-307
537-542
575-579
588-593
601-611
Missing Residue
D %
Residue
B
nsP4FL
Flex
Ring
Motif F
Thumb
NTD
Fingers
Palm
HLH
C
RdRp
D
NTD

774

Figure 2. Structure dynamics of RRV nsP4 in solution. (A) The HDX profile of the recombinant constructs of RRV: nsP4FL (blue), RdRp (red) and NTD (green), in solution. On the graph, the deuteration levels expressed in percentage (D%) are plotted against the nsP4 residue numbers (aa 1-611). The residues missing in the resolved structure are represented as gray line boxes, and their positions are indicated. The longest unresolved region in RdRp (residues 135-184) is designated as “flex”. (B-D) The color-coded ternary structures of Robetta homology models of (B) nsP4FL, (C) RdRp and (D) NTD, according to the D% by HDX. These models were generated to fill the missing residues not covered in the current crystal sturcture. The HDX-protection profiles of (A) are color-mapped to these homology models at single-amino-acid resolutions to guide the visual analyses and 3D comparisons between nsP4FL and RdRp and NTD. nsP4FL is annotated with subdomains, motifs and highlighted regions (Ring Motif F and Flex) as a reference to both RdRp and NTD. The missing residues are presented as spheres in these homology models. The residues coloured in black are those not covered by mass spectrometry. Bar at the bottom: D% is visualized in the visible light spectrum, where the red end represents high D% and the blue end represents low D%.


A
RRV
NV
EV71
CSFV
B
G
F
A
B
C
D
E
RRV
GETV
SINV
CHIKV
ONNV
VEEV
BFV
SPDV
NV
CSFV
BVDV
HCV
PV
EMCV
EV71
COXV

791

Figure 3. Comparison of the structure of nsP4 of RRV with those of other viral RdRps. (A) The RRV $RdRp^{SER}$ and its top structural homologs, including Norwalk virus (NV, PDB 5TSN), Enterovirus-71 (EV71, PDB 6KWQ) and classical swine fever virus (CSFV, PDB 5Y6R). Structures were compared based on Dali and structural alignment. Conserved polymerase motifs are marked by colors as follows: motif A: cyan; motif B: navy, motif C: red; motif D: brown; motif E: yellow; motif F: black and motif G: forest green. The conserved aspartate residues in motifs C and A are displayed as Cα spheres, while the conserved phenylalanine residue in motif E is shown in stick format. (B) The conservation of RdRps of +RNA virus based on primary sequence and secondary structure presented through Espript*. The sequence and secondary structure of RRV nsP4 (genus *Alphavirus,* family *Togaviridae*) was compared to those of different alphaviruses (RRV-to-SPDV) and members of the *Caliviridae* (NV)*, Flaviviridae* (CSFV-to-HCV) and *Picornaviridae* (PV-to-COXV) families. The secondary structure is labeled on top of the alignment and shows only alpha helixes (spring), beta strands (arrow) and turns (TT). Conserved motifs A-G are boxed and colored according to (A). [*Espript coloring: The red-boxed columns are the most conserved, while the blue-boxed columns classify amino acid conservation based on their similar functional groups. *Espript label: *Togaviridae* - Ross River virus (RRV), Getah virus (GETV), Sindbis virus (SINV), chikungunya virus (CHIKV), o'nyong'nyong virus (ONNV), Venezuela equine encephalitis virus (VEEV), Barmah Forest virus (BFV), and salmon pancreas disease virus (SPDV); *Caliviridae* - Norwalk virus (NV); *Flaviviridae* - classical swine fever virus (CSFV), bovine viral diarrhea virus (BVDV), and hepatitis C virus (HCV); *Picornaviridae* - poliovirus (PV), encephalomyocarditis virus (EMCV), enterovirus-71 (EV71), and coxsackievirus (COXV)].

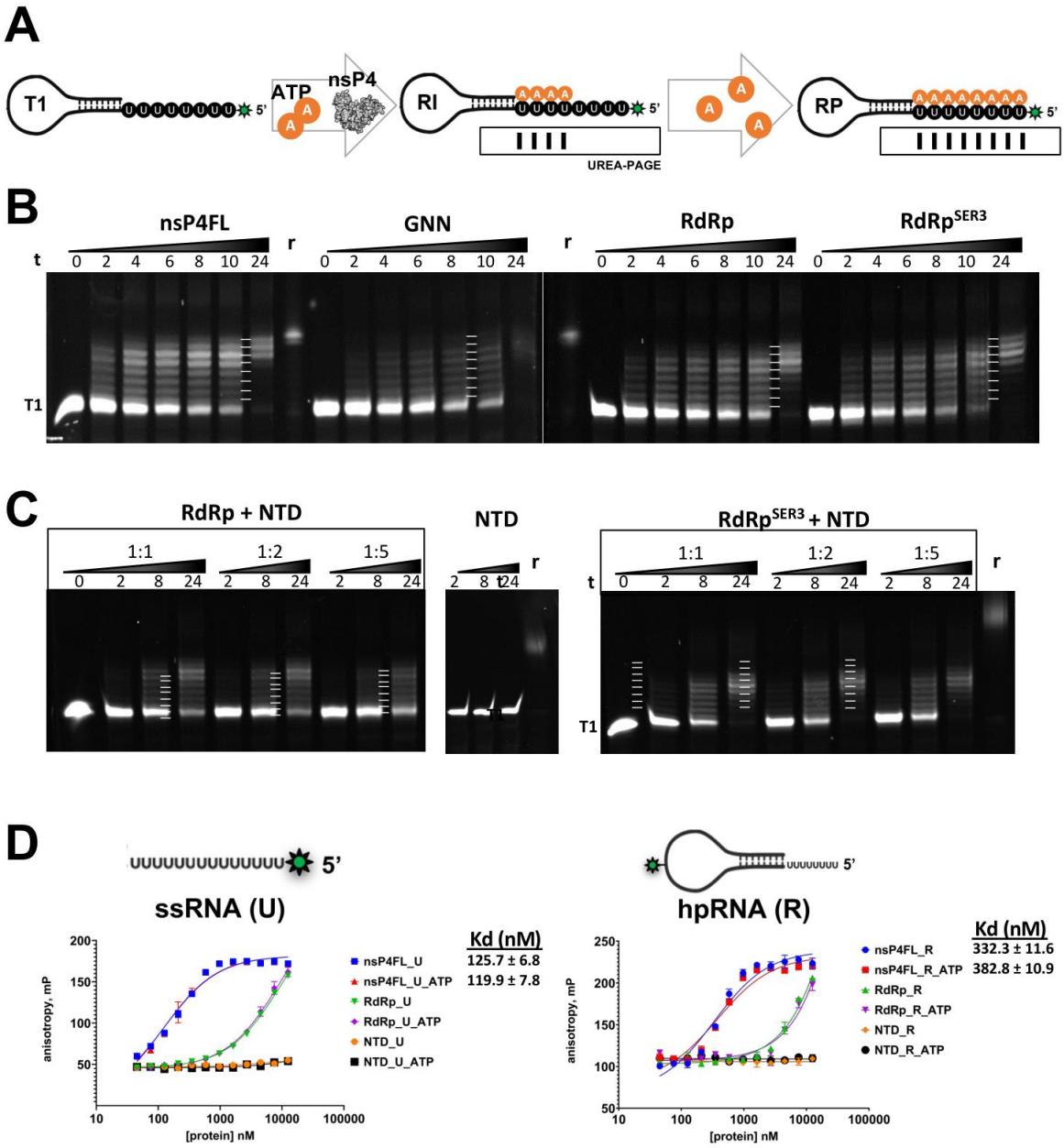


Figure 4. RNA polymerase activities of nsP4FL, RdRp and RdRp^{SER3}. (A) Schematic presentation of the RNA polymerase reaction. Extension of fluorescently labeled (green star on the 5' end) RNA template T1 by recombinant nsP4 protein (gray particle) results in incorporation of A-residue (orange circle) from ATP into the RNA intermediate (RI; snapshot of one possible RI) and RNA product (RP), both of which are shown below with expected band migration corresponding to the number of A-residue added to T1 on a UREA-PAGE gel. (B-C) The biochemical characterization of recombinant nsP4 proteins using discontinuous time-point gel-based polymerase assay performed for 0-24 h. The images of 17.5% denaturing UREA-PAGE

gels are annotated with T1, time-point (t), products made by dengue virus RdRp used as a positive control (r) and names of the tested recombinant nsP4 proteins (nsP4FL, polymerase weakened mutant of nsP4FL: GNN, RdRp, $RdRp^{SER3}$ and NTD). For the polymerase reactions between RdRp or $RdRp^{SER3}$ and NTD, the stoichiometric molar ratio is shown in (B-C). The incorporation of each base is marked with a short white line shown between the last 2 time points. (D) The fluorescent polarization between different recombinant nsP4 proteins (nsP4FL, RdRp and NTD) measured in anisotropy was plotted against the log_{10} concentration scale in nanomolar (nM) range. ssRNA and hpRNA used for the analysis are shown above corresponding graph titles, and a fluorescent label is shown as a green star on RNA 5' end. All FP experiments were conducted in duplicate to generate binding affinities (K_d) labeled at the graph sample legend.

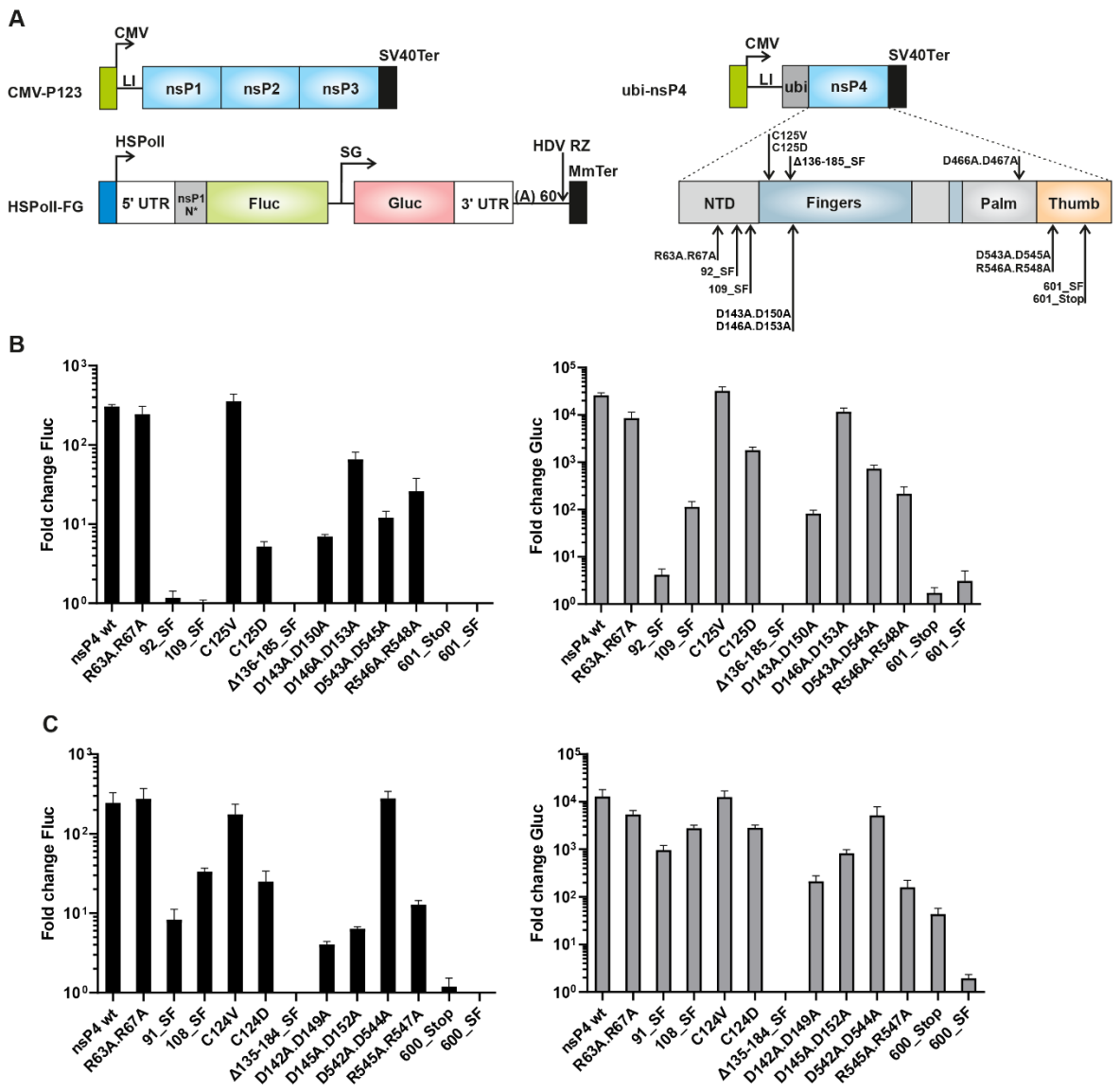


Figure 5. Effects of mutations in nsP4 on the activities of RRV and SINV *trans*-replicases. (A) The transreplication system is composed of two plasmids (CMV-P123 and CMV-ubi-nsP4) for the expression of RRV and SINV replicase and a plasmid HSPoII-FG for the production of replication-competent RNA containing firefly luciferase (Fluc) and *Gaussia* luciferase (Gluc) reporters. In this assay, the increase in expression levels of Fluc and Gluc represented the production of full-length genomic RNA and SG RNA, respectively. HSPoII, a truncated promoter (residues −211 to −1) for human RNA polymerase I; 5′ UTR, full length 5′ UTR of an alphavirus; nsP1 N* - region encoding the N-terminal 77 (RRV) or 114 (SINV) amino acid residues of nsP1; SG - SG promoter spanning (with respect to termination codon of nsP4) from position -79 to the end of intergenic region; 3′ UTR, truncated (last 110 residues) 3′ UTR of an

alphavirus; HDV RZ - antisense strand ribozyme of hepatitis delta virus; MmTer, a terminator for RNA polymerase I in mice. CMV, immediate early promoter of human cytomegalovirus; LI, leader sequence of the herpes simplex virus thymidine kinase gene with artificial intron; SV40Ter, simian virus 40 late polyadenylation region; Ubi-sequence encoding for human ubiquitin. Positions of mutations in nsP4 (residues numbered as in nsP4 of RRV) are indicated. (B, C) U2OS cells in 12-well plates were cotransfected with matching combinations of CMV-P123 and CMV-ubi-nsP4 or its mutant versions (in a 1:1 molar ratio) and the corresponding HSPolI-FG plasmids. As negative controls, CMV-ubi-nsP4^{GAA}, which encodes nsP4 lacking RNA polymerase activity, was used instead of CMV-ubi-nsP4. Cells were incubated at 37°C and lysed 18 h post transfection. Fluc and Gluc activities produced by wt and mutant replicases were normalized to the nsP4^{GAA} controls. Values obtained for nsP4^{GAA} controls were taken as 1. Means + standard deviation of three independent experiments are shown. (B) Effects of introduced mutations on RRV replicase activity. (C) Effects of introduced mutations on SINV replicase activity.

Table 1. Data collection, phasing and refinement statistics.

	SeM-RdRp^{SER3}	RdRp^{SER3}
Data Collection	Se-SAD phasing	Native
Space group	P 2_1 2_1 2_1	P 2_1 2_1 2_1
Unit cell dimensions		
a, b, c (Å)	64.152 68.575 100.978	64.728 68.196 99.406
α, β, γ (°)	90 90 90	90 90 90
Wavelength	0.9792 (peak)	0.9998 (native)
Resolution range (Å)	46.85 - 2.83 (2.931 - 2.83)	32.36 - 2.6 (2.693 - 2.6)
Total reflections	296954 (29212)	147593 (15292)
Unique reflections	11129 (1107)	14041 (1387)
Mean I/sigma(I)	17.19 (0.30)	23.93 (1.84)
Completeness (%)	99.93 (99.82)	99.82 (99.93)
Multiplicity	26.7 (26.3)	10.5 (11.0)
R_{merge}[b]	-	0.06609 (1.415)
R_{meas}[c]	-	0.07009 (1.487)
R_{pim}[d]	-	0.02275 (0.4524)
$CC_{1/2}$[e]	-	0.978 (0.667)
FOM	0.70 (0.66)	-
Phase Error (°)	35.60 (40.21)	-
Refinement Statistics		
Reflections used in refinement		14044 (1387)
Reflections used for R-free		668 (74)
R_{work}[f]		0.2655 (0.3891)
R_{free}[g]		0.3268 (0.4431)
Number of nonhydrogen atoms		3115
macromolecules		3112
solvent (water)		3
Protein residues		401
RMSD (bonds)		0.006
RMSD (angles)		0.76
Ramachandran Statistics		
Ramachandran favored (%)		97.14
Ramachandran allowed (%)		2.34
Ramachandran outliers (%)		0.52
Rotamer Outliers (%)		1.48
Clashscore		11.82
Average B-factor		97.63
macromolecules		97.68
solvent		46.75

a. The values presented in parentheses are of the highest resolution shell.

b. $R_{merge} = \sum |I_{obs} - I_{avg}| / \sum I_{avg}$

c. $R_{meas} = \sum hkl \sqrt{\frac{n}{n-1}} \sum_{i=1}^{n} |I_i(hkl) - \bar{I}(hkl)| / \sum_{hkl} \sum_{i=1}^{n} I_i(hkl)$,
$where\ R_{meas}\ is\ the\ precision\ indicator\ of\ individual\ observation\ for\ unmerged\ data.$

d. $R_{pim} = \sum hkl \sqrt{\frac{1}{n} - 1} \sum_{i=1}^{n} |I_i(hkl) - \bar{I}(hkl)| / \sum_{hkl} \sum_{i=1}^{n} I_i(hkl)$, $where\ R_{pim}\ is\ the\ precision\ indicator\ for\ the\ merged\ data.$

e. $Correlation\ Coefficient, CC = \sum(x - \langle x \rangle)(y - \langle y \rangle) / \left[\sum(x - \langle x \rangle)^2 \sum(y - \langle y \rangle)^2\right]^{1/2}, where\ CC_{1/2}\ is\ of\ two\ half\ datasets.$

f. $R_{work} = \sum |F_{obs} - F_{calc}| / \sum F_{obs}$

g. $R_{free} = \sum |F_{obs} - F_{calc}| / \sum I_{calc}$,$where\ R_{free}\ was\ randomly\ sampled\ from\ 5\%\ reflections.$

869