Structural insights into Rift Valley fever virus replication machinery

Zhongzhou Chen (chenzhongzhou@cau.edu.cn)
China Agriculture University https://orcid.org/0000-0003-1319-9664
Hong-Wei Wang
Tsinghua University https://orcid.org/0000-0001-9494-8780
Xue Wang
China Agriculture University
Cuixia Hu
Tsinghua University
Jia Wang
Tsinghua University
Xiaofei Dong
China Agriculture University
Jie Xu
Tsinghua University
Xiaorong Li
China Agricultural University
Manfeng Zhang
Genome Analysis Laboratory of the Ministry of Agriculture, Agricultural Genomics Institute at Shenzhen, Chinese Academy of Agricultural Sciences
Hongyun Lu
China Agriculture University
Wei Ye
Fourth Military Medical University
Fanglin Zhang
Fourth Military Medical University
Wei Wu
China Agricultural University

Article

Keywords: Rift Valley fever virus, RNA-dependent RNA polymerase, structure, replication

Posted Date: April 15th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-398449/v1

Manuscript title: Structural insights into Rift Valley fever virus replication machinery

Xue Wang[1,4], Cuixia Hu[2,4], Wei Ye[3], Jia Wang[2], Xiaofei Dong[1], Jie Xu[2], Xiaorong Li[1], Manfeng Zhang[1], Hongyun Lu[1], Fanglin Zhang[3], Wei Wu[1], Hong-Wei Wang[2,*] and Zhongzhou Chen[1,*]

[1] State Key Laboratory of Agrobiotechnology and Beijing Advanced Innovation Center for Food Nutrition and Human Health, College of Biological Sciences, China Agricultural University, Beijing, China, 100193

[2] Ministry of Education Key Laboratory of Protein Science, Tsinghua-Peking Joint Center for Life Sciences, Beijing Advanced Innovation Center for Structural Biology, School of Life Sciences, Tsinghua University, Beijing, China, 100084

[3] Department of Microbiology, School of Preclinical Medicine, Fourth Military Medical University, Xi'an, China

[4] These authors contributed equally: Xue Wang, Cuixia Hu.

* To whom correspondence should be addressed:
Zhongzhou Chen
E-mail: chenzhongzhou@cau.edu.cn
Hong-Wei Wang
E-mail: hongweiwang@tsinghua.edu.cn (H.-W.W.)

Abstract

Rift Valley fever virus (RVFV) belongs to the order *Bunyavirales* and is the type species of genus *Phlebovirus*, which accounts for over 50% of family *Phenuiviridae* species. RNA-dependent RNA polymerase (L protein) is responsible for facilitating the replication and transcription of the virus. We report two cryo-EM RVFV L protein structures at 3.6 Å and 3.8 Å resolution in the presence and absence of RNA, respectively. In this first L protein structure of genus Phlebovirus, viral RNA induces considerable conformational changes of the polymerase. The RVFV L protein priming loop is distinctly different from those of other L proteins and undergoes large movements related to its replication elongation role. Structural and biochemical analyses indicate that a single template can initiate RNA replication, which is notably enhanced by 5' viral RNA. These findings advance our understanding of RNA synthesis mechanism and provide a basis for antiviral inhibitor development.

KEYWORDS: Rift Valley fever virus; RNA-dependent RNA polymerase; structure; replication

INTRODUCTION

The order *Bunyavirales*, negative-sense, single-stranded RNA virus, is divided into twelve families (*Arenaviridae*, *Cruliviridae*, *Fimoviridae*, *Hantaviridae*, *Leishbuviridae*, *Mypoviridae*, *Nairoviridae*, *Peribunyaviridae*, *Phasmaviridae*, *Phenuiviridae*, *Tospoviridae*, and *Wupedeviridae*) (https://talk.ictvonline.org/taxonomy/), including over 300 identified species [1]. The order *Bunyavirales* can infect plants, animals, and humans, posing a serious threat to agriculture, animal husbandry, human life, and property. For example, Tomato spotted wilt virus can infect more than 1,000 plant species, Hantaan orthohantavirus and Lassa mammarenavirus cause severe hemorrhagic fevers with a mortality rate of up to 50%, Crimean-Congo hemorrhagic fever virus also causes severe hemorrhagic fever in humans with mortality rates of up to 30% [2,3]. Rift Valley fever virus (RVFV) belongs to the order *Bunyavirales*, family *Phenuiviridae*, genus *Phlebovirus* (https://talk.ictvonline.org/taxonomy/), and is one of the most prominent and well-known bunyaviruses [4]. The genus *Phlebovirus* includes 60 species and accounts for more than half of the family *Phenuiviridae* species (117 species). RVFV, the type species of this genus, is a zoonosis, leading to severe infections in animals and humans and significant economic losses.

RVFV was first isolated in Kenya in 1931 by Daubney et al. through inoculating lambs with the serum from a moribund sheep [5]. Rift Valley fever (RVF) is widespread in Africa, and multiple outbreaks have been recorded in Africa and the Arabian Peninsula in the 21st century. Recent outbreaks in a French overseas department and some seropositive and imported cases in Turkey, Libya, and China have raised concerns that RVF could extend to other parts of the world, which poses a serious threat to global health and economics [6-11]. RVFV is transmitted among ruminants by bites from mosquitos mainly of the *Aedes* and *Culex* genera, or by direct contact with body fluids of viremic animals [12]. RVFV is classified as a category A high priority pathogen by the National Institute of Allergy and Infectious Diseases and virus handling requires high- or maximum-containment facilities, such as a biosafety level 3 enhanced (BSL-3E) or BSL-4 laboratory [13]. RVFV leads to severe disease in livestock, with up fatality rates of up to 70% in young animals and as high as 20%–30% in adult ruminants. In addition, infected pregnant animals undergo spontaneous abortion at an alarmingly high rate, in the range of 40–100% [14]. After being infected with RVFV, humans may experience generalized fatigue, low-grade fever, headache, photophobia, and joint pains. Some patients infected with RVFV will develop encephalitis, retinopathy, or disseminated intravascular coagulation leading to hemorrhage and even death, with a case fatality rate of approximately 20% [11]. The cross-border movement of goods and animals means that there is now a threat of RVFV dissemination to Asia and Europe [15].

RVFV has a single-stranded, tripartite RNA genome, composed of L, M, and S segments. The large negative-sense L RNA segment encodes a single-polypeptide L protein and the negative-sense M RNA segment encodes the glycoprotein precursor. The S segment adopts an ambisense strategy for expressing NSs and N proteins [16]. The RVFV viral genome is wrapped by the N proteins and the conserved 3′ and 5′ ends of the genomic RNA (vRNA) are bound to L protein, resembling the influenza virus [17,18]. The L protein is a multifunctional enzyme that catalyzes genome replication and viral gene transcription, which are initiated *de novo* by prime-and-realign and a cap-snatching mechanisms, respectively [19].

To date, precautions to avoid new RVFV outbreaks mostly rely on weather monitoring, and no specific drugs or licensed vaccines are available to treat or prevent infection [20]. Moreover, no genus RNA-dependent RNA polymerase (RdRp) *Phlebovirus* structures have been reported. To understand the role of L protein in the viral life cycle and identify a feasible drug target, we solved two RVFV polymerase structures, with or without RNA. These are the first genus *Phlebovirus* L structures to be described. Cryogenic electron microscopy (cryo-EM) revealed RVFV L structures at 3.6 Å or 3.8 Å resolution, in the presence or absence of RNA, respectively, and was used to build a molecular model for the core polymerase domain. These structures guided the elucidation of the mechanism of viral replication and transcription. Using biochemical assays we verified the novel endonuclease activity and polymerase activity of the RVFV L protein. The RVFV L protein endonuclease activity has a preference toward $MnCl_2$ and has polymerase activity in the presence of only template RNA or both template and allosteric activator RNA.

RESULTS

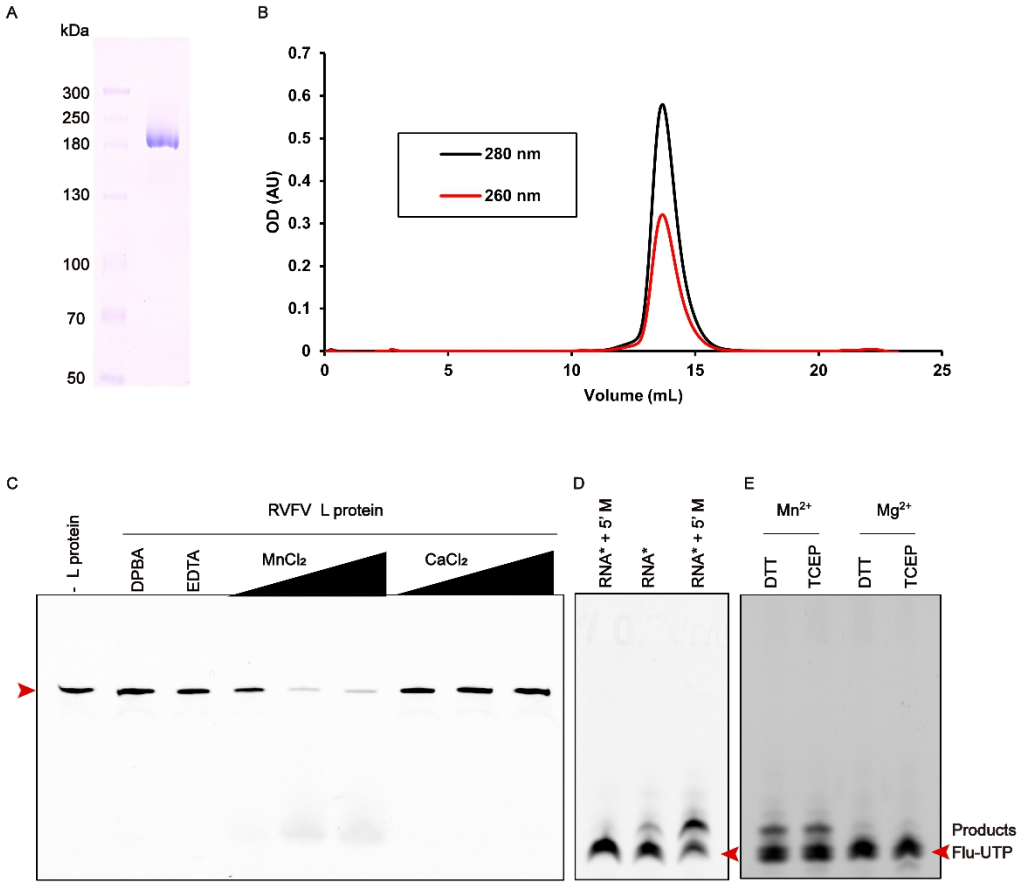


Figure 1. *In vitro* enzymatic activity assays

(A) SDS–PAGE profile of Rift Valley fever virus (RVFV) L protein.

(B) Size-exclusion chromatogram of RVFV L protein. Absorbance curves for the sample at wavelengths of 260 nm and 280 nm.

(C) Endonuclease activity assay. RVFV L protein (0.06 μM) was incubated with 0.45 μM of fluorescently labeled 30 nt PolyA RNA substrate at 30℃ for 40 min in the presence of 10, 25,

or 50 mM $MnCl_2$ or $CaCl_2$. Reactions without protein, in the presence of EDTA or the known endonuclease-specific inhibitor 2,4-dioxo-4-phenylbutanoic acid (DPBA), were negative controls. All input is shown as red arrows in Figure C, D, and E.

(D) *In vitro* polymerase activity assay. The endonuclease and polymerase inactivation site double mutant (D111A/D1133A) was added in the left lane as a negative control. Mutant (D111A) was added in the middle and right lanes. These three lanes were incubated with the conserved 5' 20 nt of the M segment or/and a 20 nt template RNA named RNA^* (5' M: 5'-ACACAAAGACGGUGCAUUAA-3', RNA^*: 5'-UGUGUUUCUGGCCACGUUGA-3'). Replication products were detected by fluorescence (fluorescein-12-UTP).

(E) Effect of metal ions (Mn^{2+} or Mg^{2+}) and reducing agent (DTT or TCEP (Tris(2-carboxyethyl)phosphine)) on *de novo* RNA synthesis by endonuclease inactivation site mutant (D111A). The assay was performed with 5 mM $MnCl_2$/$MgCl_2$ and 2.5 mM DTT/TCEP for 40 min at 30℃.

Enzymatic properties of RVFV L protein in vitro

After extensive screening, we successfully constructed and expressed the full-length L protein using the *Pichia pastoris* expression system. Full-length L protein existed as a monomer in solution (Figure 1A, Figure 1B). *In vitro* endonuclease activity experiments revealed that the L protein, but not the endonuclease site mutant (D111A), had strong endonuclease activity in the presence of $MnCl_2$ (Supplementary Figure S1A). Using 30 nt polyA as an RNA substrate, the L protein polymerase had ion-concentration dependent endonuclease activity in the presence of $MnCl_2$, but no activity in the presence of $CaCl_2$ (Figure 1C, Supplementary Figure S1) [21,22]. To visualize the *de novo* synthesis of viral RNA, we tested the ability of the purified RVFV L protein to incorporate fluorescent nucleotides *in vitro*. To prevent RNA degradation during the experiment, we used the D111A endonuclease site mutant and tested its RNA replication activity. We added fluorescein-12-UTP into the *in vitro* replication reaction system as the substrate, and monitored the RNA synthesis products by electrophoresis on 20% polyacrylamide gels containing 7 M urea. Compared with the D111A/D1133A negative control mutant, which lacks polymerization activity, a clear replication product was produced by the D111A mutant (Figure 1D).

Examination of the synthesized products revealed that most were dinucleotides, and several oligonucleotide bands were also observed. Together, these results show that RVFV L protein can produce oligonucleotides using nucleotides with large modifications, such as fluorescein, as the substrate. This differs from the Vesicular Stomatitis Virus L protein, which can only incorporate nucleotides with smaller modifications into viral RNA [23]. This result illustrates that the active center of the RVFV L protein is larger than that of the Vesicular Stomatitis Virus L protein. We further tested how Mg^{2+} and Mn^{2+} metal ions affected RVFV L protein polymerase activity in the presence of different reducing agents (DTT or TCEP). Urea PAGE analysis showed that the type of reducing agent did not significantly affect RVFV L protein polymerase activity. However, RVFV L protein polymerase activity was stronger in the presence of $MnCl_2$ than in the presence of $MgCl_2$ (Figure 1E). This result contrasts with that reported for severe fever with thrombocytopenia syndrome virus (SFTSV) L protein, the structure of which revealed an active Mg^{2+}-coupled polymerase center [21,24]. Interestingly, the RVFV L protein was also capable of replicating RNA in the presence of only one template RNA (Figure 1D). Therefore, the RVFV L protein can

incorporate a nucleotide into a nucleotide/oligonucleotide in the absence of 5' vRNA. This activity is notably enhanced by the addition of 5' vRNA. This is different from the replication mode of most bunyaviruses that need both 5’ RNA and 3’ RNA for successful replication. To explore the mechanism of this unique RVFV L protein property we decided to reveal the RVFV L protein structure.

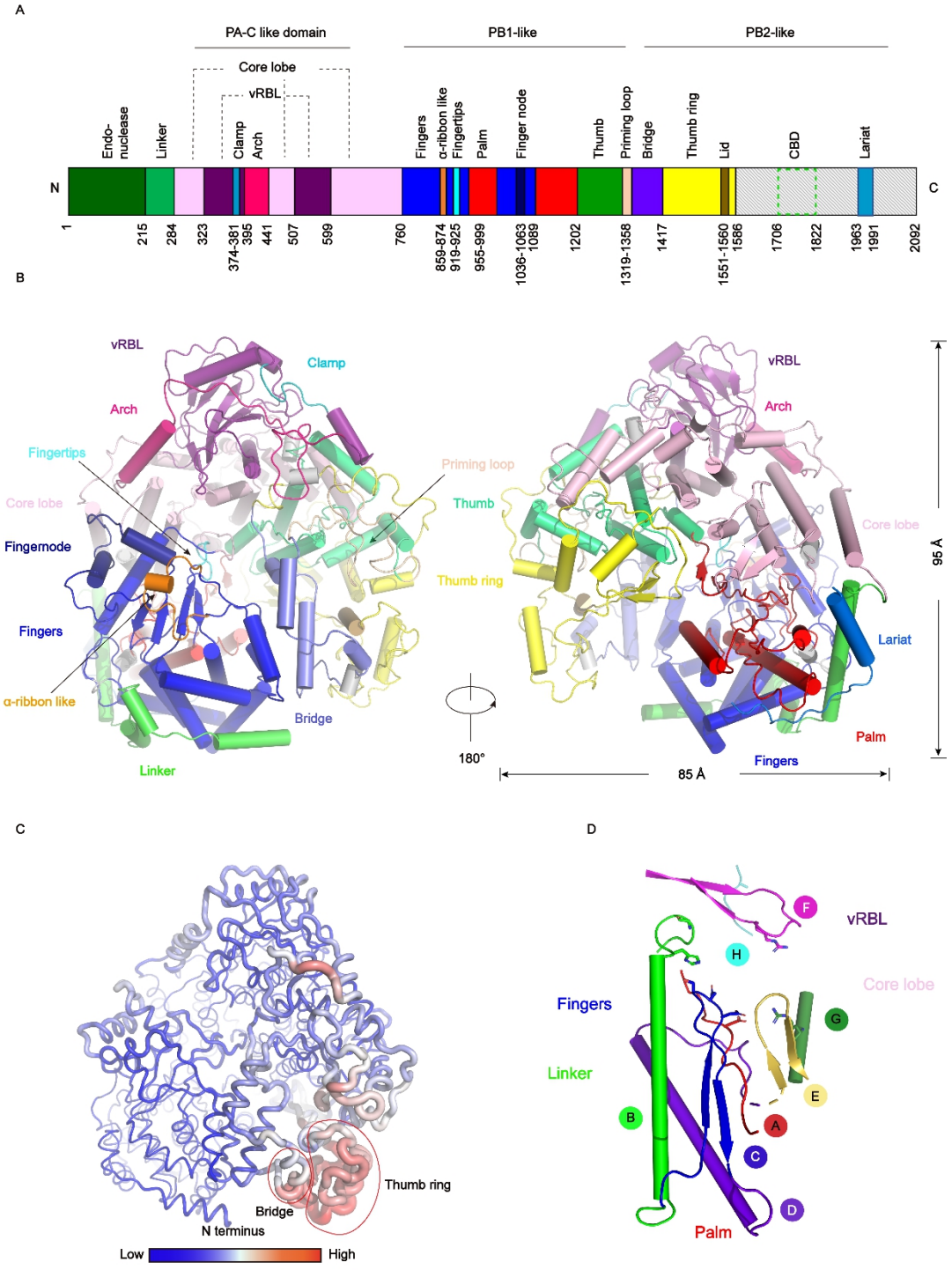


Figure 2. Overall structure of the Rift Valley fever virus (RVFV) L_{RNA} protein and conserved

motifs

(A) Schematic representation of the monomeric RVFV L_{RNA} protein domain structure.

(B) Cartoon representation of the RVFV L_{RNA} protein. The structure is colored by domains using the same color code as in (A). 3_{10} helices are colored in gray.

(C) The RVFV L_{RNA} protein B-factor map. A larger radius and red color represent high B-factor values and a smaller radius and blue color represent low B-factor values.

(D) The arrangement of the conserved RdRp motifs in the RVFV active site colored in red, green, blue, purpleblue, yelloworange, magenta, forest, and teal for motifs A–H, respectively.

Overall RVFV L protein structure

We performed single particle cryo-EM analysis of the RVFV L protein and obtained two 3D reconstructions at 3.6 and 3.84 Å resolution in the presence and absence of RNA, respectively (Supplementary Figures S2–S6, Supplementary Table S1). Our solved L_{RNA} structure had ~1400 residues, corresponding to the PA-C, PB1, and PB2 N-terminus of influenza virus [17] (Figure 2A). The L protein was co-purified with the partially dsRNA formed by the 3' and 5' conserved RNA genome ends (5' vRNA 5'-GGUGCA-3' and 3' vRNA 3'-UGUGUUUCUGGCCACGU-5') (Supplementary Figure S7), similar to that described for La Crosse encephalitis virus (LACV) [25,26]. The overall shape of the L_{RNA} structure resembled a figure-"6" (Supplementary Figure S2F). The dimensions of the RVFV L_{RNA} protein structure are about 85 Å x 75 Å x 95 Å (Figure 2B). Moreover, an additional endonuclease domain was solved in the L_{apo} structure. The RVFV L_{apo} cryo-EM density map contains the endonuclease domain, RdRp domain, and a short lariat, while the RVFV L_{RNA} contains the RdRp domain and a longer lariat.

The first domain, at the L_{apo} protein N-terminus, is the endonuclease domain (residues 1–214) (Supplementary Figure S8). This domain plays an essential role in cleaving 5' m7G caps from host mRNAs during the cap-snatching process. These free caps are subsequently used as primers for genome transcription [14,27]. Then a linker domain (residues 215–283) wraps around the core structure domain, connecting the endonuclease domain to the three core lobe domains (residues 284–322, 441–506, 599–760) (Figure 2B). The fingers domain (residues 760–858, 874–918, 925–954, 999–1035, 1063–1088) displays a "sandwich" shape, comprised of upper and lower α-helix layers with a β-sheet in the middle (Figure 2B). The catalytic center of RdRp is constituted of fingers, thumb, and palm domains. The thumb domain (residues 1202–1318) is involved in the formation of the template entry channel to help stabilize initial NTPs binding to the template. Moreover, the key residues of the thumb domain can promote template translocation following the first condensation reaction through conformational rearrangement [28]. In brief, the fingers, palm, and thumb domains (Supplementary Figure S9) probably collaborate to promote vRNA template and NTP binding. Furthermore, both L_{apo} and L_{RNA} structures have an additional C-terminal lariat, which contains an α-helix and a loop, and interacts with the linker, palm, and core lobe domains. The C-terminal lariat (residues 1963–1990) in the L_{RNA} structure is longer than that in the L_{apo} structure (Figure 2A, 2B, Supplementary Figure S8). The C-terminal lariat circles the large globular core and folds back to stabilize the fingers' structure. In summary, the active center of the L protein polymerase is relatively stable and conservative, while the structures of the thumb ring and bridge regions are flexible (Figure 2C).

Conserved motifs of RdRp

Common and unique features of viral RNA-dependent polymerase motifs (motifs A-F) that contain key residues for the enzymatic function have been identified in many RNA-dependent polymerases [29]. Like L proteins from many RNA viruses, RVFV L protein has six conserved structural motifs (motifs A-F) that cooperate with viral RNA template, NTPs, and metal ions to promote the condensation of nucleotide triphosphates (Figure 2D). RVFV L protein also has two single-stranded negative-sense RNA virus specific motifs, G and H (Figure 2D). Motif G (668-QTIARYIIME-677) is located in the core lobe domain and consists of an α-helix, the conserved Arg672 of which can participate in genome replication and transcription by interacting with the priming NTPs. Motif H (1045-HRELDI-1050) is located in the fingernode domain and consists of a loop. Motifs A to E are within the palm subdomain while Motif F (913-HICLFKKQQHGGLREIYV-930), composed of two β-strands trapping a loop, constitutes part of the fingertips (Figure 2D). Motif A (984-WTCATSDDARKWN-996) connects the palm and fingers domains, and contains one of the invariant catalytic Asp991 residues present both in DNA and RNA polymerases. The conserved Lys994 in RdRps of single-stranded negative-sense RNA viruses allows them to use manganese instead of magnesium as the cofactor ion, which is consistent with our activity assay results (Figure 1E). Trp995 of Motif A sits in a strategic position where it can maintain L protein structural stability. Motif B has a conservative Gly1087 at a position for the loop to change conformation, which allows the N-terminal loop to bind to the RNA template. Motif C is composed of a β-sheet-loop-β-sheet structure, with the conserved loop region sequence (1132-SDD-1134) forming the active center of RdRp. Motif D consists of an α-helix and a loop, forming a structural scaffold to change RdRp's conformation when a correct nucleotide is bound [30]. Motif E (1185-VMEYNSEFYF-1194) is located between the palm and thumb domains and consists of a β-hairpin structure (Figure 2D), which works in the correct positioning of the 3'OH primer end [29]. These motifs and functional regions participate in the most critical steps for the correct recognition and incorporation of ribonucleotides.

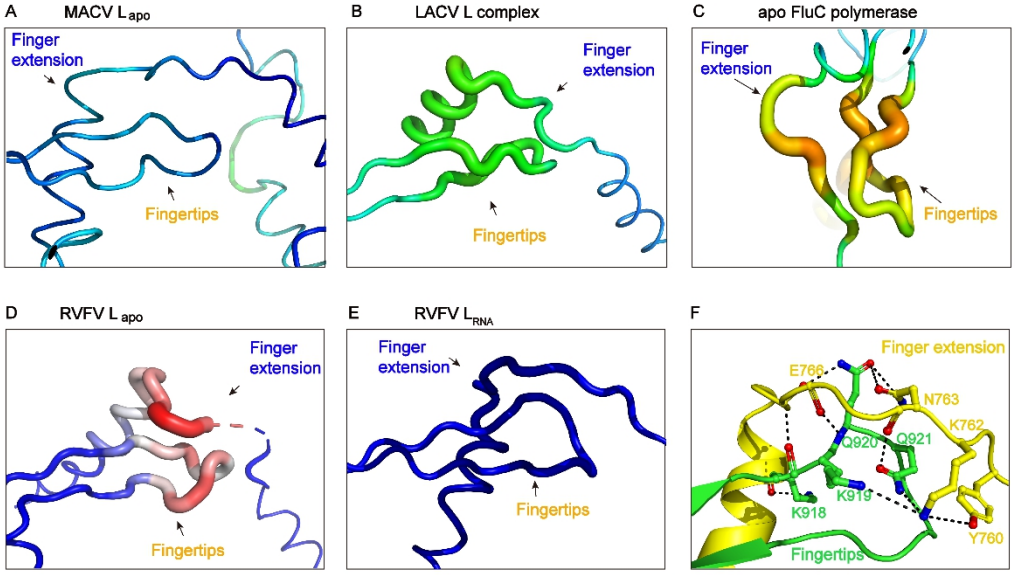


Figure 3. The fingertips loops of machupo virus (MACV) L_{apo}, LACV L complex, apo FluC polymerase, Rift Valley fever virus (RVFV) L_{apo}, and RVFV L_{RNA}

The B-factor maps for (A) MACV (PDB 6KLD), (B) LACV (PDB 6Z8K), (C) Influenza C (FluC) (PDB 5D98), (D) RVFV L_{apo}, (E) RVFV L_{RNA}. (A, B, E) Fingertips are conservative, while FluC's and LACV's fingertips are unstable, and their replication process requires 5' RNA.
(F) Interaction of residues between fingertips and finger extension. The residue on the fingertips and the finger extension are colored green and yellow, respectively.

The fingertips of RVFV L protein

The fingertips domain (Figure 3), mostly composed of the conserved motif F, extends from the fingers domain to the thumb domain, and is located directly above RdRps catalytic center. According to previous reports, during the process of NTPs binding to RdRp, the positively charged Lys and Arg side chain residues in the fingertips domain can interact with the triphosphate moiety, which also interacts with both divalent metal ions in the active center (Supplementary Figure S10)[31]. Moreover, the fingertips domain can bind 5' viral RNA by basic residues [17,25]. Based on the RVFV L protein structural analysis, we identified that the key residues (Lys918, Arg926) (Supplementary Figure S10E) for its vitality are positioned in the middle loop region, right at the junction of the template entry channel and NTP entry channel (Supplementary Figure S11). Therefore, it is reasonable to consider that the fingertips domain is crucial for binding both template and NTPs. The previously reported structure of the apo FluCPol harbors highly unstable fingertips (Figure 3C), whereas the same region in the 5' vRNA-bound FluAPol and FluBPol structure is stable [17,32]. For the LACV (order *Bunyavirales*) L protein [25], 5' vRNA induces conformational changes of the fingertips and the adjacent finger extension, further stabilizing them. Similarly, the fingertips (motif F) of RVFV L protein and the associated finger extension in RVFV apo L protein are disordered, but both are highly stable after nucleic acid addition, even though they do not directly bind to nucleic acid (Figure 3D and E and Movie S1). Moreover, the longer C-terminal lariat solved in the L_{RNA} structure further stabilized the fingers' structure. The finger

extension is oriented in a conformation that provides enough space to hold the fingertip. Residues Tyr760, Lys762, Asn763, and Glu766 (finger extension domain) are highly conserved and have especially strong interactions with the residues (Lys918, Lys919, Gln920, Gln921) in the fingertips domain (Figure 3F).

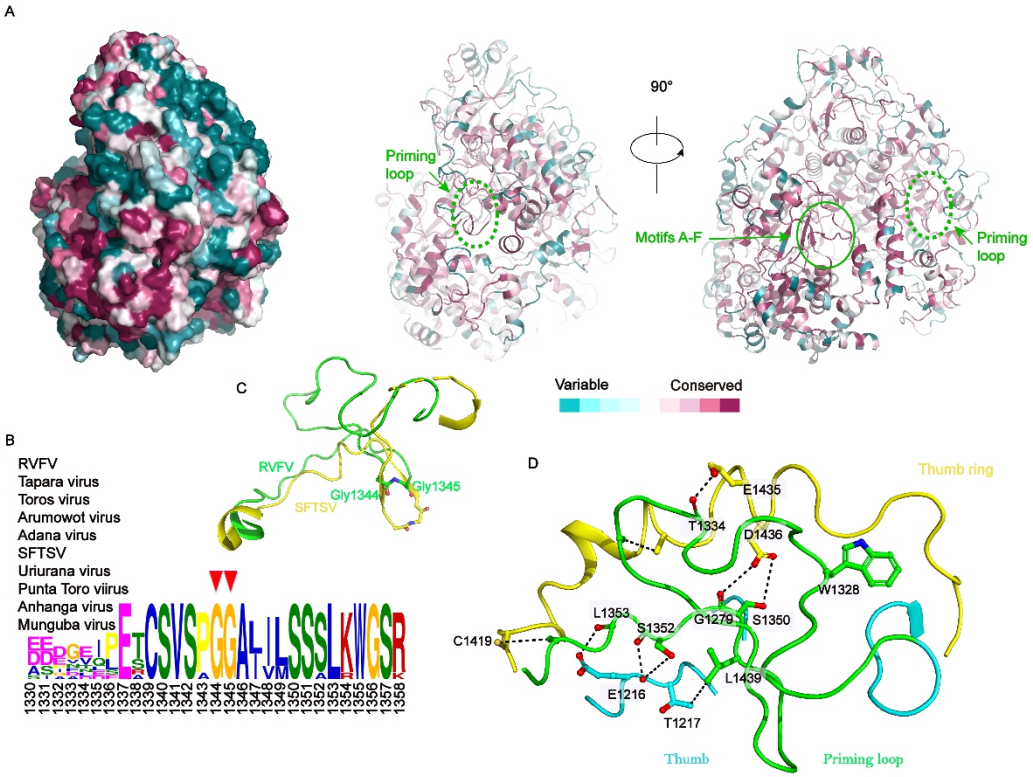


Figure 4. Conservation analysis for RVFV L protein

(A) Sequence conservation analysis. The left figure displayed in surface mode is positioned at the same view as the middle figure. Two different views of the Rift Valley fever virus (RVFV) L protein displayed in ribbon mode and colored based on the residue conservation among family *Phenuiviridae*. The bottom color code bar is from low conservation (cyan) to high conservation (purple). Figures are colored using the Consurf server (http://consurf.tau.ac.il/2016/) based on a multiple sequence alignment including the sequences of Tapara virus (YP_009346035), Toros virus (YP_009246447), Arumowot virus (YP_009010958.1), Adana virus (YP_009227127.1), SFTSV (YP_006504091.1), Uriurana virus (YP_009346036.1), Punta Toro virus (YP_009512941.1), Anhanga virus (API68876.1), and Munguba virus (YP_009346010.1).

(B) Sequence conservation in the putative priming loop.

(C) The priming loop and connecting regions are shown as ribbons colored in green (RVFV) and yellow (SFTSV and PDB 6L42). Two glycine residues (Gly1344 and Gly1345), which are important for an early stage of RNA synthesis, are shown as sticks.

(D) Structural demonstration shows the interaction sites of the priming loop (green) and thumb (cyan) and thumb ring domains (yellow).

Conservation analysis for RVFV L protein

We performed sequence alignment of L proteins from a set of ten representative viruses from the family *Phenuiviridae* and analyzed the conservation differences among the various domains [33]. The most conserved regions were the unique motifs A–F of RNA viruses (Figure 4A, Supplementary Figure S12). Based on this sequence alignment, and conservation analysis of the overall sequence of the polymerase on the Consurf website, the residues on the polymerase surface were identified as variable and the residues around the polymerase active center and the four access channels (Figure 5) were highly conserved, and corresponded to the motifs A–F (Figure 4A). The conserved residues around the channels were mainly positively charged residues, reflecting their roles in nucleic acids interaction during viral genome replication and transcription.

The priming loop (residues 1319–1358) is a fundamental element that generally stabilizes the first 'priming' nucleotide of the product during replication initiation, but is not essential for internal initiation and transcription [34]. The priming loop structure is missing in the pre-initiation structure (L_{apo} protein) (Supplementary Figure S11A). Interestingly, the addition of RNA stabilizes the conformation of the priming loop, revealing its role in replication (Supplementary Figure S11B). MEME-ChIP analysis revealed that the two glycine residues (Gly1344 and Gly1345) at the tip of the priming loop are conserved [35,36], and structural comparison with SFTSV (Figure 4B, Figure 4C) showed that these residues are essential for identifying specific primers. However, a structural comparison between RVFV and SFTSV L proteins demonstrated a distinct difference in their priming loops (Figure 4C). The RVFV L_{RNA} protein (green) priming loop is located next to the template exit tunnel and extrudes from the polymerase active center cavity to the surface of the protein, forming a compact interaction with the thumb domain and thumb ring domain (Figure 4D, Supplementary Figure S11). The priming loop of SFTSV L protein extends to the polymerase active center, so the template exit tunnel is smaller in the pre-initial state of replication (Figure 5B). The L_{RNA} protein thumb ring domain residues (Cys1419, Glu1435, Asp1436) are tightly bound with several priming loop residues (Thr1334, Ile1335, Glu1337), and the thumb domain residues (Glu1216, Thr1217, Gly1279, Phe1315) act on the other part of the priming loop residues (Trp1328, Leu1349, Ser1350, Ser1352, Leu1353) (Figure 4D). These two domains strictly regulate the conformation of the priming loop to adapt to different stages of viral replication or transcription (Figure 4D). The tight cooperation of these two domains provides a larger space for template exiting (Figure 5A). We speculate that the L_{RNA} protein is in the elongation compatible state, consistent with LACV and FluA polymerase structures [26,37] (Figure 5C, D).

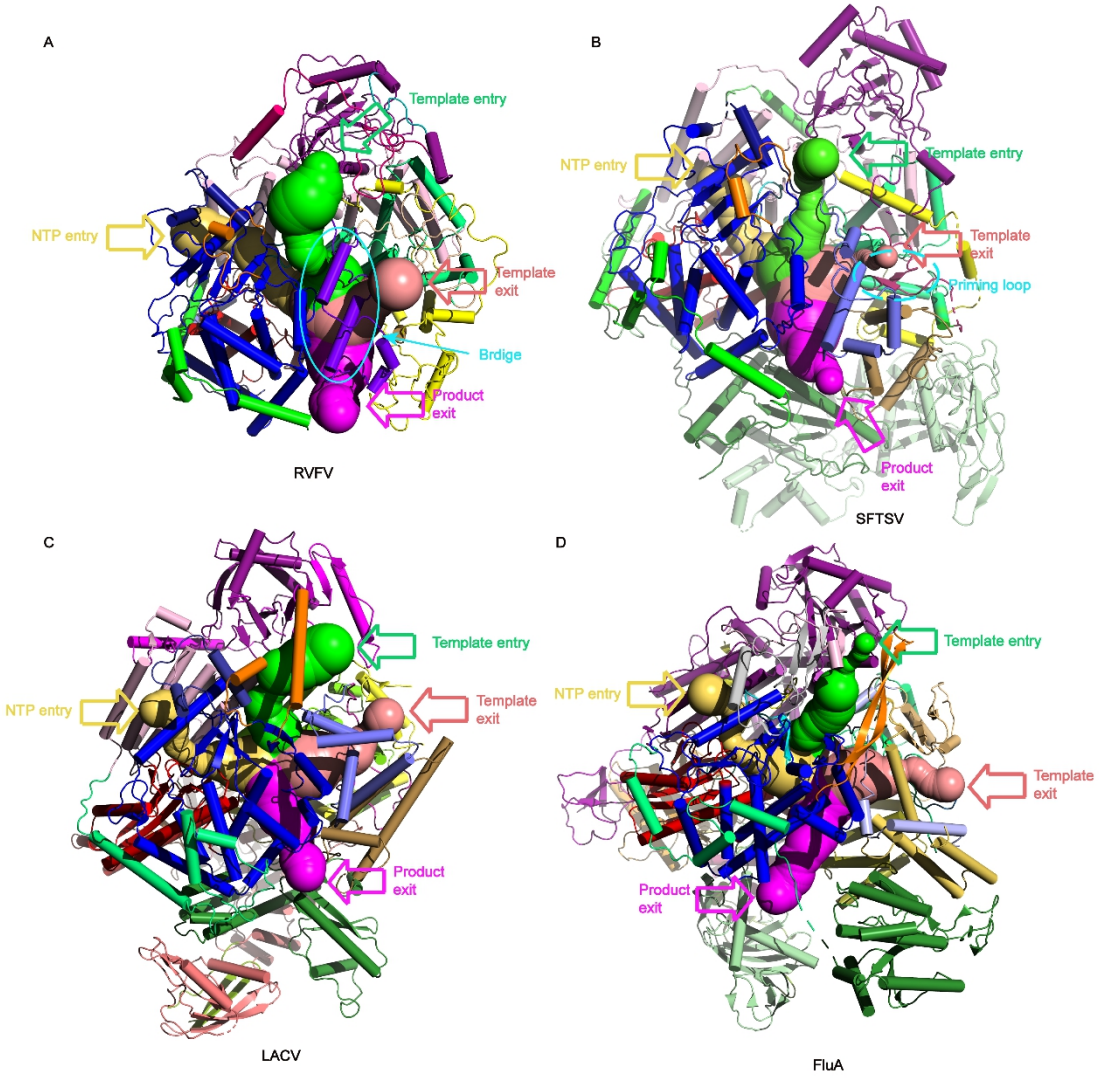


Figure 5. The entrance and exit channels of Rift Valley fever virus (RVFV), SFTSV, LACV and FluA polymerases

The template entry, NTP entry, template exit, and product exit channels of (A) RVFV L_{RNA} protein, (B) SFTSV L_{apo} protein (PDB 6L42), (C) LACV L protein (PDB 6Z8K), (D) Influenza A polymerase (FluA) (PDB 6T0V) are colored in limegreen, yelloworange, salmon, and magenta, respectively. The colors of the RVFV L domains are the same as those used in Figure 2A.

Comparison of four channels

The catalytic residues of the polymerases are at the center of an enclosed cavity connected to the exterior by four channels: an incorporated template entry channel, template exit channel, NTP entry channel, and product exit channel (Figure 5). The configurations of the four tunnels of RVFV L protein are similar to those described for SFTSV and LACV L proteins, and for Flu A polymerase (Figure 5). The template entry channel is mainly composed of vRBL, core lobe, and thumb domains. The conserved Trp513 aromatic residue, located on an α-helix of the vRBL domain adjacent to the core lobe domain, is critical for stabilizing the vRBL domain and core lobe

domain by interacting with Phe581, Pro328, Val517, Leu654, and Leu658. Residues Ser1411, Leu1412, and Ser1413 (bridge domain), located in the middle of the template entry channel, are involved in identifying the sequence-specific template entry process. Similar to that in SFTSV polymerase, the α-ribbon in RVFV polymerase is distant from the template entry channel. The NTP entry channel is mainly composed of the core lobe domain and fingers domain, and the fingers domain separates the NTP entry and template entry channels (Figure 5A). The product exit channel is surrounded by endonuclease, linker, fingers, thumb, and bridge domains. The bridge domain blocks the product exit channel from the template exit channel (Figure 5A). When RNA is added, the template exit and template entry channels become relatively spacious (Supplementary Figure S11). The domains around the template exit channel undergo conformational change triggered by vRNA and expand outward. Although the vRNA is not resolved in the L_{RNA} structure, the binding of vRNA affects protein conformation and keeps the RVFV L protein in a state of replication elongation. In contrast, the SFTSV L_{apo} template exit tunnel is narrower compared than our L_{RNA} structure (Figure 5B).

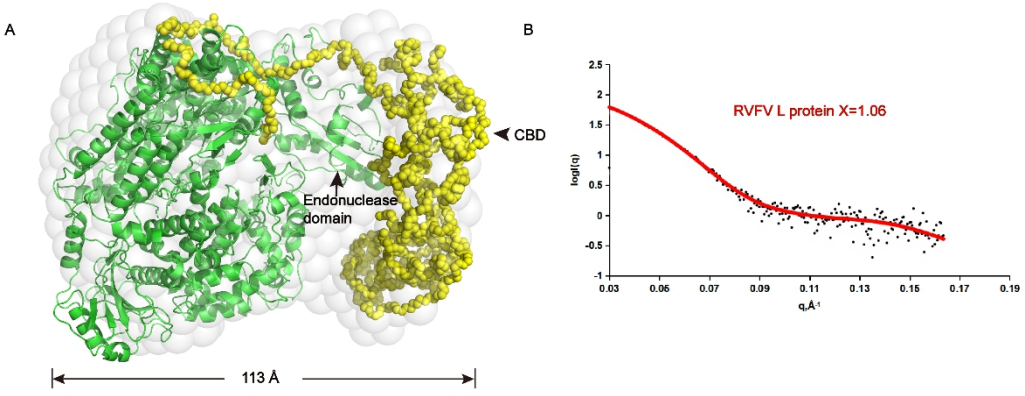


Figure 6. Superposition of low-resolution *ab initio* model and rigid body model

(A) The *ab initio* model is shown as light-gray spheres. PA-like, PB1-like, and PB2-like N-terminal domains of RVFV L protein are colored in green. The simulated PB2 C-terminal domain is colored in yellow.

(B) Comparison of SAXS experimental data and calculated scattering profiles. Experimental data are represented in black dots. The data calculated from the RVFV L protein is in red.

DISCUSSION

According to the World Health Organization, RVF is one of the top 20 global pandemic and epidemic diseases (https://www.who.int/emergencies/diseases/en/). In the 21st century, multiple RVF outbreaks have occurred, causing devastating effects on global health and economics. The RVFV polymerase, is located on the 'L' segment of the genome, and is the largest protein in the virus, with a molecular mass of approximately 238 kDa. Moreover, no RNA-dependent RNA polymerase structures of the genus *Phlebovirus* have been reported. Here, after optimizing the expression and purification of this large protein, we solved the first L protein structures of the genus *Phlebovirus* in the presence and absence of vRNA. Comparison of the RdRp core structure with those previously reported in *Bunyavirales* (SFTSV) and other two sNSRV families, including *Arenaviridae* (lassa virus and machupo virus) and Orthomyxoviridae (influenza) (Supplementary Figure S13), revealed that the RdRp core is typically conserved (Supplementary Figure S13B).

The N-terminal endonuclease domain connects to the polymerase core via a flexible linker (Supplementary Figure S13A). By contrast, the C-terminal domain is tightly integrated into the thumb subdomain of the polymerase core. All domains contribute to the formation of the channels that facilitate template entry and exit, NTP entry, and product exit (Figure 5). Our results reveal that the RVFV polymerase has compact core structure and is a flexible multidomain protein. Comparison of the L_{apo} and L_{RNA} proteins indicates that considerable conformational changes are induced upon RNA interaction (Supplementary Figure S11, Movie S1).

The segmented single strand RNA genomes of negative-sense viruses perform two distinct RNA synthetic processes: genome replication via producing a positive-sense antigenome that acts as a template to generate progeny negative-strand genomes, and transcription including capped, and polyadenylated subgenomic messenger RNAs. As in other negative-sense viruses, the vRNA in RVFV is packaged with two virally expressed proteins (N and L protein), into a ribonucleoprotein (RNP) complex that is competent for replication and transcription. RNP complexes are generally formed by nucleoprotein winding of genomic RNA to form a rope, and both termini of each N-encapsidated genomic RNA segment can bind to one copy of the viral L polymerase protein. Unlike other negative-strand RNA viruses, the RVFV RNP complex has no helical symmetry. The nucleoprotein-encapsulated RNA is shaped like a flexible rope [38].

RVFV replication and transcription occur in the cytoplasm. This is unlike the influenza virus, which vRNPs are transported into the nucleus for replication and transcription [39]. Replication is a primer-independent process. First, a replication intermediate called cRNA (complementary RNA) is synthesized, and then cRNA is used as a template for replication. In the pre-initial stage of replication, the priming loop is disordered (Supplementary Figure S11A). The priming loop undergoes large movements with the addition of template RNA (Supplementary Figure S11B), and is distinctly different from those of other L proteins of bunyaviruses (Figure 4C). Then, the priming loop recruits the first nucleotide to bind to the vRNA, and provides sufficient space in the active center for the replication product [26]. Consistent with the replication initiation mechanism in SFTSV [21], called prime and realign, initiation of replication in RVFV might occur at several positions in the 3’ vRNA template.

Our *in vitro* experiments showed that the replication product was not detected when using the 5’ and 3’ ends of any of or all of the three viral genome segments as transcription templates. The low activity is guessed that the synthesized templates (20 nt) are short, the last nucleotide A (is complementary to the detected fluorescein-12-UTP) in the template is far away from the 3’ end and the detection method is not very sensitive. Moreover, N or other cofactors might be required to form a complex such as RNP to increase polymerase activity. Therefore, we redesigned the template RNA with an A at the 3’ end, and with a U as the first base of the synthesized product at the 5’ end. This increased the amount of the fluorescent product, making it easier to detect. Surprisingly, when only one template RNA was added, the L protein was capable of replication (Figure 1D). Therefore, the RVFV L protein can incorporate a nucleotide into a nucleotide/oligonucleotide in the absence of 5' vRNA, and this activity is notably enhanced by the addition of the 5' vRNA. This is different from the replication mode of most bunyaviruses that require both 5’ vRNA and 3’ vRNA for successful replication. Interestingly, 5’ vRNA can

significantly increase the RNA polymerase activity. This is because the basic residues (Lys918 and Arg926) in the fingertips domain (Figure 3D, 3E, Supplementary Figure S10E) might bind the 5' viral RNA and stabilize the RdRp core structure.

Transcription of RVFV genomes is a primer-dependent process that involves capturing host capped RNAs using a cap-snatching mechanism. During cap-snatching, short capped fragments are snatched by cap-binding domain (CBD) and host mRNAs are cleaved at a position close to the 5' cap by the polymerase endonuclease domain. Although we successfully obtained the full-length protein, we were unable to fit the C-terminal CBD into the cryo-EM density in both structures. The CBD might need capped vRNA to stabilize the conformation. To understand the location of the CBD is in the RVFV L protein structure, we used the pure, monodisperse, and monomeric full-length RVFV L protein to perform SAXS experiments and obtained a low-resolution structure in solution (Figure 6). Using the rigid body fitting of the recently published PB2 C-terminal including CBD [24], the L_{RNA} protein structure, and the endonuclease domain of RVFV L_{apo} protein, we obtained an integrated RVFV L protein model (Figure 6A), which fit well with the SAXS envelope (Figure 6B). The integrated RVFV L protein model shows that the endonuclease and CBD domains are close to each other, which is supported by their functions: the CBD captures RNA and passes it to the endonuclease domain for cutting to be used as transcription primer.

At present, viral RdRps are promising drug targets. For example, Ribavirin, a broad-spectrum antiviral drug that selectively inhibits the RNA-dependent RNA polymerase, is active against hemorrhagic fever viruses in cell culture systems. Ribavirin has both prophylactic and therapeutic efficacy against bunyavirus, especially on genus *Phlebovirus*. Within 24 hours after infection, the combined treatment with favipiravir and ribavirin significantly increases survival outcomes and reduces virus titers in serum and tissue [40]. Using liposomes as drug carriers, a nontoxic, low-dose regimen of ribavirin had a therapeutic effect comparable to that achieved with higher, but potentially more toxic, doses of free ribavirin [41]. Mitoxantrone also alleviates severe diseases caused by RVFV in mice [42]. Moreover, the inhibitory effect of 2′-Fluoro-2′-deoxycytidine on RVF is over ten-times stronger than that of ribavirin [43]. Thus, our structures provide an important basis for developing effective antivirals for protecting humans and animals exposed to, and infected with, RVFV.

MATERIALS AND METHODS

Small scale expression

The codon-optimized sequence for RVFV (strain: ZH-501, GenBank: DQ375406.1) L protein was synthesized and cloned (using the seamless assembly cloning method) into a modified pPICZ expression vector with a C-terminal 6 × His tag without/with a GFP tag (*Eco*R I/*Nco* I) under the AOX1 promoter. Each single colony was inoculated into 5 mL of YPD medium (1% yeast extract, 2% peptone, 2% glycerol) in a 50-mL centrifuge tube and incubated at 30°C in a shaking incubator (170 rpm) until the culture reached an OD_{600} = 6 (approximately 24 hours). The cells were harvested by centrifugation at 3,000 × g for 2 minutes at 4°C. The supernatant was decanted, and the cell pellets were resuspended in approximately 5 mL of BMM medium (100 mM potassium phosphate, pH 6.0, 1.34% YNB, 1% methanol) to induce expression. Methanol was added to a final concentration of 1% every 24 hours to maintain induction. Cell pellets were analyzed for protein expression by western blot or GFP fluorescence detection. The extremely low level of wild-type L protein expression makes it impossible to obtain the protein by purification. To improve protein expression, we tried expressing several different mutants which were located away from the endonuclease active center. We found that expression of the D103A mutant protein was significantly increased, so this site was mutated in all of our proteins. In the manuscript, this protein is referred to as wild type because the mutant has endonuclease activity (Supplementary Figure S1).

Protein expression and purification

His tagged L protein was expressed in the X-33 *Pichia* strain in BMM medium at 28℃ for 48 hours after induction with 1% methanol. After pelleting, cells were resuspended in 50 mM Tris-HCl pH 8.5, 1 M NaCl, 5% glycerol, 20 mM imidazole and 5 mM benzamidine hydrochloride. The cells were lysed in the presence of 1 mM protease inhibitor (phenylmethyl sulfonyl fluoride (Invitrogen)) using a high pressure cell disruption system and the crude lysate was centrifuged at 47,000 × g for 50 minutes at 4°C to isolate the supernatant. The protein was purified from the soluble fraction by Ni affinity chromatography (GenScript), and was washed using a buffer containing 20 mM imidazole, 50 mM Tris-HCl pH 8.5, 1 M NaCl, 5% glycerol, and 5 mM benzamidine hydrochloride to eliminate non-specific proteins. The elution buffer contained 50 mM imidazole. The eluted protein was concentrated using an Amicon Ultra centrifugal filter (Millipore). The protein was further purified by size exclusion chromatography using a Superdex 200 10/300 GL (GE Healthcare) column equilibrated in 20 mM BTP (1,3-Bis [tris (hydroxymethyl) methylamino] propane) pH 8.5, 300 mM NaCl, and 2 mM β-ME. Collected protein fractions were concentrated to 2 mg/mL and stored at 4℃ for further use.

The protein tagged only with His uses a 6×His tag as the basis of *Pichia* protein expression screening. However, this method is cumbersome, the detection is not sensitive, and it requires considerable time to screen the strains expressing the protein. We improved the expression screening using a GFP-6×His Tag. The GFP fluorescence screening method was more convenient, simple to operate, and highly sensitive. GFP-positive cells were resuspended in 50 mM Tris-HCl pH 8.5, 500 mM NaCl, 5% glycerol, 20 mM imidazole, and 5 mM benzamidine hydrochloride. The lysate was centrifuged at 47,000 × g for 50 minutes to remove cell debris and was loaded onto Ni affinity chromatography (GenScript). The elution buffer was the same as the washing buffer

(20 mM imidazole, 50 mM Tris-HCl, pH 8.5, 500 mM NaCl, 5% glycerol, and 5 mM benzamidine hydrochloride) but contained an additional 50 mM imidazole. After changing the buffer to remove the imidazole, the GFP-6×His Tags were removed by the tobacco etch virus protease at a 1:1 (w/w, protease/protein) ratio at 4°C overnight. Then, the mixture was purified using Ni affinity chromatography (GenScript). Fractions containing L protein were collected and applied to a Superdex 200 10/300 GL (GE Healthcare) column equilibrated in 50 mM Tris-HCl pH 8.5, 500 mM NaCl, and 5% glycerol. The fractions were flash-frozen, and stored at -80℃. All protein purification procedures were performed at 4 °C.

Endonuclease activity assay

The FAM-5'-Poly-A RNA 30 mer was chemically synthesized (Sangon Biotech). Reactions containing 0.06 μM protein and 0.45 μM FAM-5'-Poly-A RNA were performed in a volume of 20 μl with 0.5 U/μl RNasin (Promega), 100 mM HEPES pH 7.0, 100 mM NaCl, 5 mM dithiothreitol (DTT), and divalent cations as indicated in the figure legends, and incubated at 30℃ for 40 minutes. The reaction was stopped by adding an equivalent volume of RNA loading buffer (95% formamide, 0.5 mM EDTA, 0.025% SDS, and 0.025% xylene cyanol) and heating the samples at 100℃ for 3 minutes. Products were separated by electrophoresis on denaturing 7 M urea, 20% polyacrylamide Tris-borate-EDTA gels. Fluorescence was visualized using a ChemiDoc MP (BIO-RAD).

Polymerase activity assay

For the *de novo* initiation of replication, 0.37 μM RVFV polymerase was mixed with 1 μM 20 nt 5' end of the RVFV M genome segment (5' M: 5'-ACACAAAGACGGUGCAUUAA-3') or/and 1 μM 20 nt template RNA (5'-UGUGUUUCUGGCCACGUUGA-3'), NTPs (25 μM ATP, 25 μM GTP, 25 μM CTP and 16.25 μM UTP) and 8.75 μM fluorescein-12-UTP (Sigma) in assay buffer (50 mM Tris-HCl pH7.0, 12 mM DTT, 0.25 mM $MnCl_2$, 4 mM $MgCl_2$). Reactions were incubated at 30℃ for 3 hours and samples were analysed on a 20% acrylamide, 7 M urea denaturing gels. The fluorescence was visualized in a ChemiDoc MP (BIO-RAD).

Cryo-EM sample preparation and data collection

The two vRNA oligonucleotides (5' vRNA 5′-GGUGCA-3′ and 3' vRNA 3′-UGUGUUUCUGGCCACGU-5′) were annealed by heating at 95 °C for 5 min followed by cooling down on bench at room temperature to form partially double-stranded RNA. These two oligonucleotides are conserved 3' and 5' vRNA ends of viral genome and the selection is similar to that described for La Crosse encephalitis virus (LACV) [25,26]. This partially dsRNA was then mixed with L protein in a 1.1:1 molar ratio at 4 °C for 3 hours and the complex L_{RNA} was ultimately resolved on Superdex 200 10/300 GL in a gel filtration buffer (50 mM BTP pH 8.5, 300 mM NaCl). Aliquots of 4 μL purified RVFV L_{RNA} protein at a concentration of ~0.4 mg/ml were applied to glow-discharged holey carbon Cu grids (Quantifoil R1.2/1.3) and 4 μL purified RVFV L_{apo} protein at a concentration of ~0.264 mg/ml were applied to glow-discharged homemade graphene grids [44]. Then these two kinds of grids were flash-cooled in liquid ethane using an FEI Vitrobot Mark IV. After initial screening, the well-prepared cryo-specimens were transferred onto an FEI Titan Krios transmission electron microscope operated at 300 kV for data collection. The microscope was equipped with a GIF-Quantum energy filter, which was used with a slit width of

20 eV. Images were recorded using a K3 Summit direct electron counting camera (Gatan Inc.) at a calibrated magnification of 64,000 (pixel size of 1.08 Å). The defocus range of the datasets was roughly -1.0 to -3.0 μm.

Image processing

Beam induced motion and anisotropic magnification were corrected using the MotionCor2 program [45]. The initial contrast transfer function (CTF) parameters were estimated by the Gctf program [46] at the micrograph level and the micrographs with the estimated resolution limit worse than 5 Å were discarded. The selected Gctf star file was imported into RELION 3.0.8 program [47] and all subsequent particle picking, classification, and reconstruction progress were performed in RELION 3.0.8.

For the RVFV L_{RNA} protein dataset, about 5,090,000 particles were automatically selected from about 3,700 micrographs. After two rounds of extensive 2D classification, a subset of 3,380,000 particles were selected and further subjected to 3D classification with an initial model that was produced by the RELION program and low-pass filtered to a resolution of 40 Å. Because of the large amount of particles, we chose to split the particles into 3 groups and did the 3D classification separately. We used the same procedure of 3D classification on all the 3 groups separately. During the 3D classification, we used an angular sampling interval of 7.5 degree and did not perform local angular search. After all the three groups were executed with primary 3D classification, we joined the particles into a total dataset of 1,184,890 particles to do further 3D classification. After two rounds of 3D classification, a reconstruction from 554,102 particles displayed clear features of secondary structural elements. We then performed 3D refinement from the ~550,000 particles, yielding a reconstruction at a resolution of 3.84 Å. After postprocessing, we got a final map at 3.6 Å resolution as determined by the Fourier shell correlation (FSC) 0.143 cut-off value.

For the RVFV L_{apo} protein, we totally had three datasets. For each dataset, we did primary 2D classification and first round of 3D classification as the L_{RNA} dataset by the RELION program. What is different from the L_{RNA} dataset is that we could find out the endonuclease domain density in the second round of 3D classification although the number of these particles that contain endonuclease domain density is small. From the three datasets, we finally got 303,964 endonuclease domain density containing particles together to perform 3D refinement, yielding a reconstruction at a resolution of 4.01 Å. After postprocessing, we got a final map at 3.84 Å resolution as determined by the Fourier shell correlation (FSC) 0.143 cut-off value.

Model building

The SFTSV L structure was used to guide manual building of RVFV L_{RNA} by Chainsaw [48]. Manual model building was carried out using Coot [49] and refinement of the coordinates was performed using Phenix [50]. The structure of apo is a structural analysis based on RNA-bound L as a model.

Small-angle X-ray scattering

The overall model of the full-length RVFV L protein was obtained with the solved RVFV Lapo (without the Lariat) and the simulated PB2 C-terminal domain built from the structure of SFTSV (PDB 6L42) using the SASREF [51] with the distance restriction between the linker domain and the Lariat.

Small-angle X-ray scattering measurements were performed on beamline BL19U2 at the SSRF following previously published methods [52]. Briefly, all proteins were subjected to size exclusion chromatography in a buffer containing 20 mM BTP pH 8.5, 300 mM NaCl. The BioXTAS RAW software was used for data processing. GNOM provided the pair distribution function P(r) of the particle, the maximum size D_{max} and the Porod volume. The 20 individual ab initio reconstructions were generated with DAMMIN, averaged using DAMAVER, and aligned using SUPCOMB. The structures were visualized using PyMOL.

Data availability

Structural data are available from the PDB database (http://www.rcsb.org/pdb/home/home.do) (accession numbers 7EEI and 7EFE). EM data have been deposited with the EMDB (accession number EMD-31077 and EMD-31085). All remaining relevant data are included in the manuscript and its supplementary material.

Acknowledgments:

We thank J.L. Lei, Y.J. Xu and T. Yang for their support in cryo-EM and high-performance computation in the National Protein Science Facility (Beijing) at Tsinghua University. We thank J. Wen and X.M. Li for their friendly help in cryo-EM sample screening in the National Protein Science Facility (Beijing) at Tsinghua University. We thank Y. Li, C. Liu for some cryo-EM dataset collection at Shuimu Inc. Beijing. We would like to thank the staff of beamlines BL19U2 at the Shanghai Synchrotron Radiation Facility for the excellent technical assistance.

Funding

This work was supported financially by National Key Research and Development Program of China [2018YFE0113100, 2016YFA0501100 to H.-W.W.]; National Natural Science Foundation of China [31825009 to H.-W.W. and 31872713, 32071210, and 31800629], Open Fund of the State Key Laboratory of Pathogenic Microbial Biosafety [SKLPBS1834].

Declaration of Competing Interest

The authors declare no conflict of interest.

Author Contributions

Conceptualization, Z.C. and H.W.; methodology, Z.C. and H.W.; software, X.W., C.H., Z.C. and H.W.; formal analysis, W.Y., J.W., X.D., J.X., X.L., M.Z., H.L., W.W. and F.Z.; investigation, X.W., C.H., Z.C. and H.W.; resources, Z.C. and H.W.; data curation, X.W., C.H., Z.C. and H.W.; writing—original draft preparation, X.W., C.H., Z.C. and H.W.; writing—review and editing, X.W., C.H., Z.C. and H.W.; supervision, Z.C. and H.W.; project administration, Z.C. and H.W.; funding acquisition, Z.C., H.W.

References:

1. Freiberg, A.N., Sherman, M.B., Morais, M.C., Holbrook, M.R. & Watowich, S.J. Three-dimensional organization of Rift Valley fever virus revealed by cryoelectron tomography. *J. Virol.* **82**, 10341-8 (2008).
2. Komoda, K., Narita, M., Yamashita, K., Tanaka, I. & Yao, M. Asymmetric Trimeric Ring Structure of the Nucleocapsid Protein of Tospovirus. *J. Virol.* **91**, e01002-17 (2017).
3. Reguera, J. et al. Comparative Structural and Functional Analysis of Bunyavirus and Arenavirus Cap-Snatching Endonucleases. *PLoS Pathog* **12**, e1005636 (2016).
4. Eifan, S., Schnettler, E., Dietrich, I., Kohl, A. & Blomstrom, A.L. Non-structural proteins of arthropod-borne bunyaviruses: roles and functions. *Viruses* **5**, 2447-68 (2013).
5. Salekwa, L.P., Wambura, P.N., Matiko, M.K. & Watts, D.M. Circulation of Rift Valley Fever Virus Antibody in Cattle during Inter-Epizootic/Epidemic Periods in Selected Regions of Tanzania. *Am. J. Trop. Med. Hyg.* **101**, 459-466 (2019).
6. Flick, R., et al. Rift Valley fever virus. *Curr Mol Med* **5**, 827-34 (2005).
7. Guardado-Calvo, P. et al. A glycerophospholipid-specific pocket in the RVFV class II fusion protein drives target membrane insertion. *Science* **358**, 663-667 (2017).
8. Golnar, A.J., Turell, M.J., LaBeaud, A.D., Kading, R.C. & Hamer, G.L. Predicting the mosquito species and vertebrate species involved in the theoretical transmission of Rift Valley fever virus in the United States. *PLoS Negl Trop Dis* **8**, e3163 (2014).
9. Vloet, R.P.M. et al. Transmission of Rift Valley fever virus from European-breed lambs to Culex pipiens mosquitoes. *PLoS Negl Trop Dis* **11**, e0006145 (2017).
10. Nielsen, S.S. et al. Rift Valley Fever - epidemiological update and risk of introduction into Europe. *EFSA J.* **18**, e06041 (2020).
11. Wang, Q. et al. Neutralization mechanism of human monoclonal antibodies against Rift Valley fever virus. *Nat Microbiol* **4**, 1231-1241 (2019).
12. Balenghien, T. et al. Towards a better understanding of Rift Valley fever epidemiology in the south-west of the Indian Ocean. *Vet. Res.* **44**, 78 (2013).
13. Lokugamage, N., Freiberg, A.N., Morrill, J.C. & Ikegami, T. Genetic subpopulations of Rift Valley fever virus strains ZH548 and MP-12 and recombinant MP-12 strains. *J. Virol.* **86**, 13566-75 (2012).
14. Gogovi, G.K. et al. Modeling the Tertiary Structure of the Rift Valley Fever Virus L Protein. *Molecules* **24, 1768**(2019).
15. Balkhy, H.H., et al. Rift Valley fever: an uninvited zoonosis in the Arabian peninsula. *Int J Antimicrob Agents* **21**, 153-7 (2003).
16. Terasaki, K., Murakami, S., Lokugamage, K.G. & Makino, S. Mechanism of tripartite RNA genome packaging in Rift Valley fever virus. *Proc Natl Acad Sci U S A* **108**, 804-9 (2011).
17. Pflug, A., Guilligay, D., Reich, S. & Cusack, S. Structure of influenza A polymerase bound to the viral RNA promoter. *Nature* **516**, 355-60 (2014).
18. Reich, S. et al. Structural insight into cap-snatching and RNA synthesis by influenza polymerase. *Nature* **516**, 361-6 (2014).
19. Gogrefe, N., Reindl, S., Gunther, S. & Rosenthal, M. Structure of a functional cap-binding domain in Rift Valley fever virus L protein. *PLoS Pathog* **15**, e1007829 (2019).
20. Ferron, F. et al. The hexamer structure of Rift Valley fever virus nucleoprotein suggests a

mechanism for its assembly into ribonucleoprotein complexes. *PLoS Pathog* **7**, e1002030 (2011).
21. Vogel, D. et al. Structural and functional characterization of the severe fever with thrombocytopenia syndrome virus L protein. *Nucleic Acids Res.* **48**, 5749-5765 (2020).
22. Kumar, G., Cuypers, M., Webby, R.R., Webb, T.R. & White, S.W. Structural insights into the substrate specificity of the endonuclease activity of the influenza virus cap-snatching mechanism. *Nucleic Acids Res.* **49**, 1609-1618 (2021).
23. Heinrich, B.S., Cureton, D.K., Rahmeh, A.A. & Whelan, S.P. Protein expression redirects vesicular stomatitis virus RNA synthesis to cytoplasmic inclusions. *PLoS Pathog* **6**, e1000958 (2010).
24. Wang, P. et al. Structure of severe fever with thrombocytopenia syndrome virus L protein elucidates the mechanisms of viral transcription initiation. *Nat Microbiol* **5**, 864-871 (2020).
25. Gerlach, P., Malet, H., Cusack, S. & Reguera, J. Structural Insights into Bunyavirus Replication and Its Regulation by the vRNA Promoter. *Cell* **161**, 1267-79 (2015).
26. Arragain, B. et al. Pre-initiation and elongation structures of full-length La Crosse virus polymerase reveal functionally important conformational changes. *Nat. Commun.* **11**, 3590 (2020).
27. Bouloy, M., et al. Molecular biology of rift valley Fever virus. *Open Virol J* **4**, 8-14 (2010).
28. te Velthuis, A.J. Common and unique features of viral RNA-dependent polymerases. *Cell Mol. Life Sci.* **71**, 4403-20 (2014).
29. Jacome, R., Becerra, A., Ponce de Leon, S. & Lazcano, A. Structural Analysis of Monomeric RNA-Dependent Polymerases: Evolutionary and Therapeutic Implications. *PLoS ONE* **10**, e0139001 (2015).
30. Yang, X. et al. Motif D of viral RNA-dependent RNA polymerases determines efficiency and fidelity of nucleotide addition. *Structure* **20**, 1519-27 (2012).
31. Venkataraman, S., Prasad, B. & Selvarajan, R. RNA Dependent RNA Polymerases: Insights from Structure, Function and Evolution. *Viruses* **10**, 76 (2018).
32. Thierry, E. et al. Influenza Polymerase Can Adopt an Alternative Configuration Involving a Radical Repacking of PB2 Domains. *Mol Cell* **61**, 125-37 (2016).
33. Madeira, F. et al. The EMBL-EBI search and sequence analysis tools APIs in 2019. *Nucleic Acids Res.* **47**, W636-W641 (2019).
34. Te Velthuis, A.J., Robb, N.C., Kapanidis, A.N. & Fodor, E. The role of the priming loop in Influenza A virus RNA synthesis. *Nat Microbiol* **1**, 16029 (2016).
35. Bailey, T.L., et al. Fitting a mixture model by expectation maximization to discover motifs in biopolymers. *Proc Int Conf Intell Syst Mol Biol* **2**, 28-36 (1994).
36. Liang, B. et al. Structure of the L Protein of Vesicular Stomatitis Virus from Electron Cryomicroscopy. *Cell* **162**, 314-327 (2015).
37. Wandzik, J.M. et al. A Structure-Based Model for the Complete Transcription Cycle of Influenza Polymerase. *Cell* **181**, 877-893 (2020).
38. Raymond, D.D., Piper, M.E., Gerrard, S.R. & Smith, J.L. Structure of the Rift Valley fever virus nucleocapsid protein reveals another architecture for RNA encapsidation. *Proc Natl Acad Sci U S A* **107**, 11769-11774 (2010).
39. Te Velthuis, A.J. Influenza virus RNA polymerase: insights into the mechanisms of viral

RNA synthesis. *Nat Rev Microbiol* **14**, 479-93 (2016).
40. Scharton, D. et al. Favipiravir (T-705) protects against peracute Rift Valley fever virus infection and reduces delayed-onset neurologic disease observed with ribavirin treatment. *Antiviral Res* **104**, 84-92 (2014).
41. Kende, M., Alving, C.R., Rill, W.L., Swartz, G.M., Jr. & Canonico, P.G. Enhanced efficacy of liposome-encapsulated ribavirin against Rift Valley fever virus infection in mice. *Antimicrob Agents Chemother* **27**, 903-7 (1985).
42. Lang, Y. et al. Identification and evaluation of antivirals for Rift Valley fever virus. *Vet. Microbiol.* **230**, 110-116 (2019).
43. Smee, D.F., Jung, K.H., Westover, J. & Gowen, B.B. 2'-Fluoro-2'-deoxycytidine is a broad-spectrum inhibitor of bunyaviruses in vitro and in phleboviral disease mouse models. *Antiviral Res* **160**, 48-54 (2018).
44. Zheng, L. et al. Robust ultraclean atomically thin membranes for atomic-resolution electron microscopy. *Nat. Commun.* **11**, 541 (2020).
45. Zheng, S.Q. et al. MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nat Methods* **14**, 331-332 (2017).
46. Zhang, K. Gctf: Real-time CTF determination and correction. *J. Struct. Biol.* **193**, 1-12 (2016).
47. Scheres, S.H. RELION: implementation of a Bayesian approach to cryo-EM structure determination. *J. Struct. Biol.* **180**, 519-30 (2012).
48. Stein, N. CHAINSAW: a program for mutating pdb files used as templates in molecular replacement. *J. Appl. Crystallogr.* **41**, 641-643 (2008).
49. Emsley, P., et al. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.* **60**, 2126-32 (2004).
50. Afonine, P.V. et al. Towards automated crystallographic structure refinement with phenix.refine. *Acta Crystallogr. D Biol. Crystallogr.* **68**, 352-67 (2012).
51. Petoukhov, M.V. & Svergun, D.I. Global rigid body modeling of macromolecular complexes against small-angle scattering data. *Biophys. J.* **89**, 1237-50 (2005).
52. Gao, Y. et al. Human apo-SRP72 and SRP68/72 complex structures reveal the molecular basis of protein translocation. *J Mol Cell Biol* **9**, 220-230 (2017).

Manuscript title: Structural insights into Rift Valley fever virus replication machinery

Xue Wang[1,4], Cuixia Hu[2,4], Wei Ye[3], Jia Wang[2], Xiaofei Dong[1], Jie Xu[2], Xiaorong Li[1], Manfeng Zhang[1], Hongyun Lu[1], Fanglin Zhang[3], Wei Wu[1], Hong-Wei Wang[2,*] and Zhongzhou Chen[1,*]

[1] State Key Laboratory of Agrobiotechnology and Beijing Advanced Innovation Center for Food Nutrition and Human Health, College of Biological Sciences, China Agricultural University, Beijing, China, 100193

[2] Ministry of Education Key Laboratory of Protein Science, Tsinghua-Peking Joint Center for Life Sciences, Beijing Advanced Innovation Center for Structural Biology, School of Life Sciences, Tsinghua University, Beijing, China, 100084

[3] Department of Microbiology, School of Preclinical Medicine, Fourth Military Medical University, Xi'an, China

[4] These authors contributed equally: Xue Wang, Cuixia Hu.

* To whom correspondence should be addressed:
Zhongzhou Chen
E-mail: chenzhongzhou@cau.edu.cn
Hong-Wei Wang
E-mail: hongweiwang@tsinghua.edu.cn (H.-W.W.)

Supplementary data

Supplementary Table S1. Cryo-EM data collection, refinement, and validation statistics.

	RVFV L_{RNA}	RVFV L_{apo}
Data collection and processing		
Magnification	64,000	64,000
Voltage (kV)	300	300
Electron exposure ($e^-/Å^2$)	50	50
Defocus range (μm)	-1.0 to -3.0	-1.0 to -3.0
Pixel size (Å)	1.08	1.08
Symmetry imposed	C1	C1
Final particle images (no.)	660K	303K
Map resolution (Å)	3.6	3.84
FSC threshold	0.143	0.143
Map resolution range (Å)	3.5-4.7	3.7-5.7
Refinement		
Initial model used (PDB code)	6l42	
Map sharpening *B* factor ($Å^2$)	-180	-180
Model composition		
Non-hydrogen atoms	10249	10273
Protein residues	1401	1410
B factors ($Å^2$)		
Protein	104.67	116.07
R.m.s deviation		
Bond lengths (Å)	0.009	0.009
Bond angles (°)	1.498	1.103
Validation		
MolProbity score	1.76	1.41
Clashscore	5.25	7.54
Poor rotamers (%)	0.42	0.32
Ramachandran plot		
Favored (%)	92.26	100.00
Allowed (%)	7.74	0.00
Disallowed (%)	0.00	0.00

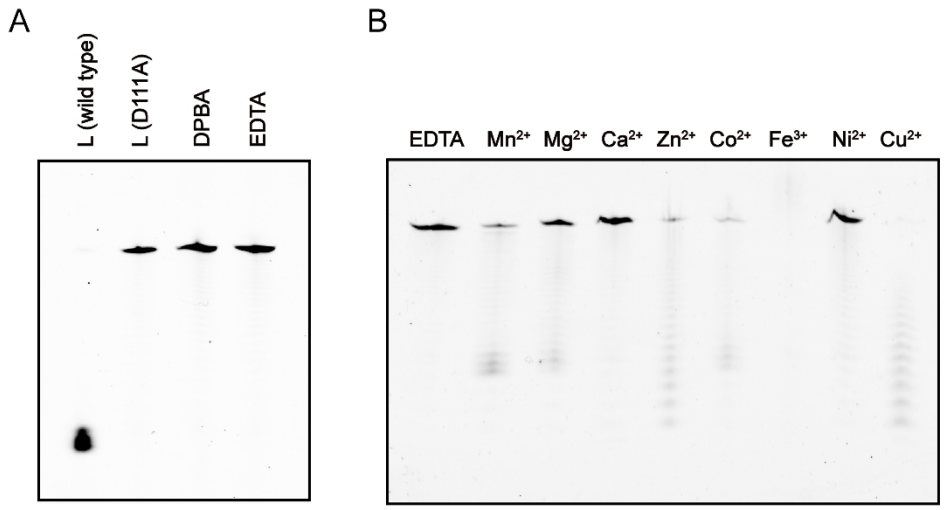


Figure S1. Endonuclease activity assay.

(A) 0.06 μM RVFV L protein was incubated at 30℃ for 40 min with 0.45 μM fluorescently labeled 30 nt PolyA RNA substrate. Reactions in the presence of RVFV L (D111A), EDTA, or DPBA were negative controls.

(B) 0.06 μM RVFV L protein was incubated at 30℃ for 40 min with 0.45 μM fluorescently labeled 30 nt PolyA RNA substrate in the presence of 25 mM $MnCl_2$, $MgCl_2$, $CaCl_2$, $ZnSO_4$, $CoCl_2$, $FeCl_3$, $NiCl_2$, or $CuSO_4$. The reaction in the presence of EDTA was negative control.

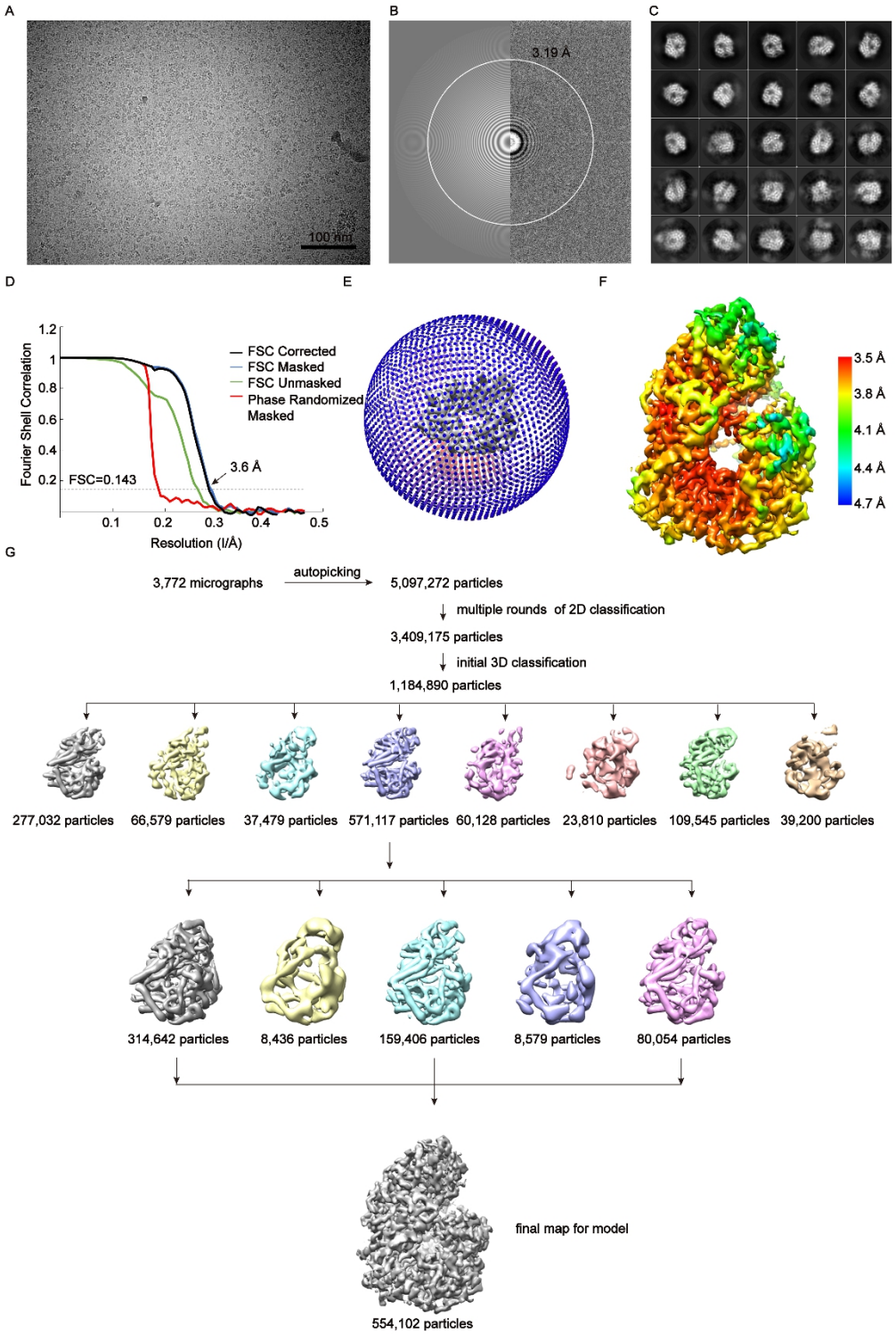


Figure S2. Cryo-EM reconstruction of RVFV L_{RNA} protein

(A) Raw image of RVFV L_{RNA} protein particles in vitreous ice recorded by K3 camera at defocus range of -1.5 to 2.5 µm. Scale bar 100 nm.

(B) Power spectrum of image (A) and the white circle indicated the spatial frequency corresponding to 3.19 Å estimated by the Gctf program.

(C) Representative 2D classes. The box size of each class was 172.8 Å.
(D) Fourier shell correlation (FSC) curves of the postprocessed 3D map, following the gold standard refinement.
(E) Euler angle distribution.
(F) Local resolution was performed with Relion3.1
(G) The flowthrough of the data processing used to get the final map.

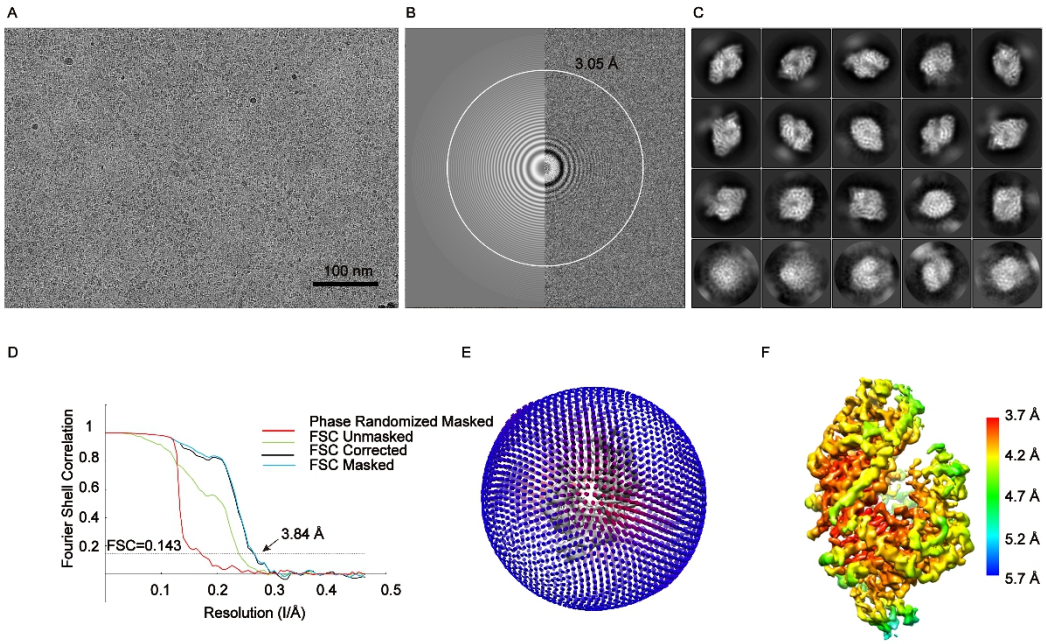


Figure S3. Cryo-EM reconstruction of RVFV L_{apo} protein

(A) Representative micrograph of RVFV L_{apo} protein in vitreous ice after MotionCorr2 at defocus of ~1.6 µm, scale bar 100 nm.

(B) Power spectrum of the image shown in (A). White circle indicated the spatial frequency corresponding to 3.05 Å resolution.

(C) 2D classification result of the RVFV L_{apo} protein.

(D) Fourier Shell Correlation (FSC) curves between the independently refined two half-maps showed the overall 3.84 Å resolution of the density map by the 0.143 cut-off.

(E) Angular distribution for particle projections visualized as a sphere around the RVFV L_{apo} final density map.

(F) Local resolution estimation result, performed by Relion3.1.

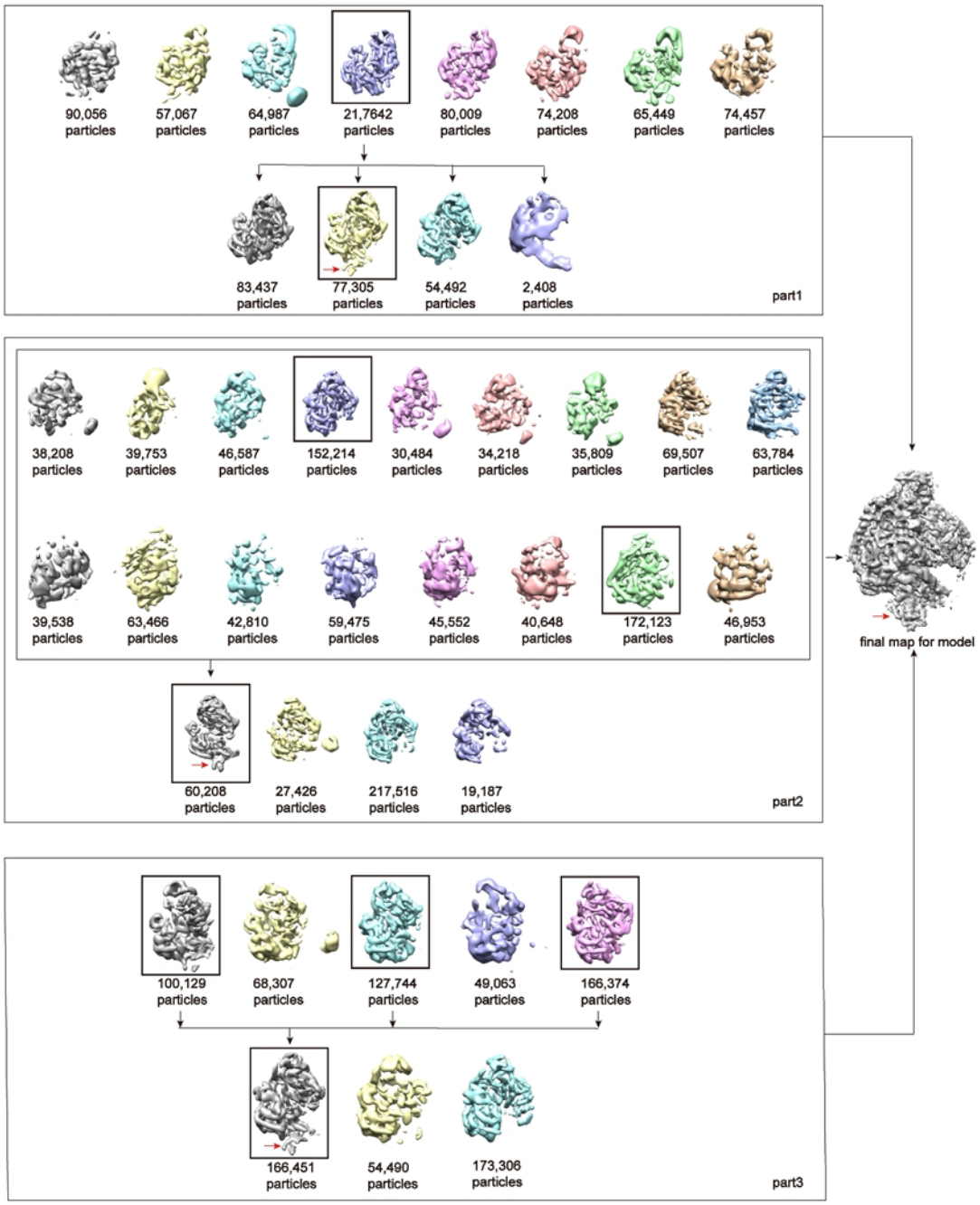
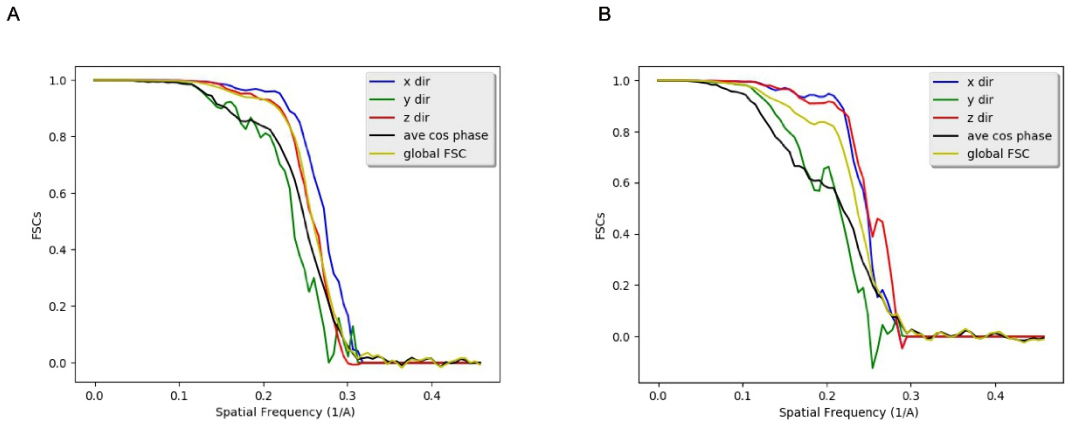


Figure S4. Cryo-EM data processing procedures of RVFV L_{apo} protein.

At the beginning of the data processing procedure, part1, we only collected about ~1,500 micrographs to test whether this sample had the potential that could be reconstructed to a high-resolution map or not. It's delightful that we discovered the L protein's N-terminal density, indicated by the red arrow, in the second time of 3D classification. However, the number of protein particles that contain the endonuclease domain was not enough to reconstruct high-resolution map for modeling. Therefore, we collected more cryo-EM data of the same grid as part1, and the processing progress of next two data collections was shown in part2 and part3, respectively. Subsequently, we got 303,964 particles to do 3D auto-refinement and postprocessing

sharpening, which resulted in a map of 3.84 Å resolution calculated by the FSC curve with 0.143 cutoff. In the end, we adopted this 3.84 Å resolution map for the model building.



71

72 **Figure S5. Directional FSC profile of (A) L_{RNA} and (B) L_{apo}**

73 3DFSC plots of the final data were analyzed via the web portal for 3DFSC (https://3dfsc.salk.edu)

74 [1].

75

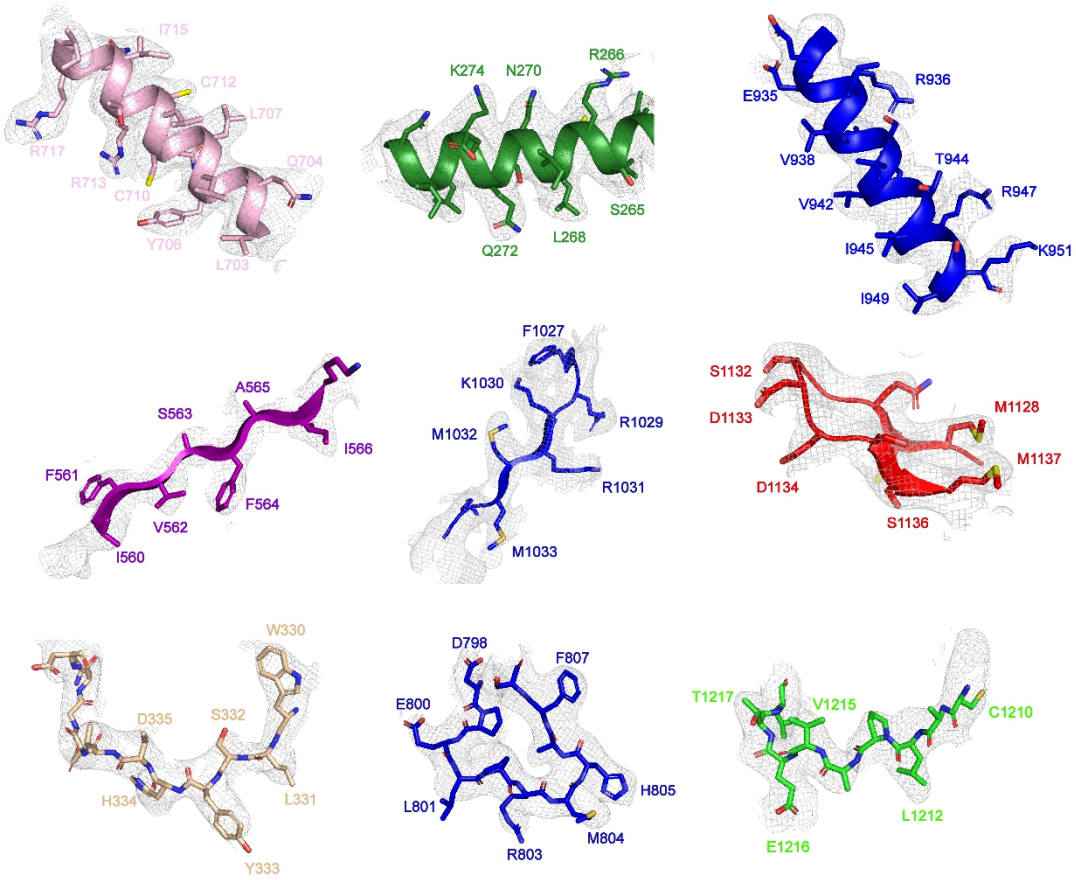


Figure S6. Representative cryo-EM map

Representative regions of the protein model, colored as Figure 2A.

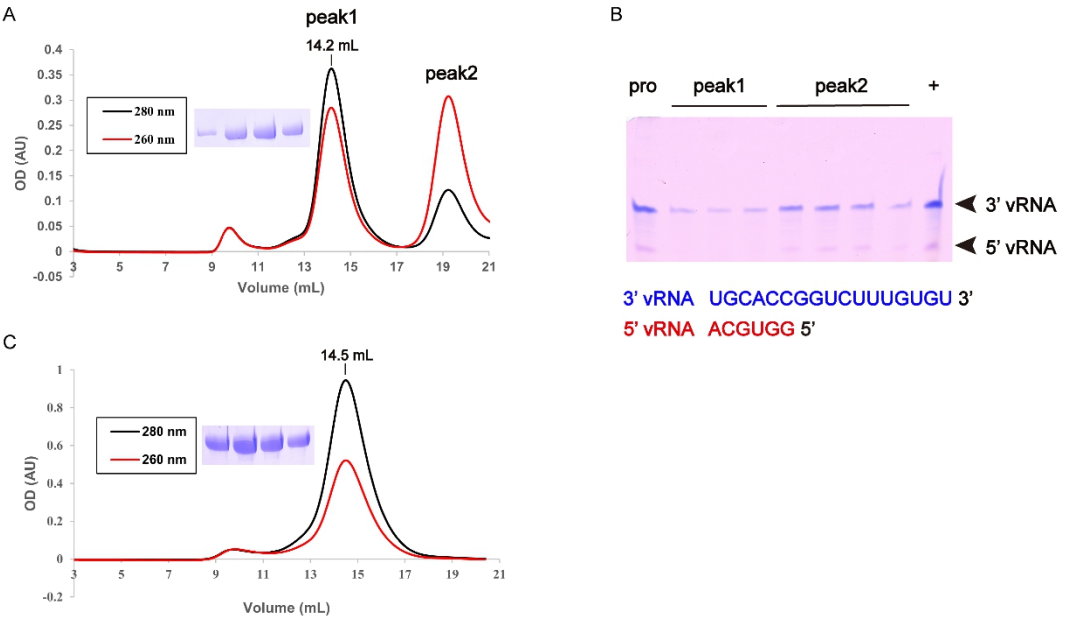
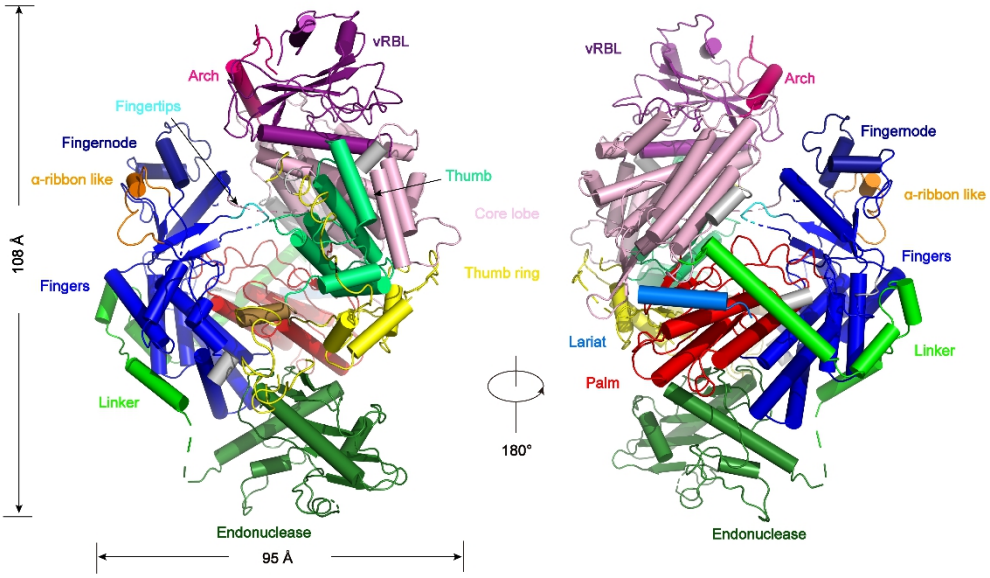


Figure S7. RNAs were co-purified with RVFV L protein

(A) Size-exclusion chromatogram (Superdex 200 10/300 GL) of RVFV L protein and RNA. The absorbance curves for the sample at wavelengths of 260 nm and 280 nm were given.

(B) RNA detection was performed on the collection corresponding to the SEC peak on the left by electrophoresis on 20% polyacrylamide gels containing 7 M urea. 3' vRNA stands for 3′-UGUGUUUCUGGCCACGU-5′, 5' vRNA is 5′-GGUGCA-3′. Pro stands for before the column. + stands for individual RNA samples.

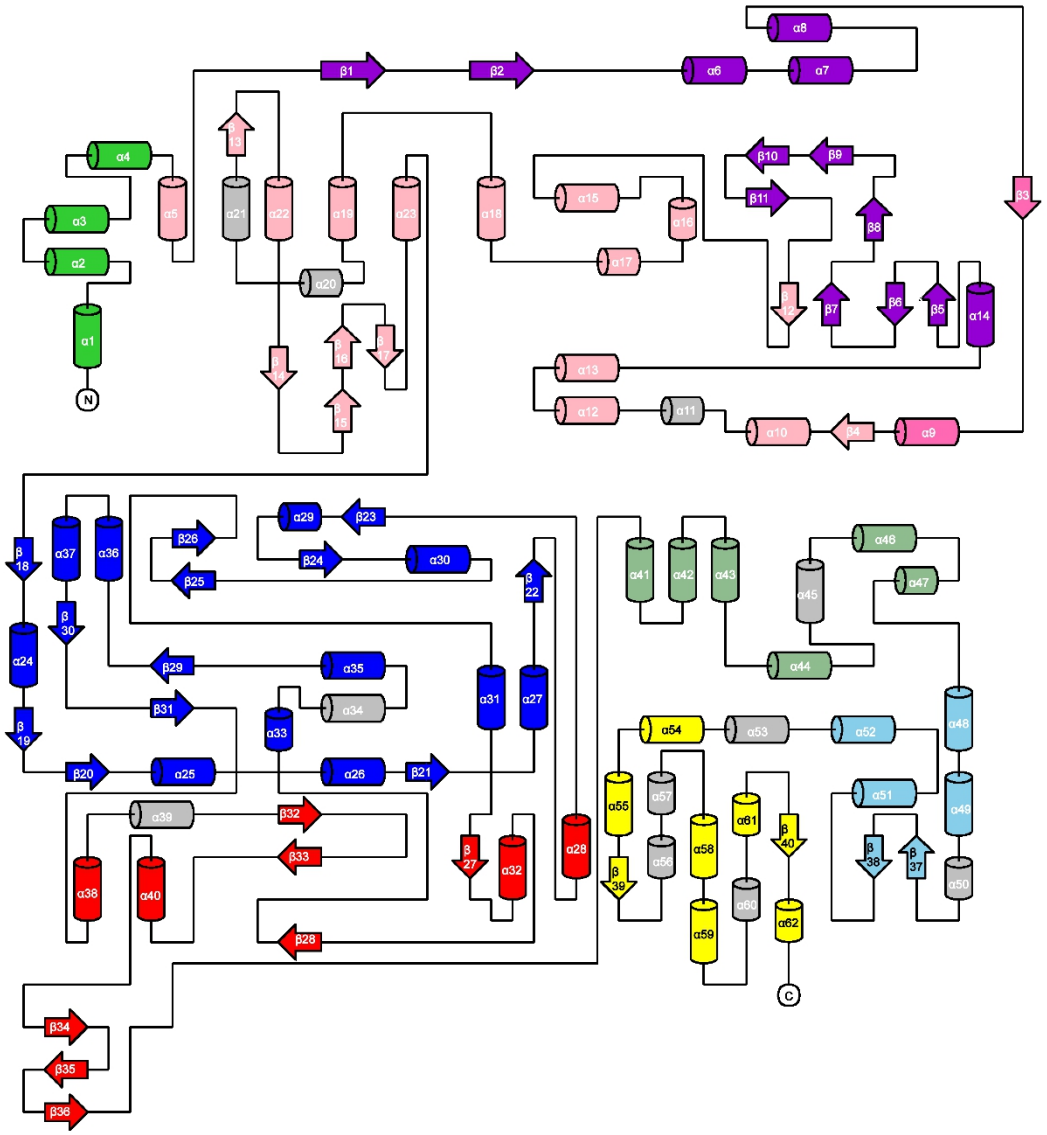
(C) Size-exclusion chromatogram (Superdex 200 10/300 GL) of RVFV L protein. The absorbance curves for the sample at wavelengths of 260 nm and 280 nm were given.



91

92 **Figure S8. Overall structure of the RVFV L_{apo} protein.**

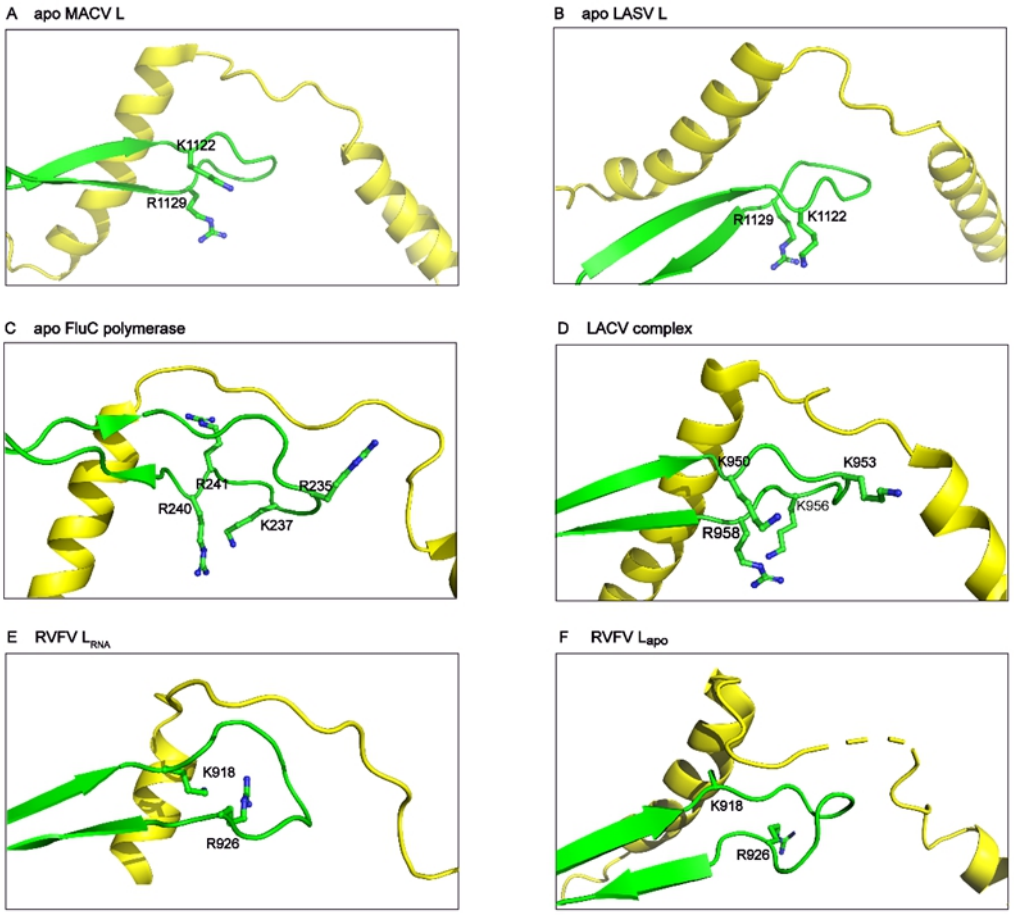
93 Overall dimension of this structure is about 95 Å x 70 Å x 108 Å. The dimension is larger than
94 that of RVFV L_{RNA} protein due to the resolving of the endonuclease domain in L_{apo} structure.



95

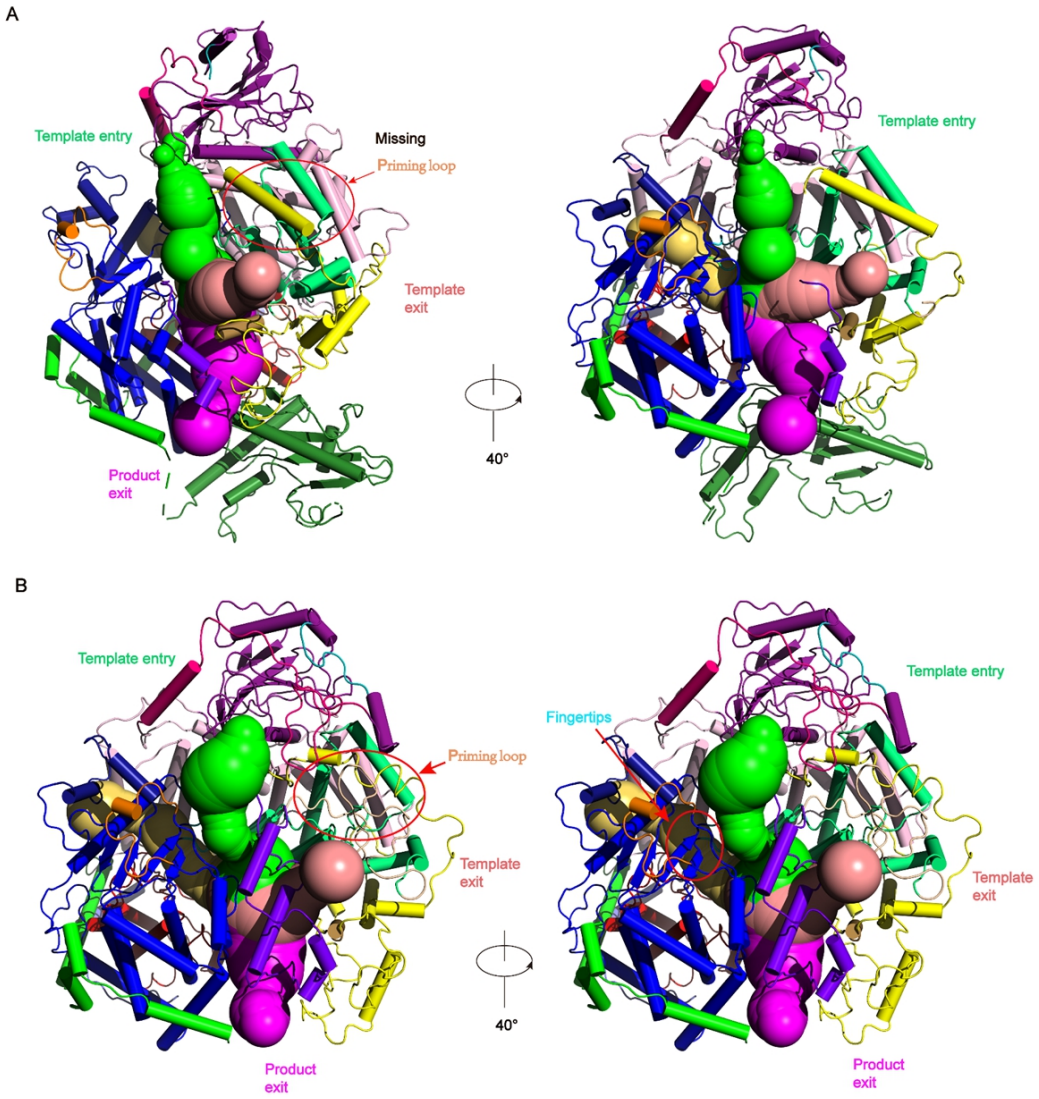
96 **Figure S9. Topology of RVFV L_{RNA} protein.**

97 3_{10} helices were colored gray. Linker domain was colored green, core lobe domain was colored
98 lightpink, vRBL domain was colored deeppurple, fingers domain was colored blue, palm domain
99 was colored red, thumb domain was colored darkseagreen, bridge domain was colored navyblue,
100 thumb ring domain was colored yellow.



101

102 **Figure S10. The structures for fingertips loops of apo MACV, apo LASV, apo FluC, LACV**
103 **complex, RVFV L_{RNA}, and RVFV L_{apo} polymerase.**
104 The basic residues on the fingertips domain were marked. The side chain of K918 in RVFV L_{apo}
105 was flexible in the density.

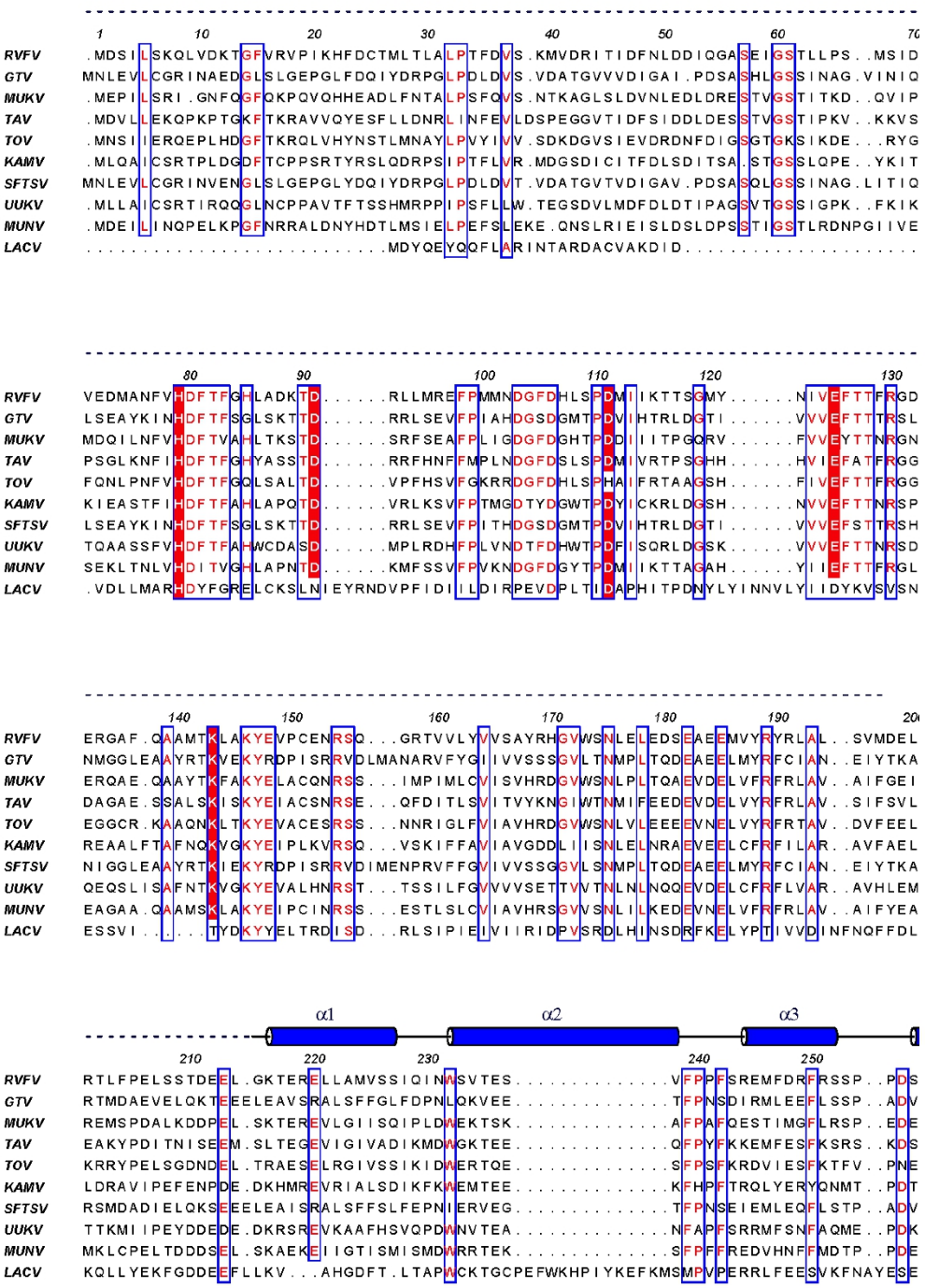


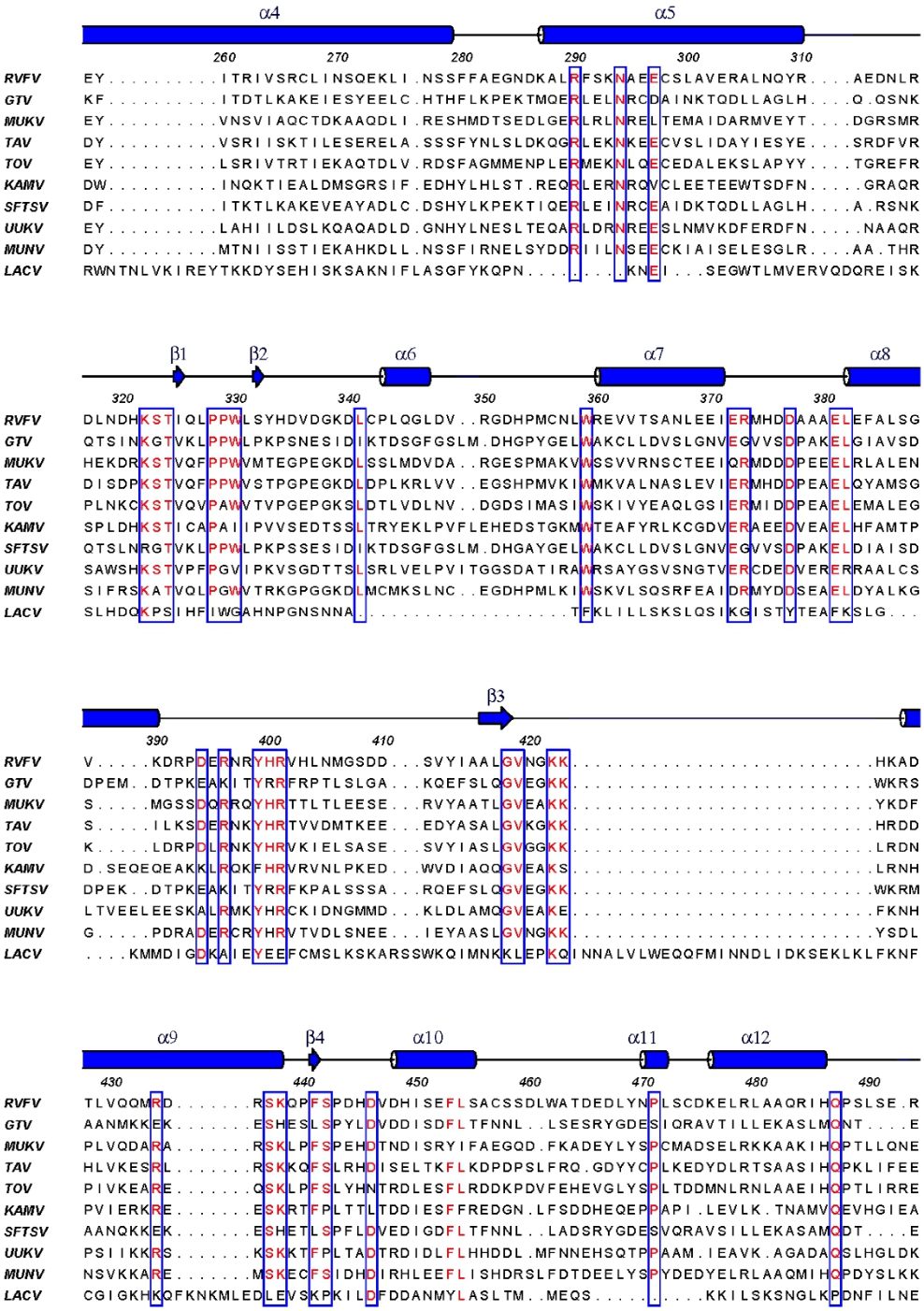
106

107 **Figure S11. The entrance and exit channels comparison of RVFV L_{apo} protein and RVFV**
108 **L_{RNA} protein**

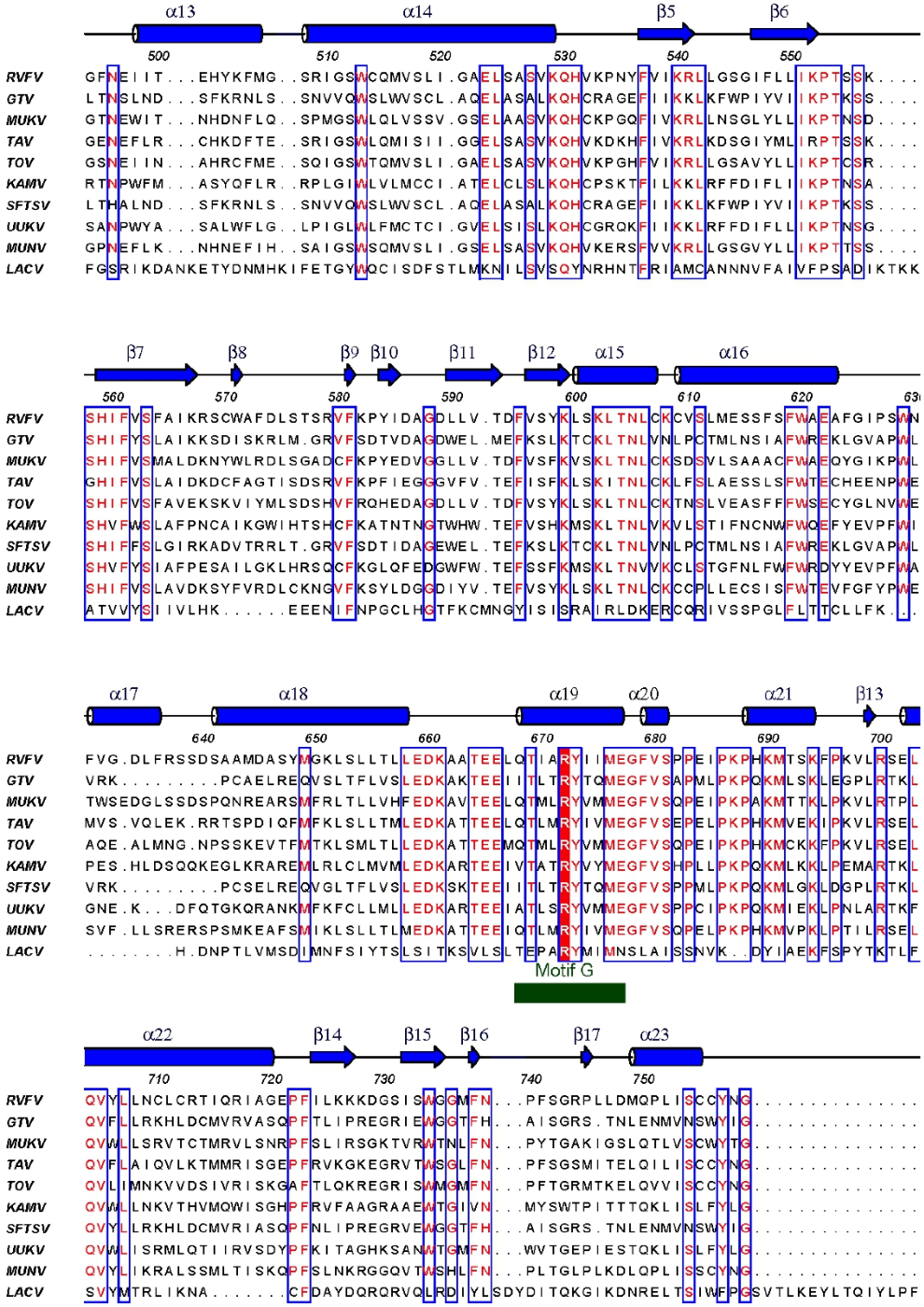
109 (A) RVFV L_{apo} protein. Note that the priming loop is missing in the structure of L_{apo}.

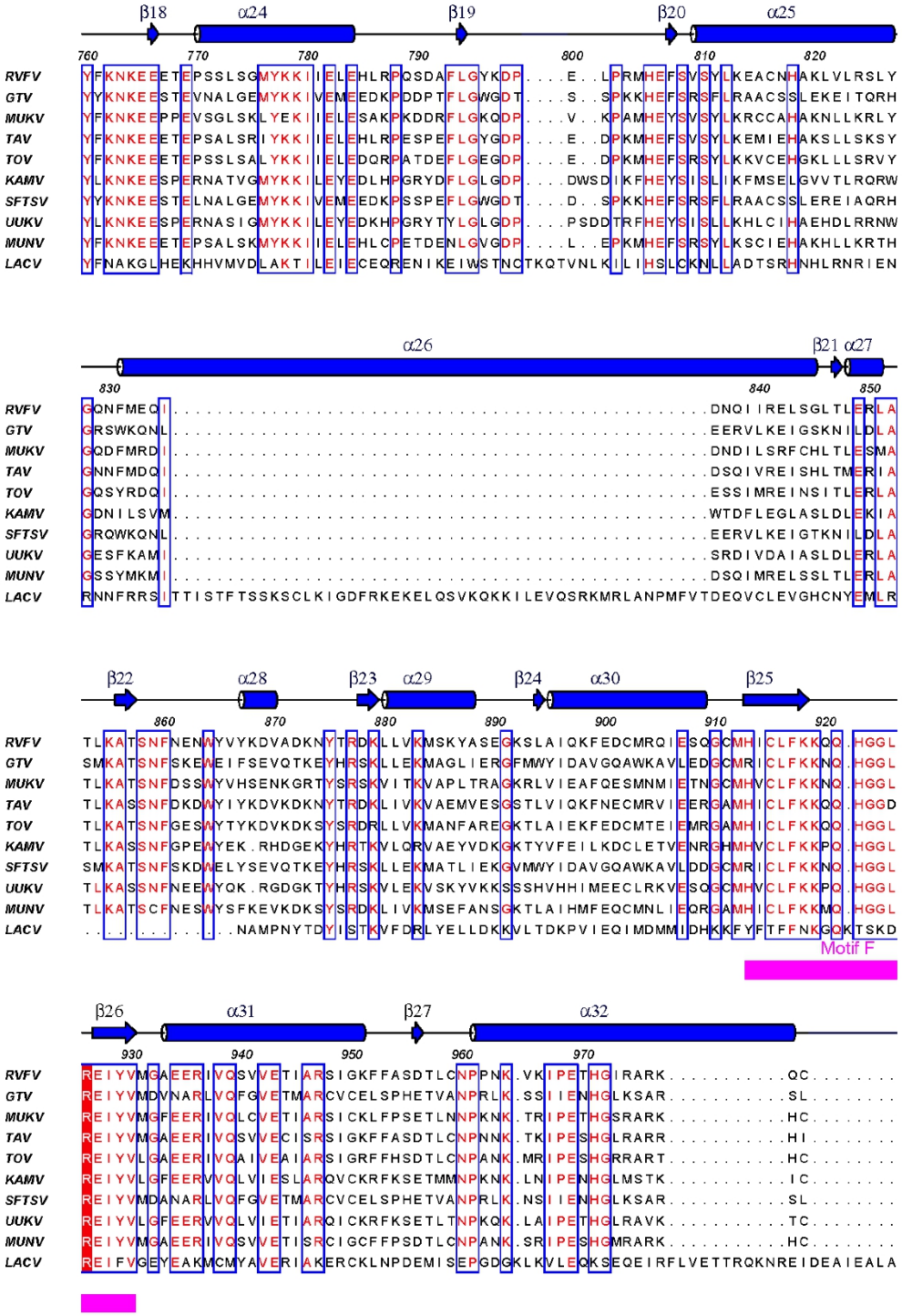
110 (B) RVFV L_{RNA} protein.

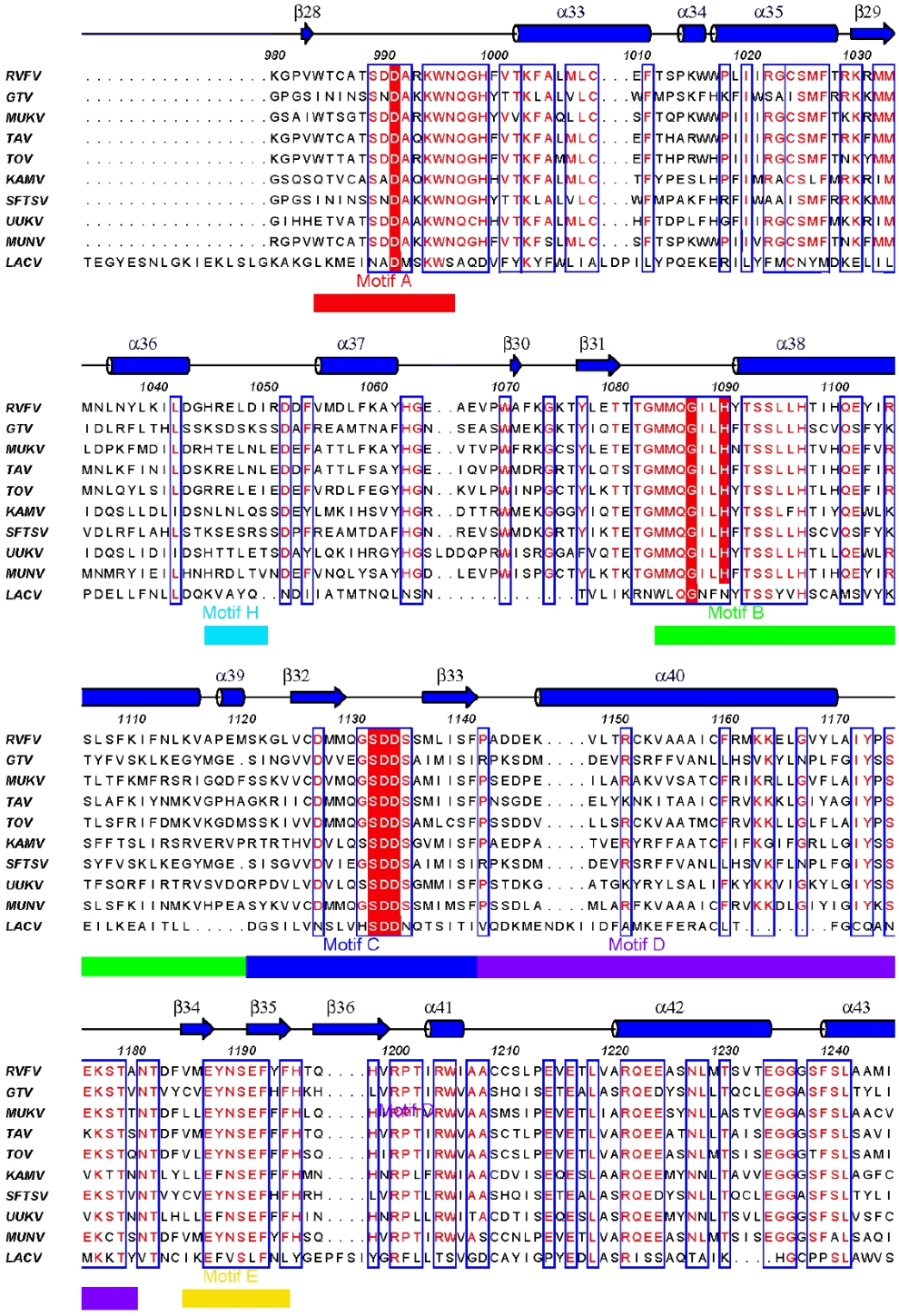

RVFV
GTV
MUKV
TAV
TOV
KAMV
SFTSV
UUKV
MUNV
LACV
α1
α2
α3


α4
α5
260 270 280 290 300 310
RVFV EY.....ITRIVSRCLINSQEKLI.NSSFFAEGNDKALRFSKNAEECSLAVERALNQYR....AEDNLR
GTV KF.....ITDTLKAKEIESYEELC.HTHFLKPEKTMQERLELNRCDAINKTQDLLAGLH....Q.QSNK
MUKV EY.....VNSVIAQCTDKAAQDLI.RESHMDTSEDLGERLRLNRELTEMAIDARMVEYT....DGRSMR
TAV DY.....VSRIISKTILESERELA.SSSFYNLSLDKQGRLEKNKEECVSLIDAYIESYE....SRDFVR
TOV EY.....LSRIVTRTIEKAQTDLV.RDSFAGMMENPLERMEKNLQECEDALEKSLAPYY....TGREFR
KAMV DW.....INQKTIEALDMSGRSIF.EDHYLHLST.REQRLERNRQVCLEETEEWTSDFN.....GRAQR
SFTSV DF.....ITKTLKAKEVEAYADLC.DSHYLKPEKTIQERLEINRCEAIDKTQDLLAGLH....A.RSNK
UUKV EY.....LAHIILDSLKQAQADLD.GNHYLNESLTEQARLDRNREESLNMVKDFERDFN.....NAAQR
MUNV DY.....MTNIISSTIEKAHKDLL.NSSFIRNELSYDDRIILNSEECKIAISELESGLR....AA.THR
LACV RWNTNLVKIREYTKKDYSEHISKSAKNIFLASGFYKQPN.....KNEI...SEGWTLMVERVQDQREISK
β1 β2 α6 α7 α8
320 330 340 350 360 370 380
RVFV DLNDHKSTIQLPPWLSYHDVDGKDLCPLQGLDV..RGDHPMCNLWREVVTSANLEEIERMHDDAAAELEFALSG
GTV QTSINKGTVKLPPWLPKPSNESIDIKTDSGFGSLM.DHGPYGELWAKCLLDVSLGNVEGVVSDPAKELGIAVSD
MUKV HEKDRKSTVQFPPWVMTEGPEGKDLSSLMDVDA..RGESPMAKVWSSVVRNSCTEEIQRMDDDPEEELRLALEN
TAV DISDPKSTVQFPPWVSTPGPEGKDLDPLKRLVV..EGSHPMVKIWMKVALNASLEVIERMHDDPEAELQYAMSG
TOV PLNKCKSTVQVPAWVTVPGEPGKSLDTLVDLNV..DGDSIMASIWSKIVYEAQLGSIERMIDDPEAELEMALEG
KAMV SPLDHKSTICAPAIIPVVSEDTSSLTRYEKLPVFLEHEDSTGKMWTEAFYRLKCGDVERAEEDVEAELHFAMTP
SFTSV QTSLNRGTVKLPPWLPKPSSESIDIKTDSGFGSLM.DHGAYGELWAKCLLDVSLGNVEGVVSDPAKELDIAISD
UUKV SAWSHKSTVPFPGVIPKVSGDTTSLSRLVELPVITGGSDATIRAWRSAYGSVSNGTVERCDEDVERERRAALCS
MUNV SIFRSKATVQLPGWVTRKGPGGKDLMCMKSLNC..EGDHPMLKIWSKVLSQSRFEAIDRMYDDSEAELDYALKG
LACV SLHDQKPSIHFIWGAHNPGNSNNA.....TFKLILLSKSLQSIKGISTYTEAFKSLG...
β3
390 400 410 420
RVFV V.....KDRPDERNRYHRVHLNMGSDD...SVYIAALGVNGKK.....HKAD
GTV DPEM..DTPKEAKITYRRFRPTLSLGA...KQEFSLQGVEGKK.....WKRS
MUKV S.....MGSSDQRRQYHRTTLTLEESE...RVYAATLGVEAKK.....YKDF
TAV S.....ILKSDERNKYHRTTVDMTKEE...EDYASALGVKGKK.....HRDD
TOV K.....LDRPDLRNKYHRVKIELSASE...SVYIASLGVGGKK.....LRDN
KAMV D.SEQEQEAKKLRQKFHRVRVNLPKED...WVDIAQQGVEAKS.....LRNH
SFTSV DPEK..DTPKEAKITYRRFKPALSSSA...RQEFSLQGVEGKK.....WKRM
UUKV LTVEELEESKALRMKYHRCKIDNGMMD...KLDLAMQGVEAKE.....FKNH
MUNV G.....PDRADERCRYHRVTVDLSNEE...IEYAASLGVNGKK.....YSDL
LACV ....KMMDIGDKAIEYEEFCMSLKSKARSSWKQIMNKKLEPKQINNALVLWEQQFMINNDLIDKSEKLKLFKNF
α9 β4 α10 α11 α12
430 440 450 460 470 480 490
RVFV TLVQQMRD.....RSKQPFSPDHDVDHISEFLSACSSDLWATDEDLYNPLSCDKELRLAAQRIHQPSLSE.R
GTV AANMKKEK.....ESHESLSPYLDVDDISDFLTFNNL..LSESRYGDESIQRAVTILLEKASLMQNT....E
MUKV PLVQDARA.....RSKLPFSPEHDTNDISRYIFAEGQD.FKADEYLYSPCMADSELRKKAAKIHQPTLLQNE
TAV HLVKESRL.....RSKKQFSLRHDIISELTKFLKDPDPSLFRQ.GDYYCPLKEDYDLRTSAASIHQPKLIFEE
TOV PIVKEARE.....QSKLPFSLYHNTRDLESFLRDDKPDVFEHEVGLYSPLTDDMNLRNLAAEIHQPTLIRRE
KAMV PVIERKRE.....ESKRTFPLTTLTDDIESFFREDGN.LFSDDHEQEPPAPI.LEVLK.TNAMVQEVHGIEA
SFTSV AANQKKEK.....ESHETLSPFLDVEDIGDFLTFNNL..LADSRYGDESVQRAVSILLEKASAMQDT....E
UUKV PSIIKKRS.....KSKKTFPLTADTRDIDLFLHHDDL.MFNNEHSQTPPAAM.IEAVK.AGADAQSLHGLDK
MUNV NSVKKARE.....MSKECFSIDHDIRHLEEFLISHDRSLFDTDEELYSPYDEDYELRLAAQMIHQPDYSLKK
LACV CGIGKHKQFKNKMLEDLEVSKPKILDFDDANMYLASLTM..MEQS.....KKILSKSNGLKPDNFILNE

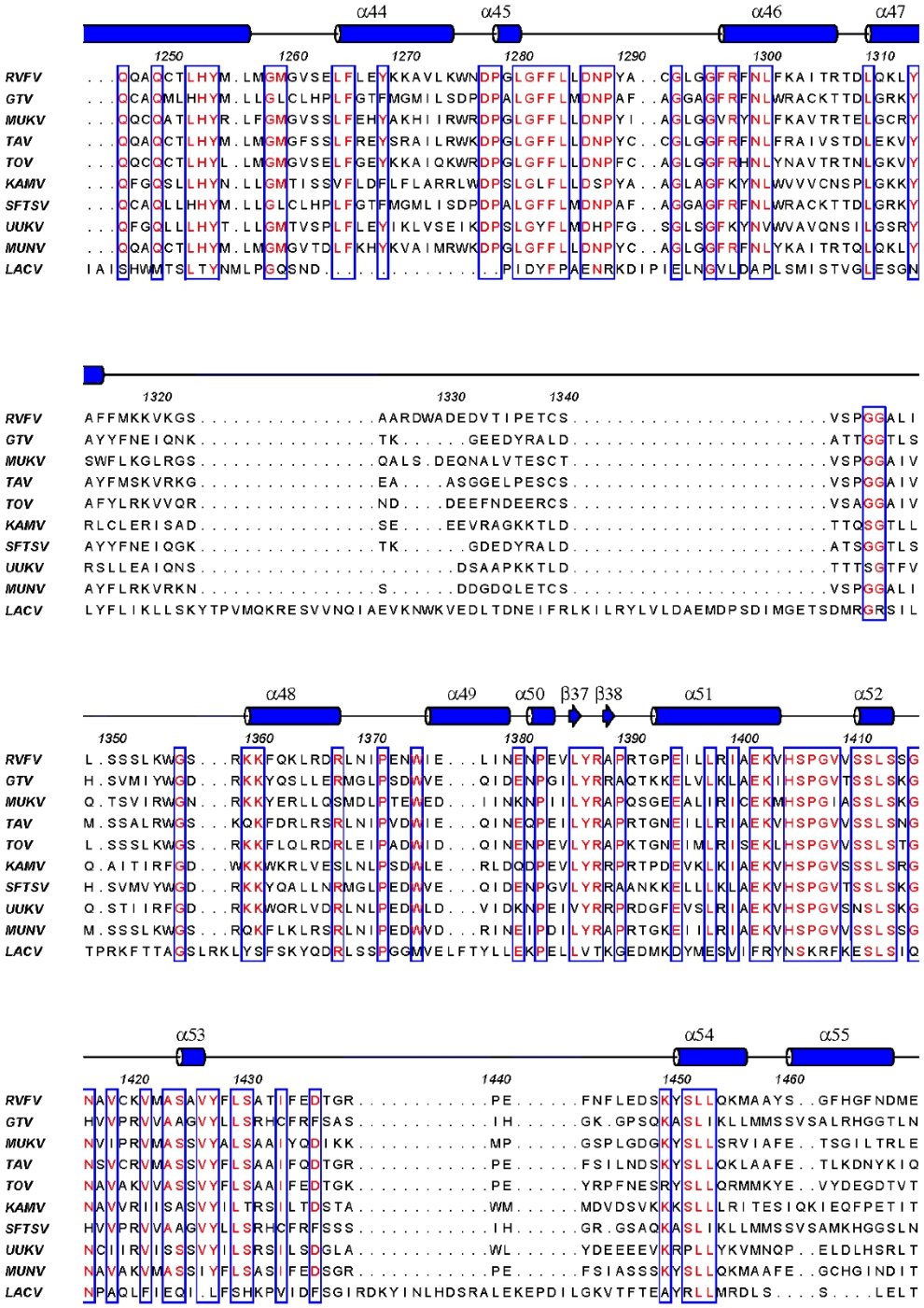
112

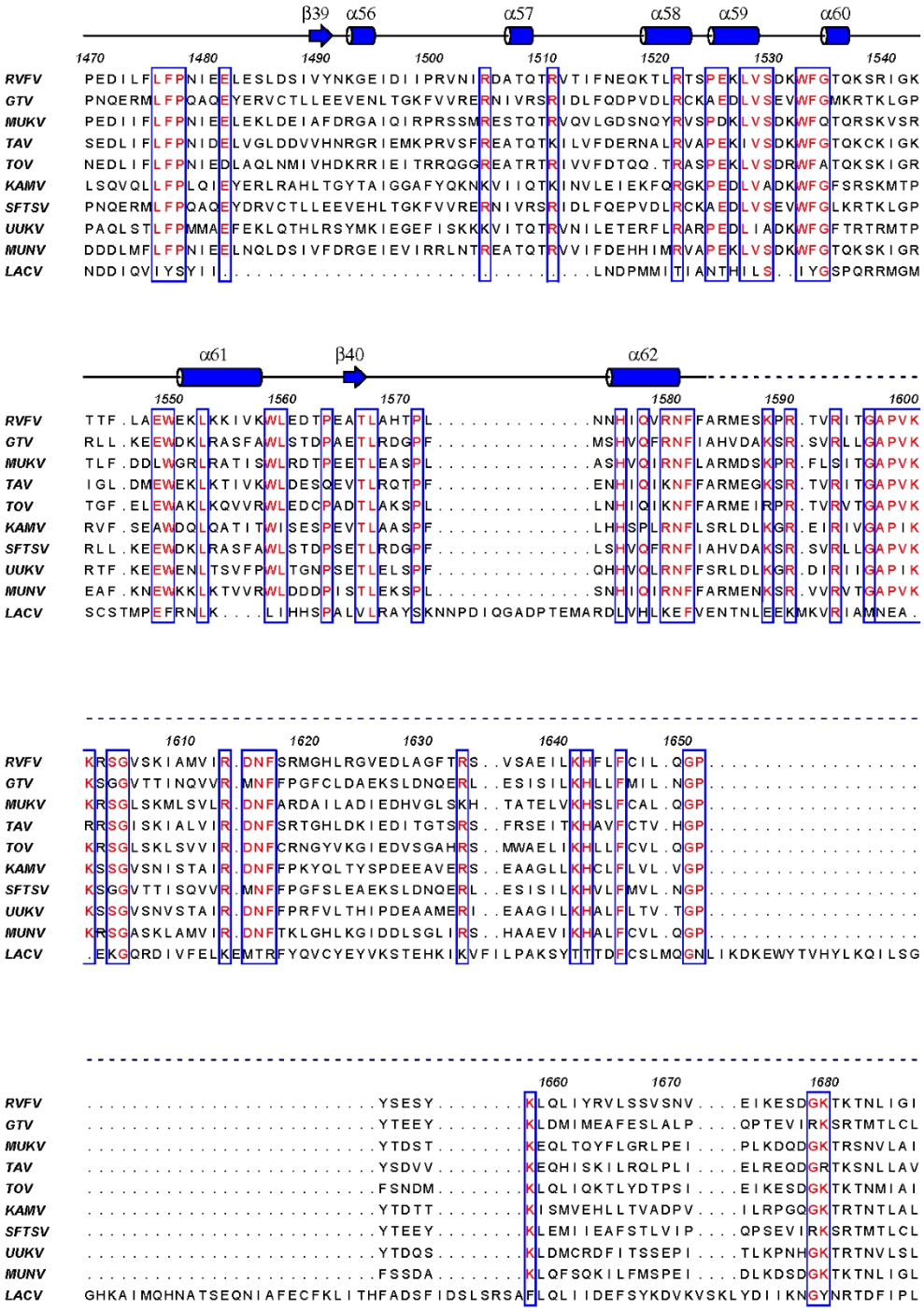

α13
α14
β5
β6
β7
β8
β9
β10
β11
β12
α15
α16
α17
α18
α19
α20
α21
β13
α22
β14
β15
β16
β17
α23
Motif G
RVFV
GTV
MUKV
TAV
TOV
KAMV
SFTSV
UUKV
MUNV
LACV

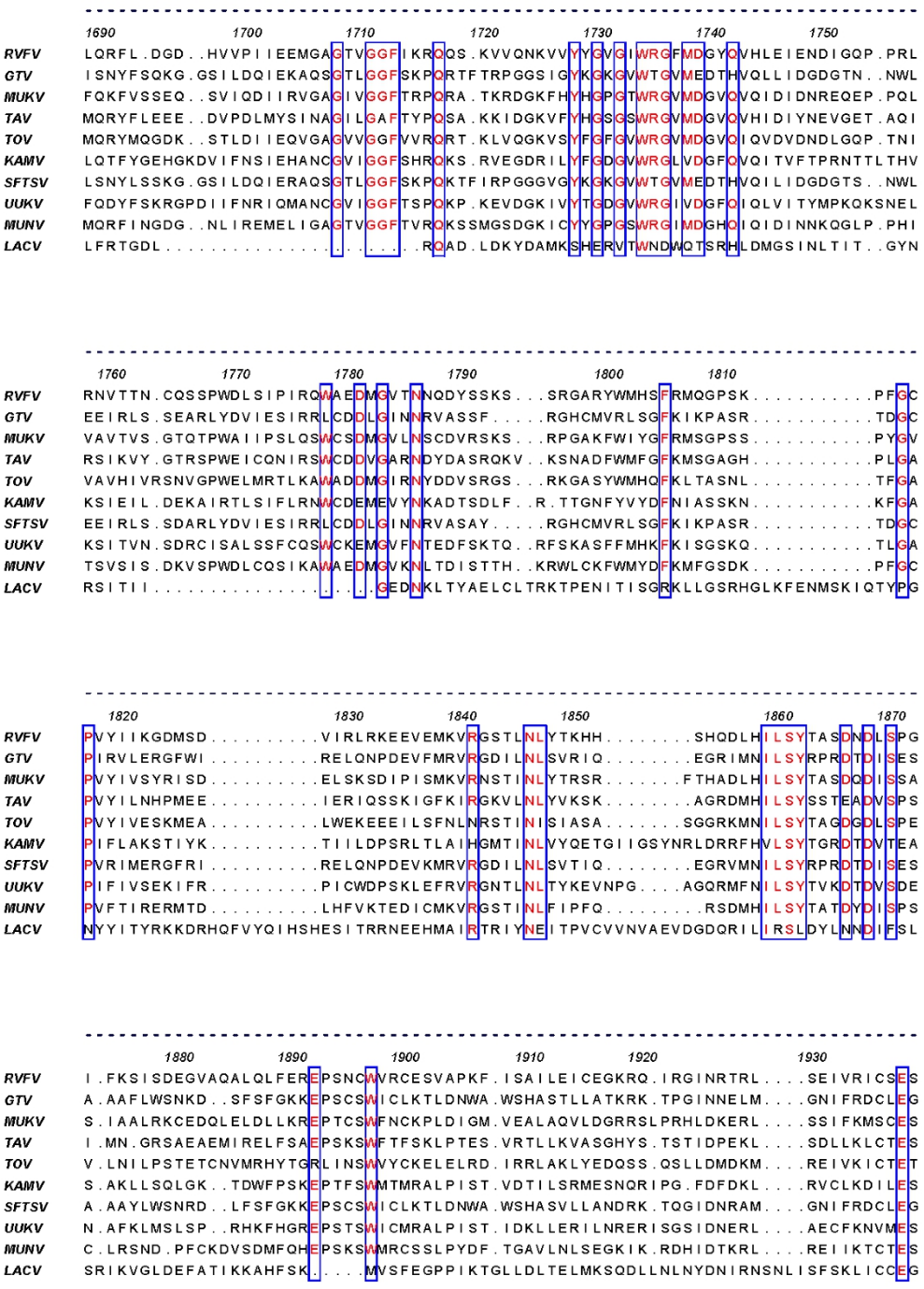

β18 α24 β19 β20 α25
760 770 780 790 800 810 820
RVFV YFKNKEEETEPSSLSGMYKKIIELEHLRPQSDAFLGYKDP....E..LPRMHEFSVSYLKEACNHAKLVLRSLY
GTV YYKNKEESTEVNALGEMYKKIVEMEEDKPDDPTFLGWGDT....S..SPKKHEFSRSFLRAACSSLEKEITQRH
MUKV YFKNKEEPPEVSGLSKLYEKIIELESAKPKDDRFLGKQDP....V..KPAMHEYSVSYLKRCCAHAKNLLKRLY
TAV YFKNKEEETEPSALSRIYKKIIELEHLRPESPEFLGYGDP....E..DPKMHEFSVSYLKEMIEHAKSLLSKSY
TOV YFKNKEEETEPSSLSALYKKIIELEDQRPATDEFLGEGDP....E..DPKMHEFSRSYLKKVCEHGKLLLSRVY
KAMV YLKNKEESPERNATVGMYKKILEYEDLHPGRYDFLGLGDP....DWSDIKFHEYSISLLKFMSELGVVTLRQRW
SFTSV YYKNKEESTELNALGEMYKKIVEMEEDKPSSPEFLGWGDT....D..SPKKHEFSRSFLRAACSSLEREIAQRH
UUKV YLKNKEESPERNASIGMYKKILEYEDKHPGRYTYLGLGDP....PSDDTRFHEYSISLLKHLCIHAEHDLRRNW
MUNV YFKNKEEETEPSALSKMYKKIIELEHLCPETDENLGVGDP....L..EPKMHEFSRSYLKSCIEHAKHLLKRTH
LACV YFNAKGLHEKHHVMVDLAKTILEIECEQRENIKEIWSTNQTKQTVNLKILIHSLCKNLLADTSRHNHLRNRIEN
α26 β21 α27
830 840 850
RVFV GQNFMEQI.....DNQIIRELSGLTLERLA
GTV GRSWKQNL.....EERVLKEIGSKNILDLA
MUKV GQDFMRDI.....DNDILSRFCHLTLESMA
TAV GNNFMDQI.....DSQIVREISHLTMERIA
TOV GQSYRDQI.....ESSIMREINSITLERLA
KAMV GDNILSVM.....WTDFLEGLASLDLEKIA
SFTSV GRQWKQNL.....EERVLKEIGTKNILDLA
UUKV GESFKAMI.....SRDIVDAIASLDLERLA
MUNV GSSYMKMI.....DSQIMRELSSLTLERLA
LACV RNNFRRSITTISTFTSSKSCLKIGDFRKEKELQSVKQKKILEVQSRKMRLANPMFVTDEQVCLEVGHCNYEMLR
β22 α28 β23 α29 β24 α30 β25
860 870 880 890 900 910 920
RVFV TLKATSNFNENWYVYKDVADKNYTRDKLLVKMSKYASEGKSLAIQKFEDCMRQIESQGCMHICLFKKQQ.HGGL
GTV SMKATSNFSKEWEIFSEVQTKEYHRSKLLEKMAGLIERGFMWYIDAVGQAWKAVLEDGCMRICLFKKNQ.HGGL
MUKV TLKATSNFDSSWYVHSENKGRTYSRSKVITKVAPLTRAGKRLVIEAFQESMNMIETNGCMHVCLFKKNQ.HGGL
TAV TLKASSNFDKDWYIYKDVKDKNYTRDKLIVKVAEMVESGGTLVIQKFNECMRVIEERGAMHICLFKKQQ.HGGD
TOV TLKATSNFGESWYTYKDVKDKSYSRDRLLVKMANFAREGKTLAIEKFEDCMTEIEMRGAMHICLFKKQQ.HGGL
KAMV TLKASSNFGPEWYEK.RHDGEKYHRTKVLQRVAEYVDKGKTYVFEILKDCLETVENRGHMHVCLFKKPQ.HGGL
SFTSV SMKATSNFSKDWELYSEVQTKEYHRSKLLEKMATLIEKGVMWYIDAVGQAWKAVLDDGCMRICLFKKNQ.HGGL
UUKV TLKASSNFNEEWYQK.RGDGKTYHRSKVLEKVSKYVKKSSSHVHHIMEECLRKVESQGCMHVCLFKKPQ.HGGL
MUNV TLKATSCFNESWYSFKEVKDKSYSRDKLIVKMSEFANSGKTLAIHMFEQCMNLIEQRGAMHICLFKKMQ.HGGL
LACV .....NAMPNYTDYISTKVFDRLYELLDKKVLTDKPVIEQIMDMMIDHKKFYFTFFNKGQKTSKD
Motif F
β26 α31 β27 α32
930 940 950 960 970
RVFV REIYVMGAEERIVQSVVETIARSIGKFFASDTLCNPPNK.VKIPETHGIRARK.....QC.....
GTV REIYVMDVNARLVQFGVETMARCVCELSPHETVANPRLK.SSIIENHGLKSAR.....SL.....
MUKV REIYVMGFEERIVQLCVETIARSICKLFPSETLNNPKNK.TRIPETHGSRARK.....HC.....
TAV REIYVMGAEERIVQSVVECISRSIGKFFASDTLCNPNNK.TKIPESHGLRARR.....HI.....
TOV REIYVLGAEERIVQAIVEAIARSIGRFFHSDTLCNPANK.MRIPESHGRRART.....HC.....
KAMV REIYVLGFEERVVQLVIESLARQVCKRFKSETMMNPKNK.LNIPENHGLMSTK.....IC.....
SFTSV REIYVMDANARLVQFGVETMARCVCELSPHETVANPRLK.NSIIENHGLKSAR.....SL.....
UUKV REIYVLGFEERVVQLVIETIARQICKRFKSETLTNPKQK.LAIPETHGLRAVK.....TC.....
MUNV REIYVMGAEERIVQSVVETISRCIGCFFPSDTLCNPANK.SRIPESHGMRARK.....HC.....
LACV REIFVGEYEAKMCMYAVERIAKERCKLNPDEMISEPGDGKLKVLEQKSEQEIRFLVETTRQKNREIDEAIEALA


β28 α33 α34 α35 β29
980 990 1000 1010 1020 1030
RVFV KGPVWTCATSDDARKWNQGHFVTKFALMLC EFTSPKWWPLIIRGCSMFTRKRMM
GTV GPGSININSSNDAKKWNQGHYTTKLALVLC WFMPSKFHKFIWSAISMFRRKKMM
MUKV GSAIWTSGTSDDARKWNQGHYVVKFAQLLC SFTQPKWWPIIIRGCSMFTKKRMM
TAV KGPVWTCATSDDAQKWNQGHFVTKFALMLC EFTHARWWPIIIRGCSMFTRKFMM
TOV KGPVWTTATSDDARKWNQGHFVTKFAMMLC EFTHPRWHPIIIRGCSMFTNKYMM
KAMV GSQSQTVCASADAQKWNQGHHVTKFALMLC TFYPESLHPFIMRACSLFMRKRIM
SFTSV GPGSININSSNDAKKWNQGHYTTKLALVLC WFMPAKFHRFIWAAISMFRRKKMM
UUKV GIHHETVATSDDAAKWNQCHHVTKFALMLC HFTDPLFHGFIIRGCSMFMKKRIM
MUNV RGPVWTCATSDDAKKWNQGHFVTKFSLMLC SFTSPKWWPIIVRGCSMFTNKFMM
LACV TEGYESNLGKIEKLSLGKAKGLKMEINADMSKWSAQDVFYKYFWLIALDPILYPQEKRILYFMCNYMDKELIL
Motif A
α36 α37 β30 β31 α38
1040 1050 1060 1070 1080 1090 1100
RVFV MNLNYLKILDGHRELDIRDDFVMDLFKAYHGE AEVPWAFKGKTYLETTTGMMQGILHYTSSLLHTIHQEYIR
GTV IDLRFLTHLSSKSDSKSSDAFREAMTNAFHGN SEASWMEKGKTYIQTETGMMQGILHFTSSLLHSCVQSFYK
MUKV LDPKFMDILDRHTELNLEDEFATTLFKAYHGE VTVPWFRKGCSYLETETGMMQGILHNTSSLLHTVHQEFVR
TAV MNLKFINILDSKRELNLEDDFATTLFSAYHGE IQVPWMDRGRTYLQTSTGMMQGILHFTSSLLHTIHQEFIR
TOV MNLQYLSILDGRRELEIEDEFVRDLFEGYHGN KVLPWINPGCTYLKTTTGMMQGILHYTSSLLHTLHQEFIR
KAMV IDQSLLDLIDSNLNLQSSDEYLMKIHSVYHGR DTTRWMEKGGGYIQTETGMMQGILHYTSSLFHTIYQEWLK
SFTSV VDLRFLAHLSTKSESRSSDPFREAMTDAFHGN REVSWMDKGRTYIKTETGMMQGILHFTSSLLHSCVQSFYK
UUKV IDQSLIDIIDSHTTLETSDAYLQKIHRGYHGSLDDQPRWISRGGAFVQTETGMMQGILHYTSSLLHTLLQEWLR
MUNV MNMRYIEILHNHRDLTVNDEFVNQLYSAYHGD LEVPWISPGCTYLKTKTGMMQGILHFTSSLLHTIHQEYIR
LACV PDELLFNLLDQKVAYQ NDIIATMTNQLNSN TVLIKRNWLQGNFNYTSSYVHSCAMSVYK
Motif H
Motif B
α39 β32 β33 α40
1110 1120 1130 1140 1150 1160 1170
RVFV SLSFKIFNLKVAPEMSKGLVCDMMQGSDDSSMLISFPADDEK VLTRCKVAAAICFRMKKELGVYLAIYPS
GTV TYFVSKLKEGYMGE SINGVVDVVEGSDDSAIMISIRPKSDM DEVRSRFFVANLLHSVKYLNPLFGIYSS
MUKV TLTFKMFRSRIGQDFSSKVVCDVMQGSDDSAMIISFPSEDPE ILARAKVVSATCFRIKRLLGVFLAIYPS
TAV SLAFKIYNMKVGPHAGKRIICDMMQGSDDSSMIISFPNSGDE ELYKNKITAAICFRVKKKLGIYAGIYPS
TOV TLSFRIFDMKVKGDMSSKIVVDMMQGSDDSAMLCSFPSSDDV LLSRCKVAATMCFRVKKLLGLFLAIYPS
KAMV SFFTSLIRSRVERVPRTRTHVDVLQSSDDSGVMISFPAEDPA TVERYRFFAATCFIFKGIFGRLLGIYSS
SFTSV SYFVSKLKEGYMGE SISGVVDVIEGSDDSAIMISIRPKSDM DEVRSRFFVANLLHSVKFLNPLFGIYSS
UUKV TFSQRFIRTRVSVDQRPDVLVDVLQSSDDSGMMISFPSTDKG ATGKYRYLSALIFKYKKVIGKYLGIYSS
MUNV SLSFKIINMKVHPEASYKVVCDMMQGSDDSSMIMSFPSSDLA MLARFKVAAAICFRVKKDLGIYIGIYKS
LACV EILKEAITLL DGSILVNSLVHSDDNQTSITIVQDKMENDKIIDFAMKEFERACLT FGCQAN
Motif C
Motif D
β34 β35 β36 α41 α42 α43
1180 1190 1200 1210 1220 1230 1240
RVFV EKSTANTDFVMEYNSEFYFHTQ HVRPTIRWIAACCSLPEVETLVARQEEASNLMTSVTEGGGSFSLAAMI
GTV EKSTVNTVYCVEYNSEFHFHKH LVRPTIRWVAASHQISETEALASRQEDYSNLLTQCLEGGASFSLTYLI
MUKV EKSTTNTDFLLEYNSEFFFHLQ HVRPTVRWVAASMSIPEVETLIARQEESYNLLASTVEGGASFSLAACV
TAV KKSTSNTDFVMEYNSEFFFHTQ HVRPTIRWVAASCTLPEVETLVARQEEATNLLTAISEGGGSFSLSAVI
TOV EKSTQNTDFVLEYNSEFFFHSQ HIRPTIRWVAASCSLPEVETLVARQEEASNLMTSISEGGGTFSLSAMI
KAMV VKTTNNTLYLLEFNSEFFFHMN HNRPLFRWIAACDVISEQESLAARQEEMYNNLTAVVEGGGSFSLAGFC
SFTSV EKSTVNTVYCVEYNSEFHFHRH LVRPTLRWIAASHQISETEALASRQEDYSNLLTQCLEGGASFSLTYLI
UUKV VKSTNNTLHLLEFNSEFFFHIN HNRPLLRWITACDTISEQESLASRQEEMYNNLTSVLEGGGSFSLVSFC
MUNV EKCTSNTDFVMEYNSEFYFHSQ HVRPTIRWVASCCNLPEVETLVARQEEASNLMTSISEGGGSFALSAQI
LACV MKKTYVTNCIKEFVSLFNLYGEPFSIYGRFLLTSVGDCAYIGPYEDLASRISSAQTAIK HGCPPSLAWVS
Motif E

115


α44
α45
α46
α47
1250
1260
1270
1280
1290
1300
1310
RVFV
GTV
MUKV
TAV
TOV
KAMV
SFTSV
UUKV
MUNV
LACV
1320
1330
1340
α48
α49
α50
β37
β38
α51
α52
1350
1360
1370
1380
1390
1400
1410
α53
α54
α55
1420
1430
1440
1450
1460


β39 α56 α57 α58 α59 α60
1470 1480 1490 1500 1510 1520 1530 1540
RVFV PEDILFLFPNIEELESLDSIVYNKGEIDIIPRVNIRDATQTRVTIFNEQKTLRTSPEKLVSDKWFGTQKSRIGK
GTV PNQERMLFPQAQEYERVCTLLEEVENLTGKFVVRERNIVRSRIDLFQDPVDLRCKAEDLVSEVWFGMKRTKLGP
MUKV PEDIIFLFPNIEELEKLDEIAFDRGAIQIRPRSSMRESTQTRVQVLGDSNQYRVSPDKLVSDKWFQTQRSKVSR
TAV SEDLIFLFPNIDELVGLDDVVHNRGRIEMKPRVSFREATQTKILVFDERNALRVAPEKIVSDKWFGTQKCKIGK
TOV NEDLIFLFPNIEDLAQLNMIVHDKRRIEITRRQGGREATRTRIVVFDTQQ.TRASPEKLVSDRWFATQKSKIGR
KAMV LSQVQLLFPLQIEYERLRAHLTGYTAIGGAFYQKNKVIIQTKINVLEIEKFQRGKPEDLVADKWFGFSRSKMTP
SFTSV PNQERMLFPQAQEYDRVCTLLEEVEHLTGKFVVRERNIVRSRIDLFQEPVDLRCKAEDLVSEVWFGLKRTKLGP
UUKV PAQLSTLFPMMAEFEKLQTHLRSYMKIEGEFISKKKVITQTRVNILETERFLRARPEDLIADKWFGFTRTRMTP
MUNV DDDLMFLFPNIEELNQLDSIVFDRGEIEVIRRLNTREATQTRVVIFDEHHIMRVAPEKLVSDKWFGTQKSKIGR
LACV NDDIQVIYSYII.....LNDPMMITIANTHILS..IYGSPQRRMGM
α61 β40 α62
1550 1560 1570 1580 1590 1600
RVFV TTF.LAEWEKLKKIVKWLEDTPEATLAHTPL.....NNHIQVRNFFARMESKPR.TVRITGAPVK
GTV RLL.KEEWDKLRASFAWLSTDPAETLRDGPF.....MSHVQFRNFIAHVDAKSR.SVRLLGAPVK
MUKV TLF.DDLWGRLRATISWLRDTPEETLEASPL.....ASHVQIRNFLARMDSKPR.FLSITGAPVK
TAV IGL.DMEWEKLKTIVKWLDESQEVTLRQTPF.....ENHIQIKNFFARMEGKSR.TVRITGAPVK
TOV TGF.ELEWAKLKQVVRWLEDCPADTLAKSPL.....LNHIQIKNFFARMEIRPR.TVRVTGAPVK
KAMV RVF.SEAWDQLQATITWISESPEVTLAASPF.....LHHSPLRNFLSRLDLKGR.EIRIVGAPIK
SFTSV RLL.KEEWDKLRASFAWLSTDPSETLRDGPF.....LSHVQFRNFIAHVDAKSR.SVRLLGAPVK
UUKV RTF.KEEWENLTSVFPWLTGNPSETLELSPF.....QHHVQLRNFFSRLDLKGR.DIRIIGAPIK
MUNV EAF.KNEWKKLKTVVRWLDDDPISTLEKSPL.....SNHIQIRNFFARMENKSR.VVRVTGAPVK
LACV SCSTMPEFRNLK....LIHHSPALVLRAYSKNNPDIQGADPTEMARDLVHLKEFVENTNLEEKMKVRIAMNEA.
1610 1620 1630 1640 1650
RVFV KRSGVSKIAMVIR.DNFSRMGHLRGVEDLAGFTRS..VSAEILKHFLFCIL.QGP.....
GTV KSGGVTTINQVVR.MNFFPGFCLDAEKSLDNQERL..ESISILKHLLFMIL.NGP.....
MUKV KRSGLSKMLSVLR.DNFARDAILADIEDHVGLSKH..TATELVKHSLFCAL.QGP.....
TAV RRSGISKIALVIR.DNFSRTGHLDKIEDITGTSRS..FRSEITKHAVFCTV.HGP.....
TOV KRSGLSKLSVVIR.DNFCRNGYVKGIEDVSGAHRS..MWAELIKHLLFCVL.QGP.....
KAMV KSSGVSNISTAIR.DNFFPKYQLTYSPDEEAVERS..EAAGLLKHCLFLVL.VGP.....
SFTSV KSGGVTTISQVVR.MNFFPGFSLEAEKSLDNQERL..ESISILKHVLFMVL.NGP.....
UUKV KSSGVSNVSTAIR.DNFFPRFVLTHIPDEAAMERI..EAAGILKHALFLTV.TGP.....
MUNV KRSGASKLAMVIR.DNFTKLGHLKGIDDLSGLIRS..HAAEVIKHALFCVL.QGP.....
LACV .EKGQRDIVFELKEMTRFYQVCYEYVKSTEHKIKVFILPAKSYTTTDFCSLMQGNLIKDKEWYTVHYLKQILSG
1660 1670 1680
RVFV .....YSESY.....KLQLIYRVLSSVSNV....EIKESDGKTKTNLIGI
GTV .....YTEEY.....KLDMIMEAFESLALP....QPTEVIRKSRTMTLCL
MUKV .....YTDST.....KEQLTQYFLGRLPEI....PLKDQDGKTRSNVLAI
TAV .....YSDVV.....KEQHISKILRQLPLI....ELREQDGRTKSNLLAV
TOV .....FSNDM.....KLQLIQKTLYDTPSI....EIKESDGKTKTNMIAI
KAMV .....YTDTT.....KISMVEHLLTVADPV....ILRPGQGKTRTNTLAL
SFTSV .....YTEEY.....KLEMIIEAFSTLVIP....QPSEVIRKSRTMTLCL
UUKV .....YTDQS.....KLDMCRDFITSSEPI....TLKPNHGKTRTNVLSL
MUNV .....FSSDA.....KLQFSQKILFMSPEI....DLKDSDGKTKTNLIGL
LACV GHKAIMQHNATSEQNIAFECFKLITHFADSFIDSLSRSAFLQLIIDEFSYKDVKVSKLYDIIKNGYNRTDFIPL


1690 1700 1710 1720 1730 1740 1750
RVFV LQRFL.DGD..HVVPIIEEMGAGTVGGFIKRQQS.KVVQNKVVYYGVGIWRGFMDGYQVHLEIENDIGQP.PRL
GTV ISNYFSQKG.GSILDQIEKAQSGTLGGFSKPQRTFTRPGGSIGYKGKGVWTGVMEDTHVQLLIDGDGTN..NWL
MUKV FQKFVSSEQ..SVIQDIIRVGAGIVGGFTRPQRA.TKRDGKFHYHGPGTWRGVMDGVQVQIDIDNREQEP.PQL
TAV MQRYFLEEE..DVPDLMYSINAGILGAFTYPQSA.KKIDGKVFYHGSGSWRGVMDGVQVHIDIYNEVGET.AQI
TOV MQRYMQGDK..STLDIIEQVGAGVVGGFVVRQRT.KLVQGKVSYFGFGVWRGVMDGVQIQVDVDNDLGQP.TNI
KAMV LQTFYGEHGKDVIFNSIEHANCGVIGGFSHRQKS.RVEGDRILYFGDGVWRGLVDGFQVQITVFTPRNTTLTHV
SFTSV LSNYLSSKG.GSILDQIERAQSGTLGGFSKPQKTFIRPGGGVGYKGKGVWTGVMEDTHVQILIDGDGTS..NWL
UUKV FQDYFSKRGPDIIFNRIQMANCGVIGGFTSPQKP.KEVDGKIVYTGDGVWRGIVDGFQIQLVITYMPKQKSNEL
MUNV MQRFINGDG..NLIREMELIGAGTVGGFTVRQKSSMGSDGKICYYGPGSWRGIMDGHQIQIDINNKQGLP.PHI
LACV LFRTGDL..............RQAD.LDKYDAMKSHERVTWNDWQTSRHLDMGSINLTIT..GYN

1760 1770 1780 1790 1800 1810
RVFV RNVTTN.CQSSPWDLSIPIRQWAEDMGVTNNQDYSSKS...SRGARYWMHSFRMQGPSK..........PFGC
GTV EEIRLS.SEARLYDVIESIRRLCDDLGINNRVASSF.....RGHCMVRLSGFKIKPASR..........TDGC
MUKV VAVTVS.GTQTPWAIIPSLQSWCSDMGVLNSCDVRSKS...RPGAKFWIYGFRMSGPSS..........PYGV
TAV RSIKVY.GTRSPWEICQNIRSWCDDVGARNDYDASRQKV..KSNADFWMFGFKMSGAGH..........PLGA
TOV VAVHIVRSNVGPWELMRTLKAWADDMGIRNYDDVSRGS...RKGASYWMHQFKLTASNL..........TFGA
KAMV KSIEIL.DEKAIRTLSIFLRNWCDEMEVYNKADTSDLF..R.TTGNFYVYDFNIASSKN..........KFGA
SFTSV EEIRLS.SDARLYDVIESIRRLCDDLGINNRVASAY.....RGHCMVRLSGFKIKPASR..........TDGC
UUKV KSITVN.SDRCISALSSFCQSWCKEMGVFNTEDFSKTQ..RFSKASFFMHKFKISGSKQ..........TLGA
MUNV TSVSIS.DKVSPWDLCQSIKAWAEDMGVKNLTDISTTH..KRWLCKFWMYDFKMFGSDK..........PFGC
LACV RSITII..............GEDNKLTYAELCLTRKTPENITISGRKLLGSRHGLKFENMSKIQTYPG

1820 1830 1840 1850 1860 1870
RVFV PVYIIKGDMSD.........VIRLRKEEVEMKVRGSTLNLYTKHH.....SHQDLHILSYTASDNDLSPG
GTV PIRVLERGFWI.........RELQNPDEVFMRVRGDILNLSVRIQ.....EGRIMNILSYRPRDTDISES
MUKV PVYIVSYRISD.........ELSKSDIPISMKVRNSTINLYTRSR.....FTHADLHILSYTASDQDISSA
TAV PVYILNHPMEE.........IERIQSSKIGFKIRGKVLNLYVKSK.....AGRDMHILSYSSTEADVSPS
TOV PVYIVESKMEA.........LWEKEEEILSFNLNRSTINISIASA.....SGGRKMNILSYTAGDGDLSPE
KAMV PIFLAKSTIYK.........TIILDPSRLTLAIHGMTINLVYQETGIIGSYNRLDRRFHVLSYTGRDTDVTEA
SFTSV PVRIMERGFRI.........RELQNPDEVKMRVRGDILNLSVTIQ.....EGRVMNILSYRPRDTDISES
UUKV PIFIVSEKIFR.........PICWDPSKLEFRVRGNTLNLTYKEVNPG....AGQRMFNILSYTVKDTDVSDE
MUNV PVFTIRERMTD.........LHFVKTEDICMKVRGSTINLFIPFQ.....RSDMHILSYTATDYDISPS
LACV NYYITYRKKDRHQFVYQIHSHESITRRNEEHMAIRTRIYNEITPVCVVNVAEVDGDQRILIRSLDYLNNDIFSL

1880 1890 1900 1910 1920 1930
RVFV I.FKSISDEGVAQALQLFERESPSNCWVRCESVAPKF.ISAILEICEGKRQ.IRGINRTRL....SEIVRICSES
GTV A.AAFLWSNKD..SFSFGKKESPSCSWICLKTLDNWA.WSHASTLLATKRK.TPGINNELM....GNIFRDCLEG
MUKV S.IAALRKCEDQLELDLLKRESPTCSWFNCKPLDIGM.VEALAQVLDGRRSLPRHLDKERL....SSIFKMSCES
TAV I.MN.GRSAEAEMIRELFSAESPSKSWFTFSKLPTES.VRTLLKVASGHYS.TSTIDPEKL....SDLLKLCTES
TOV V.LNILPSTETCNVMRHYTGRLINSWVYCKELELRD.IRRLAKLYEDQSS.QSLLDMDKM....REIVKICTET
KAMV S.AKLLSQLGK..TDWFPSKEPTFSWMTMRALPIST.VDTILSRMESNQRIPG.FDFDKL....RVCLKDILES
SFTSV A.AAYLWSNRD..LFSFGKKESPSCSWICLKTLDNWA.WSHASVLLANDRK.TQGIDNRAM....GNIFRDCLEG
UUKV N.AFKLMSLSP..RHKFHGRESPSTSWICMRALPIST.IDKLLERILNRERISGSIDNERL....AECFKNVMES
MUNV C.LRSND.PFCKDVSDMFQHESPSKSWMRCSSLPYDF.TGAVLNLSEGKIK.RDHIDTKRL....REIIKTCTES
LACV SRIKVGLDEFATIKKAHFSK.....MVSFEGPPIKTGLLDLTELMKSQDLLNLNYDNIRNSNLISFSKLICCEG

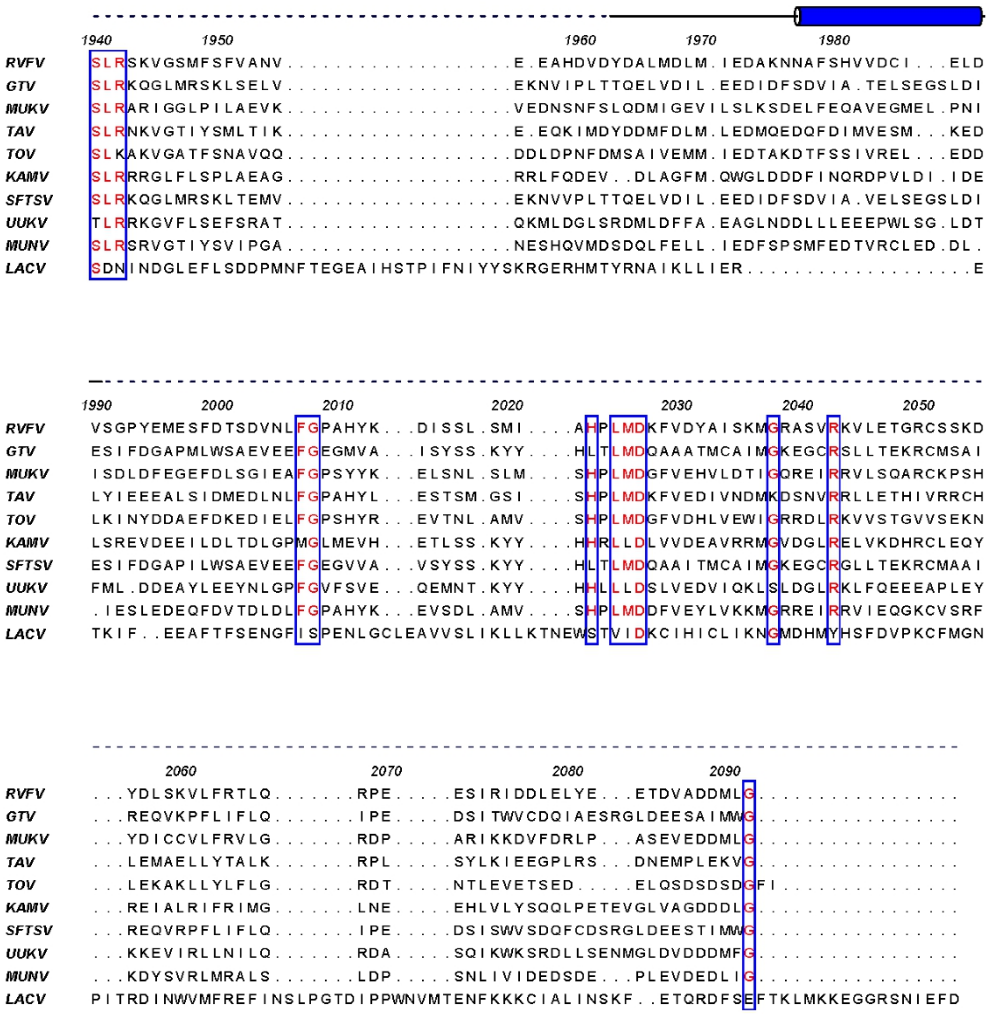


Figure S12. Amino acid sequence alignment of L proteins.

The amino acid sequences of L Proteins from RVFV (DQ375406), GTV (YP_009666941), Mukawa virus (YP_009666332), Tapara virus (YP_009346035), Toros virus (YP_009246447), KAMV (YP_009449450), SFTSV (YP_006504091.1), UUKV (NP_941973), Munguba virus (YP_009346010), and LACV (EF485038.1) were aligned using Clustal Omega [2]. Secondary structures for RVFV L_{RNA} protein and eight conserved motifs were displayed.

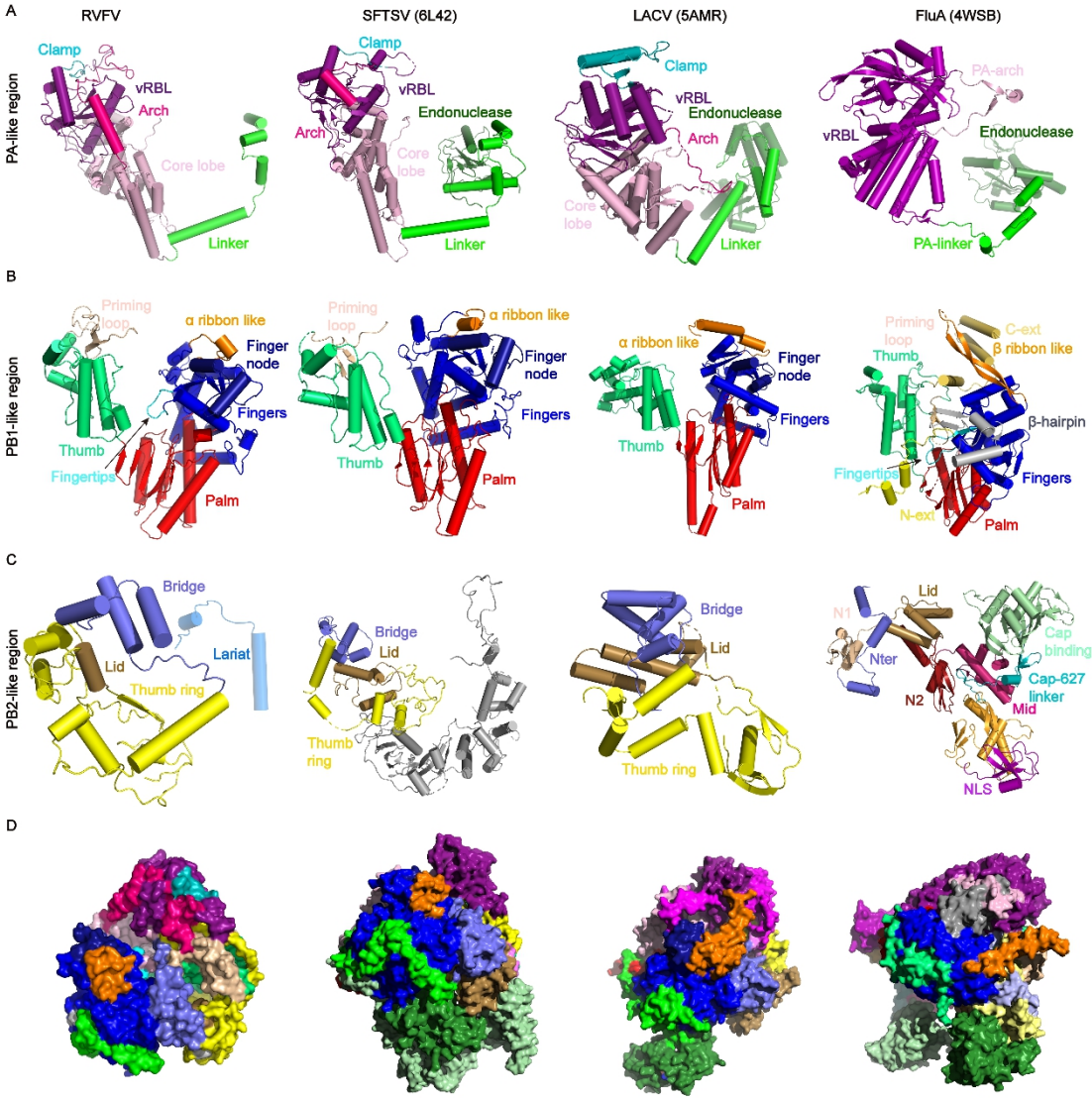


Figure S13. Comparison of each corresponding domain of RVFV, SFTSV, LACV, and FluA polymerase.

For overall structure comparison, parts of structures of RVFV L_{RNA} protein, SFTSV apo-L protein (PDB 6L42), LACV L protein complex (PDB 5AMR), and Influenza A polymerases (PDB 4WSB) were shown side by side: (A) regions corresponding to influenza virus PA, (B) PB1-like regions, (C) PB2-like regions and (D) overall architectures. Colors correspond to Figure 2A.

Movie S1. Changes in the overall conformation of polymerase caused by the presence and absence of RNA

References:

1. Tan, Y.Z. et al. Addressing preferred specimen orientation in single-particle cryo-EM through tilting. *Nat Methods* **14**, 793-796 (2017).
2. Madeira, F. et al. The EMBL-EBI search and sequence analysis tools APIs in 2019. *Nucleic Acids Res.* **47**, W636-W641 (2019).

Figures

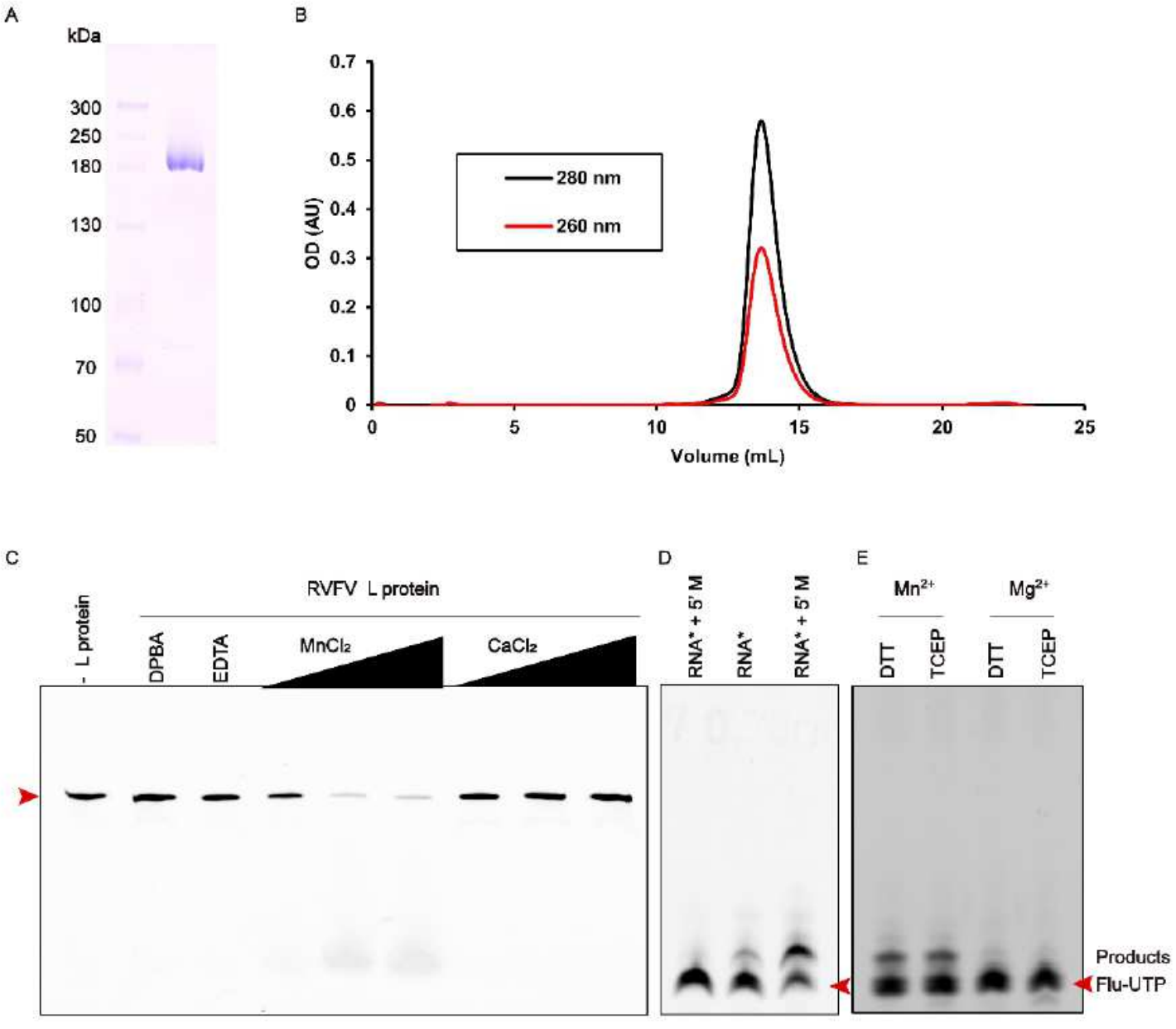


Figure 1

In vitro enzymatic activity assays (A) SDS–PAGE profile of Rift Valley fever virus (RVFV) L protein. (B) Size-exclusion chromatogram of RVFV L protein. Absorbance curves for the sample at wavelengths of 260 nm and 280 nm. (C) Endonuclease activity assay. RVFV L protein (0.06 μM) was incubated with 0.45 μM of fluorescently labeled 30 nt PolyA RNA substrate at 30□ for 40 min in the presence of 10, 25, or 50 mM MnCl2 or CaCl2. Reactions without protein, in the presence of EDTA or the known endonuclease-specific inhibitor 2,4-dioxo-4-phenylbutanoic acid (DPBA), were negative controls. All input is shown as red arrows in Figure C, D, and E. (D) In vitro polymerase activity assay. The endonuclease and polymerase inactivation site double mutant (D111A/D1133A) was added in the left lane as a negative control. Mutant

(D111A) was added in the middle and right lanes. These three lanes were incubated with the conserved 5’ 20 nt of the M segment or/and a 20 nt template RNA named RNA* (5’ M: 5’-ACACAAAGACGGUGCAUUAA-3’, RNA*: 5’-UGUGUUUCUGGCCACGUUGA-3’). Replication products were detected by fluorescence (fluorescein-12-UTP). (E) Effect of metal ions (Mn2+ or Mg2+) and reducing agent (DTT or TCEP (Tris(2-carboxyethyl)phosphine)) on de novo RNA synthesis by endonuclease inactivation site mutant (D111A). The assay was performed with 5 mM MnCl2/MgCl2 and 2.5 mM DTT/TCEP for 40 min at 30℃.

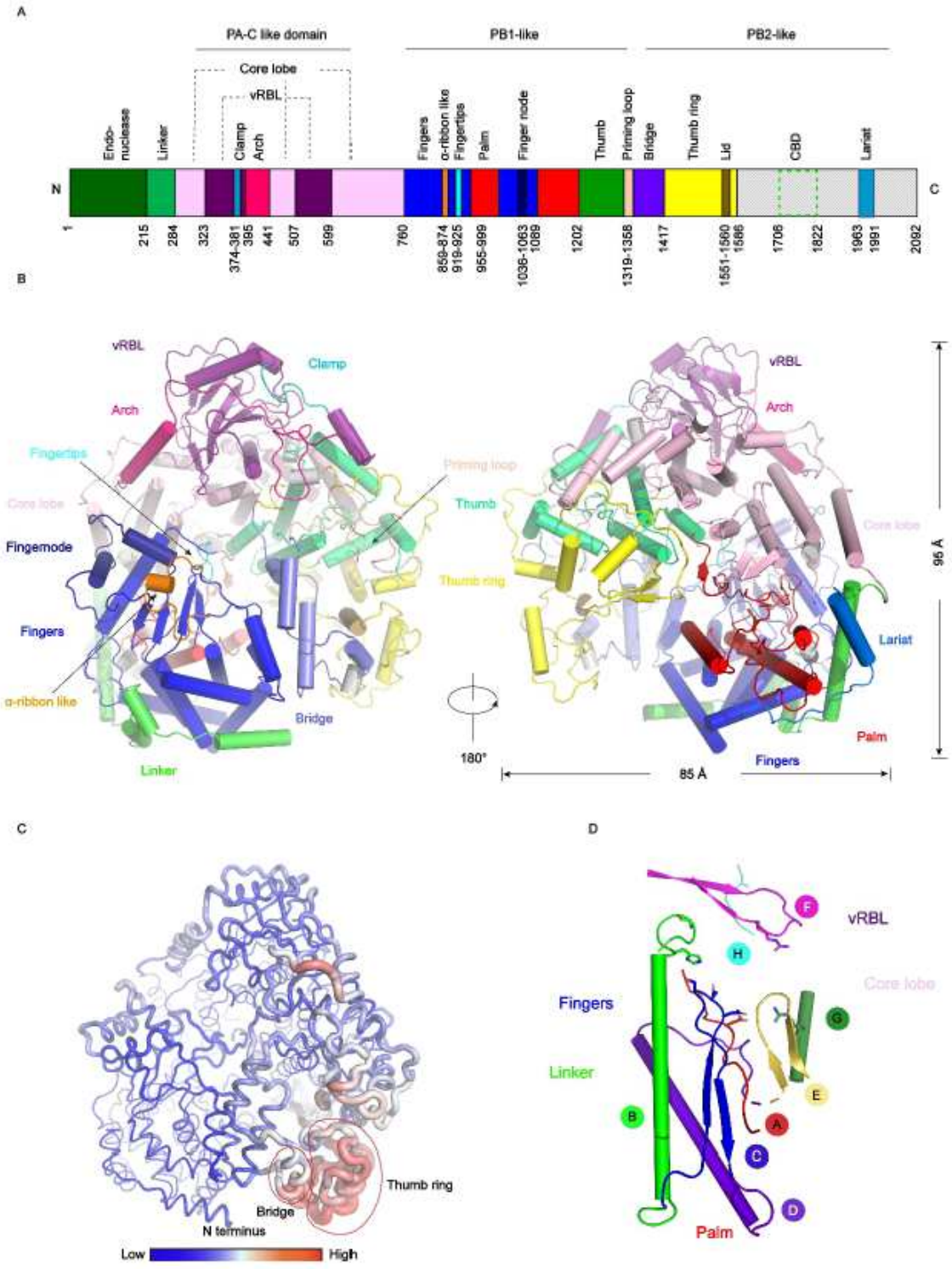


Figure 2

Overall structure of the Rift Valley fever virus (RVFV) LRNA protein and conserved motifs (A) Schematic representation of the monomeric RVFV LRNA protein domain structure. (B) Cartoon representation of the RVFV LRNA protein. The structure is colored by domains using the same color code as in (A). 310 helices are colored in gray. (C) The RVFV LRNA protein B-factor map. A larger radius and red color represent high B-factor values and a smaller radius and blue color represent low B-factor values. (D) The arrangement of the conserved RdRp motifs in the RVFV active site colored in red, green, blue, purpleblue, yelloworange, magenta, forest, and teal for motifs A–H, respectively.

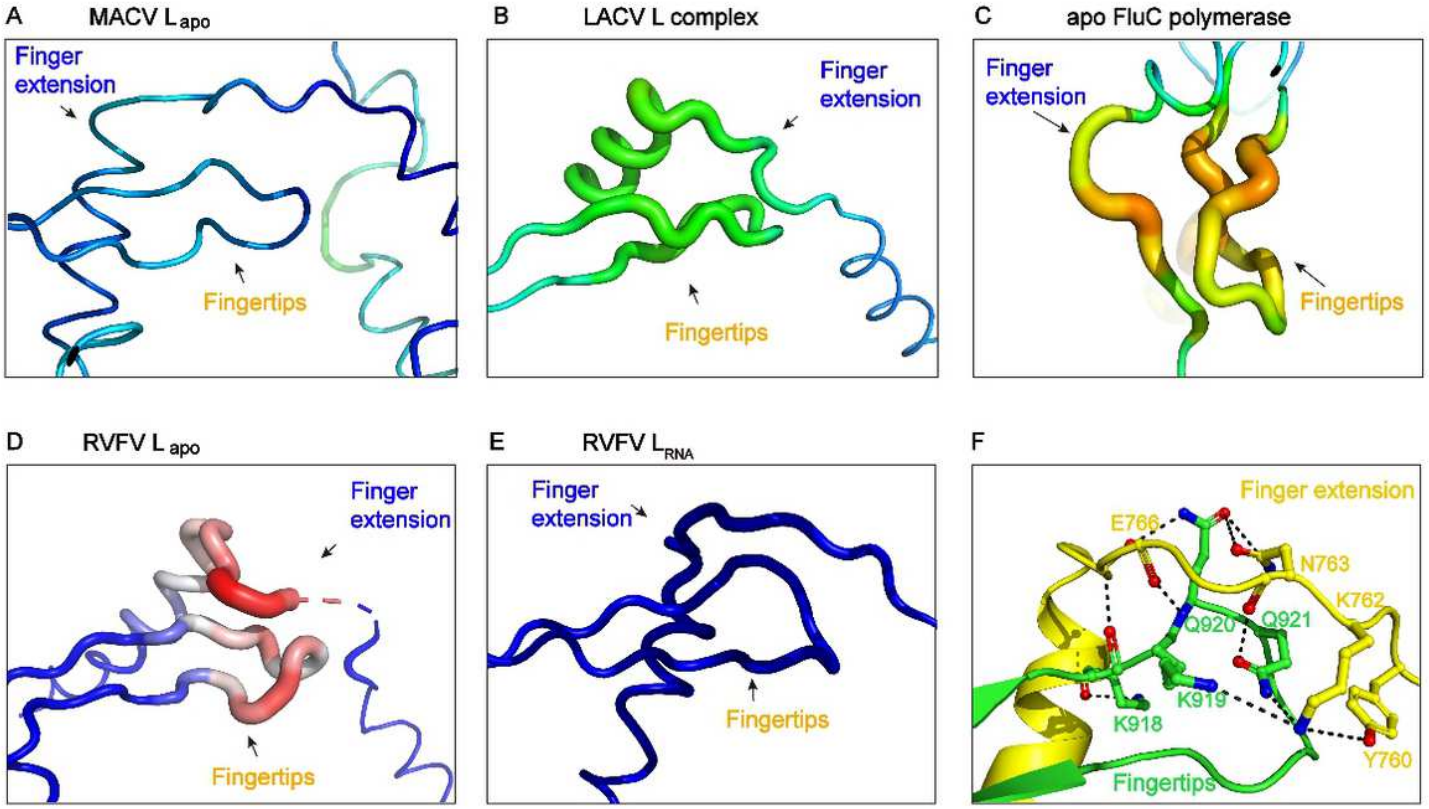


Figure 3

The fingertips loops of machupo virus (MACV) Lapo, LACV L complex, apo FluC polymerase, Rift Valley fever virus (RVFV) Lapo, and RVFV LRNA The B-factor maps for (A) MACV (PDB 6KLD), (B) LACV (PDB 6Z8K), (C) Influenza C (FluC) (PDB 5D98), (D) RVFV Lapo, (E) RVFV LRNA. (A, B, E) Fingertips are conservative, while FluC's and LACV's fingertips are unstable, and their replication process requires 5' RNA. (F) Interaction of residues between fingertips and finger extension. The residue on the fingertips and the finger extension are colored green and yellow, respectively.

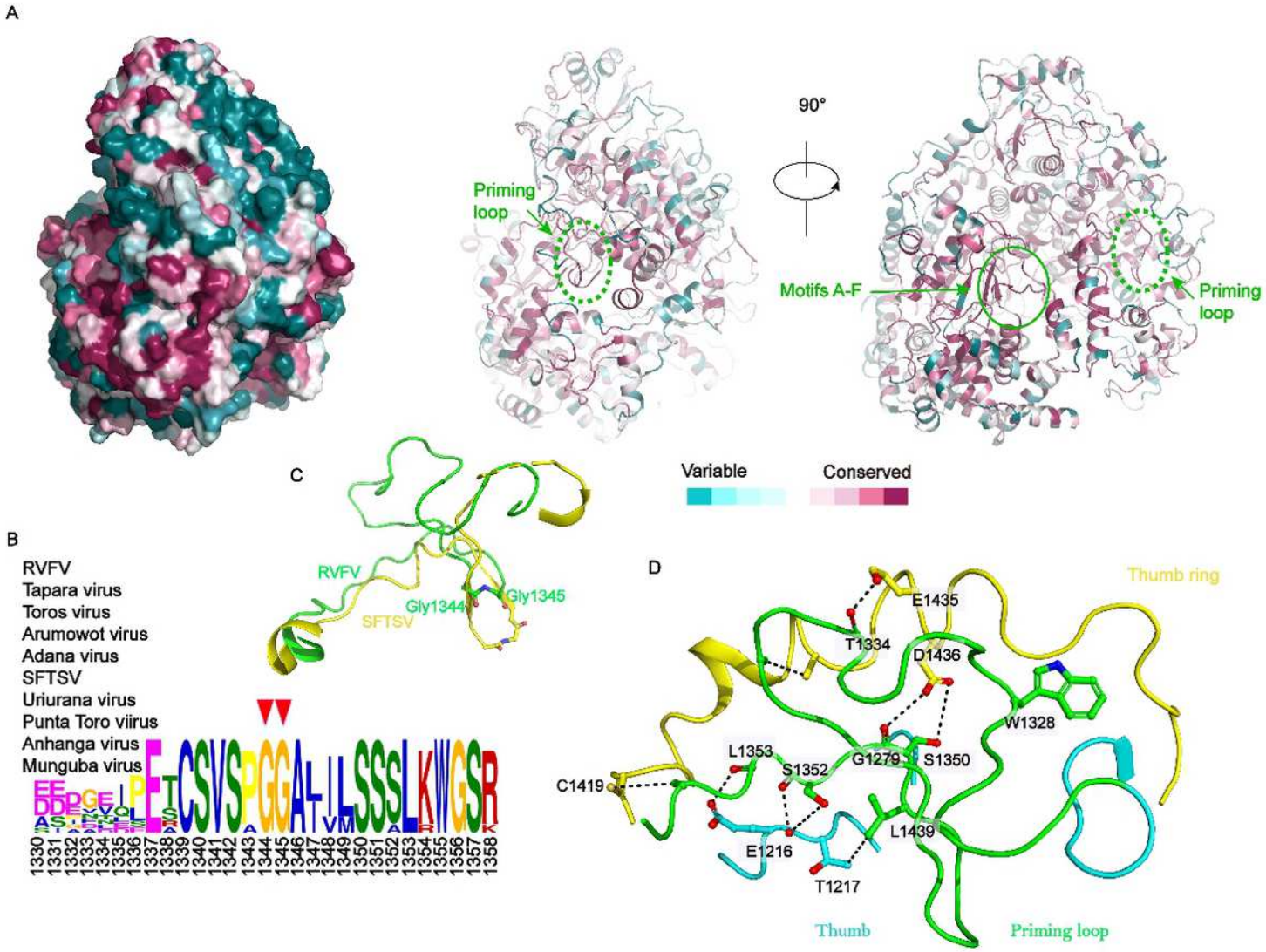


Figure 4

Conservation analysis for RVFV L protein (A) Sequence conservation analysis. The left figure displayed in surface mode is positioned at the same view as the middle figure. Two different views of the Rift Valley fever virus (RVFV) L protein displayed in ribbon mode and colored based on the residue conservation among family Phenuiviridae. The bottom color code bar is from low conservation (cyan) to high conservation (purple). Figures are colored using the Consurf server (http://consurf.tau.ac.il/2016/) based on a multiple sequence alignment including the sequences of Tapara virus (YP_009346035), Toros virus (YP_009246447), Arumowot virus (YP_009010958.1), Adana virus (YP_009227127.1), SFTSV (YP_006504091.1), Uriurana virus (YP_009346036.1), Punta Toro virus (YP_009512941.1), Anhanga virus (API68876.1), and Munguba virus (YP_009346010.1). (B) Sequence conservation in the putative priming loop. (C) The priming loop and connecting regions are shown as ribbons colored in green (RVFV) and yellow (SFTSV and PDB 6L42). Two glycine residues (Gly1344 and Gly1345), which are important for an early stage of RNA synthesis, are shown as sticks. (D) Structural demonstration shows the interaction sites of the priming loop (green) and thumb (cyan) and thumb ring domains (yellow).

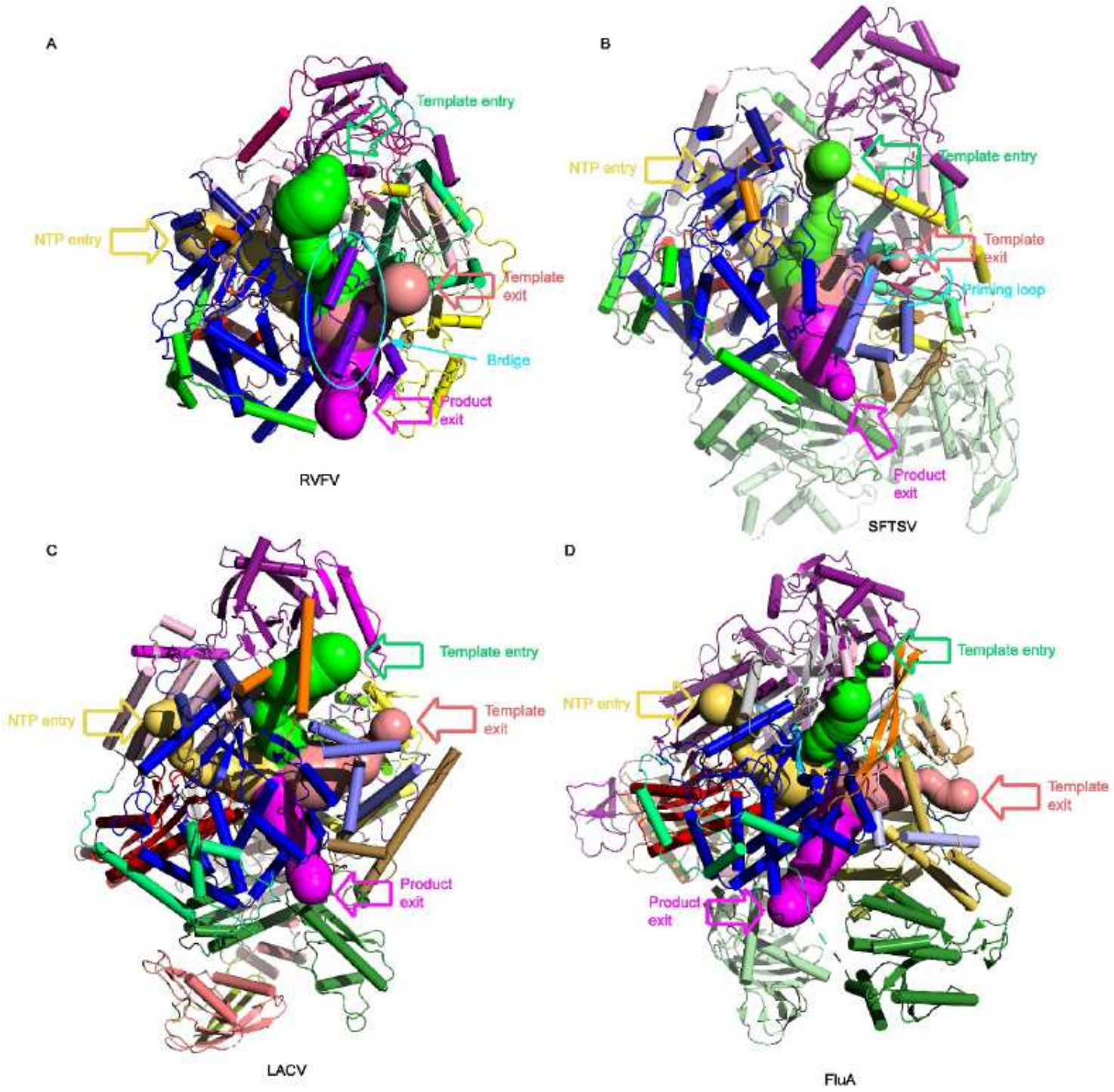


Figure 5

The entrance and exit channels of Rift Valley fever virus (RVFV), SFTSV, LACV and FluA polymerases The template entry, NTP entry, template exit, and product exit channels of (A) RVFV LRNA protein, (B) SFTSV Lapo protein (PDB 6L42), (C) LACV L protein (PDB 6Z8K), (D) Influenza A polymerase (FluA) (PDB 6T0V) are colored in limegreen, yelloworange, salmon, and magenta, respectively. The colors of the RVFV L domains are the same as those used in Figure 2A.

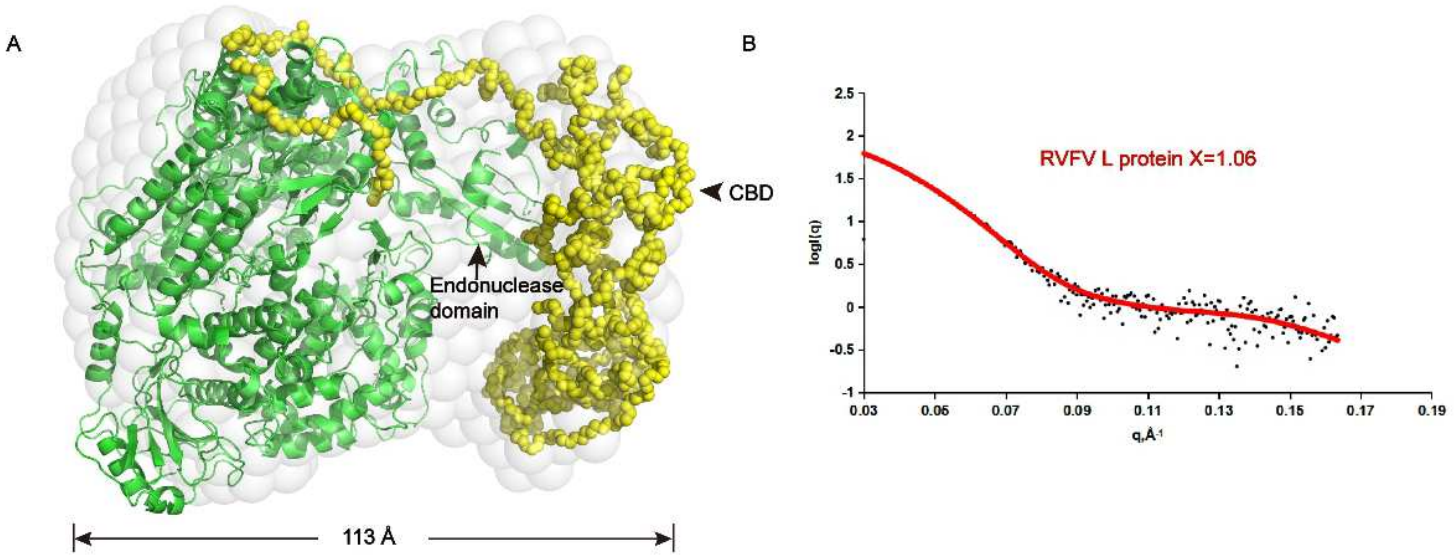


Figure 6

Superposition of low-resolution ab initio model and rigid body model (A) The ab initio model is shown as light-gray spheres. PA-like, PB1-like, and PB2-like N-terminal domains of RVFV L protein are colored in green. The simulated PB2 C-terminal domain is colored in yellow. (B) Comparison of SAXS experimental data and calculated scattering profiles. Experimental data are represented in black dots. The data calculated from the RVFV L protein is in red.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- RVFVLvalidationreportsall.pdf
- movieS1.mp4