Structural insights into RNA recognition by the Chikungunya virus nsP2 helicase

Yee-Song Law[a,b], Age Utt[c], Yaw Bia Tan[a,b], Jie Zheng[d,1], Sainan Wang[c], Ming Wei Chen[b,e], Patrick R. Griffin[d], Andres Merits[c,2], and Dahai Luo[a,b,2]

[a]Lee Kong Chian School of Medicine, Nanyang Technological University, Singapore 636921; [b]NTU Institute of Structural Biology, Nanyang Technological University, Singapore 636921; [c]Institute of Technology, University of Tartu, 50411 Tartu, Estonia; [d]Department of Molecular Medicine, The Scripps Research Institute, Jupiter, FL 33458; and [e]School of Biological Sciences, Nanyang Technological University, Singapore 637551

Chikungunya virus (CHIKV) is transmitted to humans through mosquitoes and causes Chikungunya fever. Nonstructural protein 2 (nsP2) exhibits the protease and RNA helicase activities that are required for viral RNA replication and transcription. Unlike for the C-terminal protease, the structure of the N-terminal RNA helicase (nsP2h) has not been determined. Here, we report the crystal structure of the nsP2h bound to the conserved 3′-end 14 nucleotides of the CHIKV genome and the nonhydrolyzable transition-state nucleotide analog ADP-AlF$_4$. Overall, the structural analysis revealed that nsP2h adopts a uniquely folded N-terminal domain followed by a superfamily 1 RNA helicase fold. The conserved helicase motifs establish polar contacts with the RNA backbone. There are three hydrophobic residues (Y161, F164, and F287) which form stacking interactions with RNA bases and thereby bend the RNA backbone. An F287A substitution that disrupted these stacking interactions increased the basal ATPase activity but decreased the RNA binding affinity. Furthermore, the F287A substitution reduced viral infectivity by attenuating subgenomic RNA synthesis. Replication of the mutant virus was restored by pseudoreversion (A287V) or adaptive mutations in the RecA2 helicase domain (T358S or V410I). Y161A and/or F164A substitutions, which were designed to disrupt the interactions with the RNA molecule, did not affect the ATPase activity but completely abolished the replication and transcription of viral RNA and the infectivity of CHIKV. Our study sheds light on the roles of the RNA helicase region in viral replication and provides insights that might be applicable to alphaviruses and other RNA viruses in general.

Chikungunya virus | nonstructural protein 2 | superfamily 1 helicase | ATP hydrolysis | Alphavirus

Chikungunya virus (CHIKV) causes acute febrile illness termed Chikungunya fever, which often leads to painful and long-lasting arthritic sequelae (1–4). CHIKV belongs to the genus *Alphavirus*, which includes many human pathogens. "Old World" alphaviruses, including CHIKV, o'nyong'nyong virus, Sindbis virus (SINV), and Semliki Forest virus (SFV), usually cause rashes and arthritis, and Venezuelan equine encephalitis virus and Eastern equine encephalitis virus are "New World" alphaviruses and tend to cause encephalitis. These pathogenic alphaviruses have spread to many countries and have affected many millions of people worldwide, and unfortunately there are currently no vaccines or specific antiviral drugs for the prevention or treatment of any human infections caused by alphaviruses, including CHIKV.

CHIKV has a 12-kb positive-sense RNA (+ssRNA) genome containing a 5′-methylguanylate cap, a 3′-polyadenylated tail, and two ORFs. The first ORF is translated into precursors of four nonstructural proteins (nsPs) and the P1234 and P123 polyproteins. Both P123 and mature individual nsPs are involved in viral RNA replication and transcription. Each nsP has its own distinct enzymatic and nonenzymatic functions as follows: nsP1 displays the methyl- and guanylyltransferase activities required for cap synthesis and plasma membrane-anchoring ability (5, 6); nsP2 contains an N-terminal RNA helicase with both nucleotide

triphosphatase (NTPase) and RNA triphosphatase activities and a cysteine protease at the C-terminal region that is responsible for polyprotein processing (7, 8); nsP3 shows ADP ribosyl-binding and hydrolase activities and is used for the initiation of viral genome replication (9, 10); and nsP4 exhibits RNA-dependent RNA polymerase activity (11, 12). These nsPs together with host factors form replication complexes (RCs) that catalyze the synthesis of a negative-sense (−) antigenome and subsequently new genomic and subgenomic (SG) RNAs (13). The SG RNA serves as the mRNA for structural proteins, capsid, and glycoproteins (E1, E2, and E3) required for progeny virus assembly (14).

The CHIKV nsP2 plays a vital role in viral infection. In infected cells, the majority of the alphavirus nsP2 proteins are not bound to RCs located at the cell membranes but rather widely distributed between the cytoplasm and nucleus (15). The distribution of nsP2 in the nucleus inhibits antiviral signaling by hijacking the host's cellular transcriptional machinery and inducing cytopathic effects in vertebrate cells (16–18). As the largest alphavirus protein, nsP2 has multiple functional domains. Specifically, the CHIKV nsP2 possesses 5′–3′ RNA helicase and RNA annealing activities (8), and the C-terminal region (amino acids 471–798) has protease and methyltransferase-like (MTL)

Significance

The multifunctional helicase-protease protein nsP2 of Chikungunya virus (CHIKV) plays vital roles in viral RNA replication and virus–host interactions. We determined the crystal structure of the helicase part of the CHIKV nsP2 (nsP2h) bound to the conserved 3′ end of the genomic RNA and the nucleotide analog ADP-AlF$_4$. The structure of this ternary complex revealed the molecular basis for viral RNA recognition and ATP hydrolysis by the nsP2h. Unique hydrophobic protein–RNA interactions play essential roles in viral RNA replication. These findings expand our knowledge of CHIKV and related alphaviruses and might also have broad implications for antiviral and vaccine developments against pathogenic alphaviruses.

domains (19). Although the structures of the C-terminal protease region of alphavirus nsP2 (nsP2p) (20, 21) and uncleaved P23 precursor protein without the N-terminal helicase part of the nsP2 (nsP2h) (22) have been determined, the CHIKV nsP2h structure remains elusive.

The alphavirus nsP2 helicase is categorized into the superfamily 1 (SF1) helicases based on conserved motifs (23). SF1, which is the largest among the six helicase superfamilies, has a conserved helicase core that consists of two similar Rec-A-like domains, which work in parallel as a motor for DNA or RNA binding and NTP hydrolysis (24). To date, many crystal structures of SF1 helicases originating from different organisms, such as the bacterial DNA helicases UvrD (25) and PcrA (26), the eukaryotic RNA helicase Upf1 (27, 28), and the helicases of RNA viruses such as helicases from the tomato mosaic virus (ToMV) (29), the Middle East respiratory syndrome coronavirus (MERS-CoV) (30), and the equine arteritis virus (EAV), have been resolved (31). However, the structure of the helicase of ToMV (ToMV-hel) remains the only resolved crystal structure of a helicase from a virus belonging to the large alphavirus-like superfamily (29, 32). Although CHIKV also belongs to this superfamily, nsP2h shares less than 20% sequence identity with ToMV-hel.

In this study, we report a ternary complex of the CHIKV nsP2h bound to a 14-mer ssRNA (RNA14), which contains the sequence of the 3′ end of the CHIKV genome, and the ATP analog ADP-AlF₄, which is commonly used as a transition-state analog for ATP hydrolysis in structural biology (33, 34). The effects of mutations of the amino acid residues interacting with RNA14 and ADP-AlF₄ were analyzed to elucidate the significance of these interactions on the enzymatic function of nsP2h and on their potential roles in virus RNA replication. Together, our findings aid the understanding of the alphavirus replication process and provide valuable information for the development of antiviral treatments or vaccines against CHIKV and other pathogenic alphaviruses.

Results

Overall Structure of the Ternary Complex nsP2h-RNA14-ADP-AlF₄. To better understand the multiple functions of the CHIKV nsP2 N-terminal helicase, we determined the crystal structure of nsP2h (amino acid residues 1–465 of nsP2) in complex with a 14-mer ssRNA (RNA14; 5′-GUUUUUAAUAUUUC-3′) and a non-hydrolyzable transition-state nucleotide analog, ADP-AlF₄, which we designated as the nsP2h-RNA14-ADP-AlF₄ ternary complex (Fig. 1, *SI Appendix*, Fig. S1, and Movie S1). The structure was determined by multiwavelength anomalous diffraction (MAD) using selenomethionine (SeMet)-labeled nsP2h proteins. The crystal structure was then refined with a native dataset to a resolution of 2.00 Å (*SI Appendix*, Table S1). Briefly, two nsP2h molecules were built in the asymmetric unit (one from residues 1–463 and the other from residues 1–464) with only nine missing internal surface residues: residues 56–64 in the first nsP2h molecule and residues 55–63 in the second nsP2h molecule. The nsP2h-RNA14-ADP-AlF₄ complex was refined to obtain a good-quality structure with R_{work} and R_{free} values of 17.7% and 21.5%, respectively (*SI Appendix*, Table S1 and Movie S1). The final model includes seven ribonucleotide residues of RNA14 (residues 2–8, 5′-UUUUUAA-’3), an additional phosphate of the ninth ribonucleotide residue, and the transition-state nucleotide analog ADP-AlF₄ (Fig. 1 *B* and *C*).

The nsP2h structure is composed of multiple domains, including a small but uniquely folded N-terminal domain (NTD), the SF1 familiar accessory domains 1B and the Stalk α-helix, the two similar and parallel Rec-A-like domains (designated RecA1 and RecA2 domains), and an extended α-helical connector (Fig. 1). The NTD-Stalk-1B formed a top cover for the ssRNA-binding groove by interacting with both RecA1 and the 3′ half of the bound RNA (Fig. 1 *B–D*). The nsP2h helicase core consisted of the RecA1 (green) and RecA2 (cyan) domains, with RNA-binding motifs lying on top of the two domains and an ATPase active site between the two domains (Fig. 1 *B, C,* and *F*).

The connector, which serves as an extension of the helicase, connected the bottom part of both RecA1 and RecA2 domains to form the ATP entry tunnel.

Conserved and Unique Modes of RNA Recognition. The crystal structure revealed a unique mode of RNA recognition by the nsP2h (Fig. 1 *D* and *E* and Movie S2). Similar to the Upf1: ssRNA complex (28), the U₂–A₈ nucleotides of RNA14 were bound in the 5′-to-3′ direction in the central groove of the nsP2h. The ssRNA interacted with the conserved RNA backbone recognition motifs on top of the helicase core domains, such as Ia, Ic, III, IIIa, IV, V, and Va (Fig. 1*F* and *SI Appendix*, Fig. S2). K211 located within motif Ia was in contact with phosphates of U₆ and U₇ (Fig. 1 *D* and *E*). T233, D235, and S236 located within motif Ic interacted with phosphates and riboses of the last three ribonucleotides (U₆–A₈). Notably, W363 and R361 located within motif IV were in contact with phosphates of the first ribonucleotide (U₂) and ribonucleotides U₄ and U₅, respectively. T378 from motif V interacted with the phosphate of U₅, and motif V played a role in unwinding duplexes with a 5′-to-3′ polarity (35). N240 located between motifs Ic and II as well as N400 and H409 located between motifs V and VI showed interactions with the phosphates and riboses of U₃, U₄, U₆, and A₇. Most interestingly, unique stacking interactions were found between the nsP2h and the bound ssRNA. The aromatic side chain of F287 (motif III) intercalated between RNA bases U₄ and U₅ and separated the 5′ end of RNA U₂–U₄, which interacted with the RecA2 domain, and the 3′end of RNA U₅–A₈, which interacted with the RecA1 domain (Fig. 1 *D* and *E*). The side chains of Y161 and F164 from domain 1B, which were perpendicular to each other, formed stacking interactions with the bases of U₆ and A₇, respectively, and these interactions introduced a bend in the bound RNA (Fig. 1 *D* and *E*) and disrupted the A-form conformation of the bound RNA (Fig. 1 *A* and *B*).

Transition State of ATP Hydrolysis Mimicked by ADP-AlF₄. ADP-AlF₄ was tightly bound at the ATPase active site between the RecA1 and RecA2 domains, and the ADP adenine base was held above the connector (Fig. 1 *B* and *C* and Movie S3). In detail, the side chain ε-amino group of K192 (motif I) coordinated the β- and γ-phosphate groups directly as well as the α-phosphate group via a water molecule-mediated hydrogen bond. The presence of Mg^{2+} ions in the helicase protein has been shown to be essential for RNA helicase activity (8). Coherently, the α- and β-phosphate groups of ADP were coordinated by K192 and S193 within motif I, R312 within motif IIIa, and divalent metal ion (Mg^{2+}; black sphere). As shown in Fig. 1*F*, Mg^{2+} ions were in contact with β- and γ-phosphate groups, S193 (motif I), and three water molecule-mediated interactions with N213 (motif Ia), D252 (motif II), and E253 (motif II). The AlF₄ mimicked the transition state of the γ-phosphate group of the nucleotide analog by adopting a tetrahedral bipyramidal configuration with a water molecule above and an ADP molecule below (Fig. 1*F*). This transition state was stabilized by extensive polar interactions with K192 (motif I), Q283 (motif III), R312 (motif IIIa), and R416 (motif VI) and by one water molecule-mediated interaction (E253, motif II). Q283, the most conserved residue in motif III, acted as a γ-phosphate sensor and played a role in the coupling of ATP hydrolysis, RNA binding, and helicase activities. The ADP was located on top of the connector, and its adenine base was sandwiched between R311 (motif IIIa) and W450 (connector). The connector connected the helicase region to the protease domain and might have played a role in closing the ATP binding cleft.

Structural Dynamics of nsP2h upon Binding to RNA and ADP-AlF₄. To investigate how the binding of RNA and ADP-AlF₄ affected the conformational dynamics of nsP2h in solution, we performed a hydrogen/deuterium exchange MS (HDX-MS) analysis. RNA binding alone resulted in mild protection, as indicated by a slightly slower HDX between nsP2h and the solvent, which was limited to amino acid residues 218–238 (between motif Ia and
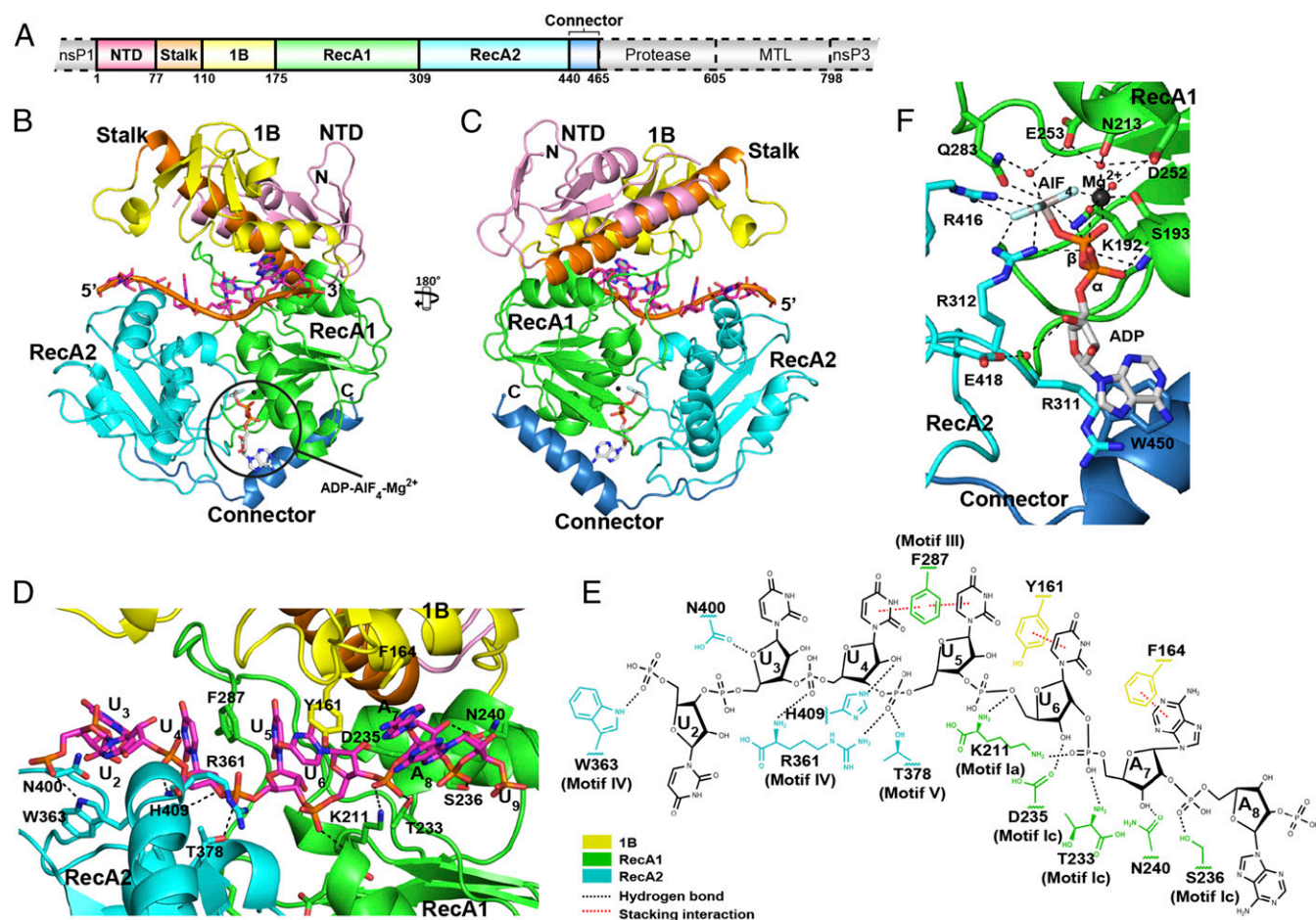
Fig. 1. Overall structure of the nsP2h-RNA14-ADP-AlF$_4$ ternary complex. (*A*) Schematic representation of the domain arrangement of nsP2h (amino acid residues 1–465 of nsP2) with the NTD in pink, the stalk domain in orange, domain 1B in yellow, the RecA1 domain in green, the RecA2 domain in cyan, and the connector in sky blue. (*B* and *C*) Ribbon diagrams of the nsP2h bound to RNA14 (5′-GUUUUUAAUAUUUC-3′; magenta sticks) and ADP-AlF$_4$ (red and blue sticks) before (*B*) and after (*C*) a 180° rotation. The bound ADP-AlF$_4$-Mg^{2+} is circled. (*D*) Detailed view of the binding of RNA14 to nsP2h. Y161 and F164 in domain 1B show stacking interactions with U$_6$ and A$_7$. K211, T233, D235, S236, N240, and F287 in RecA1 domain show interactions with U$_4$–A$_8$. R361, W363, N400, and H409 in RecA2 domain show interactions with U$_{2-5}$. (*E*) Schematic representation of the contacts between nsP2h and eight ribonucleotides. The residue numbers and colors of the domains are labeled according to nsP2h in D. (*F*) Magnified view of RecA1 (green) and RecA2 (cyan) domains bound to ADP-AlF$_4$. The Mg^{2+} ions (black) and water molecules (red) are shown as spheres. The protein main chains are simplified as ribbons. The ribonucleotides are represented by sticks. The main-chain atoms and side chains involved in the coordination of ribonucleotides are represented by sticks with single-letter codes, and the residue numbers and motif codes are shown. The black and red dashed lines represent the hydrogen and stacking interactions, respectively.

motif Ic) and amino acids 379–391 (between motif V and motif Va) (Fig. 2*A* and *SI Appendix*, Fig. S3*A*). These results suggested that RNA binding only triggered local conformational rearrangements in the absence of the nucleotide analog ADP-AlF$_4$. If both RNA14 and ADP-AlF$_4$ were present, nsP2h exhibited more conformational changes because all conserved helicase motifs in both the RecA1 and RecA2 domains were engaged in ligand binding. Stronger protection against solvent exchange was observed in the sequences within motifs I, Ia, III, IV, V, Va, and VI (Fig. 2*B* and *SI Appendix*, Fig. S3*B*). The region (amino acids184–216) between motif I and motif Ia exhibited protection against solvent exchange. The greatest protection against deuterium exchange was mapped to the p-loop sequence of motif I (Fig. 2*B* and *SI Appendix*, Fig. S5). In contrast, the sequence (amino acids 420–460) within the connector was the only region that showed some increased deuterium incorporation, likely due to the closed ATP binding cleft. It is also noteworthy that no significant conformational changes were observed in the N-terminal accessory domains in the ternary complex, including the NTD, stalk, and domain 1B (Fig. 2*B*).

To complement the HDX-MS findings, a limited proteolysis experiment was performed, and the results showed that almost

all apo nsP2h (Fig. 2*C*; lane 4) and RNA14-bound proteins (lane 5) were digested into smaller fragments by trypsin after incubation for 0.5 h. The resulting digested fragment distributions suggested that RNA binding had no effect on the stimulation of nsP2h conformational changes. In contrast, nsP2h was more resistant to trypsin digestion after incubation for 0.5 h (lane 6) and 2 h (lane 9) in the presence of both RNA14 and ADP-AlF$_4$. nsP2h bound to RNA14 and ADP-AlF$_4$ was completely digested into smaller fragments only after incubation with trypsin for 12 h (lane 12), which indicated that nsP2h exhibited a more stable conformational rearrangement after binding to RNA14 and ADP-AlF$_4$ (Fig. 2*C*). These findings were also in agreement with the results from the thermal shift assays. The melting temperature (T*m*) of nsP2h increased only by 0.4 °C in the presence of RNA14 and by 22.1 °C in the presence of both RNA14 and ADP-AlF$_4$ (Fig. 2*D*). Taken together, these results demonstrated that the ternary complex nsP2h-RNA14-ADP-AlF$_4$ was more thermostable and resistant against trypsin digestion and deuterium exchange compared with apo nsP2h or nsP2h–RNA14 (Fig. 2 *A*–*D*). The bindings of RNA14 and ADP-AlF$_4$ might trigger closed the conformational rearrangements in the
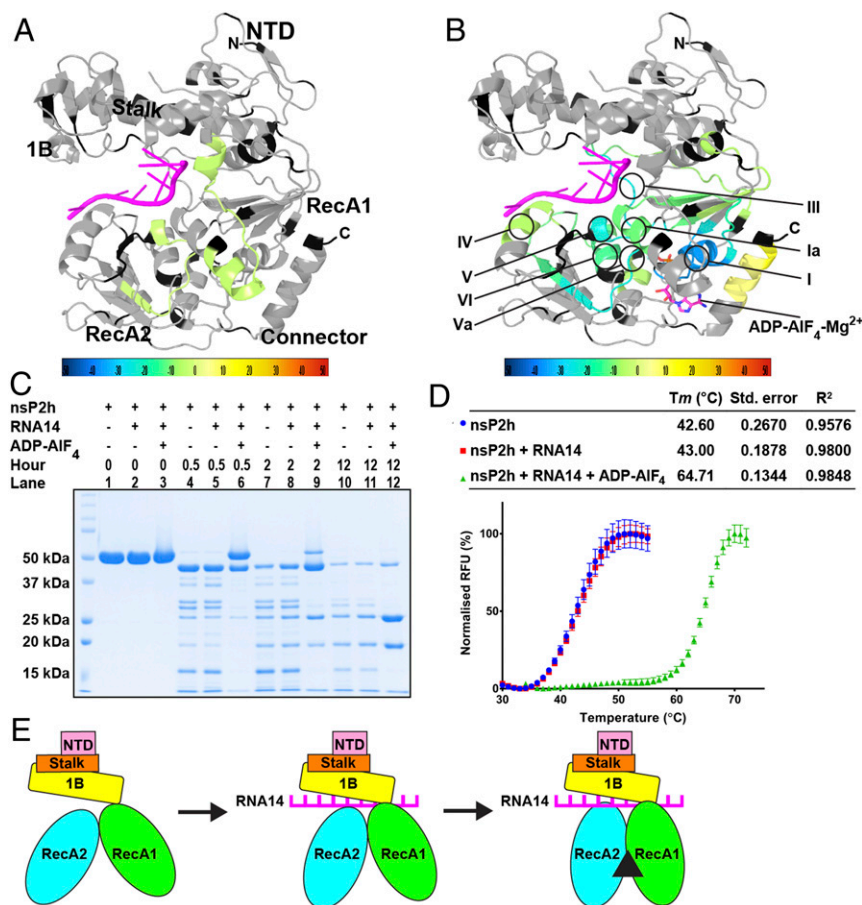
Fig. 2. Comparison view of the conformational dynamics of nsP2h induced by RNA14 and RNA14-ADP-AlF$_4$. The ribbon color diagram indicates the stabilization and destabilization of HDX. The level of deuterium differences displayed as a color gradient is located at the bottom of the ribbon color diagrams of (*A*) nsP2h bound to RNA14 and (*B*) nsP2h bound to RNA14-ADP-AlF$_4$. The ribonucleotides are shown as a magenta cartoon. The light gray color indicates regions with no statistically significant changes between the compared conditions, and the black color shows regions with no sequence coverage. The domains arrangement of *A* and *B* are the same. The conserved motifs with stronger protection against solvent exchange are circled. (*C*) Limited proteolysis of nsP2h RNA and RNA14-ADP-AlF$_4$ using trypsin at different time points. (*D*) Thermal shift assays representing the T*m* profiles of nsP2h against nsP2h with RNA14 and nsP2h with RNA14-ADP-AlF$_4$. The largest and smallest means in each dataset were normalized and set to 100% and 0%, respectively. The error bars represent the SEMs from the triplicate measurements. (*E*) Cartoon model of nsP2h conformational changes upon binding to RNA14 and ADP-AlF$_4$. RNA14 and ADP-AlF$_4$ are shown as a magenta ladder and black triangle, respectively. (*E*, *Left*) Free nsP2h adopts an open conformation. (*E*, *Middle*) Upon RNA14 binding, nsP2h shows limited conformational changes on the top of the RecA1 and RecA2 domains. (*E*, *Right*) In the presence of RNA14 and ADP-AlF$_4$, nsP2h displays a closed conformation, in which both the N-terminal region and the helicase region are engaged in binding to RNA and ATP.

nsP2h helicase core, which resembles the transition state of ATP hydrolysis (Fig. 2*E*).

Dissecting the Coupling of RNA Binding and ATP Hydrolysis by Mutagenesis. To reveal the functional consequence of the coordinated binding of RNA14 and ADP-AlF$_4$, we performed a structure-guided mutagenesis study focusing on the stacking interactions and NTP-binding sites. The Y161A or F164A substitution was introduced to weaken the hydrophobic interaction between RNA and the N-terminal accessory domains of the helicase core. The obtained mutant proteins largely retained the WT ATP hydrolysis activity (Fig. 3*A*), which suggested that the stacking interactions between the RNA and the helicase accessory domain 1B have little effect on the NTPase activity of the helicase core. Subsequently, W450 in the connector, which establishes the stacking interaction with the ADP adenine base, was mutated to Ala (W450A). Remarkably, the W450A mutation reduced the nsP2h solubility and yield, which indicated that W450 in the connector is essential for maintenance of the integrity of the catalytic site (*SI Appendix*, Fig. S4). Furthermore, compared with WT nsP2h, the protein with the W450A substitution showed lower catalytic efficiency with a k_{cat}/K_m value of 0.12 s^{-1}·µM^{-1} (Fig. 3*A*). In sharp contrast, the F287A mutation, which affected another residue involved in the stacking interactions, doubled the k_{cat} value (ATP molecules hydrolyzed per second per molecule of nsP2h) of the enzyme. The mutant protein also showed approximately twofold less effectiveness with a k_{cat}/K_m value of 0.21 s^{-1}·µM^{-1} and less binding affinity with a K_d value of 760.0 nM compared with WT nsP2h (Fig. 3 *A* and *C*). Furthermore, we observed a twofold increase in the K_d values of both nsP2h and nsP2h-F287A in the presence of ADP-AlF$_4$. A similar observation has also been reported for the transition-state analog ADP-AlF$_4$ of the SF1 RNA helicase Upf1 and the SF2

RNA helicase hepatitis C virus NS3 (28, 36). The results suggested that the RNA conformation trapped in the ternary complex was restrained by nsP2h upon ATP binding (Fig. 3*C*). As expected, the mutations of two well-studied conserved residues related to NTP binding, K192A within motif I and Q283A within motif III, completely abolished the ATPase activity (37). The changes in the protein stability of these mutants upon RNA-ADP-AlF$_4$ binding were also in agreement with the ATP hydrolysis data (*SI Appendix*, Fig. S5). Proteins harboring mutations in amino acid residues with aromatic rings that were involved in the stacking interactions with RNA (Y161A, F164A, and F287A) responded well and became significantly stabilized to a level comparable to that of the WT protein, as demonstrated with an increase of ~20 to 22 °C in their T*m* values. Due to weakened ATP binding, the proteins with ATPase site mutations (K192A and Q283A) were stabilized to a much lesser extent than the WT protein, and their T*m* values increased by ~4 to 5 °C (*SI Appendix*, Fig. S5).

We subsequently showed that ssRNA exerted little stimulatory effect on the ATPase activity of the CHIKV nsP2h (Fig. 3*B*), which agrees with the results of earlier studies (7, 8). At concentrations of 1 or 2 µM, RNA14 was unable to significantly stimulate the ATPase activity of WT nsP2h or its mutant F287A, which confirmed that RNA binding alone did not induce significant conformational changes to the helicase/ATPase (Fig. 3*B* and *SI Appendix*, Fig. S6).

Essential Roles of nsP2h–RNA Interactions on Viral RNA Replication. To examine how viral fitness might be affected by these mutations, viral RNA replication (Fig. 4*A*) was analyzed using the CHIKV *trans*-replicase system developed by Utt et al. (38). Briefly, the cytomegalovirus promoter-based expression plasmid CMV-P1234 was used to express P1234, whereas an RNA template containing firefly luciferase (Fluc) and *Gaussia* luciferase
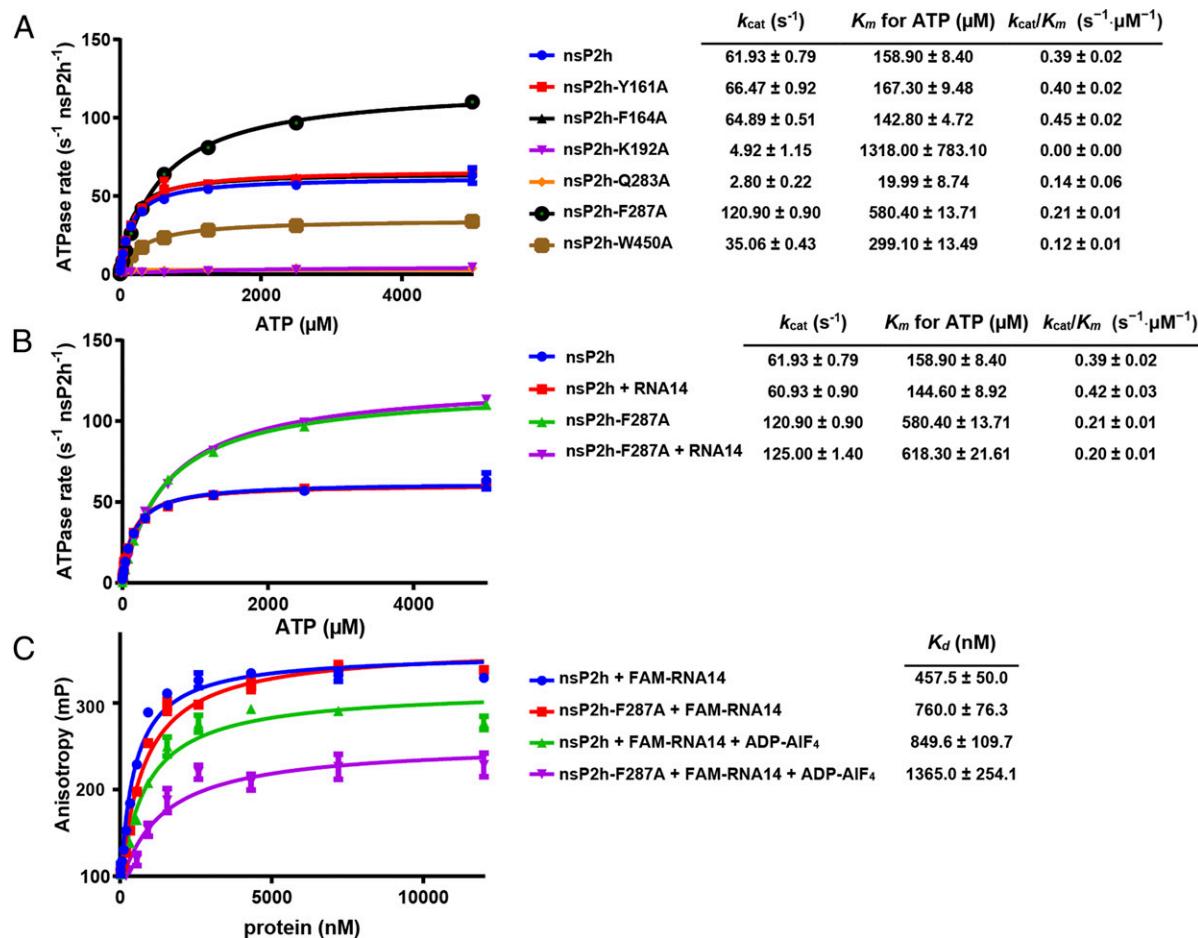
Fig. 3. Effects of mutations on the enzymatic activities of nsP2h. (*A*) ATP hydrolysis by nsP2h and mutants at varying ATP concentrations. (*B*) RNA-stimulated ATP hydrolysis by nsP2h and nsP2h-F287A at varying ATP concentrations and 1 μM RNA. The dataset of nsP2h and nsP2h-F287A without RNA stimulation from A was used as a control to compare with the dataset of RNA-stimulated ATP hydrolysis. (*C*) RNA-binding analysis of nsP2h and nsP2h-F287A through fluorescence anisotropy. All enzymatic assays were performed in triplicate, and the data are presented as the means and standard derivations.

(Gluc) reporters was produced using the human RNA polymerase I promoter (pol I-Fluc-Gluc; Fig. 4*A*). Six individual residues were mutated to Ala on the CMV-P1234 construct. As shown in Fig. 4 *B* and *C*, the Fluc and Gluc expression levels (corresponding to the synthesis of full-length and SG RNAs, respectively) of replicases containing K192A or Q283A mutations did not exceed the background levels. These results were expected because the catalytic site is critical for NTP hydrolysis (Fig. 3*A*) and important for viral replication (37). Interestingly, the same results were observed for replicases harboring the Y161A, F164A, or Y161A and F164A substitutions. In contrast, the ATPase activities of nsP2h-Y161A and nsP2h-F164A were similar to that of WT nsP2h (Fig. 3*A*). This finding suggested that although the stacking interactions were not critical for ATPase activity they might be needed for helicase activity, and thus the corresponding residues are critical for both genomic RNA and SG RNA synthesis. Similarly, no increase in the Fluc signal was observed in the presence of replicase with the W450A mutation, and only a minor increase (~0.1% of that achieved with the WT replicase) in Gluc activity was detected. Interestingly, the replicase containing the F287A substitution increased the Fluc expression level to levels comparable to those found with the WT replicase but showed a significantly reduced ability to increase Gluc expression, which suggested that the F287 residue might play a specific role in activating the promoter of SG RNA synthesis (Fig. 4*C*).

We subsequently evaluated the effects of mutations in template-RNA sequences. For this analysis, U_5 and A_7 of RNA14, which

originate from the 3′ end of the CHIKV genome, were selected because these are the most active ribonucleotides in terms of interaction with residues of nsP2h (Fig. 1 *D* and *E*). The position of pol I-Fluc-Gluc corresponding to the U_5 residue was mutated to adenine (U_5A), or alternatively the position of A_7 was mutated to uracil (A_7U). As shown in Fig. 4 *D* and *E*, the Fluc and Gluc expression levels from the RNA corresponding to pol I-Fluc-Gluc containing the U_5A substitution were greatly and significantly reduced, and no similar effect was detected for the pol I-Fluc-Gluc containing A_7U. U_5 exhibits a stacking interaction with F287 (motif III) and hydrogen bonds with R361 and T378 (motif V; Fig. 1 *D* and *E*), and thus the mutation of U_5 might disrupt these interactions. Taken together, the results showed that the residue corresponding to U_5 of RNA14 was critical for CHIKV RNA replication.

The effects of helicase mutations on the ability of RNA, transcribed by cellular RNA polymerase II using a plasmid containing the cDNA of the virus under the CMV promoter, to initiate replication and infectious virus formation (hereafter "infectivity of RNA") were examined through an infectious center assay (ICA). The results revealed that the infectivity of RNA of all CHIKV constructs containing the above-listed mutations was below the detection limit (<2 pfu/μg DNA) (Fig. 4*F*). For constructs harboring alanine residues at the Y161, F164, K192, Q283, or W450 positions, this result was expected because it correlated to the lack of genomic RNA synthesis observed in the *trans*-replicase assays (Fig. 4*B*). The lethal nature of these substitutions was further confirmed by the lack of infectious virus progeny and capsid protein expression in transfected cells (Fig. 4
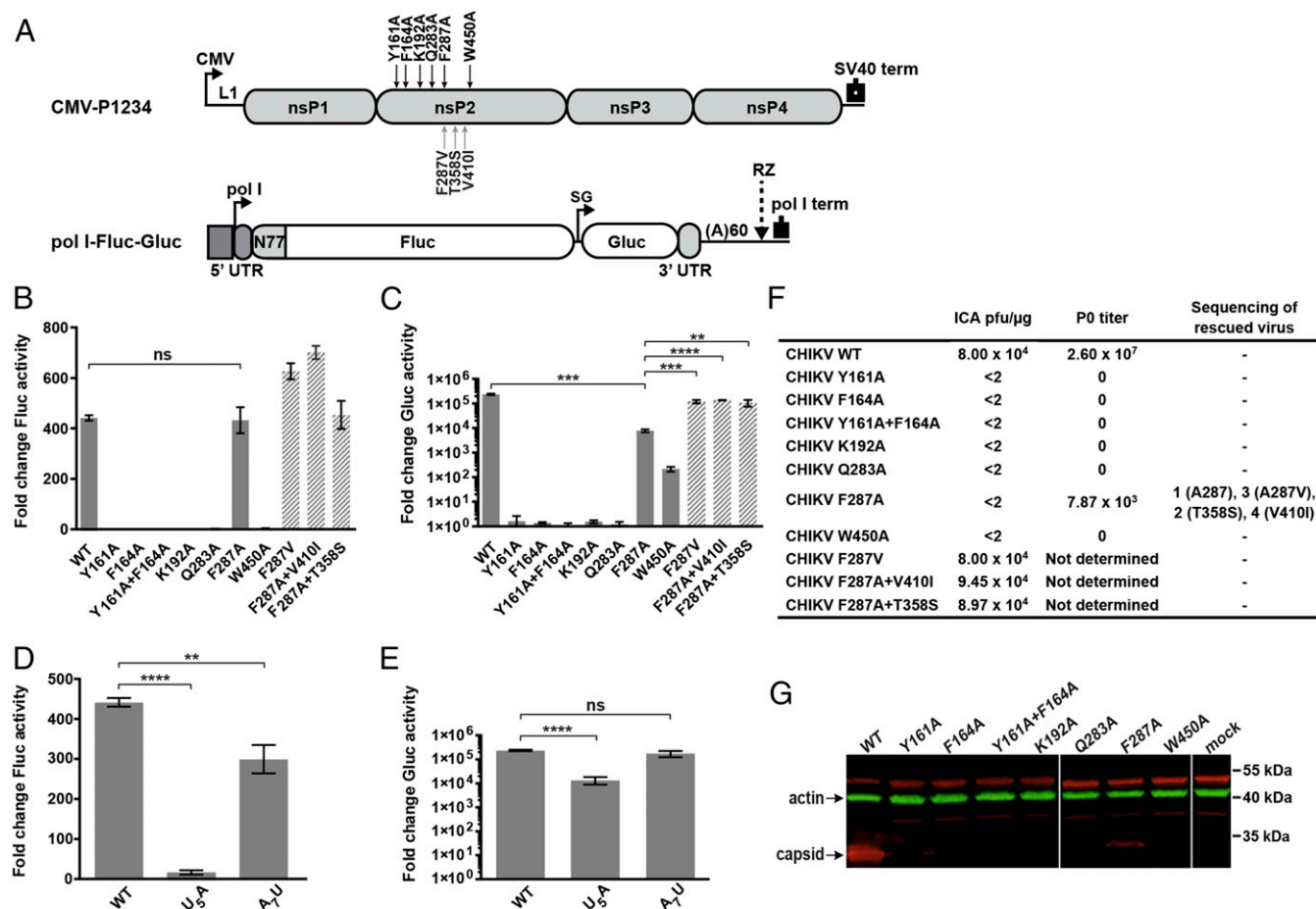
Fig. 4. Effects of mutations on the activity of CHIKV *trans*-replicase and on the rescue, propagation, and capsid protein expression of the corresponding viruses. (*A*) Schematic presentation of CMV-P1234 and pol I-Fluc-Gluc constructs. CMV, CMV promoter; LI, leader sequence from the thymidine kinase of herpes simplex virus; pol I, promoter for human RNA polymerase I; 5′ and 3′ UTRs, CHIKV 5′- and 3′-untranslated regions; N77, sequence encoding the N-terminal fragment of nsP1 (amino acids 1–77); SG, CHIKV SG promoter; RZ, antisense strand ribozyme from hepatitis delta virus; SV40 term, late poly-adenylation signal from simian virus 40. Position of structure-guided mutations (black arrows) and adaptive mutations (gray arrows) detected in the progeny of the virus harboring the F287A substitution are shown above and below the drawing of CMV-P1234, respectively. (*B* and *C*) Effects of mutations on CHIKV replicase activity. U2OS cells were cotransfected with pol I-Fluc-Gluc and CMV-P1234 (WT) and its mutant variants harboring the Y161A, F164A, Y161A+ F164A, K192A, Q283A, F287A, W450A, F287V, F287A+V410I, and F287A+T358S substitutions in nsP2. Control cells were transfected with the pol I-Fluc-Gluc and CMV-P1234 harboring mutations resulting in the DD466-467AA substitution (GAA) in nsP4. The cells were lysed at 18 h posttransfection, and the Fluc (*B*) and Gluc (*C*) activities generated by the different replicases were normalized to those measured with the inactive (GAA) control. Each column represents an average from three independent experiments, and the error bars represent the SDs. **$P < 0.01$, ***$P < 0.001$, ****$P < 0.0001$; ns, not significant (Student's unpaired t test). (*D* and *E*) Effects of mutation in the CHIKV RNA template on its replication by WT CHIKV replicase. U2OS cells were cotransfected with CMV-P1234 and pol I-Fluc-Gluc (WT) or its mutants with U_5A or A_7U substitutions, and the analysis was performed and the data are presented as described above. **$P < 0.01$, ****$P < 0.0001$; ns, not significant (Student's unpaired t test). (*F*) Effects of mutations in the helicase region of nsP2 on the efficiency of viral rescue and on the properties of rescued viruses. (*G*) Capsid protein expression in BHK-21 cells transfected with plasmids containing icDNA of WT CHIKV and its original mutant variants. The band corresponding to actin, which was used as a loading control, is shown in green. The capsid protein is shown in red and is located below the actin protein. In addition, an unspecific protein band above that of actin was detected by anti-CHIKV capsid antibodies.

F and *G*). In agreement with the *trans*-replicase data, the virus containing the F287A substitution was rescued, even though it reached a strongly diminished titer (7.87×10^3 pfu/mL) and showed reduced synthesis of capsid protein (Fig. 4 *F* and *G*). To resolve the discrepancy between the ICA data and the production of infectious progeny, the genomes of the rescued viruses were analyzed for reversions, pseudoreversions, and/or occurrences of secondary compensatory mutations. Sequencing of the 2,815-bp region (residues 2,100–4,915 of the CHIKV genome) covering all nsP2 sequences and flanking regions of nsP1 and nsP3 from 10 independent clones revealed no potential second-site adaptive mutations outside of the helicase domain, which suggested that the compensation was limited within the nsP2h. One clone out of 10 did not have any mutations with the exception of the original F287A, which indicated that the genome containing F287A was sufficiently active to be preserved in a

pool of replicating RNAs. However, the A287 residue was clearly not optimal because nine clones contained pseudoreversions or adaptive changes. Three clones had some version of pseudoreversion through introduction of the A287-to-Val (A287V) substitution (*SI Appendix*, Fig. S7*A*). Although F287A substitution was preserved in the rest of clones, four of them contained the V410-to-Ile (V410I) substitution, and remaining clones contained the T358-to-Ser (T358S) substitution (*SI Appendix*, Fig. S7 *B* and *C*). Taken together, our data indicated that the F287A mutation attenuated the infectivity of CHIKV, but the rescued virus possessed adaptive abilities.

To confirm the functional significance of the revealed mutations, these were introduced into both the CMV-P1234 and infectious cDNA constructs harboring the F287A substitution and assayed as described above. All three substitutions restored the infectivity of RNA of the infectious clone to a level similar to

that of WT CHIKV (Fig. 4F); the only differences were that the viruses with the pseudoreversion (Val287 residue) or the F287A+V410I mutations formed smaller plaques than the WT viruses or viruses harboring the F287A+T358S substitutions. The effects of all three mutations on *trans*-replicase were also nearly identical: The pseudoreversion and both compensatory changes exerted minor or no effects on the ability of the mutant *trans*-replicase to increase the expression of the Fluc marker, and the ability to increase the expression of Gluc was restored to levels similar to that of the WT construct (Fig. 4 B and C). These effects were highly consistent with the nature of the defect caused by the F287A substitution and demonstrated that the mutant virus restored its ability to express high levels of SG RNA through pseudoreversion, which likely increased the hydrophobic properties because Val does not possess an aromatic ring. Alternatively, and somewhat more surprisingly, SG RNA synthesis was also rescued by compensatory changes in the RecA2 domain that occurred at residues located close to the positions involved in RNA binding (*SI Appendix*, Fig. S7).

Structural Comparison with Other SF1 RNA Helicases. As shown in Fig. 5, *SI Appendix*, Fig. S8, and Movie S4, the Dali server revealed four top hits of SF1 RNA helicases: ToMV-hel [Protein Data Bank (PDB) ID code 3WRX; Dali Z-score: 27.1], human Upf1 (PDB ID code 2XZO; Dali Z-score: 20.5), MERS-CoV nsp13 (PDB ID code 5WWP; Dali Z-score: 19.8), and EAV nsp10 (PDB ID code 4N0N; Dali Z-score: 18.2) (39). The conserved RNA helicase domain arrangements of these top hits of SF1 RNA helicases shared slightly higher similarity to nsP2h (Fig. 5A), but their accessory domains were quite different. The structure of ToMV-hel (Fig. 5B) bound to a nucleotide analog was the most similar to that of nsP2h. Because the structure was not bound to RNA, the interaction and structural conformation between RNA and ToMV-hel are unknown. The human Upf1 (Fig. 5C) and EAV nsp10 (*SI Appendix*, Fig. S8B) were bound to ssRNA and ssDNA, respectively, on top of the helicase core domains, and these interactions were mediated by accessory domains, similarly to the findings obtained for nsP2h. However, the majority of contacts in ssRNA-Upf1 or ssDNA-EAV nsp10 were only made through backbone phosphates and nonspecific protein–base interactions. Unlike Upf1 and EAV nsp10, the RNA–nsP2h interface consisted of three additional stacking interactions between nsP2h and the ssRNA between conserved hydrophobic residues from the N-terminal region of the helicase and RNA bases. Furthermore, in contrast to the nsP2h, none of the top hit RNA helicases contained an extended α-helical connector close to the catalytic site. The W450 in the connector, the adenine base of ADP, and R311 in motif IIIa formed stacking interactions (Fig. 1 B and C). As shown above, disruptions of any stacking interactions and residues in the connector affected the ability of nsP2 to support viral RNA replication (Figs. 3 and 4), which suggested that the stacking interactions and the connector are critical for the replication of CHIKV. Taken together, the results from the structural comparison of nsP2h with similar SF1 helicases further support the assertion that the RNA recognition mode of nsP2h is unique.

Discussion

The CHIKV nsP2 plays multiple biological roles that are essential for viral replication and interaction with the host, but there is no available structural information for its N-terminal region. In this study, we determined the crystal structure of CHIKV nsP2h bound to the conserved 3′ end of genomic RNA (RNA14) with the sequence 5′-GUUUUUAAUAUUUC-3′ and ADP-AlF$_4$, which is the transition-state analog of ATP hydrolysis. This crystal structure of the alphavirus nsP2 helicase region contains the NTD, the domain 1B, the stalk, the RecA1 domain, the RecA2 domain, and the connector (Fig. 1A). The accessory NTD and the connector are relatively small but uniquely wrapped around the conserved helicase core. The RNA14 and ADP-AlF$_4$ collaboratively stabilized the helicase region by modulating conformational rearrangements of the nsP2h. We observed unique

hydrophobic stacking interactions between the nsP2h and the RNA and proved that these interactions were essential for viral replication through a structure-guided mutagenesis analysis.

The published evidence suggests that the functionality of nsP2 depends on the precise sequence of its NTD. Here, we showed that the extreme end of the NTD (amino acids 1–14) adopted an ordered hairpin structure (*SI Appendix*, Fig. S9) after cleavage between the nsP1 and nsP2 of CHIKV (R^{P4}A^{P3}G^{P2}A^{P1}↓G$^{P1'}$I$^{P2'}$I$^{P3'}$E$^{P4'}$) (40). Previous studies have indicated that the insertion of GFP between residues 8 and 11 of the SINV nsP2 (41) and the substitutions of residues 3–7 of the SFV nsP2 (40) affect ns-polyprotein processing and RNA infectivity. Moreover, the CHIKV nsP2 without the NTD lacks NTPase activity (8), which further supports possibility that disruption of the hydrogen interaction of the hairpin loop and deletion of the NTD affect the functionality of alphavirus nsP2 proteins. As shown in the nsP2h structure, the NTD was not directly involved in RNA14 binding and might be involved in the recognition of other viral proteins or protein junctions. Based on this activity, the NTD helps nsP2 regulate ns-polyprotein processing and coordinate the viral RNA replication process (19, 40). Domain 1B, which is an accessory domain, was previously described as a human Upf1 protein (27). However, according to the Dali server, the structure of Upf1, a well-studied SF1 helicase, shares only a low level of structural similarity with nsP2h, as shown in Fig. 5. Upf1 consists of an N-terminal cysteine-histidine–rich domain, a domain 1B connected to RecA1 by two α-helices, and an α-helix denoted as
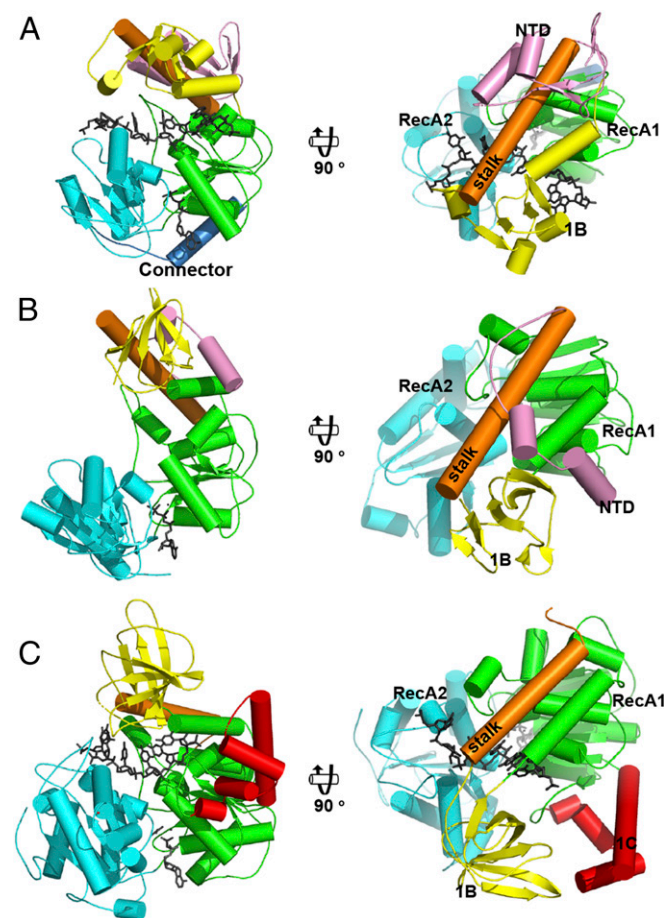

Fig. 5. Structural comparison of the nsP2h-RNA14-ADP-AlF$_4$ with SF1 helicases. (A) nsP2h (PDB ID code 6JIM), (B) ToMV-hel (PDB ID code 3WRX), and (C) Upf1 helicase-RNA complex (PDB ID code 2XZO). The ribonucleotides and the ADP-AlF$_4$ are colored in gray. The 1C domain of Upf1 is shown in red, and the domains of nsP2h are colored as in Fig. 1.

domain 1C inserted into RecA1. Domain 1B of Upf1 stabilizes the RNA-binding conformation during the transition state (28). In the nsP2h-RNA14-ADP-AlF₄ structure, domain 1B played a role in stabilizing the conformational rearrangements of nsP2h in the presence of RNA (Fig. 2), which aligned with the stacking interactions found for the Y161 and F164 residues in domain 1B with RNA14. Although the ATPase activities of the nsP2h-Y161A and nsP2h-F164A mutants remained intact (Fig. 3*A*), the activities of replicases harboring these substitutions were completely abolished (Fig. 4 *B* and *C*). The findings might suggest that the Y161A and F164A mutations in domain 1B might directly destabilize the alphavirus (+) ssRNA binding that is essential for negative-sense (−) ssRNA synthesis and thereby abolish viral RNA replication.

Similar to Y161 and F164 in domain 1B, F287 from motif III also exhibited stacking interactions with two bases of RNA14. Although F287 is located within motif III of SF1 helicases, the residue is strictly conserved as a phenylalanine among alphaviruses, whereas in other SF1 helicases, residues with smaller hydrophobic side chains, such as leucine and valine, are more common (42). Unlike inactive replicases harboring the Y161A or F164A substitution, the replicase containing the F287A substitution displayed active genomic RNA synthesis, even though its ability to perform SG RNA synthesis was strongly (∼30-fold) reduced. This finding suggested that the F287 residue might play a specific role in stabilizing the interaction with alphavirus ssRNA that is essential for the synthesis of SG RNA. Based on the findings, CHIKV containing the F287A substitution was attenuated and pseudoreverted or accumulated adaptive mutations (Fig. 4*F*) that increased the ability of its replicase to synthesize SG RNA and restored the infectivity of the mutant viral genome. Interestingly, the accumulated adaptive mutations were only found within the following conserved helicase SF1 motifs: A287V (pseudoreversion) located within motif III, T358S located within motif IV, and V410I located near motif VI. Motifs III, IV, and VI are the important contact sites between nucleic acids and NTP binding (42), and these mutations might aid the capture of (−) ssRNA for SG RNA synthesis.

The nsP2 of Old World alphaviruses displays more cytotoxic properties than nsP2 of New World alphaviruses (43, 44). Point mutations in the protease region Old World alphavirus nsP2, such as mutations of the P718 residue and insertion of five amino acid residues after residue 647 of CHIKV nsP2, reduce viral replication (45). Such attenuating changes trigger the accumulation of adaptive mutations, which complicates the design of live attenuated viruses for vaccine production (46, 47). The attenuated CHIKV containing the F287A substitution coupled with an adaptive mutation (Fig. 4*F*) might also have the potential to be a new platform for the development of live attenuated alphavirus vaccines. In addition, information about the structure of nsP2h can be used for the construction of additional attenuating mutations that, similar to very recently described variable peptide loop mutations in the protease region of CHIKV nsP2 (44), have no negative effect on viral RNA replication. Furthermore, the nsP2h bound to a nucleotide analog significantly reduced the RNA-binding affinity (Fig. 3*C*), providing a great opportunity to design a rational drug for inhibiting or attenuating viral replication. Therefore, these strategies could provide a new therapeutic option for controlling the disease in the future.

nsP2h with no C-terminal protease region lacks RNA helicase activity (8). This finding raises a question regarding the roles of the C-terminal protease in helping the function of the N-terminal helicase part of nsP2. Although the structures of the N and C termini of CHIKV nsP2 (21) are available, the conformational arrangement of the full-length nsP2 remains elusive. Because nsP2 is a nonprocessive helicase (8), it is believed that a more stable nsP2-RNA conformation is needed to modulate RNA helicase activity. Das et al. (8) suggested that the MTL domain from the C-terminal part of nsP2 could strengthen RNA–nsP2 binding. In addition, GFP inserted between the nsP2

N-terminal helicase and C-terminal protease regions did not significantly affect viral replication (38), which suggested that the two domains were connected with a flexible linker that gives the domain a notable level of freedom to move relative to each other. Thus, it would not be surprising if RNA–full-length nsP2 has a distinct and stable conformational arrangement. Therefore, further investigations are needed to study the structure of RNA–full-length nsP2 and the functional cross-talk within domains. In addition, similar to the studies on SINV uncleaved P23 precursor protein without nsP2h (22), there will be considerable interest to resolve the structure of uncleaved P23 of CHIKV, and such studies will provide essential information about the functioning of the replicase proteins in their precleavage forms and allow a deeper understanding of the process of active replicase formation.

Materials and Methods

Construction of CHIKV Expression, Viral, and *Trans*-Replicase Constructs. The region encoding CHIKV nsP2h (amino acids 436–1,000 of P1234; GenBank accession no. KC149887.1) was amplified by PCR using the primers shown in *SI Appendix*, Table S2. Ligation-independent cloning (LIC) was used to clone the obtained nsP2h fragment into a modified pSUMO-LIC expression vector encoding a cleavable N-terminal His-SUMO tag. The T4 DNA polymerase-treated amplicon and BsaI-treated vector were incubated for 1 h for annealing before transformation. The CHIKV infectious cDNA (icDNA) clone CMV-ICRES1 and the *trans*-replicase system vectors CMV-P1234 and pol I-Fluc-Gluc have been previously described (10, 38, 48). Based on the information of the CHIKV nsP2h ternary complex, point mutations were introduced on CHIKV icDNA, CMV-P1234, pol I-Fluc-Gluc, and nsP2h expression constructs using site-directed PCR-based mutagenesis and subcloning.

Production of Recombinant Proteins. The expression and purification of proteins of interest were performed as described previously by Phoo et al. (49). Expression vectors encoding WT and mutant nsP2h were used for the transformation of *Escherichia coli* Rosetta 2 (DE3) strain. The bacteria were cultured at 37 °C in LB supplemented with antibiotics and induced with 1 mM isopropyl-1-thio-β-D-galactopyranoside (IPTG) at 18 °C overnight. For the production of SeMet-labeled CHIKV nsP2h, the *E. coli* B834 (DE3) cells were transformed using the expression vector containing the sequence encoding WT nsP2h, and SeMet-labeled CHIKV-nsP2h was obtained by culturing the obtained cells at 37 °C in SelenoMethionine Medium Complete (Molecular Dimension) supplemented with antibiotics and methionine (Molecular Dimension) until the log phase. The bacterial culture was harvested and washed three times with autoclaved distilled water. The washed bacterial pellet was resuspended in warm SelenoMethionine Medium Complete supplemented with antibiotics and SeMet (Molecular Dimension), and the cultures were grown for 1 to 2 h at 37 °C and then induced with 1 mM IPTG at 18 °C overnight. The overnight bacterial cultures were harvested and lysed, and the recombinant protein was purified using HisPur Ni-NTA Resin (Thermo Fisher Scientific). The eluted proteins were digested with SUMO protease, and the obtained untagged proteins were further purified through HiTrap Heparin HP affinity columns (GE Healthcare Life Sciences) and HiLoad 16/600 Superdex 75 columns (GE Healthcare Life Sciences). All purified proteins in gel filtration buffer (20 mM Hepes, pH 7.5, 150 mM NaCl, 5% glycerol, and 2 mM DTT) were finally aliquoted and stored at −80 °C.

Crystallization and Data Collection. The nsP2h complex with a RNA14 (5′-GUUUUUAAUAUUUC-3′) at a protein:RNA molar ratio of 1:1.2 and ADP-AlF₄ (1.5 mM ADP, 15 mM NaF, and 7.5 mM AlCl₃) were added to buffer containing 25 mM Hepes, pH 7.5, 150 mM NaCl, 5% glycerol, and 2 mM DTT. The protein complex solution was mixed equally with a reservoir solution containing 100 mM Hepes, pH 7.5, and 12% PEG 8000 using the hanging-drop vapor-diffusion method to obtain a concentration of 3 mg/mL. Rod-shaped crystals grew within 1 d at 20 °C. The crystals were dehydrated overnight by soaking in dehydration buffer (100 mM Hepes, pH 7.5, and 20% PEG 8000) containing 20% glycerol cryoprotectant, picked up using nylon loops (Hampton Research), flash-frozen in liquid N₂, and mounted. The diffraction intensities of nsP2h-RNA14-ADP-AlF₄ and SeMet-labeled nsP2h-RNA14-ADP-AlF₄ were collected with the Taiwan Proton Source beamline 05A (TPS-05A) at the National Synchrotron Radiation Research Center. All of the data were indexed, integrated, and scaled by HKL2000 (50) and XDS (51).

Structure Determination. The structure of the nsP2h complex was determined using the MAD method (52). AutoSol (PHENIX) and AutoBuild (PHENIX) (53) were used to calculate the initial phases and for automatic model building,

respectively. The structures were then determined by molecular replacement using the initial phases of the nsP2h complex as the search model with Phaser-MR (PHENIX) (53). All constructed models were refined using Coot (54) and phenix.refine (PHENIX) (53).

Hydrogen/Deuterium Exchange. This method was previously described by Zheng et al. (55). For HDX-MS analysis, an nsP2h protein solution with a concentration of 10 μM was mixed with RNA14 (5′-GUUUUUAAUAUUUC-3′) at a protein:RNA molar ratio of 1:1.2 with or without ADP-AlF₄ (1.5 mM ADP, 15 mM NaF, and 7.5 mM AlCl₃) in HDX H₂O buffer containing 25 mM Hepes, pH 7.5, 150 mM NaCl, and 2 mM DTT. The mixtures were incubated for 1 h at 4 °C to reach the equilibrium states and then subjected to HDX analysis. All of the sample handling and peptide separation procedures were performed at 4 °C. One differential HDX reaction contained 5 μL of nsP2h-RNA14 or nsP2h-RNA14-ADP-AlF₄ and 20 μL of D₂O-containing HDX buffer. The reaction was at 4 °C for different time points (0, 10, 30, 60, 300, 900, or 3,600 s), and the complexes were then quenched by mixing with 25 μL of ice-cold solution containing 4 M guanidine HCl and 1% trifluoroacetic acid. Upon injection, the samples were passed through an immobilized pepsin column (2 mm × 2 cm) at 200 μL·min⁻¹, and the digested peptides were captured on a 2-mm × 1-cm C₈ trap column (Agilent) and desalted. The peptides were separated across a 2.1-mm × 5-cm C₁₈ column (1.9-μm Hypersil Gold; Thermo Fisher) with a linear gradient of 4 to 40% CH₃CN and 0.3% formic acid over 5 min. The sample handling, protein digestion, and peptide separation procedures were conducted at 4 °C. Mass spectrometric data were acquired using an Orbitrap mass spectrometer (Q Exactive; Thermo Fisher) with a measured resolving power of 65,000 at m/z 400. The HDX analyses were performed in triplicate using single preparations of each protein–ligand complex. The intensity-weighted mean m/z centroid value of each peptide envelope was calculated and subsequently converted into a percentage of deuterium incorporation. Corrections for back-exchange were made based on an estimated deuterium recovery of 70% and accounting for the known 80% deuterium content of the deuterium exchange buffer. When comparing the two samples, the perturbation % D was determined by calculating the difference between the two samples. HDX Workbench colors each peptide according to the smooth color gradient HDX perturbation key (% D) shown in each indicated figure. The differences in % D ranging from −5 to 5% are considered nonsignificant and are colored in gray according to the HDX perturbation key (56). In addition, unpaired Student's t tests were used to reveal the statistically significant ($P < 0.05$) differences between samples at each time point. One time point with a P value less than 0.05 was present for each peptide in the dataset, further confirming that the difference was significant. For data rendering, the HDX data from all overlapping peptides were consolidated into individual amino acid values using a residue averaging approach. Briefly, for each residue, the deuterium incorporation values and peptide lengths from all overlapping peptides were assembled. A weighting function that gave shorter and longer peptides a heavier or lesser weight, respectively, was applied. Each of the weighted deuterium incorporation values was then averaged to produce a single value for each amino acid. The initial two residues of each peptide, as well as prolines, were omitted from the calculations. This approach is similar to that previously described (57).

Limited Proteolysis Study. Each reaction mixture contained 20 μM nsP2h in assay buffer (20 mM Hepes, pH 7.5, 150 mM NaCl, 5% glycerol, and 2 mM DTT). The reactions were performed in the presence and absence of 30 μM RNA14 and ADP-AlF₄ (1.5 mM ADP, 15 mM NaF, and 7.5 mM AlCl₃) at room temperature for 0, 0.5, 2, and 12 h. The reaction was terminated by the addition of Laemmli sample buffer (100 mM Tris·HCl, pH 6.8, 4% SDS, 20% glycerol, 200 mM DTT, and 0.2% bromophenol blue), and the reaction products were analyzed by SDS/PAGE.

Fluorescence-Based Thermal Shift Assay. Each reaction contained 5 μM protein, 6 μM RNA14, and 20× SYPRO Orange (Invitrogen) in assay buffer (20 mM Hepes, pH 7.5, 150 mM NaCl, 5% glycerol, and 2 mM DTT). The reaction mixture was transferred to a High-Profile 96-well PCR plate (Bio-Rad). After the PCR plate was sealed with MicroSeal 'B' PCR plate sealing film (Bio-Rad), thermal scanning was performed from 25 °C to 95 °C at an increment rate of 1 °C per 10 s using a CFX96 Touch Real-Time PCR Detection System (Bio-Rad) with the SYBR excitation/emission wavelengths of 492/517 nm. The fluorescence intensity was recorded at 10-s intervals. The curve fitting and Tm data were analyzed using Prism 7 (GraphPad).

RNA-Binding Assay. The method previously described by Chen et al. (11) was used. The 5′-end fluorescein (FAM)-labeled RNA14 (FAM-RNA14; GenScript)

was used in the fluorescence polarization (FP) experiments. FAM-RNA14 and ADP-AlF₄ were diluted using assay buffer (25 mM Hepes, pH 7.5, 50 mM NaCl, 5 mM MgCl₂, 50 mM TCEP, pH 7.5, and 0.01% Triton X-100), and nsP2h and nsP2h-F287A were diluted in assay buffer to obtain a series of 14 concentrations. The FP experiment was performed in a 384-well plate (Greiner). A total volume of 30 μL contained 0 to 2,000 nM protein and 5 nM FAM-RNA14. The reaction was performed in the presence or absence of ADP-AlF₄ (1.5 mM ADP, 15 mM NaF, and 7.5 mM AlCl₃) and equilibrated at room temperature for 20 min. The FP was measured with excitation and emission wavelengths of 485 nm and 510 nm, respectively, using a Synergy H1 plate reader (BioTek). The FP values are reported in millipolarization units and were analyzed using nonlinear regression with Prism 7 (GraphPad).

NADH Coupled ATPase Activity Assay. The method previously described by Luo et al. (58) was used. The assay was performed in a 96-well half-area clear flat-bottom microplate (Corning). Each reaction contained 2.5 nM protein of interest, 1 μM RNA14, and coupled assay mix (1 mM NADH, 100 U/mL lactic dehydrogenase, 500 U/mL pyruvate kinase, and 2.5 mM phosphoenolpyruvic acid) which was diluted in ATPase assay buffer (25 mM MOPS, pH 7.4, 150 mM KCl, 10 mM MgCl₂, 2 mM DTT, and 0.01% Triton X-100). The reaction mixture was incubated for 1 to 2 h at room temperature before the addition of ATP. For determination of the kinetic constants, all of the components were fixed, with the exception of the amount of ATP, which was varied. The activity of the proteins was assayed spectrophotometrically at 340 nM using a Tecan Spark 10M microplate reader (Tecan).

Virus Rescue and Analysis. Virus rescue in BHK-21 cells was performed as previously described (59). An ICA was performed as previously described (45) with the exception that the cells were transfected with 5 μg of endotoxin-free plasmids CMV-ICRES1 and its mutant containing the Y161A, F164A, Y161A+F164A, K192A, Q283A, F287A, W450A, F287V, F287A+V410I, and F287A+T358S substitutions. The virus stocks of WT CHIKV and the mutant CHIKV variants were collected at 24 h posttransfection and at 48 h posttransfection, respectively. The obtained stocks were clarified by centrifugation at 3,000 × g for 10 min, and the virus titers were determined by plaque titration on BHK-21 cells. All of the assays were repeated three times. The presence of reversions or secondary compensatory changes was verified as previously described (40).

Analysis of CHIKV Replicase Activity in Cell Culture. The CHIKV $trans$-replicase assay was performed as described previously (38). Briefly, U2OS cells grown on 12-well plates were cotransfected with 1 μg of pol I-Fluc-Gluc (or its mutants) and 1 μg of CMV-P1234 (or its mutant variants) using Lipofectamine LTX (Thermo Fisher Scientific) reagent according to the manufacturer's instructions. The transfected cells were incubated at 37 °C for 18 h. Subsequently, 4 × 10⁵ transfected cells were lysed in 250 μL of lysis buffer (Promega), and 4 μL of lysate corresponding to 6,400 transfected cells was used for the measurement of Fluc and Gluc activities using the Dual-Luciferase-Reporter assay and a GloMax SIS luminometer (Promega). The control cells were transfected with pol I-Fluc-Gluc plasmid and CMV-P1234ᴳᴬᴬ control plasmid that encodes an inactive CHIKV replicase (due to a mutation in the nsP4 catalytic site). All Fluc and Gluc activities measured from the experimental cells were normalized to these obtained for the control cells to reveal whether and to what extent the WT or mutant replicases were able to activate the expression of the reporter, which served as proxy for positive-strand RNA synthesis by the analyzed replicases. All of the assays were repeated at least three times.

Western Blotting. Cells transfected with WT icDNA or icDNA plasmids harboring the Y161A, F164A, Y161A+F164A, K192A, Q283A, F287A, or W450A substitutions were collected at the same time as the corresponding virus stocks. Briefly, the infected cells were washed with 1 mL of PBS, lysed with Laemmli sample buffer, and then boiled for 10 min. Lysate corresponding to 50,000 transfected cells was loaded into a well of a 10% polyacrylamide gel. The proteins were separated by SDS/PAGE, transferred to polyvinylidene difluoride membranes, and detected using antibodies against CHIKV capsid protein (in-house), and the detection of β-actin (sc-47778; Santa Cruz Biotechnology) was used as a loading control. The membranes were then incubated with appropriate secondary antibodies conjugated to fluorescent labels (LI-COR), and the proteins were visualized using the LI-COR Odyssey Fc imaging system.

1. Couturier E, et al. (2012) Impaired quality of life after chikungunya virus infection: A 2-year follow-up study. *Rheumatology (Oxford)* 51:1315–1322.
2. Schilte C, et al. (2013) Chikungunya virus-associated long-term arthralgia: A 36-month prospective longitudinal study. *PLoS Negl Trop Dis* 7:e2137.
3. Burt FJ, Rolph MS, Rulli NE, Mahalingam S, Heise MT (2012) Chikungunya: A re-emerging virus. *Lancet* 379:662–671.
4. Queyriaux B, et al. (2008) Clinical burden of chikungunya virus infection. *Lancet Infect Dis* 8:2–3.
5. Ahola T, Laakkonen P, Vihinen H, Kääriäinen L (1997) Critical residues of Semliki forest virus RNA capping enzyme involved in methyltransferase and guanylyltransferase-like activities. *J Virol* 71:392–397.
6. Spuul P, et al. (2007) Role of the amphipathic peptide of Semliki forest virus replicase protein nsP1 in membrane association and virus replication. *J Virol* 81:872–883.
7. Karpe YA, Aher PP, Lole KS (2011) NTPase and 5′-RNA triphosphatase activities of Chikungunya virus nsP2 protein. *PLoS One* 6:e22336.
8. Das PK, Merits A, Lulla A (2014) Functional cross-talk between distant domains of chikungunya virus non-structural protein 2 is decisive for its RNA-modulating activity. *J Biol Chem* 289:5635–5653.
9. LaStarza MW, Lemm JA, Rice CM (1994) Genetic analysis of the nsP3 region of Sindbis virus: Evidence for roles in minus-strand and subgenomic RNA synthesis. *J Virol* 68:5781–5791.
10. Abraham R, et al. (2018) ADP-ribosyl-binding and hydrolase activities of the alphavirus nsP3 macrodomain are critical for initiation of virus replication. *Proc Natl Acad Sci USA* 115:E10457–E10466.
11. Chen MW, et al. (2017) Chikungunya virus nsP4 RNA-dependent RNA polymerase core domain displays detergent-sensitive primer extension and terminal adenylyltransferase activities. *Antiviral Res* 143:38–47.
12. Tomar S, Hardy RW, Smith JL, Kuhn RJ (2006) Catalytic core of alphavirus nonstructural protein nsP4 possesses terminal adenylyltransferase activity. *J Virol* 80:9962–9969.
13. Schwartz O, Albert ML (2010) Biology and pathogenesis of chikungunya virus. *Nat Rev Microbiol* 8:491–500.
14. Strauss JH, Strauss EG (1994) The alphaviruses: Gene expression, replication, and evolution. *Microbiol Rev* 58:491–562.
15. Frolova EI, Gorchakov R, Pereboeva L, Atasheva S, Frolov I (2010) Functional Sindbis virus replicative complexes are formed at the plasma membrane. *J Virol* 84:11679–11695.
16. Fros JJ, et al. (2010) Chikungunya virus nonstructural protein 2 inhibits type I/II interferon-stimulated JAK-STAT signaling. *J Virol* 84:10877–10887.
17. Akhrymuk I, Kulemzin SV, Frolova EI (2012) Evasion of the innate immune response: The Old World alphavirus nsP2 protein induces rapid degradation of Rpb1, a catalytic subunit of RNA polymerase II. *J Virol* 86:7180–7191.
18. Akhrymuk I, Frolov I, Frolova EI (2018) Sindbis virus infection causes cell death by nsP2-induced transcriptional shutoff or by nsP3-dependent translational shutoff. *J Virol* 92:e01388-18.
19. Rausalu K, et al. (2016) Chikungunya virus infectivity, RNA replication and non-structural polyprotein processing depend on the nsP2 protease's active site cysteine residue. *Sci Rep* 6:37124.
20. Russo AT, White MA, Watowich SJ (2006) The crystal structure of the Venezuelan equine encephalitis alphavirus nsP2 protease. *Structure* 14:1449–1458.
21. Narwal M, et al. (2018) Crystal structure of chikungunya virus nsP2 cysteine protease reveals a putative flexible loop blocking its active site. *Int J Biol Macromol* 116:451–462.
22. Shin G, et al. (2012) Structural and functional insights into alphavirus polyprotein processing and pathogenesis. *Proc Natl Acad Sci USA* 109:16534–16539.
23. Gorbalenya AE, Koonin EV (1993) Helicases: Amino acid sequence comparisons and structure-function relationships. *Curr Opin Struct Biol* 3:419–429.
24. Singleton MR, Dillingham MS, Wigley DB (2007) Structure and mechanism of helicases and nucleic acid translocases. *Annu Rev Biochem* 76:23–50.
25. Lee JY, Yang W (2006) UvrD helicase unwinds DNA one base pair at a time by a two-part power stroke. *Cell* 127:1349–1360.
26. Velankar SS, Soultanas P, Dillingham MS, Subramanya HS, Wigley DB (1999) Crystal structures of complexes of PcrA DNA helicase with a DNA substrate indicate an inchworm mechanism. *Cell* 97:75–84.
27. Cheng Z, Muhlrad D, Lim MK, Parker R, Song H (2007) Structural and functional insights into the human Upf1 helicase core. *EMBO J* 26:253–264.
28. Chakrabarti S, et al. (2011) Molecular mechanisms for the RNA-dependent ATPase activity of Upf1 and its regulation by Upf2. *Mol Cell* 41:693–703.
29. Nishikiori M, et al. (2012) Crystal structure of the superfamily 1 helicase from tomato mosaic virus. *J Virol* 86:7565–7576.
30. Hao W, et al. (2017) Crystal structure of Middle East respiratory syndrome coronavirus helicase. *PLoS Pathog* 13:e1006474.
31. Deng Z, et al. (2014) Structural basis for the regulatory function of a complex zinc-binding domain in a replicative arterivirus helicase resembling a nonsense-mediated mRNA decay helicase. *Nucleic Acids Res* 42:3464–3477.
32. Koonin EV, Dolja VV, Krupovic M (2015) Origins and evolution of viruses of eukaryotes: The ultimate modularity. *Virology* 479–480:2–25.
33. Bagshaw C (2001) ATP analogues at a glance. *J Cell Sci* 114:459–460.
34. Miles RD, Gorrell A, Ferry JG (2002) Evidence for a transition state analog, MgADP-aluminum fluoride-acetate, in acetate kinase from Methanosarcina thermophila. *J Biol Chem* 277:22547–22552.
35. Myong S, Bruno MM, Pyle AM, Ha T (2007) Spring-loaded mechanism of DNA unwinding by hepatitis C virus NS3 helicase. *Science* 317:513–516.
36. Gu M, Rice CM (2010) Three conformational snapshots of the hepatitis C virus NS3 helicase reveal a ratchet translocation mechanism. *Proc Natl Acad Sci USA* 107:521–528.
37. Rikkonen M, Peränen J, Kääriäinen L (1994) ATPase and GTPase activities associated with Semliki forest virus nonstructural protein nsP2. *J Virol* 68:5804–5810.
38. Utt A, et al. (2016) Versatile trans-replication systems for chikungunya virus allow functional analysis and tagging of every replicase protein. *PLoS One* 11:e0151616.
39. Holm L, Rosenstrom P (2010) Dali server: Conservation mapping in 3D. *Nucleic Acids Res* 38:W545–W549.
40. Lulla V, et al. (2018) Timeliness of proteolytic events is prerequisite for efficient functioning of the alphaviral replicase. *J Virol* 92:e00151-18.
41. Atasheva S, Gorchakov R, English R, Frolov I, Frolova E (2007) Development of Sindbis viruses encoding nsP2/GFP chimeric proteins and their application for studying nsP2 functioning. *J Virol* 81:5046–5057.
42. Fairman-Williams ME, Guenther U-P, Jankowsky E (2010) SF1 and SF2 helicases: Family matters. *Curr Opin Struct Biol* 20:313–324.
43. Garmashova N, et al. (2007) The old world and new world alphaviruses use different virus-specific proteins for induction of transcriptional shutoff. *J Virol* 81:2472–2484.
44. Akhrymuk I, Lukash T, Frolov I, Frolova EI (2019) Novel mutations in nsP2 abolish chikungunya virus-induced transcriptional shutoff and make the virus less cytopathic without affecting its replication rates. *J Virol* 93:e02062-18.
45. Utt A, et al. (2015) Mutations conferring a non-cytotoxic phenotype on chikungunya virus replicons compromise enzymatic properties of non-structural protein 2. *J virol* 89:3145–3162.
46. Kim DY, et al. (2011) Design of chimeric alphaviruses with a programmed, attenuated, cell type-restricted phenotype. *J Virol* 85:4363–4376.
47. Frolov I, et al. (1999) Selection of RNA replicons capable of persistent noncytopathic replication in mammalian cells. *J Virol* 73:3854–3865.
48. Hallengärd D, et al. (2014) Novel attenuated Chikungunya vaccine candidates elicit protective immunity in C57BL/6 mice. *J Virol* 88:2858–2866.
49. Phoo WW, et al. (2016) Structure of the NS2B-NS3 protease from Zika virus after self-cleavage. *Nat Commun* 7:13410.
50. Otwinowski Z, Minor W (1997) [20] Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol* 276:307–326.
51. Kabsch W (2006) Integration, scaling, space-group assignment and post refinement. *International Tables for Crystallography Volume F: Crystallography of Biological Macromolecules* (Springer, New York), pp 218–225.
52. Hendrickson WA, Horton JR, LeMaster DM (1990) Selenomethionyl proteins produced for analysis by multiwavelength anomalous diffraction (MAD): A vehicle for direct determination of three-dimensional structure. *EMBO J* 9:1665–1672.
53. Adams PD, et al. (2010) PHENIX: A comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr* 66:213–221.
54. Emsley P, Cowtan K (2004) Coot: Model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* 60:2126–2132.
55. Zheng J, et al. (2017) HDX reveals the conformational dynamics of DNA sequence specific VDR co-activator interactions. *Nat Commun* 8:923.
56. Pascal BD, et al. (2012) HDX workbench: Software for the analysis of H/D exchange MS data. *J Am Soc Mass Spectrom* 23:1512–1521.
57. Keppel TR, Weis DD (2015) Mapping residual structure in intrinsically disordered proteins at residue resolution using millisecond hydrogen/deuterium exchange and residue averaging. *J Am Soc Mass Spectrom* 26:547–554.
58. Luo D, et al. (2011) Structural insights into RNA recognition by RIG-I. *Cell* 147:409–422.
59. Ülper L, Sarand I, Rausalu K, Merits A (2008) Construction, properties, and potential application of infectious plasmids containing Semliki forest virus full-length cDNA with an inserted intron. *J Virol Methods* 148:265–270.

MICROBIOLOGY

BIOPHYSICS AND COMPUTATIONAL BIOLOGY