

ARTICLE

Open Access

# Structure of infective Getah virus at 2.8 Å resolution determined by cryo-electron microscopy

Aojie Wang<sup>1</sup>, Feng Zhou<sup>2,3</sup>, Congcong Liu<sup>1,4</sup>, Dongsheng Gao<sup>2</sup>, Ruxi Qi<sup>1</sup>, Yiheng Yin<sup>5</sup>, Sheng Liu<sup>1</sup>, Yuanzhu Gao<sup>1</sup>, Lutang Fu<sup>1</sup>, Yinhe Xia<sup>2</sup>, Yawei Xu<sup>5</sup>, Chuanqing Wang<sup>2</sup>✉ and Zheng Liu<sup>1,5</sup>✉

## Abstract

Getah virus (GETV), a member of the genus *alphavirus*, is a mosquito-borne pathogen that can cause pyrexia and reproductive losses in animals. Although antibodies to GETV have been found in over 10% of healthy people, there are no reports of clinical symptoms associated with GETV. The biological and pathological properties of GETV are largely unknown and antiviral or vaccine treatments against GETV are still unavailable due to a lack of knowledge of the structure of the GETV virion. Here, we present the structure of infective GETV at a resolution of 2.8 Å with the atomic models of the capsid protein and the envelope glycoproteins E1 and E2. We have identified numerous glycosylation and S-acylation sites in E1 and E2. The surface-exposed glycans indicate a possible impact on viral immune evasion and host cell invasion. The S-acylation sites might be involved in stabilizing the transmembrane assembly of E1 and E2. In addition, a cholesterol and a phospholipid molecule are observed in a transmembrane hydrophobic pocket, together with two more cholesterols surrounding the pocket. The cholesterol and phospholipid stabilize the hydrophobic pocket in the viral envelope membrane. The structural information will assist structure-based antiviral and vaccine screening, design, and optimization.

## Introduction

Getah virus (GETV) is a mosquito-borne arbovirus, and belongs to the Semliki Forest group of the *Alphavirus* genus within the *Togaviridae* family<sup>1</sup>. Alongside GETV, members in the Semliki Forest Group include Chikungunya virus (CHIKV), Semliki Forest virus (SFV), Mayaro virus (MAYV), Una virus (UNAV), Bebaru virus (BEBV), and O'nyong-nyong virus (ONNV). Among them, CHIKV, ONNV, and MAYV have been reported to cause severe and mortally dangerous infectious diseases in human<sup>2–4</sup>.

GETV was first isolated in Malaysia in 1955 from *Culex gelidus* mosquitoes<sup>5</sup>, and was found to have a worldwide

distribution. The primary hosts of GETV include pigs, cattle, foxes, and horses. The first outbreak of GETV among racehorses occurred in Japan in 1978 and outbreaks have re-emerged several times since then in Japan and India<sup>6,7</sup>. The infected horses exhibited pyrexia, urticarial rash on various portions of the body, and edema of the hind legs<sup>8</sup>. GETV has been isolated from newborn piglets that died of acute disease and dead fetuses removed from infected sows in Japan<sup>9,10</sup>. The emergence of GETV in China was found in 2017 in swine farms, resulting in the death of the newborn piglets 5–10 days after birth and in pregnant sows exhibiting stillbirths or fetal mummies<sup>11</sup>. GETV was also reported to infect beef cattle<sup>12</sup>, blue foxes<sup>13</sup>, and wild boars<sup>14</sup>. Neutralizing antibodies of GETV have been detected in various vertebrate species, suggesting that domesticated animals can act as reservoir hosts<sup>15</sup>. Serological investigation of GETV revealed that a widespread infection of GETV occurred in farms across China, and GETV was responsible for significant economic damage<sup>11</sup>. Although the pathogenicity

Correspondence: Chuanqing Wang (wchuanq@163.com) or Zheng Liu (liuz3@sustech.edu.cn)

<sup>1</sup>Cryo-electron Microscopy Center, Southern University of Science and Technology, Shenzhen, Guangdong, China

<sup>2</sup>College of Animal Science and Veterinary Medicine, Henan Agricultural University, Zhengzhou, Henan, China

Full list of author information is available at the end of the article

These authors contributed equally: Aojie Wang, Feng Zhou, Congcong Liu

© The Author(s) 2022



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

of GETV in humans has not yet been identified, seroepidemiologic investigations have shown that some individuals with a febrile illness have significantly higher GETV antibody titers than in healthy people<sup>16</sup>, suggesting an association of GETV with human diseases. Currently, there is no specific antiviral treatment for animals infected with GETV. Furthermore, the physiological, pathogenic, and epidemic properties of GETV are largely unknown. Considering lessons learned from the ongoing global pandemic of coronavirus disease 2019 (COVID-19) that resulted from the outbreak of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), GETV presents a potential risk of becoming a zoonosis. Thus, we should plan and prepare for the possibility of GETV transmission to humans.

Like a typical alphavirus, GETV is a lipid-enveloped, positive-sense single-stranded RNA virus<sup>1</sup>. Mature virions of alphaviruses are spherical particles with a diameter of ~70 nm. The 11-kb genome of GETV encodes two polyproteins. Among them, one polyprotein consists of four non-structural proteins (nsP1–nsP4), and another polyprotein is composed of five structural proteins, including the capsid, E3, E2, 6K, and E1 in order from the N-terminal to the C-terminal<sup>17</sup>. The structures of alphaviruses have been well studied by using cryo-electron microscopy (cryo-EM). These include the structures of Barmah Forest virus (BFV), Eastern equine encephalitis viruses (EEEV), Western equine encephalitis viruses (WEEV), Venezuelan equine encephalitis virus (VEEV), Sindbis virus (SINV), CHIKV, and MAYV, ranging from 3.5 to 13 Å in resolution; there are also structures of CHIKV with its receptor MXRA8<sup>18–26</sup>. These structures reveal a typical architecture for alphavirus organization: viral RNA occupies the center of the particle and extends to ~140 Å radially, and the RNA is surrounded by the capsid protein shell ranging between 140 and 200 Å radially; the lipid membrane shell (200–255 Å radially) separates the capsid and the outer glycoprotein shell, and the outer shell and spikes protruding outward (255–350 Å radially) are formed by the E1 and E2 glycoproteins. In addition, high-resolution maps have assisted the construction of atomic models for the capsid and the E1 and E2 glycoproteins<sup>20,22,24,25</sup>. Moreover, Mxra8, the receptor of CHIKV, was found to bind into a cleft created by two adjacent CHIKV E1–E2 heterodimers<sup>26</sup>.

In a previous study, we have isolated a new GETV strain, the V1 strain, from pregnant sows that had abortions. We have demonstrated that the V1 strain of GETV had a strong cytopathic effect<sup>27</sup>. In this study, we utilized mouse models to investigate the infectiousness and pathogenicity of mature GETV. To gain structural insight into GETV, we determined the structure of the mature GETV-V1 virion at 2.8 Å resolution by cryo-EM, the highest resolution for an alphavirus to date.

## Results

### GETV is a mosquito-borne arbovirus

GETV appears to be maintained in a natural cycle between mosquitoes and various vertebrate hosts<sup>28</sup>. There is also suspicion that GETV can be transmitted directly between horses, likely via aerosols or direct contact<sup>8</sup>. There is no clear evidence concerning how GETV is transmitted by mosquitoes. Therefore, we first investigated the transmission and infectiousness of GETV using a mouse model (Supplementary Fig. S1). As shown in Supplementary Fig. S1a, mice (2-day old) were inoculated oronasally with 10 µL ( $10^6$  TCID<sub>50</sub>/mL, 50% tissue culture infectious dose) of the GETV-V1 strain, and were then housed in a cage with their GETV-free littermates. Total RNA was extracted from tissues, including spleen, lung, cerebral cortex, and various lymph nodes at 5 days post-inoculation (DPI), and subjected to RT-PCR for GETV detection (Supplementary Fig. S1g, h). The uninfected mice remained negative for GETV (Supplementary Fig. S1a). When male adult inoculated mice (2-month old) were housed with their uninfected male littermates (Supplementary Fig. S1b), all mice tested GETV-positive at 5 DPI. Multiple wounds were found in various parts of their body, indicating that the transmission of GETV occurred through scratching and biting.

We also conducted experiments based on a special device we designed to connect two cages via a pipe. We placed screens at the two ends of the pipe to separate the mice and to allow mosquitoes to fly between the cages freely (Supplementary Fig. S1c). In a control test shown in Supplementary Fig. S1d, GETV-free mice in the two cages lived with mosquitoes and tested negative after 5 days, indicating that the mosquitoes used in our study were GETV-free. Mice were oronasally inoculated with GETV, and then placed with their noninoculated littermates in the left-hand cage. Additional GETV-free littermates were placed in the right-hand cage, and mosquitoes were introduced into the device (Supplementary Fig. S1e). At 5 DPI, all of the mice in the left-hand cage were GETV-positive, and 15% of the mice in the right-hand cage were also positive. In another control test, no mosquitoes were introduced, all of the noninoculated mice in the right-hand cage remained negative after five days (Supplementary Fig. S1f). These results confirmed that GETV is a mosquito-borne arbovirus, not an airborne virus. GETV was mainly spread by mosquito biting, traveling in the blood between mice and mosquitos. GETV can also transmit between mouses, most likely through direct contact with blood and body fluids.

### Pathogenicity of the GETV-V1 strain

GETV has been reported to be associated with reproductive losses in pregnant sows<sup>11</sup>. In the present study, we investigated the pathogenicity of GETV in pregnant mice.

Mice were inoculated oronasally with 100  $\mu$ L ( $10^6$  TCID<sub>50</sub>/mL) of the GETV-V1 strain in early- (embryonic day 6, E6), middle- (E10), and late-gestation (E14). The inoculated E6 group mice displayed severe abortion (Supplementary Fig. S2a) and delivered fetal mummies (Supplementary Fig. S2b, f). The E10 group delivered mainly mummies or stillbirths but also delivered a small proportion of healthy pups (Supplementary Fig. S2a–f). The E14 mice delivered a mixture of mummies, stillbirths, and live infants, although some newborn mice were weak (Supplementary Fig. S2a–f). In contrast, no birth defects were detected in the mice inoculated with DMEM at E6 (Supplementary Fig. S2e, f). Surprisingly, although GETV-infected pregnant mice displayed serious reproductive losses, no other disease symptom was discovered; both miscarriage and maternity mice appeared normal. These results suggested that placental and fetal infection occurs when pregnant mice are infected with GETV, eventually causing fetal damage, such as abortion, mummies, or stillbirths.

Next, we investigated the pathogenicity of GETV in newborn mice. 3-day-old mice were inoculated oronasally with 10  $\mu$ L ( $10^6$  TCID<sub>50</sub>/mL) of the GETV-V1 strain, and clinical signs were observed daily. At 4 DPI, the inoculated mice displayed mobility impairments in pelvic limbs (Fig. 1a and Supplementary Video S1). By 8 DPI, all inoculated mice displayed paralysis, mainly in pelvic limbs. Nearly 20% of inoculated mice died at 6 DPI (2 days after their paralysis manifestation), and all of the inoculated mice died by 12 DPI (Fig. 1b). On the contrary, no mice exhibited paralysis or death in the control group (Fig. 1a, b). Immunohistochemical staining revealed that abundant GETV antigens were distributed in the hippocampal dentate gyrus of the brain and in the spinal cord under lumbar vertebrae in the inoculated newborn mice (Fig. 1c). Interestingly, no such clinical signs were observed when 7-day-old newborn and 2-month-old adult mice were inoculated with GETV, and GETV was not detected in the brain or spinal cord (Supplementary Fig. S3). This suggests that GETV is only transmitted into the brain before the blood-brain barrier is established<sup>29</sup>, and that the mobility impairments in pelvic limbs are possibly caused by viral infection in the nervous system.

### Cryo-EM structure of GETV

To gain structural insight into GETV, we determined the structure of the infective GETV virion by cryo-EM. Virion particles were purified and characterized by polyacrylamide gel electrophoresis (Supplementary Fig. S4), by mass spectrometry (Supplementary Figs. S5 and S6), by negative-stain electron microscopy (Supplementary Fig. S7), and cryo-EM (Supplementary Fig. S8). The whole structure of GETV was determined at 4.1 Å resolution using Relion3.1 and JSPR softwares (Supplementary

Fig. S9). To overcome the heterogeneity of the ~70 nm virus, a block-based reconstruction method was adopted (Supplementary Figs. S9 and S10)<sup>30,31</sup>. Three blocks (5/3/2-fold symmetry axis) of GETV cryo-EM maps at 2.81/2.92/2.85 Å were acquired, respectively (Supplementary Figs. S10, S11 and Table S1). The density maps of the capsid in the pentamers were sufficient to observe the bulk residues. However, the densities for the capsid hexamers were still intermittent. The quality of this map was improved upon the averaging of 15 equivalent capsid densities from the three and twofold symmetry axis blocks. After averaging, the correlation coefficient between the densities of pentamer and hexamer was 0.92 (Supplementary Fig. S12). By combining these three high-resolution densities, the overall structure of GETV was obtained (Fig. 1d, e). Similar to other alphaviruses, GETV has an icosahedral symmetry ( $T=4$ ), comprising 60 quasi-threefold symmetric trimers (Q-trimer) and 20 icosahedral threefold symmetric trimers (I-trimer)<sup>22</sup>. There are 240 capsid proteins connected to the corresponding E2 proteins. Each asymmetric unit (ASU, Fig. 1f, g) consists of four E1–E2–capsid heterotrimers (Fig. 1h), with three heterotrimers forming one unabridged Q-trimer connected to one heterotrimers from the I-trimer. Interacting regions between the GETV structural proteins are presented in Supplementary Fig. S13 and listed in Supplementary Table S2.

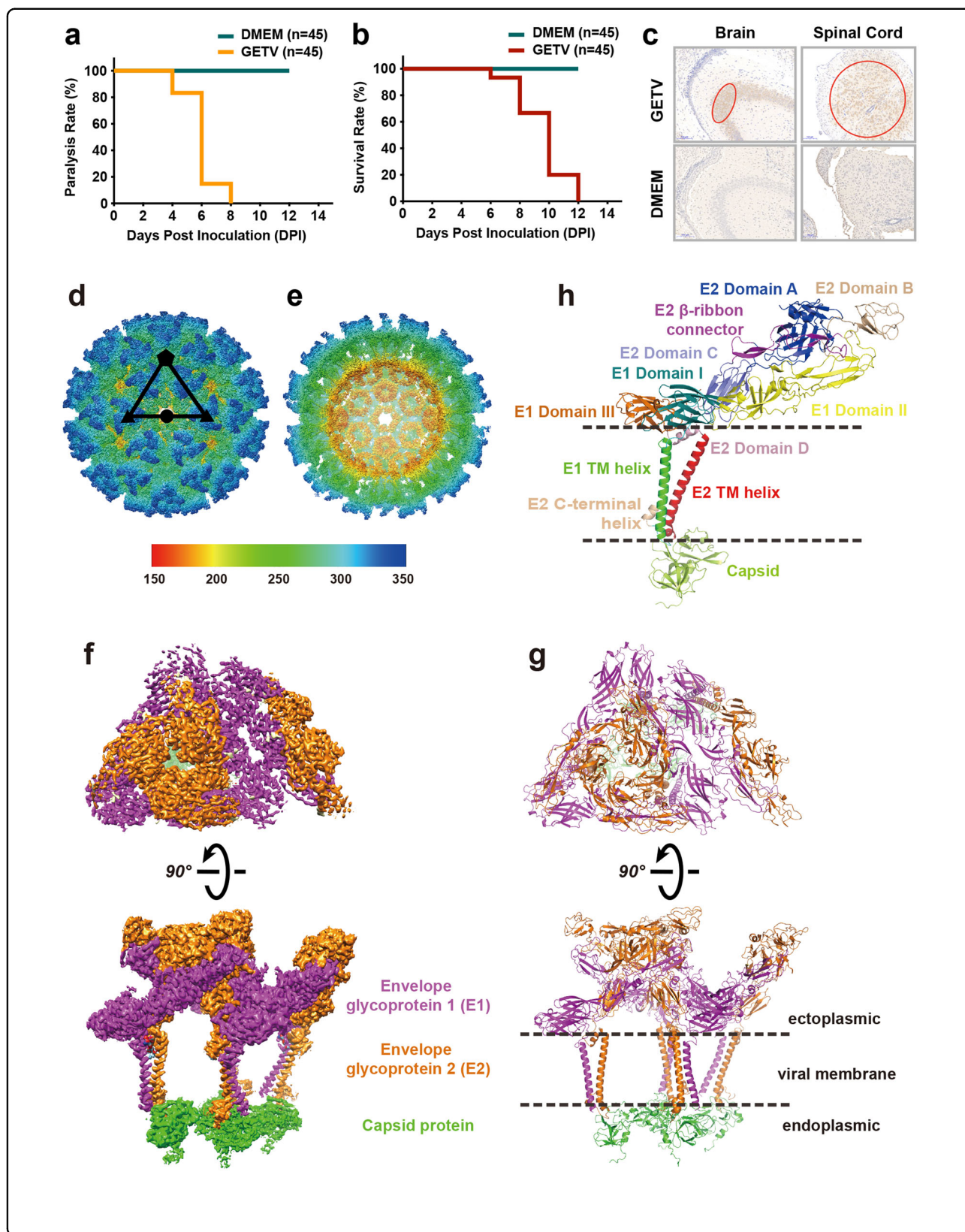
The atomic model of the E1–E2–capsid heterotrimers includes the full-length sequences of E1 and E2, as well as residues 111–268 of the capsid protein. The density for residues 1–110 of the pentamer capsid was not visible in the cryo-EM map, suggesting a flexible architecture in the core of the virion. Notably, we subjected the sequence to the recently- released AlphaFold 2<sup>32</sup>, and only flexible loop structures were predicted.

The majority of the E1 ectodomain of GETV is divided into 3 subdomains (Fig. 1h): domain I (residues 1–36, 130–164, and 274–281), domain II (residues 37–129 and 165–273; of which residues 86–96 form a hydrophobic fusion loop that inserts into the canyon formed by E2's A and B subdomains), and domain III (residues 282–379)<sup>33</sup>. Each of E1's three domains has a secondary structure of antiparallel  $\beta$  strands. The remaining ectodomain amino acids (residues 380–401) comprise the stem-loop, which connects domain III and the transmembrane (TM) helix. The E2 ectodomain comprises three subdomains (Fig. 1h): domain A (residues 1–143), domain B (173–242), and domain C (270–344)<sup>34</sup>. In addition, E2's residues 144–172 and 243–269 form the  $\beta$ -ribbon.

### Molecular interactions in the E1–E2–capsid heterotrimer

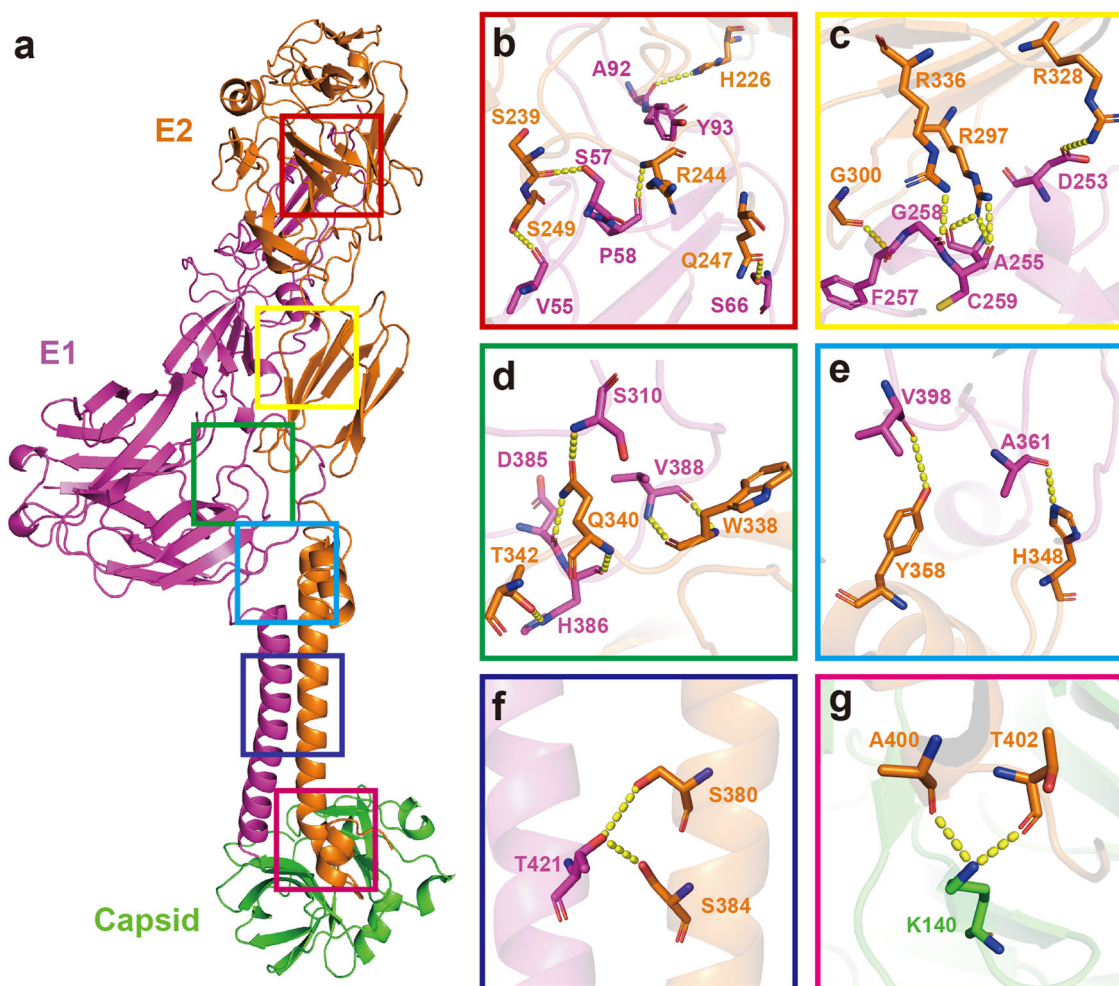
The E1, E2, and capsid proteins form the fundamental heterotrimer (Fig. 1h). There are five regions where the E1 and E2 proteins interact (Fig. 2a). In the first region, E1's





(see figure on previous page)

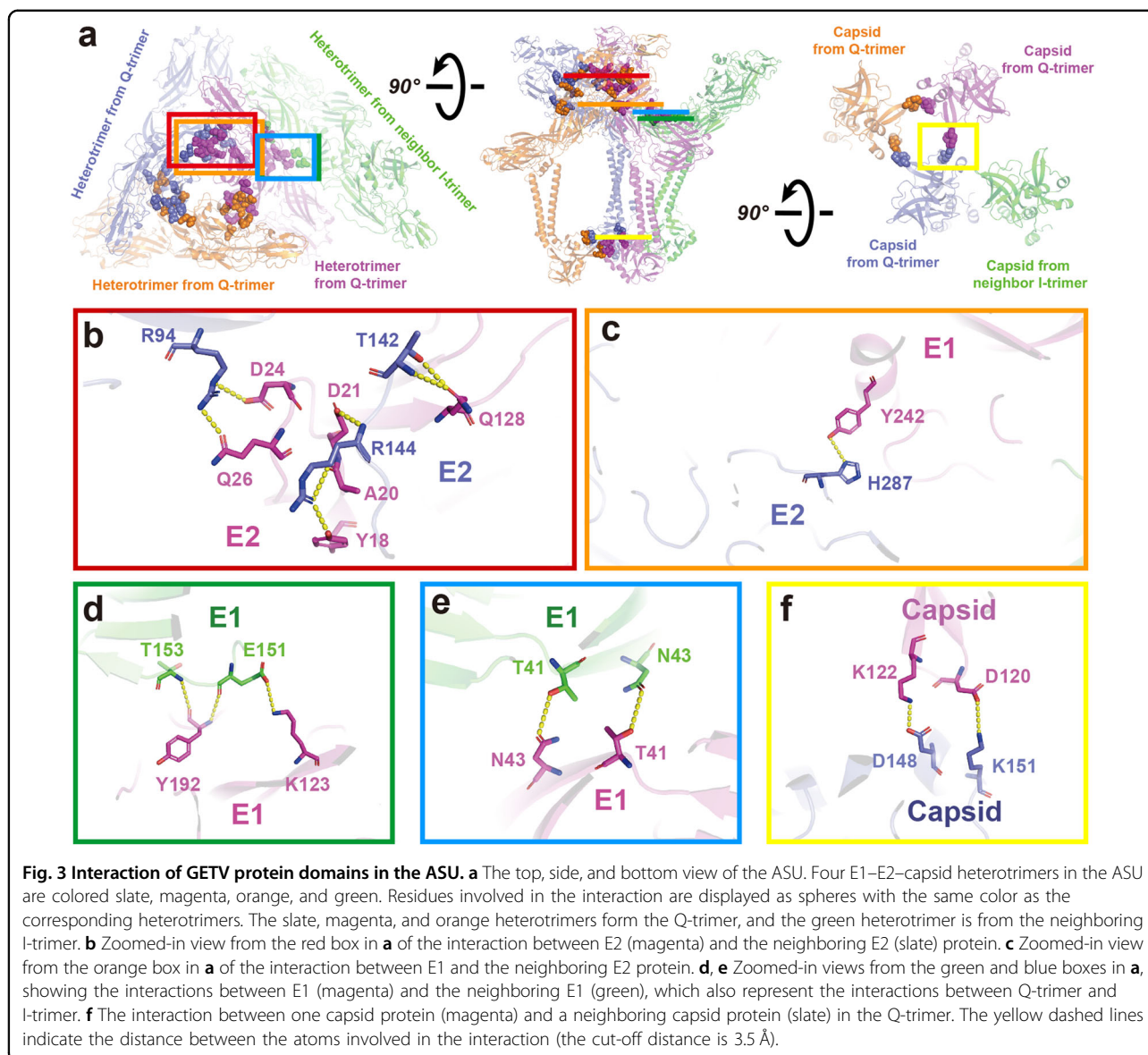
**Fig. 1 Overall structure of the infectious GETV virion determined by cryo-EM at 2.8 Å resolution.** **a–c** Ninety 3-day-old mice were randomly divided into two groups and inoculated oronasally with strain GETV-V1 suspension or an equal amount of DMEM. **a** Motor disabilities displayed in GETV-inoculated mice at 4 DPI; by 8 DPI, all of the GETV-inoculated mice displayed paralysis. **b** Mice inoculated with GETV died at 6 DPI, and all of the inoculated mice were dead by 12 DPI. **c** Tissue from hippocampal dentate gyrus in the brain and spinal cord under lumbar vertebrae were subjected for immuno-histochemistry analysis. There were abundant GETV antigens distributed in the brain and spinal cord under lumbar vertebrae in the inoculated newborn mice (highlighted in red circles). **d–h** Cryo-EM structure of GETV at 2.8 Å resolution. **d** GETV virion showing the external surface with assigned symmetry axes. **e** Central cross-section of the GETV density map. **f, g** Density map and atomic model of one asymmetric unit. The E1, E2, and capsid proteins are colored separately in magenta, orange and green. **h** Atomic model of E1–E2–capsid heterotrimer. Different subdomains of E1 (Domain I, II, III, and TM helix) and E2 (Domain A, B, C, D, TM helix, and C-terminal helix) are shown, following the previous definition for alphaviruses<sup>33,34</sup>.



**Fig. 2 Interaction of GETV protein subdomains in the E1–E2–Capsid heterotrimer.** **a** Distribution of interaction regions between E1 and E2 and between E2 and the capsid protein. **b–f** Zoomed-in views from **a** of the interaction regions between E1 and E2. **g** Zoomed-in view from **a** of the interaction region between E2 and the capsid protein. The yellow dashed lines indicate the distance between the atoms involved in the interaction (the cut-off distance is 3.5 Å).

V55, S57, P58, and S66 residues interact with the E2's residues S239, R244, Q247, and S249 (Fig. 2b). This first region is located near the tip of E1's domain II and the E2's  $\beta$ -ribbon, which is similar to that of other alphaviruses (e.g., E1 S57 with E2 H164 in SINV<sup>22</sup>; and E1 S57

with E2 H170 in CHIKV<sup>34</sup>). Other residues in this interaction region include E1's A92 and Y93 and E2's H226. These three residues are located at the interface between E1's fusion loop and the E2 domain B, similar to reports for SFV<sup>35</sup> and CHIKV<sup>34</sup>.



In the second interaction region, E1's D253, A255, F257, G258, and C259 residues interact with E2's R297, G300, R328, and R336 residues (Fig. 2c). These interactions occur near the tail of E1's domain II and E2's domain C. In the third interaction region — located at E1's stem-loop, the domain III loop and the tail loop of E2 domain C — E1's S310, D385, H386, and V388 residues interact with E2's W338, Q340, and T342 residues (Fig. 2d). Together with the second interaction region, these two areas may be required for E1/E2 dimerization. The E1 and E2 proteins pulled together through these interaction areas formed near the stem-loop of E1.

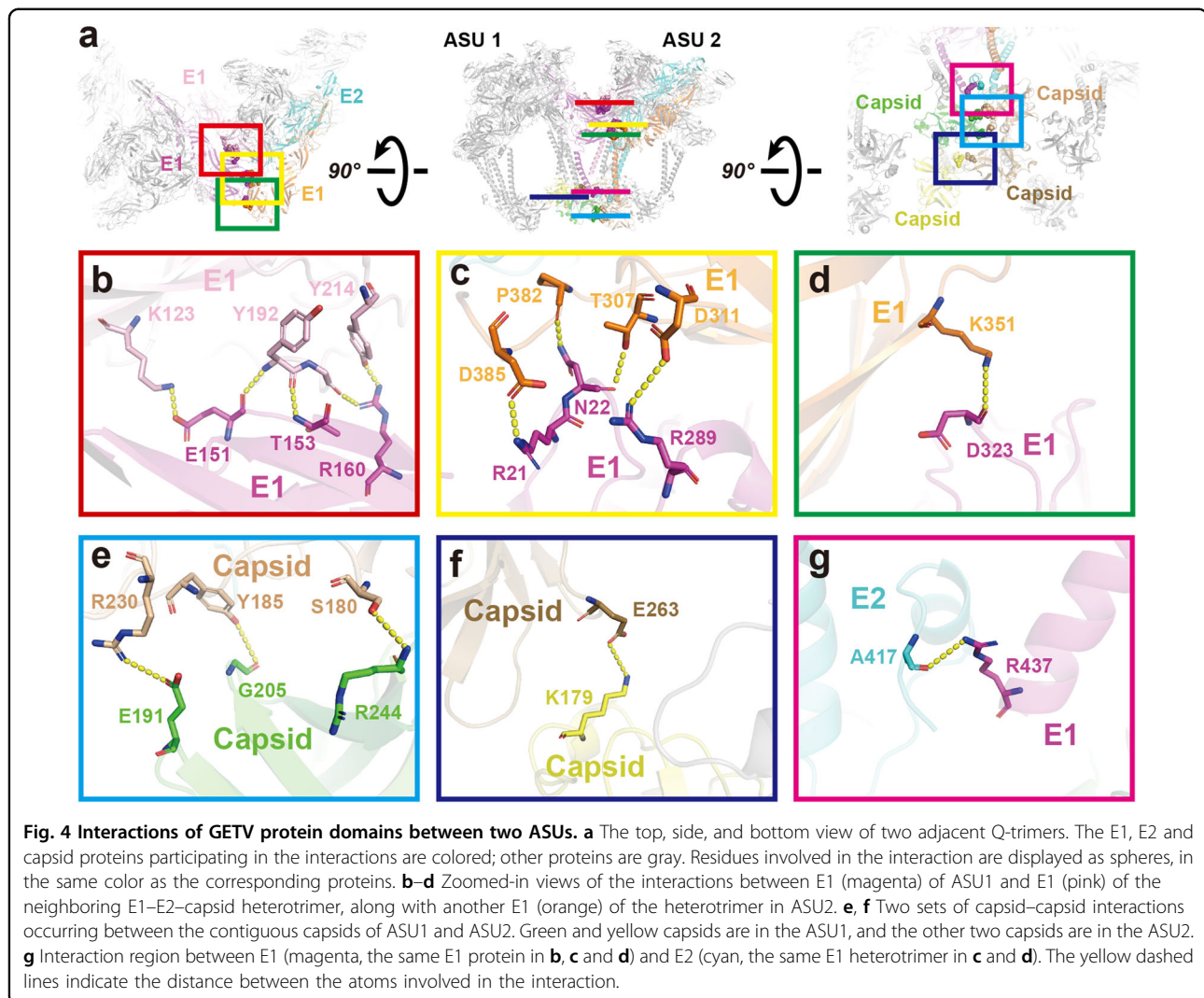
The fourth interaction region is located at E1's stem-loop, the domain II loop, and E2 domain D, near the hydrophobic pocket (Fig. 2e). E1's A361 and V398 residues interact with E2's H348 and Y358. This region is thought to be functionally important for the stability of a

hydrophobic pocket (see subsection below). The fifth interaction region involves E1's T421 and E2's S380 and S384 residues (Fig. 2f). This region is located between the E1 and E2 TM helices and the interactions stabilize these two helices. There is also an interaction region between E2 and the capsid protein that is responsible for fixing the E1–E2 heterodimer on the capsid. E2's A400, T402 residues interact with capsid protein's K140 residue (Fig. 2g). Nearly half of these residues (14/37) described above are conserved in alphaviruses (Supplementary Figs. S14–S16).

#### Intra- and Inter-ASU interactions increase virion stability

The E1–E2–capsid heterotrimer forms the Q-trimer and I-trimer. The ASU is formed by one Q-trimer and one-third of an I-trimer. Within the Q-trimer, there are three regions for inter-molecular interactions (Fig. 3a–c,





f). The E2–E2 homo-dimeric interaction occurs between the middle-region of domain A in the first E2 protein and a region of the second E2 protein comprising the joint of domains A and C; this region is positioned near by the tail of the  $\beta$ -ribbon domain (Fig. 3b). The second interaction is mediated by the helix in E1's domain II and the E2's domain C to stabilize the Q-trimer (Fig. 3c). Residues K151 and D148 in one capsid interact with D120 and K122 in the neighboring capsid (Fig. 3f). In addition, E1 of the I-trimer interact with the adjacent E1 of the Q-trimer at two regions (Fig. 3d, e). E1's T153 and E151 in domain I of the I-trimer interact with residues Y192 and K123 at the loop of E1's domain II in the Q-trimer (Fig. 3d). Residues T41 and N43 located at the  $\beta$  sheets of domain II in two E1s symmetrically interact with each other (Fig. 3e).

Six regions form contacts between two ASUs (Fig. 4a). The E1 subunit of the Q-trimer from ASU1 interacts with

an adjacent Q-trimer from AUS2 in three regions (Fig. 4b–d). The first interaction region is located between E1's domain I in ASU1 and E1's domain II in ASU2. Multiple residues, including K123, Y192, Y214 and E151, T153, and R160 are involved (Fig. 4b). The second interaction region is located between one E1's domain I (residues R21, N22, and R289) and another E1's domain III (residues D385, P382, T307, and D311) (Fig. 4c). The third interaction region is located between two E1 proteins, where residues D323 and K351 locate in the E1's domain III from two adjacent ASUs (Fig. 4d). Residues R437 and A417 are located between the tail of E1 and the C-terminal helix of E2 (Fig. 4g). Except for those interactions between the envelope glycoproteins, two sets of capsid–capsid interactions between two ASUs were also observed (Fig. 4e, f). Interestingly, most of these interacting residue pairs occur between acidic and basic residues. For instance, E151 and K123 (Fig. 4b), R21 and D385, R289 and D311 (Fig. 4c),

D323 and K351 (Fig. 4d), E191 and R230 (Fig. 4e), and K179 and E263 (Fig. 4f). Such electrostatic attractions increase the stability of protein–protein interaction.

We summarize all of the protein–protein interactions in Supplementary Table S3, including hydrogen bonds, salt bridges, and van der Waals contacts. There are a total of 47 interactions, among which 28 have been described in earlier studies, and 19 of which are first-identified in the present study. Taken together, the protein–protein interactions of the E1–E2–capsid heterotrimer, within one ASU, and between ASUs stabilize the structural assembly of the GETV, as well as other alphaviruses.

### Glycosylation sites in the E1–E2 ectodomain are surface-exposed

Viral envelope proteins have evolved to be extensively glycosylated with versatile functions ranging from immune evasion by glycan shielding to enhancement of cell infection<sup>36</sup>. Glycosylation sites have been reported in CHIKV<sup>34</sup>, EEEV<sup>24</sup>, SINV<sup>22</sup>, and MAYV<sup>25</sup>. We first predicted the *N*-glycosylation and *O*-glycosylation sites on GETV E1/E2 using NetNGlyc<sup>37</sup>, NetOGlyc<sup>38</sup>, GlycoMine<sup>39</sup>, and YinOYang<sup>37</sup>. A list of the predicted glycosylation sites is provided in Supplementary Table S4. We next utilized a mass spectrometry method to identify glycosylation sites in the GETV E1 and E2 glycoproteins (Supplementary Figs. S17 and S18), and a list of the detected sites is given in Supplementary Table S4. In the 2.8 Å resolution cryo-EM density map, we sought the densities next to the residues which were not assigned to proteins. A total of eight densities were observed: four in E1, including S66, N141, N270, and N335. Another four in E2, T155, N200, T264, and N262 (Fig. 5 and Supplementary Fig. S19). Four densities for glycans were modeled and fitted associated with N141 and N270 of E1, and N200 and N262 of E2 (Fig. 5d–g and Supplementary Fig. S19a–d). Although protruding densities adjacent to S66, T155, T264, and N335 were strong, these did not support precise fitting for glycan atomic models (Fig. 5h–k and Supplementary Fig. S19e–h).

The *N*-glycosylation of E1 N141 and E2 N262 have also been reported in the MAYV structure, for which glycan models were built<sup>25</sup>. In our GETV structure, densities associated with E1 N141 and E2 N262 were both clearly visible, allowing the fitting of atomic models for multiple hexosamine and hexose monomers. The proximal portion of the density map associated with E2 N262 accommodated three carbohydrates, which were modeled as classical *N*-glycans NAG–NAG–BMA (NAG: *N*-Acetylglucosamine and BMA:  $\beta$ -D-mannose, Fig. 5g). At the position of the third carbohydrate BMA, one branch density was observed, in which a 1-6MAN ( $\alpha$ -(1-6)-D-mannose) monomer was modeled and fitted (Fig. 5g). A similar model of NAG–NAG was fitted in the density

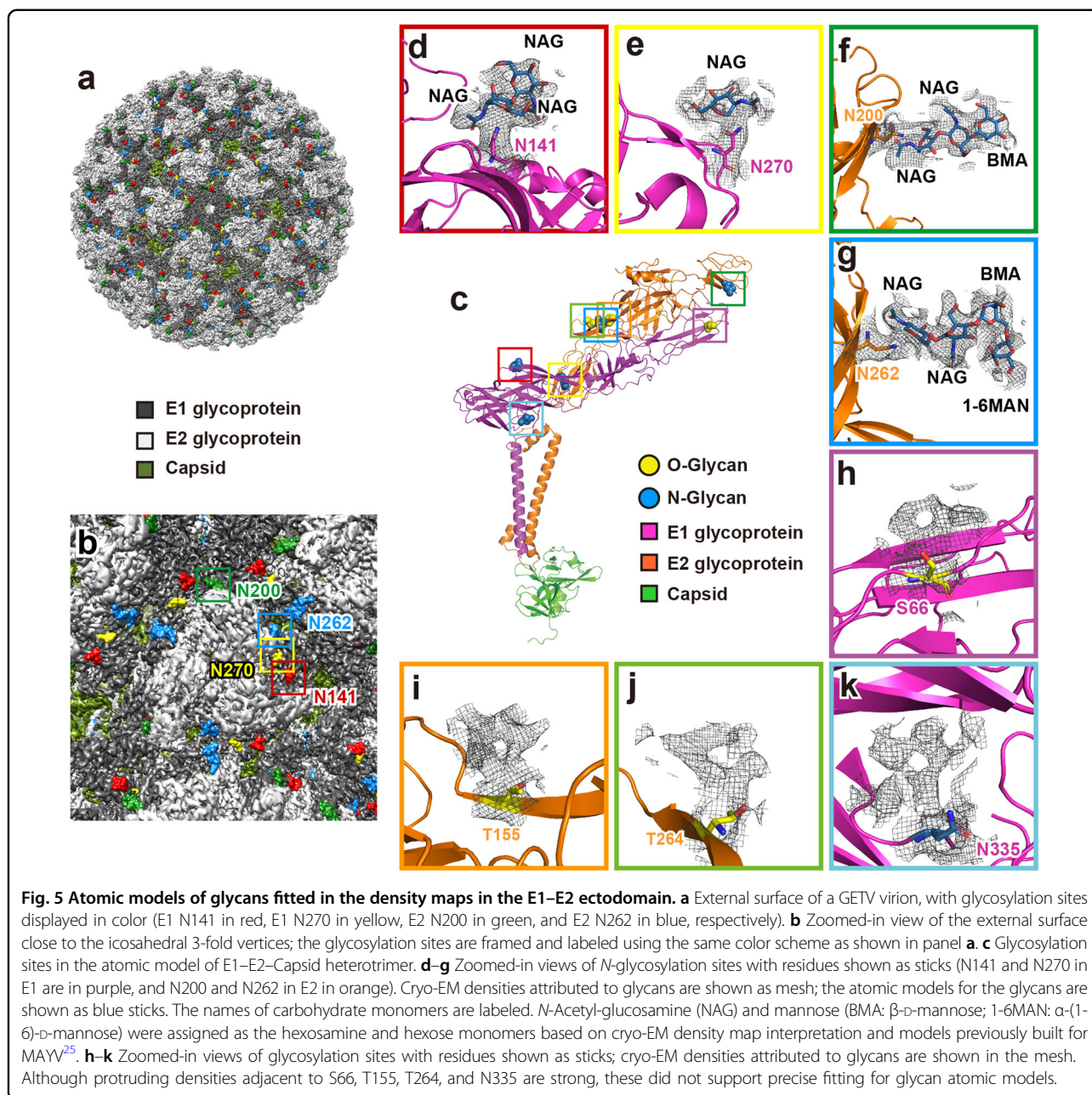
proximal to E1 N141 (Fig. 5d). E2 N200 has a density that also fitted the typical *N*-glycans, NAG–NAG–BMA (Fig. 5f). The E1 N270 has a less defined density compared to E1 N141, E2 N200 or N262 where only one NAG monomer could be modeled (Fig. 5e). All four glycan moieties were surface-exposed (Fig. 5a, b), and they were accessible on the GETV surface close to the trimer/trimer interface (Fig. 5b). This surface exposure of glycosylation sites suggests their potential roles in evading detection by the host immune system, and/or in promoting attachment to the host cells to enhance viral uptake.

### S-acylation sites in the E1 and E2 TM helices

S-acylation is a post-translational modification that attaches fatty acids, often palmitic acid, to cysteine residues; this modification occurs on both peripheral and integral membrane proteins<sup>40</sup>. These fatty acids increase a protein's hydrophobicity and also facilitate interactions with lipid bilayers, and S-acylation is known to function directly in modulating proteins' stability<sup>40</sup>. Protein acylation was originally discovered in the vesicular stomatitis virus<sup>41</sup>. Over the past 40 years, numerous viruses have been reported to undergo S-acylation, such as human immunodeficiency virus (HIV), SARS-CoV, influenza A virus, and hepatitis C virus (HCV)<sup>42</sup>. For alphaviruses, both SFV and SINV have been reported to contain S-acylation sites<sup>43–46</sup>. Using radioactivity derived from [<sup>3</sup>H]palmitic acid and [<sup>3</sup>H]stearic acid labeling, Veit et al., reported that SFV E1 contains mainly stearic acid, while E2 is acylated primarily with palmitic acid<sup>43</sup>. In a later study, Kordyukova et al. reported that SFV E1 is stoichiometrically acylated with stearic acid, whereas E2 contains palmitic acid and stearic acid at a ratio of 3:1, indicating that stearic acid was attached to only one cysteine<sup>47</sup>. In the 6 Å resolution cryo-EM structure of BFV, density was found next to residue C394 in the E2 TM helix; however, no atomic model containing a fatty acid was constructed<sup>19</sup>.

In the present study, we first predicted the possible S-acylation sites using the predictors CSS-Palm and GPS-Palm<sup>48,49</sup>. A list of predicted S-acylation sites is given in Supplementary Table S5. Next, we utilized liquid chromatography–mass spectrometry to identify S-acylation sites in the GETV E1 and E2 proteins (Supplementary Figs. S20–S22), and a list of the sites is stated in Supplementary Table S5. Finally, in the 2.8 Å resolution GETV map, we examined the densities next to the residues. A total of five such densities were identified: one at E1 C433, and four in E2 (C385, C395, C415, and C416) (Fig. 6). According to the mass spectrometry analysis, stearic acids were attached to C385 and C395 in E2, while the other three cysteines, C433 in E1, C415 and C416 in E2, were acylated with palmitic acids. Therefore, we accommodated the atomic models of stearic acid in the



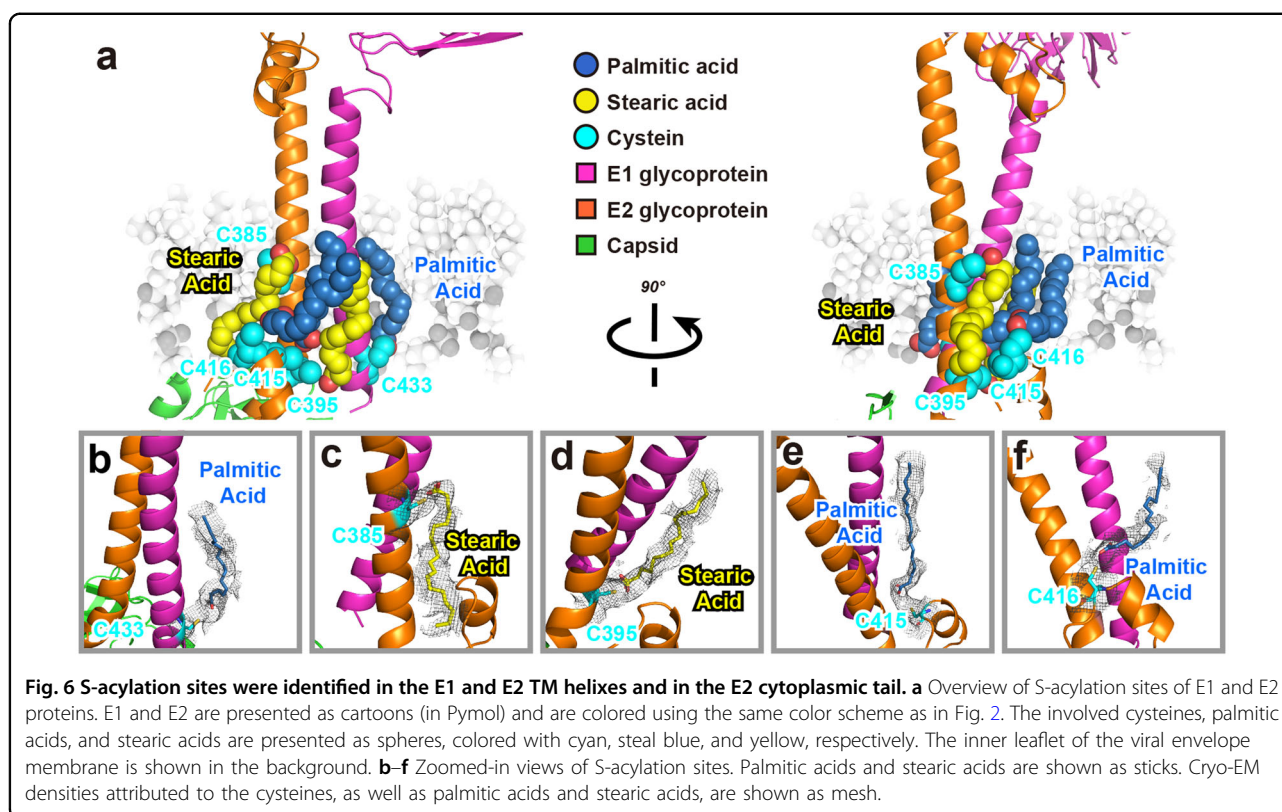


straight proximal portion of the density next to E2 C385 and C395 (Fig. 6c, d), and fitted atomic models of palmitic acid in the densities adjacent to E1 C433 and E2 C415 and C416 (Fig. 6b, e, f). Compare the primary sequence between GETV and SFV, although the five cysteines were identical, the sequences around E1 C433 (Supplementary Fig. S14) and around E2 C395 were not fully conserved (Supplementary Fig. S15), suggesting that there could be a species diversification of S-acylation between GETV and SFV. As illustrated in Fig. 6a and Supplementary Fig. S23, the five highly or absolutely conserved C-terminal S-acylation sites strongly anchored the E1 and E2 TM

helices into the inner leaflet of the lipid bilayer of the viral envelope membrane. Considering the large portion of both the E1 and E2 ectodomain (Fig. 1h), it is likely that S-acylation in the TM helices stabilizes the E1/E2 glycoproteins.

#### Cholesterol molecules and a phospholipid molecule in the hydrophobic pocket increase the stability of the E1-E2 heterodimer

Previous studies have demonstrated that a hydrophobic pocket is essential for alphavirus assembly at the outer leaflet of the envelope membrane<sup>22,25</sup>. The hydrophobic

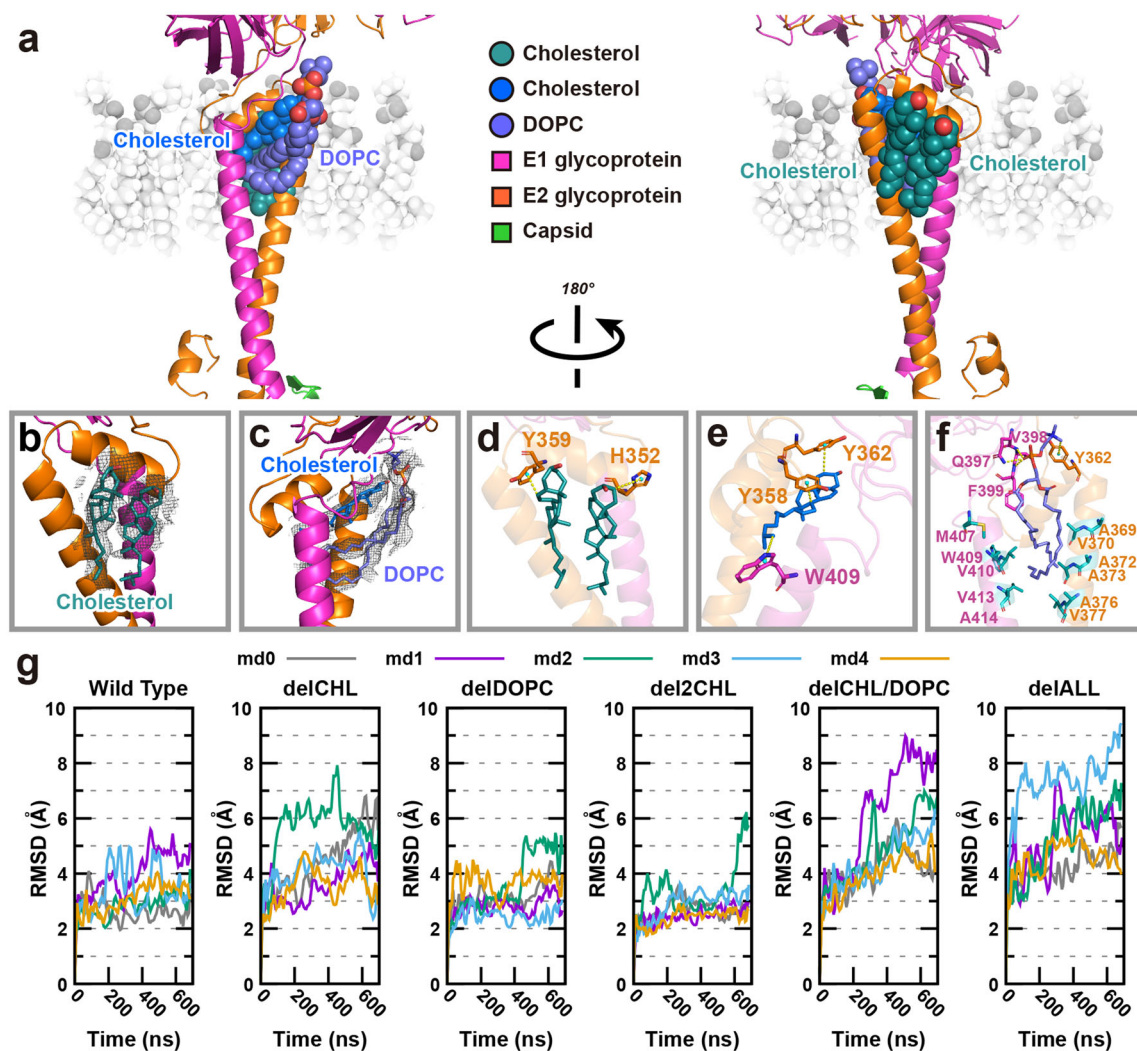


pocket comprises the TM helix from E1, and the TM helix and domain D of E2 (Fig. 1h). In the 3.5-Å-resolution SINV structure, a “pocket factor” was identified; this was predicted to be a hydrophobic phospholipid tail<sup>22</sup>. In the 4.4-Å-resolution MAYV structure, a long density was identified in the hydrophobic pocket, and a C18 hydrocarbon lipid molecule, octadecane, was fitted into the density<sup>25</sup>. Moreover, pocket factors have been characterized in other viruses, such as Zika virus<sup>50</sup> and Dengue virus<sup>51</sup>.

In our GETV structure, extra densities were clearly visible in the cavity formed by the E1 and E2 TM helices, as well as in the region immediately adjacent to the helices (Fig. 7a–c). Two cholesterol molecules were modeled and fitted into the densities adjacent to the helices (Fig. 7b), while one cholesterol was fitted in one density in the cavity (Fig. 7c). Cholesterol is the most abundant sterol in mammals, it is a major constituent of the plasma membrane and plasma lipoproteins. A previous study revealed that nearly 30% of the lipid class composition of baby hamster kidney cells (BHK, one of the most commonly used cell lines for expression of alphaviruses, and for GETV in the present study) represents cholesterol, and more than 40% of the envelope membrane of SFV particles produced in BHK cells are cholesterol<sup>52</sup>. Another density in the cavity displayed a “Y” shape, and a phospholipid molecule was modeled as “DOPC” (dioleoyl-

phosphatidylcholine, Fig. 7c), with a hydrophilic “head” containing a phosphatidylcholine, and two hydrophobic “tails” derived from oleic acid, a mono-unsaturated omega-9 fatty acid. Both densities that were fitted with two oleic acid hydrocarbon tails had a kink that can accommodate the mono-unsaturated bond (Fig. 7c).

We subsequently analyzed the interactions between E1/E2 and cholesterol/DOPC in the hydrophobic pocket. Cholesterol molecules contain four fused carbon rings referred to as A, B, C, and D rings. The principal cholesterol–protein interactions occur through CH– $\pi$  hydrogen bonds. CH– $\pi$  bonds involve a soft acid CH and a soft base  $\pi$ -electron system<sup>53</sup>. As shown in Fig. 7d, in the cholesterol at the left, the B ring interacts with the aromatic ring of Tyr-359 in subdomain D, with mean distances of 4.0 Å. The A ring in the right cholesterol contacted with the aromatic ring of His-352, with a distance of 4.4 Å. This is consistent with the establishment of the cut-off values (4.5 Å) for the distance of the donor carbon atoms to the geometric center of the  $\pi$ -acceptor system<sup>53</sup>. Figure 7e displayed the cholesterol in the hydrophobic cavity, and the A and D rings interacted with Tyr-362 and Tyr-358, with mean distances of 4.3 Å and 3.8 Å, respectively. A CH– $\pi$  hydrogen bond between a methyl group of DOPC and the aromatic ring of Tyr-362 were identified, with a distance of 4.5 Å. Moreover, hydrogen bonds between the phosphate group and Gln-



**Fig. 7** Cholesterol molecules and a phospholipid DOPC in the hydrophobic pocket. **a** Overview of the distribution of cholesterol and a phospholipid molecule DOPC. Cholesterol (blue) and DOPC (slate) are in the hydrophobic pocket formed by the E1 TM helix, the E2 TM helix, and E2 domain D. The other two cholesterol (deep teal) are positioned on the other side of domain D. E1 and E2 are presented as cartoons. The involved cholesterol and DOPC are presented as spheres, and the outer leaflet of the viral envelope membrane is shown in the background. **b, c** Zoomed-in views of cholesterol and DOPC molecules, which are shown as sticks. Cryo-EM densities attributed to the cholesterol and phospholipid are shown as mesh. **d–f** Zoomed-in views of the interactions between cholesterol or DOPC and E1 or E2. The aromatic residues positioned near the cholesterol or DOPC are shown as sticks with the same color as E1 or E2. The hydrophobic residues positioned near the tail of DOPC are shown as sticks. The center of mass of aromatic ring is shown as nb\_spheres, colored in cyan. The yellow dashed lines represent the hydrogen bonds. A portion of the hydrogen bonds is CH- $\pi$  hydrogen bonds between the soft acids CH and a soft base  $\pi$ -electron system. The remaining portion is common hydrogen bonds between the phosphate group of DOPC and Q397, V398 or F399. **g** Time evolutions from MD simulation analysis for the C $\alpha$ -RMSD of E1 (295–435) and E2 (269–422) for systems “Wild Type”, “delCHL”, “delDOPC”, “del2CHL”, “delCHL/DOPC”, and “delALL”. All curves were smoothed with the Bezier method, implemented in gnuplot.

397/Val-398/Phe-399 were confirmed. Beyond that, multiple hydrophobic residues were localized around the hydrophobic tails of DOPC. All of these interactions fixed DOPC in the hydrophobic pocket (Fig. 7f).

To further investigate how the cholesterol and DOPC molecules regulate the E1–E2 complex, we calculated the root mean square deviation of the C $\alpha$  atom (C $\alpha$ -RMSD) of E1–E2 against time in the six simulated systems using

molecular dynamics (MD) simulations (Supplementary Videos S2–S7). As shown in Fig. 7g, the RMSD evolves differently in the six systems, with the overall values from the “delCHL/DOPC” (both cholesterol and DOPC in the pocket were removed) and the “delALL” (all three cholesterol and DOPC were removed) being the most unstable. The RMSD values increased drastically higher in these two systems with the magnitude reaching as high as



9 Å, whereas those in the “delCHL” (only cholesterol in the pocket was removed) system, the “delDOPC (only DOPC in the pocket was removed) system, and the “del2CHL” (two cholesterols surrounding the pocket were removed) system being similar to or only slightly higher than that in the Wild Type system. This indicates that the structure of the E1–E2 complex changes significantly whenever the pocket cholesterol and DOPC ligands were removed together, suggesting a cooperative role between the two pocket ligands in stabilizing the E1–E2 complex. Additionally, we calculated the number of native contacts within the E1–E2 complex during the evolution. As shown in Supplementary Fig. S24, the numbers of native contacts of the “delCHL/DOPC” and “delALL” systems overall decreased to under  $4.9 \times 10^4$ , which was lower than those of the Wild Type system that decreased to  $\sim 5.0 \times 10^4$ . The systems “delCHL”, “delDOPC”, and “del2CHL” were similar to or slightly lower than those of the Wild Type system. These results show that the E1–E2 complex in the systems “delCHL/DOPC” and “delALL” lost more native contacts during simulation compared to other systems, indicating a larger structural change when the pocket cholesterol and DOPC ligands were removed together. The findings were quite similar to those presented in the C $\alpha$ -RMSD analysis, which together suggests a cooperative role between the two pocket ligands in stabilizing the E1–E2 complex.

## Discussion

Viral glycosylation has broad roles in viral pathobiology, including mediating protein folding and trafficking, facilitating viral budding and release, directing immune evasion by glycan shielding, and shaping viral infection and tropism<sup>36</sup>. For alphaviruses, glycosylation is known to be a key determinant of cellular tropism and virulence: a lack of glycosylation in Ross River virus (RRV) renders virions unable to induce alpha/beta interferon production in myeloid dendritic cells, the professional antigen-presenting cells that rapidly respond to viruses, suggesting that viral glycans may negatively regulate antiviral responses<sup>54</sup>. Glycosylation has also been demonstrated to function in alphavirus infection and replication: a lack of glycosylation of the E1 protein in SINV severely impairs viral replication<sup>55</sup>. Further, a study showed that mutations of the glycosylation sites in the Salmonid alphavirus E1 and E2 proteins severely altered the virulence and production of an infectious virus<sup>56</sup>.

Previous structural analyses have revealed glycosylation sites in alphaviruses. A crystal structure of the mature E3–E2–E1 glycoprotein complexes from CHIKV revealed three N-linked glycans: N12 in E3, N263 in E2, and N141 in E1<sup>34</sup>. The cryo-EM structure of MAYV identified two glycosylation sites: N141 in E1 and N262 in E2, with at least five carbohydrates modeled as NAG–NAG–BMA–MAN–MAN

at the E2 N262 site and with the canonical NAG–NAG–BMA glycan sequence modeled at the E1 N141 site<sup>25</sup>. Both the E1 N141 and E2 N262 sites are conserved in CHIKV, MAYV, SFV, RRV and GETV. Glycosylation sites were also identified in SINV by cryo-EM, including N139 and N245 in E1, N283 in E2, and N14 in E3, all of which displayed extra density next to an asparagine residue, although no glycan models were built<sup>22</sup>. In the present study, six new glycosylated residues — E1 S66, N270, and N335 along with E2 T155, T264, and N200 — were identified, and we built two more atomic models of the glycans (N270 and N200) precisely into the 2.8 Å cryo-EM density map (Fig. 5).

Protein S-acylation is a post-translational lipid modification wherein fatty acids, usually palmitic acid, are attached to cysteine (S-palmitoylation). There are also known examples of modifications of serine and threonine residues (O-palmitoylation)<sup>42</sup>. S-acylated proteins are typically membrane proteins, S-acylation enhances protein hydrophobicity and impacts protein subcellular localization, trafficking, stability, and interactions with other proteins<sup>57</sup>. S-acylation of viral proteins is known to be involved in virus assembly and infection. For example, the Spike protein of SARS-CoV, and quite recently, of SARS-CoV-2, have been shown to be palmitoylated; and these modifications have been associated with the cell–cell fusion process known to be essential for viral infectivity<sup>57,58</sup>. For alphaviruses, both E1 and E2 glycoproteins of SFV have been reported to contain S-acylation sites: E1 contains mainly stearic acid, while E2 is acylated primarily with palmitic acid<sup>43</sup>. Palmitoylation of cysteine residue in the E2 protein of SFV was demonstrated as an essential determinant for viral budding<sup>44</sup>. The E1 and E2 protein of SINV also have palmitoylation sites, and site-directed mutagenesis of cysteines in E1 and E2 abolished fatty acid attachment, and result in aberrant viral assembly and particle formation, slow replication at early times postinfection, and increased sensitivity to detergents<sup>45,46</sup>. Those findings supported an earlier study, which indicated that mutations in the SINV E2 glycoprotein led to defects in palmitic acid attachment, as well as defects in virus assembly and budding<sup>59</sup>. In the cryo-EM structure determination of BFV at 6 Å resolution, density was detected next to one cysteine residue in the E2 protein (C394); however, no atomic model containing a fatty acid was constructed<sup>19</sup>. In another cryo-EM study of VEEV at 4.4 Å resolution, three conserved E2 cysteine residues (C396, C416, and C417) were mapped near the lipid head groups of the inner membrane; note that no density attributable to S-acylated fatty acids was found<sup>20</sup>.

Cholesterol is an essential structural component of cell membranes and can also serve as a structural component of viral envelope membranes. The membrane of mammalian-derived SFV virions comprises more than

40% cholesterol<sup>52</sup>. Interestingly, the cholesterol content in the envelope of MAYV particles obtained from mammalian cells is 10-fold higher in cholesterol content as compared to viruses produced in mosquito cells<sup>60</sup>. A similar discrepancy in cholesterol content was found in SINV grown in mammalian vs in insect cells<sup>61</sup>. In all studies, BHK cells were used as the mammalian host cells; these are the same cells we used in the present study to propagate and purify GETV. Thus, it was not surprising that cholesterol was present in the GETV envelope membrane. It was however surprising that we detected one cholesterol molecule in the hydrophobic pocket and two cholesterol molecules surrounding the TM helices of the E1 and E2 proteins, which indicated that cholesterol directly interacts with the E1 and E2 proteins (Fig. 7 and Supplementary Fig. S25). Several studies have demonstrated the impacts of cholesterol binding to viral proteins: cholesterol directly interacts with the cytoplasmic tail of the HIV glycoprotein gp41, and cholesterol is necessary for driving lateral self-aggregation and clustering of the envelope glycoproteins, therefore facilitating HIV fusion with the host cell and consequent viral entry<sup>62</sup>. A solid-state NMR study has revealed that two cholesterol molecules bind to the influenza M2 protein. Furthermore, cholesterol-mediated M2 clustering to the budding virus induced the necessary curvature for membrane scission<sup>63</sup>. Cholesterols also interact with the glycoprotein hemagglutinin of the influenza virus, which is essential for virus replication and affects virus assembly<sup>64</sup>.

Several studies have also addressed the potential impacts of cholesterol present in viral envelope membranes, depletion of cholesterol from envelopes has been shown to abolish the infectivity of influenza virus, human herpesvirus, herpes simplex virus, hepatitis C virus, and HIV<sup>65–68</sup>. For the alphaviruses CHIKV, SFV, SINV, and MAYV, cholesterol is required for viral entry to host cells<sup>69–72</sup>. After entry, viral RNA is released and replicated; this process also depends on intracellular membranes, and cholesterol impacts on viral replication: SINV replication was faster and reached higher titer in cholesterol-rich fibroblasts, and virions produced in cholesterol-rich fibroblasts are more infectious than viral particles produced from normal human fibroblasts<sup>73</sup>. Similarly, drugs that induced intracellular cholesterol accumulation and affected cholesterol biosynthesis, conferred strong inhibition against CHIKV replication<sup>74</sup>. Notably, gene expression profiling of cells infected with the M1 strain of GETV (which was isolated from *Culex* mosquitoes and is known to infect horses and pigs<sup>75</sup>), demonstrated that more than 60% of genes associated with the cholesterol biosynthesis pathway are downregulated after GETV M1 infection<sup>76</sup>. Taken together, cholesterol is not only a molecular factor that essential for alphavirus entry, replication, budding, and exit from host

cells, but also a structural factor contributing to the stability of viral hydrophobic pocket in the envelope membrane.

## Materials and methods

### Virus infection

All protocols were approved by the Institutional Animal Care and Use Committee of Shanghai Tenth People's Hospital. Mice were healthy and had no serum neutralizing antibody against GETV before the experiment. Newborn mice (2-day old or 3-day old) were inoculated oronasally with 10  $\mu$ L ( $10^6$  TCID<sub>50</sub>/mL, 50% tissue culture infectious dose) GETV-V1 strain. Pregnant mice were inoculated oronasally with 100  $\mu$ L ( $10^6$  TCID<sub>50</sub>/mL) GETV-V1 strain in early-gestation (E6), middle-gestation (E10), and late-gestation (E14).

### Virus production and purification

Mature and infective GETV virions were handled in BSL2 facilities at Cryo-Electron Microscopy Center, Southern University of Science and Technology, and at College of Animal Veterinary Medicine, Henan Agricultural University. GETV-V1 strain (GenBank Sequence Accession: KY399029.1) was incubated in monolayers Baby Hamster Kidney Fibroblast Cells (BHK-21, ATCC CCL-10) at an approximate multiplicity of infection (MOI) of 0.1 for 30 min to allow viruses to bind and enter into the BHK cells. The supernatant was removed and replaced with a fresh culture medium (DMEM with 2% FBS and 1% PS). GETV was proliferated in 10 T175 culture flasks and harvested 36 h postinfection. The cell debris of the supernatant was removed at 10,000 $\times$  g for 1 h (Sorvall LYNX 6000 Superspeed Centrifuge). The GETV particles were concentrated by centrifuging at 80,000 $\times$  g for 1.5 h in Ultracentrifuge (Beckman Optima XPN-100) with a type 45 Ti rotor at 4 °C. After soaking in 30 mL phosphate-buffered saline (PBS, pH 7.47) for 4 h, the virus suspension was initially purified via a 20% (w/v) sucrose cushion at 80,000 $\times$  g for 1.5 h in a type SW32 Ti rotor at 4 °C. The virus pellet was gently resuspended in 2 mL PBS buffer for 4 h and loaded onto a linear 20%–50% (w/v) sucrose density gradient at 100,000 $\times$  g for 2 h in an SW41 Ti rotor at 4 °C. The light scattering band corresponding to the virus particles was collected and resuspended in PBS. The integrity and purity of the GETV particles were examined by SDS-PAGE and mass spectrometry, and by negative-stain electron microscopy.

### Cryo-EM sample preparation and data acquisition

A 5  $\mu$ L virus sample was added to the glow-discharged 400-mesh grid covered with amorphous nickel-titanium alloy film (R2/1) at 100% humidity and 6 °C, blotted with filter paper and frozen by plunging into liquid ethane using a Vitrobot Mark IV system (Thermo Fisher Scientific Inc.). Cryo-EM data were collected on

a 300 kV Titan Krios microscope (Thermo Fisher Scientific Inc.) equipped with a BioContinuum Imaging Filter (Gatan Inc.). The images were recorded on a K3 summit direct detection camera (Gatan Inc.) using SerileM software for automated image acquisition. The images were recorded at a nominal magnification of 105,000 $\times$  in super-resolution mode, yielding a calibrated pixel size of 0.83 Å. Each exposure was dose-fractionated into 32 frames leading to a total dose of 40 e<sup>-</sup>/Å<sup>2</sup>. The final defocus range of the micrographs was -0.8 to -1.5  $\mu$ m.

### Image processing

A total of 16,894 movie stacks of the GETV were collected. The beam-induced drift was corrected using MotionCor2<sup>77</sup>. The contrast transfer function (CTF) parameters were estimated using CTFFIND4<sup>78</sup>. In total, 171,059 particles were extracted using the EMAN2<sup>79</sup>. The 4 $\times$  binned particle images were subjected to reference-free 2D classification in RELION-3<sup>80</sup>. After several rounds of reference-free 2D classification, a subset of 144,254 particles was isolated for 3D classification and reconstruction. Two maps at 8.94 Å resolution and 9.54 Å resolution were obtained with 56,573 and 43,653 good particles, respectively. The coordinates were used to extract the 2 $\times$  binned particles for 3D classification using RELION-3 with I4 symmetry imposed, which resulted in two maps at 6.64 Å resolution and 6.69 Å resolution obtained from 43,775 and 62,415 good particles, respectively. The rotational and translational parameters of each particle were further refined with jalign program from JSPR, which enabled the reconstruction of a better map at 4.1 Å resolution using the j3dr program in JSPR<sup>81</sup>. To overcome the defocus gradient and the heterogeneity of the ~70 nm virus, block-based reconstruction was performed to further improve the resolution<sup>30</sup>. Three types of blocks (five trimers near the icosahedral fivefold axis, four trimers near the icosahedral-three-fold axis and two trimers near the icosahedral twofold axis) were selected, 3D classified and refined separately. Maps at 3.05 Å, 3.28 Å, and 3.14 Å, were obtained using unbinned particles. At this stage, CTF refinement was carried out to better estimate the local defocus value for each block, a final round of 3D refinement generated density maps at a resolution of 2.81 Å, 2.92 Å, and 2.85 Å, respectively, as estimated by the gold-standard Fourier shell correlation (FSC) cut-off value of 0.143 (Supplementary Figs. S9 and S10). The local resolution distribution of the final reconstruction was assessed using ResMap<sup>82</sup> (Supplementary Fig. S10).

### Model building and refinement

An asymmetric subunit of the VEEV model (PDB ID: 3J0C) was used as a template and rigid-body fitted into the

2.8 Å GETV map using UCSF Chimera<sup>83</sup>. The amino acid sequences of VEEV were then mutated to GETV and an initial 3D atomic model of E1–E2–Capsid was built using Coot<sup>84</sup>. The resulting GETV coordinates were then used as a starting point to a flexible refinement through the PHENIX<sup>85</sup>. The refinement cycle was repeated, and the quality of the final 3D atomic model of E1–E2–Capsid was evaluated using MolProbity<sup>86</sup>. The cryo-EM data collection and processing, block-based reconstruction, model building, and refinement statistics are summarized in Supplementary Table S1.

### Quantitative N-glycosylation analysis by LC-MS/MS

E1 and E2 gel bands were excised for LC-MS/MS for quantitative N-glycosylation analysis. Protein bands were dissolved by 200  $\mu$ L of 8 M urea with 10 mM DTT and digested by 100  $\mu$ L of 25 mM ammonium bicarbonate containing 0.01  $\mu$ g/ $\mu$ L trypsin before being collected and lyophilized in a sterile centrifuge. The dried polypeptide was dissolved in 40 mM NH<sub>4</sub>HCO<sub>3</sub> prepared with <sup>18</sup>O. Two units of rPNGase F were defined as the amount of enzyme that digests 100  $\mu$ g polypeptide. The lyophilized peptide fractions were resuspended in ddH<sub>2</sub>O containing 0.1% formic acid. The online Chromatography separation was performed on the EASY-nLC1200 system. The trapping and desalting procedure was carried out with 20  $\mu$ L solvent A (0.1% formic acid). DDA mass spectrum techniques were used to acquire tandem MS data on a Thermo Fisher Q Exactive Mass Spectrometer fitted with a Nano Flex ion source. For a full mass spectrometry survey scan, the target value was 3  $\times$  10<sup>6</sup> and the scan ranged from 350 to 2000 m/z at a resolution of 70,000. For the MS2 scan, only spectra with a charge state of 2–5 were selected for fragmentation by higher-energy collision dissociation. N-linked glycans can be released from the glycoprotein with the enzyme N-glycosidase F rPNGase F. Asparagine residues can cause a mass increase of 2.9882 Da in “heavy” water and identified the N-glycosylation sites by LC-MS/MS peptide mapping. The MS data were analyzed for protein modification using PEAKS Studio 8.5.

### Quantitative S-acylation analysis by LC-MS/MS

E1 and E2 gel bands were excised for LC-MS/MS for quantitative S-acylation analysis. 100  $\mu$ L of the decolorizing solution was added and placed at room temperature to decolor overnight. 100% acetonitrile was added until the gel mass turned white. 10  $\mu$ L 0.01  $\mu$ g/ $\mu$ L trypsin was added to the gel to fully absorb by icing until it became transparent. The peptides were analyzed using Easy-nLC nanoflow HPLC system connected to Orbitrap Fusion mass spectrometer (Thermo Fisher Scientific, San Jose, CA, USA). A total of 1  $\mu$ g of each sample was loaded onto a Thermo Scientific EASY column (two columns) using



an autosampler at a flow rate of 200 nL/min. The sequential separation of peptides on a Thermo Scientific EASY trap column (100  $\mu\text{m} \times 2\text{ cm}$ , 5  $\mu\text{m}$ , 100  $\text{\AA}$ , C18) and analytical column (75  $\mu\text{m} \times 25\text{ cm}$ , 5  $\mu\text{m}$ , 100  $\text{\AA}$ , C18) was accomplished using a segmented gradient from 5% to 28% Solvent B (0.1% formic acid in 100% ACN) for 40 min, followed by 28%–90% Solvent B for 2 min and then 90% Solvent B for 18 min. The column was re-equilibrated to its initial highly aqueous solvent composition before each round of analysis. The mass spectrometer was operated in positive ion mode, and MS spectra were acquired over a range of 375–1500  $m/z$ . The resolving powers of the MS scan and MS/MS scan at 200  $m/z$  for the Orbitrap Fusion were set as 120,000 and 60,000, respectively. The Data Dependent Mode was Top Speed, Cycle Time was 3 s and ions were fragmented through higher-energy collisional dissociation. The maximum ion injection times were set at 50 ms for the survey scan and 105 ms for the MS/MS scans, and the automatic gain control target values for the Master scan modes were set to  $4e^5$  and for MS/MS was set to  $1e^5$ . The dynamic exclusion duration was 40 s. S-acylation modification occurs on CKST amino acids, for which the molecular weight change after modification is 238.23/266.26 Da (palmitic acid/stearic acid). The MS data were analyzed for protein modification using Proteome Discoverer 2.4.

### MD simulation study of cholesterol and DOPC regulation of E1–E2

The solved cryo-EM structure of the full-length E1–E2 complex was first truncated to keep only the amino acids of 396–438 for E1 and 340–403 for E2 as the starting protein model for MD simulations. The truncated protein (denoted as E1–E2), together with the ligands at three different compositions, was then packed with the lipid bilayer using the CHARMM-GUI tool<sup>87</sup>. The resulting systems simulated are E1–E2 with: (1) all the cholesterol and DOPC molecules retained (denoted as Wild Type); (2) removal of the pocket cholesterol only (denoted as delCHL); (3) removal of the pocket DOPC only (denoted as delDOPC); (4) removal of both cholesterol and DOPC in the pocket (denoted as delCHL/DOPC); (5) removal of the two cholesterol surround the pocket (denoted as del2CHL); and (6) all 3 cholesterol and DOPC molecule removed (denoted as delALL). The main lipid components with the same mixed lipid ratio in both inner and outer leaflets (DPPC/DOPC/POPC/DPEPE/POPE/DPPS/POPS/Chol = 7.5:7.5:15:5:7:5:7:46) were used in constructing the membrane model following the reported lipidomic data<sup>52</sup>. The sphingomyelin was approximatively modeled by the POPC due to the lack of the force field parameters for sphingomyelin and the identical hydrophilic phosphorylcholine head group, similar conformation and charge distribution between

the two molecules. In order to obtain a consistent protein–membrane packing conformation with the cryo-EM density data, the vector from the residue Val-410 of E1 to Lys-394 of E2 was aligned with the normal axis of the lipid bilayer when packing systems. Next, the constructed protein–ligand–membrane systems were submitted to the H++ web server<sup>88</sup> to determine the protonation states of the protein under the membrane environment in order to approximate a simulation pH of 7.0. As a result, the amino acids His-355 in E1 and His-352 in E2 were simulated as protonated. The protein was diagonally positioned in the  $x$ – $y$  plane in the simulation box filled with the TIP3P water model<sup>89</sup> to save space, leading to a box size of  $\sim 87\text{ \AA} \times 87\text{ \AA} \times 122\text{ \AA}$ . The final systems were neutralized with the sodium ion and extra NaCl was added to model the experimental salt concentration of 0.15 M. All MD runs were performed using AMBER 19 package<sup>90</sup> with the ff14SB protein force field<sup>91</sup> and the LIPID17 lipid force field<sup>92</sup>. The van de Waals cutoff was set to 10  $\text{\AA}$  and PME was used for long-range electrostatic energy calculation. Minimization, heating and holding processes were carried out before the production run. All production simulations were performed at 310 K under the NPT ensemble with the integration step of 2 fs and an anisotropic pressure coupling. Five independent runs for each system of Wild Type, delCHL, delDOPC, delCHL/DOPC, del2CHL, or delALL were carried out to improve samplings, which led to a total of 30 simulations. Production trajectories of 700 ns were generated for all simulations. All figures and animations were prepared using gnuplot 5.4 and VMD<sup>93</sup>.

### Acknowledgements

We thank all staff members of the Cryo-electron Microscopy Center, Southern University of Science and Technology, for their assistance with data collection. This work was supported by the National Natural Science Foundation of China (81870246 and 82070329 to Z.L.), and by Guangdong Basic and Applied Basic Research Foundation (2019A1515110090 to R.Q.).

### Author details

<sup>1</sup>Cryo-electron Microscopy Center, Southern University of Science and Technology, Shenzhen, Guangdong, China. <sup>2</sup>College of Animal Science and Veterinary Medicine, Henan Agricultural University, Zhengzhou, Henan, China. <sup>3</sup>Henan Dairy Herd Improvement Center, Zhengzhou, Henan, China. <sup>4</sup>Institute for Hepatology, National Clinical Research Center for Infectious Disease, Shenzhen Third People's Hospital, Shenzhen, Guangdong, China. <sup>5</sup>Department of Cardiology, Shanghai Tenth People's Hospital, and Pan-Vascular Research Institute, Heart, Lung, and Blood Center, Tongji University School of Medicine, Shanghai, China

### Author contributions

Conceptualization, Z.L. and C.W.; Methodology, Z.L., C.W., C.L., and A.W.; Investigation, A.W., F.Z., C.L., D.G., R.Q., Y.Y., S.L., Y.G., L.F., and Y.X.; Resources, Z.L., Y.X., and C.W.; Writing (original draft), Z.L., A.W., C.L., R.Q., and C.W.; Writing (review & editing), Z.L.; Supervision, Z.L. and C.W.; Funding Acquisition, Z.L.

### Data availability

The atomic coordinates of E1–E2–Capsid of the GETV have been deposited in the Protein Data Bank under the accession code 7FD2. The cryo-EM maps of the envelope glycoprotein and capsid core, and the clipped virus at 5-fold, 3-fold and 2-fold have been deposited in the Electron Microscopy Data Bank

under accession codes EMD-31533, EMD-31614, EMD-31615, and EMD-31617, respectively.

#### Conflict of interest

The authors declare no competing interests.

#### Publisher's note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41421-022-00374-6>.

Received: 14 September 2021 Accepted: 3 January 2022

Published online: 11 February 2022

#### References

- Lvov, D. K., Shchelkanov, M. Y., Alkhovsky, S. V. & Deryabin, P. G. In *Zoonotic Viruses in Northern Eurasia* (eds Dmitry Konstantinovich Lvov, Mikhail Yurievich Shchelkanov, Sergey Vladimirovich Alkhovsky, & Petr Grigorievich Deryabin) 135–392 (Academic Press, 2015).
- Hua, C. & Combe, B. Chikungunya Virus-Associated Disease. *Curr. Rheumatol. Rep.* **19**, 69 (2017).
- Rezza, G., Chen, R. & Weaver, S. C. O'nyong-nyong fever: a neglected mosquito-borne viral disease. *Pathog. Glob. Health* **111**, 271–275 (2017).
- Tesh, R. B. et al. Mayaro virus disease: an emerging mosquito-borne zoonosis in tropical South America. *Clin. Infect. Dis.* **28**, 67–73 (1999).
- Berge, T. O. International catalogue of arboviruses, including certain other viruses of vertebrates. [Washington]: U.S. Dept. of Health, Education, and Welfare, Public Health Service, 278–279 (1975).
- Bannai, H. et al. Epizootiological Investigation of Getah Virus Infection among Racehorses in Japan in 2014. *J. Clin. Microbiol.* **53**, 2286–2291 (2015).
- Brown, C. M. & Timoney, P. J. Getah virus infection of Indian horses. *Trop. Anim. Health Prod.* **30**, 241–252 (1998).
- Kamada, M. et al. Equine Getah virus infection: isolation of the virus from racehorses during an enzootic in Japan. *Am. J. Trop. Med. Hyg.* **29**, 984–988 (1980).
- Yago, K., Hagiwara, S., Kawamura, H. & Narita, M. A fatal case in newborn piglets with Getah virus infection: isolation of the virus. *Nihon Juigaku Zasshi* **49**, 989–994 (1987).
- Shibata, I., Hatano, Y., Nishimura, M., Suzuki, G. & Inaba, Y. Isolation of Getah virus from dead fetuses extracted from a naturally infected sow in Japan. *Vet. Microbiol.* **27**, 385–391 (1991).
- Yang, T. et al. An outbreak of Getah virus infection among pigs in China, 2017. *Transbound. Emerg. Dis.* **65**, 632–637 (2018).
- Liu, H. et al. First isolation and characterization of Getah virus from cattle in northeastern China. *BMC Vet. Res.* **15**, 320 (2019).
- Shi, N., Li, L. X., Lu, R. G., Yan, X. J. & Liu, H. Highly Pathogenic Swine Getah Virus in Blue Foxes, Eastern China, 2017. *Emerg. Infect. Dis.* **25**, 1252–1254 (2019).
- Sugiyama, I. et al. Serological survey of arthropod-borne viruses among wild boars in Japan. *J. Vet. Med. Sci.* **71**, 1059–1061 (2009).
- Li, Y. et al. Serological Survey of Getah Virus in Domestic Animals in Yunnan Province, China. *Vector Borne Zoonotic Dis.* **19**, 59–61 (2019).
- Li, X. D., Qiu, F. X., Yang, H., Rao, Y. N. & Calisher, C. H. Isolation of Getah virus from mosquitoes collected on Hainan Island, China, and results of a serosurvey. *Southeast Asian J. Trop. Med. Public Health* **23**, 730–734 (1992).
- Rayner, J. O., Dryga, S. A. & Kamrud, K. I. Alphavirus vectors and vaccination. *Rev. Med. Virol.* **12**, 279–296 (2002).
- Sherman, M. B. & Weaver, S. C. Structure of the recombinant alphavirus Western equine encephalitis virus revealed by cryoelectron microscopy. *J. Virol.* **84**, 9775–9782 (2010).
- Kostyuchenko, V. A. et al. The structure of barmah forest virus as revealed by cryo-electron microscopy at a 6-angstrom resolution has detailed transmembrane protein architecture and interactions. *J. Virol.* **85**, 9327–9333 (2011).
- Zhang, R. et al. 4.4 Å cryo-EM structure of an enveloped alphavirus Venezuelan equine encephalitis virus. *EMBO J.* **30**, 3854–3863 (2011).
- Yap, M. L. et al. Structural studies of Chikungunya virus maturation. *Proc. Natl. Acad. Sci. USA* **114**, 13703–13707 (2017).
- Chen, L. et al. Implication for alphavirus host-cell entry and assembly indicated by a 3.5 Å resolution cryo-EM structure. *Nat. Commun.* **9**, 5326 (2018).
- Chen, C. L. et al. Cryo-EM structure of eastern equine encephalitis virus in complex with heparan sulfate analogues. *Proc. Natl. Acad. Sci. USA* **117**, 8890–8899 (2020).
- Hasan, S. S. et al. Cryo-EM Structures of Eastern Equine Encephalitis Virus Reveal Mechanisms of Virus Disassembly and Antibody Neutralization. *Cell Rep.* **25**, 3136–3147 (2018).
- Ribeiro-Filho, H. V. et al. Cryo-EM structure of the mature and infective Mayaro virus at 4.4 Å resolution reveals features of arthritogenic alphaviruses. *Nat. Commun.* **12**, 3038 (2021).
- Basore, K. et al. Cryo-EM Structure of Chikungunya Virus in Complex with the Mxra8 Receptor. *Cell* **177**, 1725–1737 (2019).
- Zhou, F. et al. Isolation and phylogenetic analysis of Getah virus from a commercial modified live vaccine against porcine reproductive and respiratory syndrome virus. *Mol. Cell Probes* **53**, 101650 (2020).
- Fukunaga, Y., Kumanomido, T. & Kamada, M. Getah virus as an equine pathogen. *Vet. Clin. North Am. Equine Pr.* **16**, 605–617 (2000).
- Oliver, K. R., Scallan, M. F., Dyson, H. & Fazakerley, J. K. Susceptibility to a neurotropic virus and its changing distribution in the developing brain is a function of CNS maturity. *J. Neurovirol.* **3**, 38–48 (1997).
- Zhu, D. et al. Pushing the resolution limit by correcting the Ewald sphere effect in single-particle Cryo-EM reconstructions. *Nat. Commun.* **9**, 1552 (2018).
- Wang, N. et al. Architecture of African swine fever virus and implications for viral assembly. *Science* **366**, 640–644 (2019).
- Jumper, J. et al. Highly accurate protein structure prediction with AlphaFold. *Nature*, **596**, 583–589 (2021).
- Lescar, J. et al. The Fusion glycoprotein shell of Semliki Forest virus: an icosahedral assembly primed for fusogenic activation at endosomal pH. *Cell* **105**, 137–148 (2001).
- Voss, J. E. et al. Glycoprotein organization of Chikungunya virus particles revealed by X-ray crystallography. *Nature* **468**, 709–712 (2010).
- Roussel, A. et al. Structure and interactions at the viral surface of the envelope protein E1 of Semliki Forest virus. *Structure* **14**, 75–86 (2006).
- Watanabe, Y., Bowden, T. A., Wilson, I. A. & Crispin, M. Exploitation of glycosylation in enveloped virus pathobiology. *Biochim. Biophys. Acta Gen. Subj.* **1863**, 1480–1497 (2019).
- Gupta, R. & Brunak, S. Prediction of glycosylation across the human proteome and the correlation to protein function. *Pac. Symp. Biocomput.* **7**, 310–322 (2002).
- Steenfot, C. et al. Precision mapping of the human O-GalNAc glycoproteome through SimpleCell technology. *EMBO J.* **32**, 1478–1488 (2013).
- Li, F. et al. GlycoMine: a machine learning-based approach for predicting N-, C- and O-linked glycosylation in the human proteome. *Bioinformatics* **31**, 1411–1419 (2015).
- Linder, M. E. & Deschenes, R. J. Palmitoylation: policing protein stability and traffic. *Nat. Rev. Mol. Cell Biol.* **8**, 74–84 (2007).
- Schmidt, M. F. & Schlesinger, M. J. Fatty acid binding to vesicular stomatitis virus glycoprotein: a new type of post-translational modification of the viral glycoprotein. *Cell* **17**, 813–819 (1979).
- Veit, M. Palmitoylation of virus proteins. *Biol. Cell* **104**, 493–515 (2012).
- Veit, M., Reverey, H. & Schmidt, M. F. Cytoplasmic tail length influences fatty acid selection for acylation of viral glycoproteins. *Biochem. J.* **318**, 163–172 (1996).
- Zhao, H., Lindqvist, B., Garoff, H., von Bonsdorff, C. H. & Liljeström, P. A tyrosine-based motif in the cytoplasmic domain of the alphavirus envelope protein is essential for budding. *EMBO J.* **13**, 4204–4211 (1994).
- Ivanova, L. & Schlesinger, M. J. Site-directed mutations in the Sindbis virus E2 glycoprotein identify palmitoylation sites and affect virus budding. *J. Virol.* **67**, 2546–2551 (1993).
- Ryan, C., Ivanova, L. & Schlesinger, M. J. Effects of site-directed mutations of transmembrane cysteines in sindbis virus E1 and E2 glycoproteins on palmitoylation and virus replication. *Virology* **249**, 62–67 (1998).
- Kordyukova, L. V., Serebryakova, M. V., Baratova, L. A. & Veit, M. Site-specific attachment of palmitate or stearate to cytoplasmic versus transmembrane cysteines is a common feature of viral spike proteins. *Virology* **398**, 49–56 (2010).

48. Ren, J. et al. CSS-Palm 2.0: an updated software for palmitoylation sites prediction. *Protein Eng. Des. Sel.* **21**, 639–644 (2008).
49. Ning, W. et al. GPS-Palm: a deep learning-based graphic presentation system for the prediction of S-palmitoylation sites in proteins. *Brief. Bioinform.* **22**, 1836–1847 (2021).
50. DiNunno, N. M. et al. Identification of a pocket factor that is critical to Zika virus assembly. *Nat. Commun.* **11**, 4953 (2020).
51. Modis, Y., Ogata, S., Clements, D. & Harrison, S. C. A ligand-binding pocket in the dengue virus envelope glycoprotein. *Proc. Natl. Acad. Sci. USA* **100**, 6986–6991 (2003).
52. Kalvodova, L. et al. The lipidomes of vesicular stomatitis virus, semliki forest virus, and the host plasma membrane analyzed by quantitative shotgun mass spectrometry. *J. Virol.* **83**, 7996–8003 (2009).
53. Nishio, M., Umezawa, Y., Fantini, J., Weiss, M. S. & Chakrabarti, P. CH- $\pi$  hydrogen bonds in biological macromolecules. *Phys. Chem. Chem. Phys.* **16**, 12648–12683 (2014).
54. Shabman, R. S., Rogers, K. M. & Heise, M. T. Ross River virus envelope glycans contribute to type I interferon production in myeloid dendritic cells. *J. Virol.* **82**, 12374–12383 (2008).
55. Knight, R. L., Schultz, K. L., Kent, R. J., Venkatesan, M. & Griffin, D. E. Role of N-linked glycosylation for sindbis virus infection and replication in vertebrate and invertebrate systems. *J. Virol.* **83**, 5640–5647 (2009).
56. Aksnes, I., Markussen, T., Braeen, S. & Rimstad, E. Mutation of N-glycosylation Sites in Salmonid Alphavirus (SAV) Envelope Proteins Attenuate the Virus in Cell Culture. *Viruses* **12**, 1071 (2020).
57. Wu, Z. et al. Palmitoylation of SARS-CoV-2 S protein is essential for viral infectivity. *Signal Transduct. Target. Ther.* **6**, 231 (2021).
58. McBride, C. E. & Machamer, C. E. Palmitoylation of SARS-CoV S protein is necessary for partitioning into detergent-resistant membranes and cell-cell fusion but not interaction with M protein. *Virology* **405**, 139–148 (2010).
59. Gaedigk-Nitschko, K. & Schlesinger, M. J. Site-directed mutations in Sindbis virus E2 glycoprotein's cytoplasmic domain and the 6K protein lead to similar defects in virus assembly and budding. *Virology* **183**, 206–214 (1991).
60. Sousa, I. P. Jr. et al. Envelope lipid-packing as a critical factor for the biological activity and stability of alphavirus particles isolated from mammalian and mosquito cells. *J. Biol. Chem.* **286**, 1730–1736 (2011).
61. Hafer, A., Whittlesey, R., Brown, D. T. & Hernandez, R. Differential incorporation of cholesterol by Sindbis virus grown in mammalian or insect cells. *J. Virol.* **83**, 9113–9121 (2009).
62. Nieto-Garai, J. A. et al. Cholesterol in the Viral Membrane is a Molecular Switch Governing HIV-1 Env Clustering. *Adv. Sci. (Weinh.)* **8**, 2003468 (2021).
63. Elkins, M. R. et al. Cholesterol-binding site of the influenza M2 protein in lipid bilayers from solid-state NMR. *Proc. Natl. Acad. Sci. USA* **114**, 12946–12951 (2017).
64. Hu, B., Höfer, C. T., Thiele, C. & Veit, M. Cholesterol Binding to the Transmembrane Region of a Group 2 Hemagglutinin (HA) of Influenza Virus Is Essential for Virus Replication, Affecting both Virus Assembly and HA Fusion Activity. *J. Virol.* **93**, e00555–19 (2019).
65. Sun, X. & Whittaker, G. R. Role for influenza virus envelope cholesterol in virus entry and infection. *J. Virol.* **77**, 12543–12551 (2003).
66. Huang, H. et al. Human herpesvirus 6 envelope cholesterol is required for virus entry. *J. Gen. Virol.* **87**, 277–285 (2006).
67. Campbell, S. M., Crowe, S. M. & Mak, J. Virion-associated cholesterol is critical for the maintenance of HIV-1 structure and infectivity. *Aids* **16**, 2253–2261 (2002).
68. Aizaki, H. et al. Critical role of virion-associated cholesterol and sphingolipid in hepatitis C virus infection. *J. Virol.* **82**, 5715–5724 (2008).
69. Phalen, T. & Kielian, M. Cholesterol is required for infection by Semliki Forest virus. *J. Cell Biol.* **112**, 615–623 (1991).
70. Lu, Y. E., Cassese, T. & Kielian, M. The cholesterol requirement for sindbis virus entry and exit and characterization of a spike protein region involved in cholesterol dependence. *J. Virol.* **73**, 4272–4278 (1999).
71. Carvalho, C. A. M., Silva, J. L., Oliveira, A. C. & Gomes, A. M. O. On the entry of an emerging arbovirus into host cells: Mayaro virus takes the highway to the cytoplasm through fusion with early endosomes and caveolae-derived vesicles. *PeerJ* **5**, e3245 (2017).
72. Tsetsarkin, K. A., Vanlandingham, D. L., McGee, C. E. & Higgs, S. A single mutation in chikungunya virus affects vector specificity and epidemic potential. *PLoS Pathog.* **3**, e201 (2007).
73. Ng, C. G. et al. Effect of host cell lipid metabolism on alphavirus replication, virion morphogenesis, and infectivity. *Proc. Natl. Acad. Sci. USA* **105**, 16326–16331 (2008).
74. Wicht, S. et al. Imipramine Inhibits Chikungunya Virus Replication in Human Skin Fibroblasts through Interference with Intracellular Cholesterol Trafficking. *Sci. Rep.* **7**, 3145 (2017).
75. Zhai, Y. G. et al. Complete sequence characterization of isolates of Getah virus (genus Alphavirus, family Togaviridae) from China. *J. Gen. Virol.* **89**, 1446–1456 (2008).
76. Liang, J. et al. Inhibition of the mevalonate pathway enhances cancer cell oncolysis mediated by M1 virus. *Nat. Commun.* **9**, 1524 (2018).
77. Zheng, S. Q. et al. MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nat. Methods* **14**, 331–332 (2017).
78. Rohou, A. & Grigorieff, N. CTFIND4: Fast and accurate defocus estimation from electron micrographs. *J. Struct. Biol.* **192**, 216–221 (2015).
79. Tang, G. et al. EMAN2: an extensible image processing suite for electron microscopy. *J. Struct. Biol.* **157**, 38–46 (2007).
80. Zivanov, J. et al. New tools for automated high-resolution cryo-EM structure determination in RELION-3. *Elife* **7**, e42166 (2018).
81. Guo, F. & Jiang, W. Single particle cryo-electron microscopy and 3-D reconstruction of viruses. *Methods Mol. Biol.* **1117**, 401–443 (2014).
82. Kucukelbir, A., Sigworth, F. J. & Tagare, H. D. Quantifying the local resolution of cryo-EM density maps. *Nat. Methods* **11**, 63–65 (2014).
83. Pettersen, E. F. et al. UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).
84. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D. Biol. Crystallogr.* **66**, 486–501 (2010).
85. Adams, P. D. et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D. Biol. Crystallogr.* **66**, 213–221 (2010).
86. Chen, V. B. et al. MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D. Biol. Crystallogr.* **66**, 12–21 (2010).
87. Lee, J. et al. CHARMM-GUI Membrane Builder for Complex Biological Membrane Simulations with Glycolipids and Lipoglycans. *J. Chem. Theory Comput.* **15**, 775–786 (2019).
88. Gordon, J. C. et al. H++: a server for estimating pKas and adding missing hydrogens to macromolecules. *Nucleic Acids Res.* **33**, W368–W371 (2005).
89. Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W. & Klein, M. L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **79**, 926–935 (1983).
90. Case, D. A. et al. The Amber biomolecular simulation programs. *J. Comput. Chem.* **26**, 1668–1688 (2005).
91. Maier, J. A. et al. ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *J. Chem. Theory Comput.* **11**, 3696–3713 (2015).
92. Dickson, C. J. et al. Lipid14: The Amber Lipid Force Field. *J. Chem. Theory Comput.* **10**, 865–879 (2014).
93. Humphrey, W., Dalke, A. & Schulten, K. VMD: visual molecular dynamics. *J. Mol. Graph.* **14**, 33–38 (1996).