ARTICLE

Check for updates

Dimeric Transmembrane Structure of the SARS-CoV-2 E Protein

Rongfu Zhang[1,2,6], Huajun Qin[1,6], Ramesh Prasad [3], Riqiang Fu [2], Huan-Xiang Zhou [3,4✉] & Timothy A. Cross [1,2,5✉]

The SARS-CoV-2 E protein is a transmembrane (TM) protein with its N-terminus exposed on the external surface of the virus. At debate is its oligomeric state, let alone its function. Here, the TM structure of the E protein is characterized by oriented sample and magic angle spinning solid-state NMR in lipid bilayers and refined by molecular dynamics simulations. This protein was previously found to be a pentamer, with a hydrophobic pore that appears to function as an ion channel. We identify only a front-to-front, symmetric helix-helix interface, leading to a dimeric structure that does not support channel activity. The two helices have a tilt angle of only 6°, resulting in an extended interface dominated by Leu and Val sidechains. While residues Val14-Thr35 are almost all buried in the hydrophobic region of the membrane, Asn15 lines a water-filled pocket that potentially serves as a drug-binding site. The E and other viral proteins may adopt different oligomeric states to help perform multiple functions.

[1] Department of Chemistry and Biochemistry, Florida State University, Tallahassee, FL 32306, USA. [2] National High Magnetic Field Laboratory, Tallahassee, FL 32310, USA. [3] Department of Chemistry, University of Illinois Chicago, Chicago, IL 60607, USA. [4] Department of Physics, University of Illinois Chicago, Chicago, IL 60607, USA. [5] Institute of Molecular Biophysics, Florida State University, Tallahassee, FL 32306, USA. [6]These authors contributed equally: Rongfu Zhang, Huajun Qin. ✉email: hzhou43@uic.edu; timothyacross@gmail.com

The SARS-CoV-2 virus in the past few years has resulted in over 700 million COVID-19 infections and 6.9 million deaths worldwide[1]. SARS-CoV-2 is an enveloped, positive-strand RNA virus that belongs to the coronaviridae family[2] and is closely related to the SARS-CoV-1[3] and MERS-CoV[4] viruses that caused previous epidemics. Four proteins decorate the surface of SARS-CoV-2 virion, including the membrane (M), envelope (E), nucleocapsid (N), and spike (S) proteins. The M, E, and S proteins are integral membrane proteins that are all reported to be homo-oligomers. The 75-residue E protein encodes a single transmembrane (TM) helix and has been proposed to form pentamers leading to a transmembrane pore that is capable of conducting ions[5–11]. The SARS-CoV-2 virus is synthesized in the endoplasmic reticulum Golgi intermediate compartment (ERGIC) of mammalian cells[8,10]. In the ERGIC and Golgi environment the N- and C-termini of the E protein are in the lumen and cytoplasm, respectively[12].

Alam et al.[13] presented a functional pangenomic analysis revealing that SARS-CoV-1 and SARS-CoV-2 E proteins display two conserved key features, an N-terminal region with putative ion channel activity and a C-terminal PDZ-binding motif, that play key roles in inducing an inflammasome response leading to acute respiratory distress syndrome, a major cause of death from viral infections. Verdia-Bagguena et al.[14] suggested that ion conductance and cation selectivity are potentially controlled by the charge of the lipid membranes giving additional flexibility for viral reproductive conditions. While a portion of the E protein is in the viral membrane coat, most is located in intracellular transport sites of the host including the ER, Golgi, and ERGIC sites involved in viral assembly and budding[15]. In vitro studies have shown that the removal of the E protein from recombinant coronavirus particles attenuates viral maturation and produces incompetent progenies[16–18]. Consequently, the E protein is an important viral protein with multiple functions and possibly multiple conformational states in different environments.

Several solution NMR characterizations of the SARS-CoV-1 E protein in detergent micelles resulted in pentameric structures forming a TM hydrophobic pore that has been identified as an ion channel[10,19,20]. An initial solid-state NMR (ssNMR) study of the TM domain (residues 8–38) in lipid bilayers of the SARS-CoV-2 E protein assumed this pentameric model, but added high-resolution structural detail for the TM helices (Protein Data Bank entry 7K3G)[21]. In 2021, a solution NMR study[22] of the full-length SARS-CoV-2 E protein in detergent micelles did not challenge this oligomerization state and defined the secondary structure, revealing a putative transmembrane helix comprising residues 8–43 and a cytoplasmic helix comprising residues 53–60. In addition, a one-dimensional (1D) oriented-sample (OS) ssNMR spectrum in DMPC lipids was interpreted as indicating a large tilt angle, 45°, of the transmembrane helix. Although ion

conductance through a pentameric assembly of E constructs that include the TM sequence has been demonstrated in various membrane mimetic environments using a large concentration gradient and/or membrane potential, whether the E protein has in vivo channel activity is still an open question.

For atomic-level structural studies of small membrane proteins in a native-like environment, ssNMR has clear advantages over other structural techniques such as X-ray crystallography, solution NMR, and cryo-electron microscopy (EM). Crystallization is only possible when contacts between protein molecules exist in a crystal lattice – for small TM proteins this often leads to structural distortions[23]. Cryo-EM is growing as a technique for structural characterization of large protein assembles, but is not applicable to small proteins embedded in lipid bilayers. Solution NMR would also be a good approach if it was performed with the protein in nanodisc preparations[24,25]. However, most solution NMR studies of membrane proteins have been performed in detergent micelles that can also lead to distorted structures[23]. Taken together, with a sample in an extensive lipid bilayer environment, ssNMR is a technique that is capable of high-resolution structural characterization for small membrane proteins in a native-like environment.

We report here a SARS-CoV-2 E protein TM dimeric structure in liquid-crystalline lipid bilayers that has not been previously characterized. Our extensive work on different E TM constructs using SDS-PAGE electrophoresis (Fig. 1) and ssNMR contradicts the pentameric conclusion of prior characterizations. The dimeric structure of the E TM domain (E_{12-37}) obviously cannot conduct ions as previously observed for the pentameric characterizations. Still, like the PDB 7K3G structure and multiple previous models, the tilt of the helices in our dimer in the lipid bilayer is small. However, unlike the pentameric structures where neighboring subunits have a front-to-back arrangement, our dimeric structure has the two subunits in a front-to-front, symmetric arrangement. The E protein is present in the viral coat at low copy numbers (based on indirect analyses)[26], where there is no evidence for channel formation. Potentially the dimeric structure characterized here represents the native structure of the E protein in the viral coat.

Results

TM sequence-containing E constructs form a stable dimer. We prepared a variety of E constructs (Supplementary Fig. 1), including E_{1-52}, E_{7-43}, E_{8-41}, E_{12-37}, and E_{12-65} that expressed well enough for SDS-PAGE electrophoresis (Fig. 1). All of these constructs contained the TM sequence and ran primarily as dimers in our SDS-PAGE experiments. Importantly, there was no evidence of pentamers in the SDS-solubilized environment of these gels, the frequently suggested E oligomeric state in the literature. We used a ladder that included low molecular weight
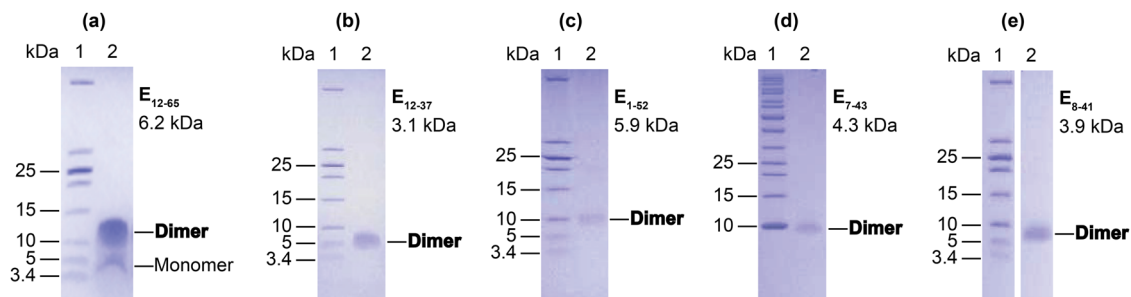
Fig. 1 SDS-PAGE gels of purified E constructs. a E_{12-65}. **b** E_{12-37}. **c** E_{1-52}. **d** E_{7-43}. **e** E_{8-41}. All of these constructs included leading Ser-Asn-Ala residues from the TEV cleavage site (Supplementary Fig. 1). The molecular weight of each construct is shown below its designation. The observed bands for monomer and dimer next to lane 2 in each gel are indicated. All display dominant bands for the dimer.
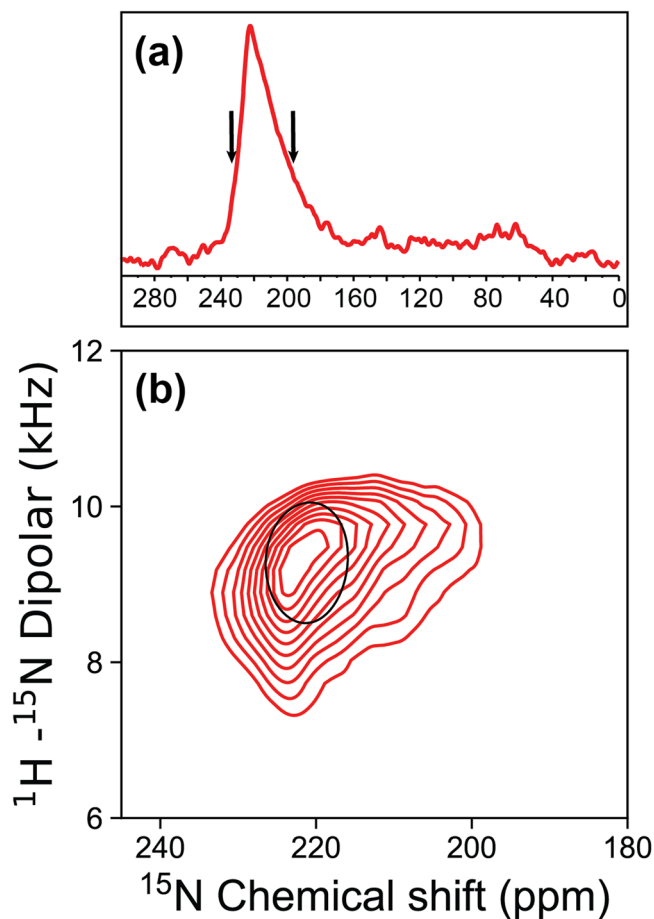
Fig. 2 ssNMR spectra of uniformly ^{15}N labeled E_{12-37} in aligned POPC/ POPG bilayers. a 1D ^{15}N spectrum. **b** 2D PISEMA spectrum of the same sample, with contours down to the spectral level as shown by the arrows on the 1D spectrum. Superimposed is a PISA wheel for a helix tilt angle of 6° with respect to the bilayer normal, calculated using a chemical shift tensor with $\sigma_{11} = 54.5$ ppm, $\sigma_{22} = 82.8$ ppm, $\sigma_{33} = 227.1$ ppm and a dipolar coupling constant of 10.4 kHz[28]. The amino acid sequence for E_{12-37} with its N-terminal SNA tag is: SNAL$_{12}$IVNSVLLF$_{20}$LAFVVFLLVT$_{30}$LAILTAL$_{37}$.

markers (down to 3.4 kDa). Without the low molecular weight markers (Supplementary Fig. 2), the dimer band could be easily misidentified as a monomer.

We proceeded to prepare oriented samples of these E constructs in POPC/POPG bilayers (4:1 molar ratio), and achieved the highest alignment efficiency with E_{12-37}. Our subsequent structural characterization thus focused on this E construct.

E_{12-37} forms a highly regular TM helix with a small tilt angle. Figure 2a displays the 1D ^{15}N ssNMR spectrum of a uniformly ^{15}N labeled E_{12-37} oriented sample with the lipid bilayer normal parallel to the magnetic field. The signal intensities spread over a very limited anisotropic ^{15}N chemical shift range centered around 220 ppm, demonstrating a well-aligned sample. The central chemical shift is close to the σ_{33} element of the backbone ^{15}N chemical shift tensor, indicating that the TM helical axis is close to the bilayer normal. The broadening on the low-field side of the spectra can be largely accounted for by the range of σ_{33} rigid-limit values (230-215 ppm) for the amino acid composition of E_{12-37}. On the other hand, the broadening on the high-field side is more substantial, down to 175 ppm, suggesting that there is a small dispersion of amide ^{15}N tensor orientations relative to the bilayer normal. A small tilt of the helical axis with respect to the bilayer

normal is required to account for this dispersion. Another contributing factor is a small variation in the normal directions of the bilayers aligned between thin glass slides, typically no more than 2-3°.

Figure 2b displays a 2D spectrum of the same sample, obtained using the polarization inversion spin exchange at the magic angle (PISEMA) pulse sequence[27] that correlates the ^{15}N anisotropic chemical shift with the ^1H-^{15}N dipolar coupling. Contours in the PISEMA spectrum are displayed only down to the level of the black arrows shown in the 1D chemical shift spectrum (Fig. 2a), due to a more limited signal-to-noise ratio in the 2D PISEMA spectrum. The contours span the range of chemical shifts from 233-200 ppm. Importantly, this range encompasses all of the resonance frequencies for the amino acid specific labeled spectra presented below.

In the ^1H-^{15}N dipolar dimension, the PISEMA contours span a range from 7.3 to 10.4 kHz. These large dipolar coupling values, limited to a narrow range, again show that the entire peptide forms a highly regular helix with a small tilt angle. There is not even a single residue at the peptide termini that has a significantly reduced anisotropic chemical shift or dipolar coupling that could result from fraying of the helix.

The backbone ^{15}N resonance frequencies in PISEMA spectra are inherently dependent on the orientations of the ^{15}N chemical shift and ^1H-^{15}N dipolar tensors relative to the magnetic field, which is parallel to the bilayer normal. The orientations of the tensors in turn are determined by the orientations of the individual peptide planes. For water-soluble helices, the angle between the peptide plane and the helical axis is approximately 12°, meaning that even when the helical axis is aligned with the magnetic field, neither the ^1H-^{15}N dipolar vector nor the ^{15}N σ_{33} element, which are both in the peptide plane, are parallel to the magnetic field. However, for TM helices the peptide plane tilt angle is reduced to 6° with respect to the helical axis, due to the repulsive interaction between the carbonyl oxygens and fatty acyl environment of the lipid bilayer compared to an attractive aqueous environment for the carbonyl oxygens[28,29]. Consequently, for a TM helix aligned with the magnetic field, the PISEMA spectrum would reflect the backbone N-H bond vectors at a fixed angle of ~6° with respect to the magnetic field. When the helix is tilted with respect to the magnetic field, the range of angles increases. For example, for a helix tilted at 6°, the angles of the N-H bond vectors with respect to the magnetic field would range from ~0° to ~12°, depending on the self-rotation of the helix. To determine the tilt and rotation of the TM helix, numerous amino acid specific ^{15}N labeled samples of E_{12-37} were prepared for OS ssNMR spectroscopy.

Figure 3 displays six sets of 2D PISEMA spectra for six E_{12-37} preparations using amino acid specific ^{15}N labeling. With these samples, 24 of the 26 residues of E_{12-37} were labeled. The residues that were not labeled were Asn15 and Ser16 near the N-terminus. All of the observed resonances fall in a very small region that overlaps with the PISEMA spectrum of the uniformly ^{15}N labeled sample. Together, the PISEMA spectra indicate that the entire E_{12-37} sequence, including Asn15 and Ser16, forms an α-helix. As illustrated in Supplementary Fig. 3 for the Val ^{15}N labeled sample, there is no spectral intensity other than those shown in the spectra of Figs. 2 and 3. Indeed, the observed resonances for the entire sequence fall in a very narrow region that overlaps with a polarity index slant angle (PISA) wheel drawn for a helix with a 6° tilt with respect to the bilayer normal. A PISA wheel presents the predicted PISEMA resonances as the peptide plane traces an ideal α-helix trajectory that has a preset tilt angle[30,31].

The chemical shift linewidths observed in our oriented samples were typically less than 5.0 ppm over the total chemical shift range of 170 ppm. The narrow linewidths document excellent
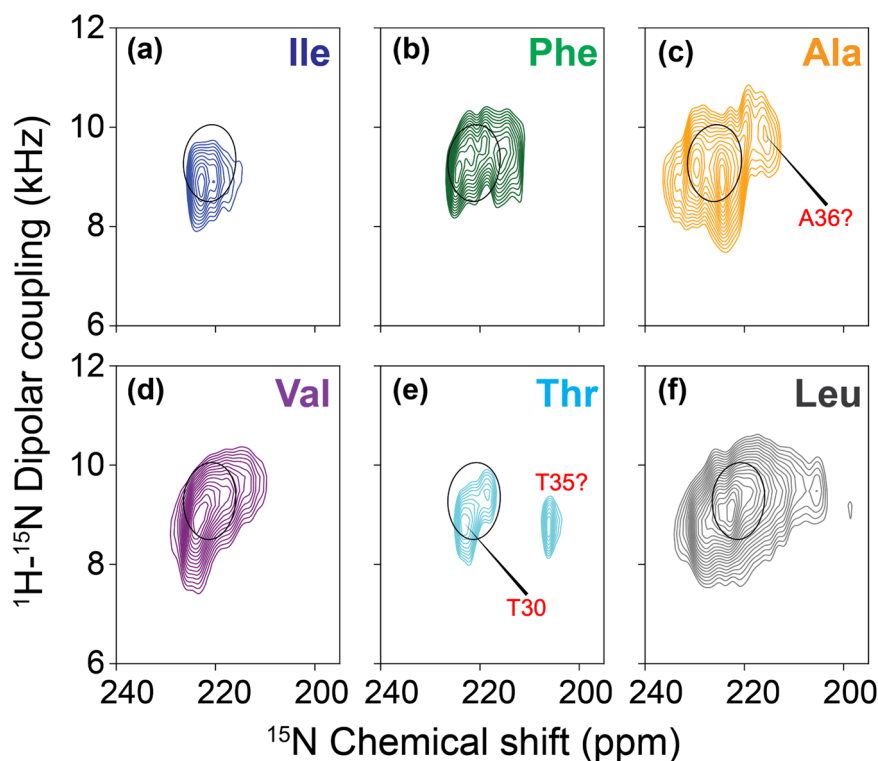
Fig. 3 PISEMA spectra of ^{15}N labeled E$_{12–37}$ peptides in POPC/POPG bilayers. a ^{15}N-labeling at Ile13 and Ile33. **b** ^{15}N-labeling at Phe20, Phe23, and Phe26. **c** ^{15}N-labeling at Ala22, Ala32, and Ala36. **d** ^{15}N-labeling at Thr30 and Thr35. **e** ^{15}N-labeling at Val14, Val17, Val24, Val25, and Val29. **f** ^{15}N-labeling at Leu12, Leu18, Leu19, Leu21, Leu27, Leu28, Leu31, Leu34, and Leu37. Superimposed on the spectra is a PISA wheel calculated at a tilt of 6°. For the Ala spectrum, the calculated PISA wheel used a somewhat larger σ_{33} of 232.1 ppm.

alignment of our samples resulting in high-resolution structural restraints. Normally, we would attempt to assign resonances to a PISA wheel pattern having 3.6 resonances per turn of a helical wheel, but the tilt of the helix here is so small that even with the excellent linewidths of the individual resonances for a given amino acid type it is not possible, with some exceptions noted below, to make such assignments. For amino acid type labeled preparations such as Val with more than two residues labeled, the dispersion is too small to generate spectral resolution of the individual resonances. However, the residues in the first and last turns of the helix may be more dynamic due to their peptide planes having a single backbone hydrogen bond, as well as their placement outside the hydrophobic region of the membrane (see below). The increased dynamics may shift the anisotropic chemical shift resonances slightly upfield and also reduce spectral intensity, such as that for Thr35. As a result, the Thr30 and Thr35 residues can be assigned to the resonances at 223 and 206 ppm, respectively. Likewise, the extra intensity in the upfield region of the Ala ^{15}N labeled spectrum can be tentatively assigned to Ala36 in the last turn of the helix. For the Ala residues, the chemical shift tensor has a σ_{33} that is somewhat greater than the other hydrophobic amino acids and consequently the entire spectral envelope is shifted downfield compared to the other residues by ~5 ppm[28].

The OS alignment efficiency for a longer construct, E$_{8–41}$, was not as good as that of E$_{12–37}$. In contrast to the PISEMA spectrum of Val ^{15}N labeled E$_{12–37}$, where all the resonances are around the PISA wheel for a 6° helix tilt, the corresponding spectrum of E$_{8–41}$ has extra intensities below 150 ppm (Supplementary Fig. 3). These upfield intensities arise from less well-oriented materials as well as the dynamic residues 8–11 and 38–41 of aligned E$_{8–41}$ that are outside the lipid bilayer, and constitute the so-called powder pattern.

E$_{12–37}$ forms a symmetric helix-helix dimer interface. Having defined the helix tilt by the OS experiments, we proceeded to define the interhelical interface by magic angle spinning (MAS) experiments. To obtain interhelical distance restraints, different amino acid type ^{13}C-labeled E$_{12–37}$ peptides were prepared and mixed in a 1:1 molar ratio. We utilized the ^{13}C-^{13}C correlations between two types of ^{13}C-labeled amino acids to identify interhelical cross peaks, as the correlations through the dipolar recoupling at different mixing times provide distance measurements for up to 8 Å between two ^{13}C nuclei[32,33].

The labeling schemes were carefully chosen based on the primary sequence and helical wheel projection. Because the E TM sequence consists of mostly long sidechain residues, there were no obvious interaction motifs (such as Gly at i and $i + 4$ or $i + 7$ positions) to guide the labeling. We first chose Val and Leu residues as labeling sites, because all 5 Val residues are located within an arc of the helical wheel that spans 140°, while the 9 Leu residues populate throughout the whole helical wheel (Supplementary Fig. 4a). This choice could give a reasonable chance of obtaining interhelical cross peaks while still narrowing down the interhelical interface.

As shown in Fig. 4a, for the ^{13}C-Leu and ^{13}C-Val mixture at a 600 ms mixing time, prominent interhelical cross peaks were observed for LCα-VCα, LCβ-VCα, LCγ-VCα, LCα-VCβ, LCβ-VCβ, and LCγ-VCβ. These cross peaks include all possible pairs of Val and Leu aliphatic carbons except for those at the branched termini, which are highly dynamic due to bond rotation. In particular, the Cα-Cα cross peak indicates a remarkably close approach between the helix backbones at the interface. When the mixing time was reduced to 300 ms (Fig. 4b), we were still able to observe the LCα-VCα cross peak, suggesting a distance of ≤ 6.5 Å[33]. With a further reduction of the mixing time to 100 ms, the interhelical cross peaks were no longer observable due to
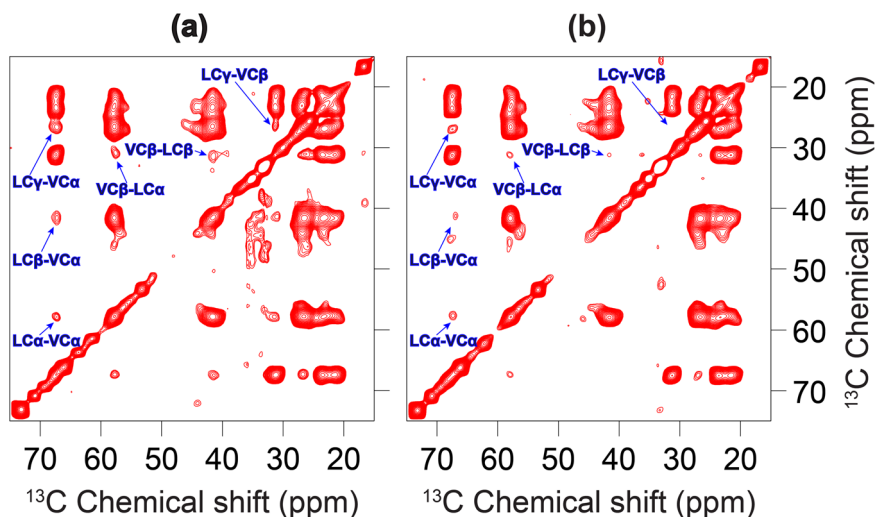
Fig. 4 ^{13}C-^{13}C correlation spectra of a 1:1 mixture of ^{13}C-Leu labeled E$_{12-37}$ and ^{13}C-Val labeled E$_{12-37}$ in POPC/POPG liposomes. Data obtained at mixing times of (**a**) 600 ms and (**b**) 300 ms. Interhelical cross peaks are assigned. In (**a**), the signal to noise (S/N) ratios for ^{13}C-Leu and ^{13}C-Val cross peaks are: LCγ-VCα, 5.7; LCβ-VCα, 2.8; and LCα-VCα, 2.5. Interhelical cross peaks with S/N ratios above 2 are usually considered strong.

insufficient time for interhelical polarization transfer (Supplementary Fig. 5).

Because there are 9 Leu and 5 Val residues in the TM sequence, at first sight it seemed not possible to unambiguously assign the cross peaks to specific Leu-Val pairs. However, the Leu-Val pairs that are most likely to be at the interhelical interface are those with one residue at position i and the other at $i \pm 1$. Indeed, the interface of a homo-oligomer must have i to $i \pm 1$ contacts. In the E TM sequence, there are two i to $i \pm 1$ Leu-Val pairs: Val17-Leu18 and Leu28-Val29. Given the strong intensities of the Leu-Val cross peaks (signal to noise ratios from 2.5 to 5.7) and the fact that at most one quarter of the mixture could contribute to any cross peak, most likely both of these Leu-Val pairs generated cross peaks.

To further narrow down the possibilities of the interhelical interface, we introduced a V25M mutation and prepared a sample mixing the ^{13}C-Val labeled mutant with the ^{13}C-Met labeled mutant (Supplementary Fig. 6). No interhelical cross peaks were observed on the ^{13}C-^{13}C correlation spectrum of this sample at a mixing time of 600 ms, suggesting that Val24 on one subunit and Val25 on the other subunit are not close and at least one of them is not in the interhelical interface.

In our third MAS sample, the ^{13}C-Leu labeled peptide was mixed with the ^{13}C-Phe labeled peptide, because in the helical wheel the 3 Phe residues are located on a 120° arc opposite to the 140° arc spanned by the 5 Val residues (Supplementary Fig. 4a). As shown in Fig. 5, no cross peaks were observed between the aliphatic carbons of Leu and Phe, and only a weak cross peak was observed between Leu Cγ and a Phe aromatic carbon. So the Cα atoms of all Leu-Phe pairs are far apart, but their long sidechains may reach each other within ~8 Å.

We used the above restraints, both positive and negative, from the ^{13}C-^{13}C correlation spectra to uniquely define the interhelical interface, by scanning for all possible helix-helix arrangements. The observation that one side of the helical wheel (the "Val" or front face) is in the interface while the opposite side (the "Phe" or back face) is not in the interface supports a front-to-front, symmetric dimer interface. Scanning over helix rotation produced only a single dimer interface, shown in Supplementary Fig. 4b, that satisfies all the restraints. In this interface, Leu21 and Val25 are at the core, and Val17, Leu18, Leu28, and Val29 are at the periphery. In contrast, the interhelical interface of a homo-
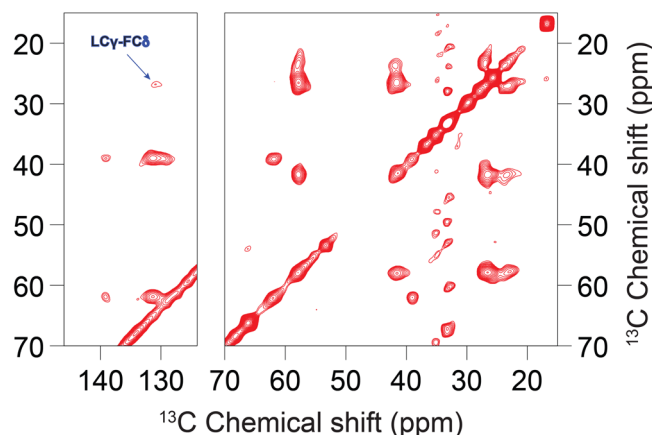
Fig. 5 ^{13}C-^{13}C correlation spectrum of an equimolar mixture of ^{13}C-Leu labeled E$_{12-37}$ and ^{13}C-Phe labeled E$_{12-37}$ in POPC/POPG liposomes at a mixing time of 600 ms.

pentamer structure is lined by two subunits in a front-to-back arrangement, with the Val face paired with the Phe face (or portions thereof). Thus unavoidably at least one of the Phe Cα atoms would be placed at the interface. For example, the solution NMR structure 5X29[19] has the Phe23 Cα at the center of the interface. Our scanning of helix rotation produced a pentamer with an interface that minimizes the burial of Phe Cα atoms (Supplementary Fig. 4c), but even in this model, Cα-Cα cross peaks for both Val24-Val25 and Phe26-Leu28 pairs would be observable. This model is very similar to the ssNMR structure 7K3G, which, in addition to the Cα-Cα contacts of Val24-Val25 (6.4 Å) and Phe26-Leu28 (7.3 Å) that would give rise to observable cross peaks, has a close Phe26 Cδ – Leu27 Cγ distance (5.1 Å) that would generate a strong cross peak. These expectations are inconsistent with our ^{13}C-^{13}C correlation spectra (Supplementary Fig. 6 and Fig. 5). The positive data for ^{13}C-^{13}C correlation between Leu-Val pairs and negative results between Leu-Phe pairs also allowed us to rule out an antiparallel alignment between the two TM helices in dimers, as such dimers could not satisfy both the positive and negative restraints.
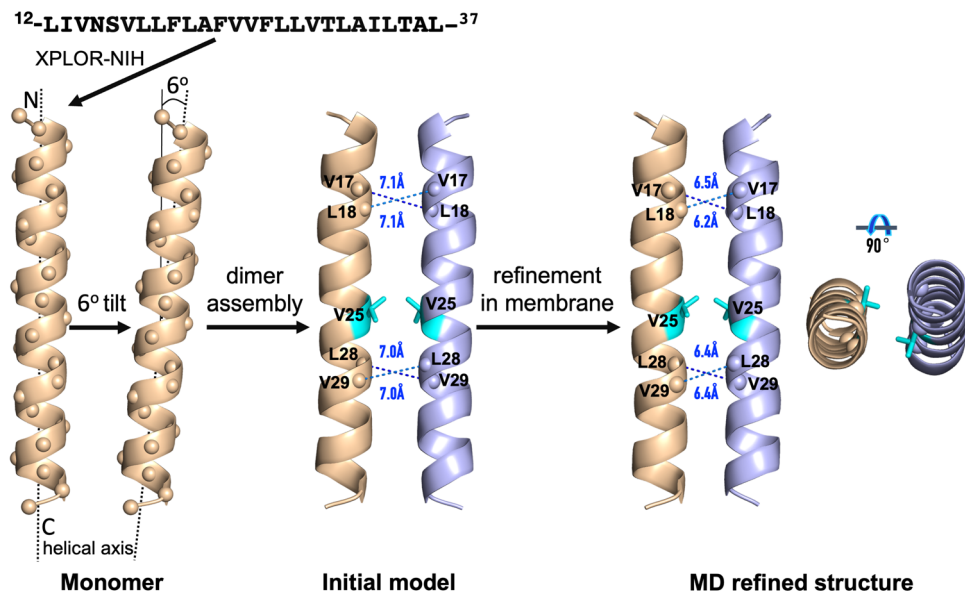
Fig. 6 Refinement of the E$_{12-37}$ dimer by restrained molecular dynamics simulations in POPC/POPG membranes.

Table 1 NMR and refinement statistics for E TM dimer structure.

	E TM Dimer
NMR distance and dihedral constraints	
Distance constraints	
Intermolecular	2 × 2
Hydrogen bonds	44 × 2
Total dihedral angle restraints	
φ	24 × 2
ψ	24 × 2
Structure statistics	
Deviations from idealized geometry	
Bond lengths (Å)	0.11 ± 0.02
Bond angles (°)	9 ± 2
Average pairwise r.m.s. deviation[a] (Å)	
Heavy	0.73 ± 0.08
Backbone	0.38 ± 0.04

[a]Pairwise r.m.s. deviation was calculated for residues 14 to 35 among 14 refined structures.

Dimer structure of E$_{12-37}$ refined in POPC/POPG membranes exposes Asn15 to a water-filled pocket for potential drug binding. Based on the well-defined helix tilt by the OS data and the well-defined dimer interface by the MAS data, we built an initial model by Xplor-NIH[34] and refined the structure by restrained molecular dynamics simulations in POPC/POPG membranes (Fig. 6). The resulting structure satisfies all the NMR restraints, including the helix tilt angle of 6°; Cα-Cα distances ≤ 6.5 Å for all Val17-Leu18 and Leu28-Val29 pairs; no Leu-Phe Cα-Cα distance < 9 Å; and an 8.0 Å distance between Phe26 Cδ and Leu28 Cγ that can account for a weak cross peak between this pair of atoms (Fig. 5). We repeated the refinement simulations 14 times; the resulting models are very close to each other, with backbone root-mean-square deviations (RMSDs) of 0.38 ± 0.04 Å from model 1 over residues Val14-Thr35 (Table 1; Supplementary Fig. 7a).

The Cα atoms of residues Val14-Thr35 are positioned in the hydrophobic region of the membrane. This hydrophobic region is well-suited for the nonpolar sidechains among these residues but may pose a problem for the polar sidechains, of which there are four: Asn15, Ser16, Thr30, and Thr35. The sidechain hydroxyls of

both Thr30 and Thr35 form hydrogen bonds with the backbone carbonyls four residues upstream (Supplementary Fig. 7b). On the other hand, Asn15 and Ser16 are close enough to the membrane surface such that water molecules push away the lipids to form hydrogen bonds with these polar sidechains (Supplementary Fig. 8a). The number of water molecules is only 3 to 7 around each Ser16 sidechain but 8 to 12 around each Asn15 sidechain, generating a water-filled pocket that may serve as a drug-binding site (Supplementary Fig. 8b). In unrestrained molecular dynamics simulations, the dimeric structure is stable, as judged by RMSD values (~3 Å; Supplementary Fig. 9) typical of those observed in simulations of many structured proteins[35].

Discussion

We have combined OS and MAS ssNMR, molecular modeling, and restrained molecular dynamics simulations to determine the dimeric structure of the SARS-CoV-2 E protein in liquid-crystalline lipid bilayers. Resonances in the PISEMA spectra are limited to a small region around 220 ppm in ^{15}N anisotropic chemical shift and 9 kHz in ^{1}H-^{15}N dipolar coupling. Such narrowly limited large values in both dimensions clearly indicate that the entire sequence of E$_{12-37}$ forms a highly regular helix with a small tilt angle. MAS ^{13}C-^{13}C dipolar restraints unequivocally define a symmetric interhelical interface, pointing to a dimer structure. The refined dimer structure of E$_{12-37}$ has a helix tilt angle of only 6°, with an extended interface dominated by Leu and Val sidechains. Residues Val14-Thr35 are positioned in the hydrophobic region of the membrane, but a pocket in the membrane surface allows the Asn15 sidechain to access water. Asn15 has been shown to be the major site for binding the amiloride class of drugs[22]. Because these molecules have a high tendency for lipid partition[36], they may bind to Asn15 from the lipid bilayer.

The first clear indication of dimers was provided by SDS-PAGE gels, where the E protein was solubilized by the detergent SDS. Although we did not directly determine the oligomeric state of E$_{12-37}$ in lipid bilayers, dipolar restraints from ^{13}C-^{13}C correlation spectra allowed us to unequivocally define a front-to-front, symmetric dimer interface. All previous solution NMR characterizations of the E protein in detergent micelles have produced pentameric structures[10,19,20]; a recent ssNMR study also assumed a pentameric model[21]. In SDS-PAGE electrophoresis, E$_{9-35}$

appeared as bands consistent with dimer, trimer, and pentamer, with higher oligomers increased at higher peptide loading levels[6]. In a blue native-PAGE gel, E_{8-65} solubilized in the detergent lyso-myristoyl phosphatidylglycerol (LMPG) migrated as a ladder of oligomers, again with a shift toward higher oligomers at increasing peptide-to-detergent ratios; the largest oligomer was assumed to be pentamer[19]. Parthasarathy et al.[11] fit their analytical ultracentrifugation data for SAR-COV E in the detergent C14 betaine to a monomer-pentamer model. Recently Somberg et al.[37] used the centerband-only detection of exchange (CODEX) technique to determine the oligomeric state of E_{8-38} in lipid bilayers. At a peptide-to-lipid ratio of 1:34, the CODEX data supported pentamer as the oligomeric state. When the peptide-to-lipid ratio was doubled, CODEX provided evidence for the clustering of pentamers. Our ^{13}C-^{13}C dipolar restraints specifically led to the elimination of pentameric models. Contrary to the small tilt angles in our dimer structure and in the pentamer structure of Mandal et al.[21], Park et al.[22] suggested a tilt angle of 45° based on a 1D OS spectrum in DMPC lipids, ostensibly to accommodate a long helix (extending to Glu8 on the N-terminal side) that was determined by solution NMR in detergent micelles. In our ssNMR structure, residues Leu11 and Ile12 are positioned in the membrane interfacial region; though the helix could extend further into the N-terminal side, these residues (M_1YSFV-$SEETGT_{11}$) would be in the extraviral aqueous environment.

The function of the SARS-CoV-2 E protein has often been described as an ion channel in a pentameric state. This putative activity is difficult to understand because the 18-residue long stretch, from Val17 to Leu34, of the TM sequence contains just one weakly hydrophilic residue, Thr30. The resulting highly hydrophobic pore would not be favorable for ion conductance or even for water exposure. Channel pores are generally enriched in polar and charged residues[38]. A recent ssNMR study suggested that opening of the pentameric E channel may involve movements of Phe20 and Phe26 from partially pore-facing to lipid-facing[39]. It is unclear why a highly hydrophobic pore would have evolved for ion conductance. Although ions may be induced to flow through such an unfavorable pore with high concentration gradients, this conductance activity is not likely to be relevant in the viral envelope where relatively few copies of the protein are found[26]. Indeed, the literature has not suggested such activity in the viral envelope.

The TM domains of E orthologs in related viruses are required for both virus assembly and release[40,41]. Alanine insertions in the TM helix of the mouse hepatitis coronavirus (MHV) A59 E protein significantly crippled virus production, and the effect was attributed to the disruption of the alignment of polar sidechains on the surface of the TM helix[40]. Polar residues Gln15, Cys23, and Thr27 in this E protein are all positioned on one face of the TM helix, and the authors suggested that they may participate in interactions with the TM domain of the M protein. Interestingly, the polar face of the MHV E protein aligns with what we have referred to as the Phe face in the SARS-CoV-2 E protein (Supplementary Fig. 4b). This face is away from the dimeric interface and thus available for interacting with other proteins; it also harbors three polar residues of the SARS-CoV-2 E protein: Asn15, Ser16, and Thr30. If this face indeed participates in interactions with the M protein, then the inhibitory effects of the amiloride class of drugs may have an alternative mechanism. Instead of acting as a channel blocker, they inhibit E-M protein binding.

Unlike many TM helical domains that are stabilized by close helix-helix packing afforded by short sidechains (Gly or Ala) at i and $i + 4$ or $i + 7$ positions, the E dimer structure determined here is stabilized by an extended interface comprising mostly of Leu and Val sidechains. This interface has some resemblance to those of leucine zippers, which are distinguished by a Leu residue at every seventh position. Each Leu sidechain at position i of one helix interacts with the corresponding Leu sidechain (at i') and two other hydrophobic sidechains at $i' - 3$ and $i' + 4$ on the opposite helix. Salt bridges between oppositely charged sidechains at $i - 4$ and $i' + 1$ or at $i + 1$ and $i' - 4$ provide further stabilization. These charged sidechains also contribute to solubility in water. The E protein has Leu residues at positions 21 and 28; Leu21 on one helix of the dimer indeed interacts with Leu18, Leu21, and Val25 on the opposite helix (Supplementary Fig. 4b), but that is where the resemblance to a leucine zipper ends. Supercoiling of leucine zippers reduces the helical periodicity to 3.5 residues per turn, whereas straight helices in TM domains have a periodicity of 3.6 residues per turn. As a result, Leu28 does not reach into the interface as deeply as Leu21 and forms hydrophobic interactions with Val25 and Val29 on the opposite helix. If Leu28 reached the same position as Leu21 in the interface, it would interact with Val25, Leu28, and Aal32 on the opposite helix.

The dimer structure of E_{12-37} was determined from samples with a 1:25 monomer-to-lipid ratio, which corresponds to more than two annuli of lipid molecules around each E dimer and leaves little chance for E dimers to encounter each other. With increasing monomer-to-lipid ratio, dimers or other small oligomers might form clusters (as found for the TM domains of the influenza M2 proton channel[42] and the SARS-CoV-2 E protein[37]) or perhaps even reorganize into higher oligomers. While this paper was under review, a preprint was posted reporting the SARS-CoV-2 E protein as a dimer in detergent micelles according to native mass spectrometry[43]. In nanodiscs, small viral proteins have been observed to populate a range of oligomeric states[44]. It may well be that in the ERGIC where the virus is assembled and the E protein is present at high concentrations[15], it forms pentamers, but in the mature virus where it is present at a low copy number[26], it only forms dimers. Viral proteins often perform multiple functions. The change in oligomeric state may facilitate that.

Materials and Methods

Gene cloning. The gene sequences of all E fragments were based on the SARS-CoV-2 isolate Wuhan-Hu-1 (NC_045512). As illustrated in Supplementary Fig. 1a, four fragments E_{12-37}, E_{1-52}, E_{7-43}, and E_{8-41} were separately cloned into the pTBMalE vector[45] following our previous work on influenza A M2 protein[46,47]. For the fifth fragment E_{12-65}, the pTBGST vector[45] was used as illustrated in Supplementary Fig. 1b. All gene constructs contained a 6 × His-tag for protein purification and a TEV cleavage site (ENLYFQSNA) for fusion protein removal. Cysteines in the E fragments were mutated to serine for E_{1-52}, E_{7-43}, and E_{8-41} and to alanine for E_{12-65}. Cloned gene constructs were all verified by DNA sequencing.

Protein expression and purification. *E. coli* BL21(DE3)RP cells were used to express all E constructs. Similar expression conditions were used for vectors containing GST and MBP fusion proteins. In short, 100 mL each of overnight Luria Broth (LB) culture was evenly divided into two 1-L LB media and grew at 37 °C with vigorous shaking to an OD_{600} of ~0.8. The cells were spun down and washed with 30 mL of sterile M9 media. They were then resuspended in 500 mL of M9 media supplemented with 100 mg of a desired ^{15}N-labeled or ^{13}C-labeled L-amino acid (Cambridge Isotope Labs, Inc. or Sigma-Aldrich, Inc.) as well as the remaining 19 unlabeled amino acids. Cultures grew at 37 °C for another 20 min and over-expression was induced for 1.5 hr with 0.4 mM IPTG. Upon over-expression, the cells were

harvested and resuspended in 60 mL of buffer (20 mM Tris-HCl, pH 8.0, 500 mM NaCl) prior to storage at −80 °C. For the expression of uniformly ^{15}N-labeled proteins, the sole nitrogen source in the M9 minimal media was replaced with ^{15}NH$_4$Cl, and the induction time was increased to 8 hr.

Cells were thawed in a 37 °C water bath prior to adding 0.25 mg/mL of lysozyme and 5 units of Benzonase nuclease (Millipore Sigma) per mL of cells. Cells were fully lysed by sonication (Fisher-100) using three cycles of 1.5 min pulsing and 1.5 min rest. 3% Empigen (Millipore Sigma) was then added and the tube was rotated at 4 °C for 2-3 hr prior to centrifugation at 10,000 g for 30 min. The supernatant was mixed with 20 mL nickel affinity resin (Qiagen) and rotated at 4 °C overnight.

The column was washed initially with buffer (20 mM Tris-HCl, pH 8.0, 500 mM NaCl, 3% Empigen, and 3 mM imidazole) and then for the second and third wash with buffer containing reduced Empigen (1.5% and 0.75%, respectively). For MBP fusion protein purification, the final wash used a buffer without Empigen (20 mM Tris-HCl pH 8.0, 500 mM NaCl). For GST, the Empigen in the final wash was replaced with DPC (20 mM Tris-HCl pH 8.0, 500 mM NaCl, 0.2% DPC), until the UV absorbance returned to baseline. The protein was eluted with buffer (20 mM Tris-HCl pH 8.0, 500 mM NaCl, and 200 mM imidazole). Generally, 160–200 mg of MBP fusion protein and ~80 mg of GST fusion protein were obtained per L of culture, respectively.

The MBP fusion protein was digested with Tobacco Etch Virus (TEV) protease at a protein-to-protease molar ratio of 1:2. The mixture was precipitated by adding 6% trifluoroacetic acid. The protein was then pelleted and washed with ddH$_2$O twice, 100 mM Tris-HCl pH 8.0 twice, and finally ddH$_2$O. The wet pellet was placed in a vacuum chamber overnight to remove any water content. The following day, methanol was added to the dried protein powder to extract the E fragment. Upon centrifugation at 15,000 g, methanol containing the E fragment was carefully removed and placed under a stream of nitrogen gas to remove the bulk methanol. Residual methanol was removed by vacuum overnight and the E fragment was ready for NMR sample preparation. The typical yield of E$_{12-37}$ was ~ 8–10 mg/L of culture.

SDS-PAGE electrophoresis. The oligomeric states of the E fragments were assessed using SDS-PAGE electrophoresis. SDS buffer (2% SDS, 100 mM DTT, 10% glycerol, 100 mM Tris-HCl, and 0.01% bromophenol blue dye) was added to each E fragment preparation. After brief vortexing, the sample was loaded onto a 12% SDS-PAGE gel (Tris-Tricine). The sample was electrophoresed at room temperature at a constant voltage of 200 V for 40 min. After completion, the SDS-PAGE gel was stained with EZ-blue (Bio-Rad). Uncropped and unedited gel images are presented in Supplementary Fig. 10a–f.

NMR sample preparation. For oriented samples, lipid films were made using 25 mg of chloroform-dissolved POPC/POPG lipids (Avanti Polar Lipids, Inc.) in a 4:1 molar ratio. Most of the chloroform was removed by passing nitrogen gas over the sample. The remaining chloroform was removed in a vacuum chamber overnight. 1 mL of trifluoroethanol (TFE) was added to 4 mg of an E fragment and 2 mL of TFE was added to 25 mg of lipid film separately. They were then mixed at a 1:25 protein-to-lipid molar ratio and the solution was adjusted to pH 7.5 by adding a few µL of 0.5 N NaOH. The mixture was spun at 90 rpm for 2-3 h and then spread evenly on 25 thin glass slides (5.7 mm × 10 mm × 0.04–0.06 mm, Matsunami Glass, Inc.). The glass slides were left in the vacuum chamber overnight to remove any trace of TFE, and then transferred to a hydration chamber (23 °C at 98% relative humidity via a saturated solution of potassium sulfate) for 2 h. All glass slides were stacked together into a

glass cell (6.3 mm × 4.3 mm × 18 mm, New Era Scientific Inc.) and sealed with a homemade plastic plug and beeswax (Hampton Research Inc.). MAS samples were prepared in a similar manner. Instead of spreading over glass slides, the hydrated proteoliposomes were packed into 3.2 mm thin-walled MAS rotors (Revolution NMR LLC).

Solid-state NMR spectroscopy. All OS ssNMR experiments were performed on a Bruker NEO 600 NMR spectrometer with a Larmor frequency of 600.13 MHz for ^{1}H. A custom-built low-electrical-field static probe with a rectangular coil was used. 1D cross-polarization and 2D PISEMA spectra were collected at 22 °C. The initial ^{1}H 90° pulse strength was set to 50 kHz. The cross-polarization contact time was set to 1000 ms. A 50 kHz B$_1$ field was used for both Lee–Goldburg irradiation and the Spinal-64 decoupling scheme during acquisition[48]. The frequency jumps used to satisfy the Lee–Goldburg off-resonance condition were set to ±40.8 kHz. The number of t1 increments was 16; the number of scans per t1 increment was 3072 to 5128 depending on sample sensitivity.

For MAS NMR, a custom-built 3.2 mm low-E double-resonance MAS probe[49] was used on a Bruker Avance 600 NMR spectrometer. The sample spinning rate was maintained at 10 kHz ± 3 Hz and the reading temperature was set to 228 K. The ^{13}C magnetization was enhanced by cross polarization with a contact time of 1 ms. The ^{1}H 90° pulse length was 3.2 µs before cross polarization. A phase alternated recoupling irradiation scheme[50], replacing the traditional dipolar assisted rotational resonance pulse sequence[51,52], was utilized with different mixing times (up to 600 ms) to identify interhelical cross peaks. The ^{13}C chemical shifts were referenced to the carbonyl carbon resonance of glycine at 178.4 ppm relative to tetramethylsilane.

Search for oligomer models. An ideal α-helix was built for the 26 Cα atoms of E$_{12-37}$ using the following parameters: 3.6 residues per turn (corresponding to a 100° increment per residue on the helical wheel), 2.3 Å radius, and 1.5 Å rise per residue. The sequence of E$_{12-37}$ was assigned to the Cα atoms on the helix, with an arbitrary initial rotation of the helix. A duplicate was then placed to the right of the first helix, with a 5.1 Å spacing between the Cα rims of the helices[53]. The second helix was also self-rotated by an angle appropriate for a given oligomeric state: 180° for a dimer and 72° for a pentamer.

The rotation angles of the two helices were scanned from 0 to 360° with a 10° interval. For each helix rotation angle, interhelical Cα-Cα distances were calculated to identify pairs that could yield ^{13}C-^{13}C cross peaks, with a cutoff of 7.2 Å[33]. Models were filtered according to the following criteria: (1) no ^{13}C-^{13}C cross peaks between residues 24 and 25; (2) no ^{13}C-^{13}C cross peaks between any Leu-Phe pairs; (3) at least one ^{13}C-^{13}C cross peak between a Leu-Val pair.

Structure refinement for E$_{12-37}$ dimer. A monomer model of E$_{12-37}$ was built as an α-helix using Xplor-NIH[34], with backbone phi and psi angles assigned in the ranges −65° to −60° and −45° to −40°, respectively. The monomer was oriented with a 6° tilt angle with respect to the z axis; a duplicate was rotated to generate a dimer with C2 rotational symmetry. The dimer was subjected to optimization by simulated annealing, with the following harmonic restraints: (1) backbone phi and psi angles, restrained to the above assigned values; (2) backbone i to $i + 4$ hydrogen bonding, with O-N distance restrained to 3.0 Å and O-H distance restrained to 2.0 Å; (3) interhelical Cα-Cα distances between Val17-Leu18, Leu28-Val29, and Val25-Val25 pairs, restrained to 6.0 Å. During the simulated annealing, the

temperature was cooled from 3500 K to 25 K in 1000 steps, and the force constants for the restraints were ramped up from 5 to 1000 kcal/mol rad^{-2} for phi and psi angles and from 2 to 30 kcal/mol Å^{-2} for distances. The annealing simulation was repeated 100 times; a dimer model was selected based on considerations of the helix tilt angle (~6°) and the distances between Val17-Leu18 and Leu28-Val29 pairs (~7 Å).

Molecular dynamics simulations for the refinement of the initial model were run in NAMD[54] with the CHARMM36 force field[55]. Using the CHARMM-GUI web server[56], the initial model was placed in a POPC/POPG bilayer (4:1 ratio; 100 lipids per leaflet). The system was solvated with 8453 TIP3P water molecules, along with Na$^+$ and Cl$^-$ ions for charge neutralization and for generating a salt concentration of 30 mM, resulting in a simulation box of 83.6 × 83.6 × 81.5 Å^3. After energy minimization (10000 cycles of conjugate gradient), simulations were run in four segments, with the first two at constant temperature and volume and the second two at constant temperature and pressure, for 125, 125, 125, and 250 ps, respectively. During the first three segments, restraints on the lipids were gradually reduced, but positional restraints on the protein backbone heavy atoms were maintained with a force constant of 10 kcal mol^{-1} Å^{-2}. In the final segment of the refinement simulation (250 ps at constant temperature and pressure), the positional restraints on the protein were removed; instead the three sets of restraints listed as (1) to (3) in the preceding paragraph were imposed, with force constants of 5 kcal mol^{-1} Å^{-2} for backbone hydrogen bonding, 10 kcal mol^{-1} Å^{-2} for interhelical distances, and 5 kcal mol^{-1} rad^{-2} for phi and psi angles (minimum at −60° for phi and −45° for psi). The tilt angles of the helices were restrained to 6° with a force constant of 10 kcal mol^{-1}rad^{-2}. One last set of restraints was imposed based on the PISEMA spectra. Because the resonances of all the residues had very small dispersions (Fig. 3), we assigned all residues of a given amino-acid type (e.g., three Phe residues) to the overall peak position of the PISEMA spectrum acquired with ^{15}N labeling of that amino-acid type (Supplementary Table 1). The exceptions were Thr35 and Ala36, which were assigned to a distinct peak position (Fig. 3). The force constants for ^{15}N anisotropic chemical shifts and ^{1}H–^{15}N dipolar couplings were 0.01 kcal mol^{-1} ppm^{-2} and 1 kcal mol^{-1} kHz^{-2}, respectively. Frames were saved at 5 ps intervals; the frame with the helix tilt angles closest to 6° was selected as the refined structure. This 250-ps segment was repeated to generate 14 models for deposition to the Protein Data Bank (PDB).

The time step was 1 fs in all the simulations. All bonds connected to hydrogens were constrained by the SHAKE algorithm[57]. Long-range electrostatic interactions were treated by the particle mesh Ewald method[58]. The cutoff distance for nonbonded interaction was 12 Å, with force switching at 10 Å for van der Waals interactions. The Langevin thermostat with a damping constant of 1 ps^{-1} was used to maintain the temperature at 310 K and the Langevin piston[59] was used to maintain the pressure at 1 atm.

To check the stability of the dimeric structure, we continued the simulations of four models from the last frame of the refinement simulations but without any restraints. These unrestrained simulations were run at constant temperature and pressure for 100 ns. Frames were saved at 100 ps intervals. RMSDs were calculated on Cα atoms of residues Val14-Thr35 using a tcl script in VMD[60], in reference to the starting structures.

Statistics and reproducibility. Fourteen models of the E TM dimer were calculated from replicate refinement MD simulations. The mean and standard deviation of the RMSDs from model 1 were calculated for the other 13 models. To verify the stability of

the dimeric structure, we also carried four replicate unrestrained simulations. The replicates demonstrate reproducibility.

Reporting summary. Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

All data generated or analyzed during this study are included in this published article (and its supplementary data files). The source data for all the plots presented in figures are deposited in GitHub at https://github.com/hzhou43/SARS-CoE, with https://doi.org/10.5281/zenodo.10015217. The structure of the E TM dimer has been deposited to the PDB (entry name 8T1U). All other data are available from the corresponding author on reasonable request.

Code availability

Data analysis procedures were described under Materials and Methods. Computer programs used were cited and publicly available. The codes to scan for all possible helix-helix arrangements in dimers and pentamers, the input files for MD simulations of the E TM dimer, and the initial and final coordinate files are deposited in GitHub at https://github.com/hzhou43/SARS2-CoE/, with https://doi.org/10.5281/zenodo.10015217.

References

1. World Health Organization https://www.who.int/emergencies/diseases/novel-coronavirus-2019 (2023).
2. Wu, F. et al. A new coronavirus associated with human respiratory disease in China. *Nature* **579**, 265–269 (2020).
3. Aronin, S. I. & Sadigh, M. Severe acute respiratory syndrome. *Conn. Med.* **68**, 207–215 (2004).
4. de Groot, R. J. et al. Commentary: Middle East Respiratory Syndrome Coronavirus (MERS-CoV): announcement of the coronavirus study group. *J. Virol.* **87**, 7790–7792 (2013).
5. Wilson, L., McKinlay, C., Gage, P. & Ewart, G. SARS coronavirus E protein forms cation-selective ion channels. *Virology* **330**, 322–331 (2004).
6. Torres, J., Wang, J., Parthasarathy, K. & Liu, D. X. The transmembrane oligomers of coronavirus protein E. *Biophys. J.* **88**, 1283–1290 (2005).
7. Liao, Y., Yuan, Q., Torres, J., Tam, J. P. & Liu, D. X. Biochemical and functional characterization of the membrane association and membrane permeabilizing activity of the severe acute respiratory syndrome coronavirus envelope protein. *Virology* **349**, 264–275 (2006).
8. Torres, J. et al. Model of a putative pore: the pentameric alpha-helical bundle of SARS coronavirus E protein in lipid bilayers. *Biophys. J.* **91**, 938–947 (2006).
9. Parthasarathy, K. et al. Structural flexibility of the pentameric SARS coronavirus envelope protein ion channel. *Biophys. J.* **95**, L39–L41 (2008).
10. Pervushin, K. et al. Structure and inhibition of the SARS coronavirus envelope protein ion channel. *PLoS Pathog.* **5**, e1000511 (2009).
11. Parthasarathy, K. et al. Expression and purification of coronavirus envelope proteins using a modified β-barrel construct. *Protein Expr. Purif.* **85**, 133–141 (2012).
12. Venkatagopalan, P., Daskalova, S. M., Lopez, L. A., Dolezal, K. A. & Hogue, B. G. Coronavirus envelope (E) protein remains at the site of assembly. *Virology* **478**, 75–85 (2015).
13. Alam, I. et al. Functional pangenome analysis shows key Features of E protein are preserved in SARS and SARS-CoV-2. *Front. Cell Infect. Microbiol.* **10**, 405 (2020).
14. Verdiá-Báguena, C. et al. Coronavirus E protein forms ion channels with functionally and structurally-involved membrane lipids. *Virology* **432**, 485–494 (2012).
15. Nieto-Torres, J. L. et al. Subcellular location and topology of severe acute respiratory syndrome coronavirus envelope protein. *Virology* **415**, 69–82 (2011).
16. DeDiego, M. L. et al. A severe acute respiratory syndrome coronavirus that lacks the E gene is attenuated in vitro and in vivo. *J. Virol.* **81**, 1701–1713 (2007).
17. Kuo, L., Hurst, K. R. & Masters, P. S. Exceptional Flexibility in the Sequence Requirements for Coronavirus Small Envelope Protein Function. *J. Virol.* **81**, 2249–2262 (2007).

18. Ortego, J., Ceriani, J. E., Patiño, C., Plana, J. & Enjuanes, L. Absence of E protein arrests transmissible gastroenteritis coronavirus maturation in the secretory pathway. *Virology* **368**, 296–308 (2007).

19. Surya, W., Li, Y. & Torres, J. Structural model of the SARS coronavirus E channel in LMPG micelles. *Biochim. Biophys. Acta Biomembr.* **1860**, 1309–1317 (2018).

20. Li, Y., Surya, W., Claudine, S. & Torres, J. Structure of a conserved golgi complex-targeting signal in coronavirus envelope proteins. *J. Biol. Chem.* **289**, 12535–12549 (2014).

21. Mandala, V. S. et al. Structure and drug binding of the SARS-CoV-2 envelope protein transmembrane domain in lipid bilayers. *Nat. Struct. Mol. Biol.* **27**, 1202–1208 (2020).

22. Park, S. H. et al. Interactions of SARS-CoV-2 envelope protein with amilorides correlate with antiviral activity. *PLoS Pathog.* **17**, e1009519 (2021).

23. Zhou, H. X. & Cross, T. A. Influences of membrane mimetic environments on membrane protein structures. *Annu. Rev. Biophys.* **42**, 361–392 (2013).

24. Laguerre, A. et al. From nanodiscs to isotropic bicelles: a procedure for solution nuclear magnetic resonance studies of detergent-sensitive integral membrane proteins. *Structure* **24**, 1830–1841 (2016).

25. Denisov, I. G. & Sligar, S. G. Nanodiscs in membrane biochemistry and biophysics. *Chem. Rev.* **117**, 4669–4713 (2017).

26. Renz, A., Widerspick, L. & Drager, A. Genome-scale metabolic model of infection with SARS-CoV-2 mutants confirms Guanylate Kinase as robust potential antiviral target. *Genes* **12**, 796 (2021).

27. Wu, C. H., Ramamoorthy, A. & Opella, S. J. High-resolution heteronuclear dipolar solid-state NMR spectroscopy. *J. Magn. Reson. A* **109**, 270–272 (1994).

28. Page, R. C., Kim, S. & Cross, T. A. Transmembrane helix uniformity examined by spectral mapping of torsion angles. *Structure* **16**, 787–797 (2008).

29. Kim, S. & Cross, T. A. 2D solid state NMR spectral simulation of 3₁₀, α, and π-helices. *J. Magn. Reson* **168**, 187–193 (2004).

30. Marassi, F. M. & Opella, S. J. A solid-state NMR index of helical membrane protein structure and topology. *J. Magn. Reson* **144**, 150–155 (2000).

31. Wang, J. et al. Imaging membrane protein helical wheels. *J. Magn. Reson* **144**, 162–167 (2000).

32. Castellani, F. et al. Structure of a protein determined by solid-state magic-angle-spinning NMR spectroscopy. *Nature* **420**, 99–102 (2002).

33. Zech, S. G., Wand, A. J. & McDermott, A. E. Protein structure determination by high-resolution solid-state NMR spectroscopy: application to microcrystalline ubiquitin. *J. Am. Chem. Soc.* **127**, 8618–8626 (2005).

34. Schwieters, C. D., Kuszewski, J. J., Tjandra, N. & Clore, G. M. The Xplor-NIH NMR molecular structure determination package. *J. Magn. Reson* **160**, 65–73 (2003).

35. Meyer, T. et al. MoDEL (Molecular Dynamics Extended Library): a database of atomistic molecular dynamics tTrajectories. *Structure* **18**, 1399–1409 (2010).

36. Simchowitz, L., Woltersdorf, O. W. Jr & Cragoe, E. J. Jr Intracellular accumulation of potent amiloride analogues by human neutrophils. *J. Biol. Chem.* **262**, 15875–15885 (1987).

37. Somberg, N. H. et al. SARS-CoV-2 Envelope protein forms clustered pentamers in lipid bilayers. *Biochemistry* **61**, 2280–2294 (2022).

38. Nugent, T. & Jones, D. T. Detecting pore-lining regions in transmembrane protein sequences. *BMC Bioinform.* **13**, 169 (2012).

39. Medeiros-Silva, J. et al. pH- and calcium-dependent aromatic network in the SARS-CoV-2 envelope protein. *J. Am. Chem. Soc.* **144**, 6839–6850 (2022).

40. Ye, Y. & Hogue, B. G. Role of the coronavirus E viroporin protein transmembrane domain in virus assembly. *J. Virol.* **81**, 3597–3607 (2007).

41. Machamer, C. E. & Youn, S. The transmembrane domain of the infectious bronchitis virus E protein is required for efficient virus release. *Adv. Exp. Med Biol.* **581**, 193–198 (2006).

42. Paulino, J., Pang, X., Hung, I., Zhou, H. X. & Cross, T. A. Influenza A M2 channel clustering at high protein/lipid ratios: viral budding implications. *Biophys. J.* **116**, 1075–1084 (2019).

43. Townsend, J. A. et al. Differences in oligomerization of the SARS-CoV-2 envelope protein, poliovirus VP4, and HIV Vpu. *bioRxiv* https://doi.org/10.1101/2023.08.18.553902 (2023).

44. Townsend, J. A. et al. Influenza AM2 channel Oligomerization is sensitive to its chemical environment. *Anal. Chem.* **93**, 16273–16281 (2021).

45. Qin, H. et al. Construction of a series of vectors for high throughput cloning and expression screening of membrane proteins from Mycobacterium tuberculosis. *BMC Biotechnol.* **8**, 51 (2008).

46. Hu, J. et al. Structural biology of transmembrane domains: efficient production and characterization of transmembrane peptides by NMR. *Protein Sci.* **16**, 2153–2165 (2007).

47. Sharma, M. et al. Insight into the mechanism of the influenza a proton channel from a structure in a lipid bilayer. *Science* **330**, 509–512 (2010).

48. Fung, B. M., Khitrin, A. K. & Ermolaev, K. An improved broadband decoupling sequence for liquid crystals and solids. *J. Magn. Reson.* **142**, 97–101 (2000).

49. Gor'kov, P. L. et al. Using low-E resonators to reduce RF heating in biological samples for static solid-state NMR up to 900 MHz. *J. Magn. Reson.* **185**, 77–93 (2007).

50. Weingarth, M., Demco, D. E., Bodenhausen, G. & Tekely, P. Improved magnetization transfer in solid-state NMR with fast magic angle spinning. *Chem. Phys. Lett.* **469**, 342–348 (2009).

51. Takegoshi, K., Nakamura, S. & Terao, T. C-13-H-1 dipolar-assisted rotational resonance in magic-angle spinning NMR. *Chem. Phys. Lett.* **344**, 631–637 (2001).

52. Takegoshi, K., Nakamura, S. & Terao, T. C-13-H-1 dipolar-driven C-13-C-13 recoupling without C-13 rf irradiation in nuclear magnetic resonance of rotating solids. *J. Chem. Phys.* **118**, 2325–2341 (2003).

53. Dai, J. & Zhou, H. X. General rules for the arrangements and gating motions of pore-lining helices in homomeric ion channels. *Nat. Commun.* **5**, 4641 (2014).

54. Phillips, J. C. et al. Scalable molecular dynamics with NAMD. *J. Comput. Chem.* **26**, 1781–1802 (2005).

55. Klauda, J. B. et al. Update of the CHARMM all-atom additive force field for lipids: validation on six lipid types. *J. Phys. Chem. B* **114**, 7830–7843 (2010).

56. Jo, S., Kim, T., Iyer, V. G. & Im, W. CHARMM-GUI: A web-based graphical user interface for CHARMM. *J. Comput. Chem.* **29**, 1859–1865 (2008).

57. Ryckaert, J.-P., Ciccotti, G. & Berendsen, H. J. C. Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *J. Comput. Phys.* **23**, 327–341 (1977).

58. Essmann, U. et al. A smooth particle mesh Ewald method. *J. Chem. Phys.* **103**, 8577–8593 (1995).

59. Feller, S. E., Zhang, Y. H., Pastor, R. W. & Brooks, B. R. Constant-pressure molecular-dynamics simulation - the Langevin Piston method. *J. Chem. Phys.* **103**, 4613–4621 (1995).

60. Humphrey, W., Dalke, A. & Schulten, K. VMD: visual molecular dynamics. *J. Mol. Graph* **14**, 33–38 (1996). 27-8.

Acknowledgements

Author contributions
T.A.C., R.F., and H.-X.Z. designed the research; R.Z., H.Q., and R.P. performed the research and analyzed the data; T.A.C., R.Z., and H.-X.Z. wrote the manuscript.

Competing interests
The authors declare no competing interests.

Additional information
Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s42003-023-05490-x.

Correspondence and requests for materials should be addressed to Huan-Xiang Zhou or Timothy A. Cross.

Peer review information *Communications Biology* thanks the anonymous reviewers for their contribution to the peer review of this work. Primary Handling Editors: Gene Chong and David Favero.

Reprints and permission information is available at http://www.nature.com/reprints

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