<table>
<thead>
<tr>
<th>Title</th>
<th>Structure of a conserved Golgi complex-targeting signal in coronavirus envelope proteins</th>
</tr>
</thead>
<tbody>
<tr>
<td>Author(s)</td>
<td>Li, Yan; Surya, Wahyu; Claudine, Stephanie; Torres, Jaume</td>
</tr>
<tr>
<td>Date</td>
<td>2014</td>
</tr>
<tr>
<td>URL</td>
<td><a href="http://hdl.handle.net/10220/25655">http://hdl.handle.net/10220/25655</a></td>
</tr>
<tr>
<td>Rights</td>
<td>© 2014 The American Society for Biochemistry and Molecular Biology. This is the author created version of a work that has been peer reviewed and accepted for publication by The Journal of Biological Chemistry, The American Society for Biochemistry and Molecular Biology. It incorporates referee’s comments but changes resulting from the publishing process, such as copyediting, structural formatting, may not be reflected in this document. The published version is available at: [<a href="http://dx.doi.org/10.1074/jbc.M114.560094">http://dx.doi.org/10.1074/jbc.M114.560094</a>].</td>
</tr>
</tbody>
</table>
Structure of a Conserved Golgi Complex-targeting Signal in Coronavirus Envelope Proteins

Yan Li¹, Wahyu Surya¹, Stephanie Claudine and Jaume Torres*

School of Biological Sciences, Nanyang Technological University, 637551, Singapore.

Running Title: Structure of a coronavirus envelope protein

¹Equal contribution
*To whom correspondence should be addressed: Jaume Torres; School of Biological Sciences, Nanyang Technological University, 60 Nanyang Drive, 637551 Singapore; Tel: +65 6316 2857; Fax: +65 6791 3856; E-mail: jtorres@ntu.edu.sg
Capsule

Background: Coronavirus envelope (CoV E) proteins have a predicted β-coil-β motif reported to target the Golgi-complex.

Result: This conserved domain forms β-structure on its own, but is α-helical in the context of full length SARS-CoV E protein.

Conclusion: This domain is potentially involved in large conformational transitions.

Significance: This is the first structural data of the extramembrane domain of any coronavirus E protein.

ABSTRACT

Coronavirus envelope (CoV E) proteins are ~100-residue polypeptides with at least one channel-forming α-helical transmembrane (TM) domain. The extramembrane C-terminal tail contains a completely conserved proline, at the center of a predicted β-coil-β motif. This hydrophobic motif has been reported to constitute a Golgi-targeting signal, or a second TM domain. However, no structural data for this, or other extramembrane domains in CoV E proteins, is available. Herein, we show that the E protein in the severe acute respiratory syndrome (SARS) virus has only one TM domain in micelles, whereas the predicted β-coil-β motif forms a short membrane-bound α–helix connected by a disordered loop to the TM domain. However, complementary results suggest that this motif is potentially poised for conformational change, or in dynamic exchange with other conformations.

INTRODUCTION

Coronaviruses (CoV; order Nidovirales, family Coronaviridae, subfamily Coronavirinae) are enveloped viruses organized into three groups (1-3): group 1 (α-coronaviruses), group 2 (β-coronaviruses) and group 3 (γ-coronaviruses). Coronaviruses have been known to cause common cold symptoms in humans, and a variety of lethal diseases in birds and mammals (4). However, in 2003, the virus responsible for the severe acute respiratory syndrome (SARS-CoV) (5) produced a near pandemic with 8,273 cases and 775 deaths (6). In 2012, a novel β-coronavirus (HCoV-EMC) (7-9) was discovered that has already led to many fatalities (10-12). The main coronavirus structural proteins are S (spike), E (envelope), M (membrane) and N (nucleocapsid), where S, E, and M are integral membrane proteins. E proteins are ~100 residues long polypeptides which are minor components in virions but are abundantly expressed inside infected cells (13). They have a short hydrophilic N-terminus, at least one predicted terminal transmembrane (TM) domain and a less hydrophobic C-terminal tail. Co-expression of E and M proteins is sufficient for formation and release of virus-like particles (VLP) (14-19), and E proteins have been proposed to participate in inducing membrane curvature, or in the scission of particles (20). Mutations at their C-terminal extramembrane domain impair viral assembly and maturation in the murine hepatitis virus (MHV) (20). In the transmissible gastroenteritis virus (TGEV), the absence of E protein resulted in a blockade of virus trafficking in the secretory pathway and prevention of virus maturation (21,22). In the case of the severe acute respiratory syndrome coronavirus (SARS-CoV), viruses lacking the E gene showed attenuation and did not grow in the central nervous system (23,24), whereas stress response genes were upregulated and cell apoptosis increased (25). These results suggested a role of E protein in both tissue tropism and pathogenicity, where modulation of stress responses contribute to viral attenuation. In fact, SARS-CoV ΔE attenuated viruses constitute promising vaccine candidates (26-28). Thus, although E proteins are not absolutely essential for in vitro or in vivo coronavirus replication, their absence is clearly deleterious. Immunofluorescence experiments have shown that SARS-CoV E has a cytoplasmically oriented C-terminus and a luminal N-terminus, i.e., an N<sub>exo</sub>C<sub>cyto</sub> orientation (13), which is consistent with the presence of a single TM domain.

SARS-CoV E protein and other CoV E proteins have channel activity in synthetic membranes (29-31). This channel activity is mediated by
formation of pentameric oligomers (32-34) and is only very mildly selective for cations (35). The only available structural data for CoV E proteins, obtained using synthetic TM peptides, is derived from the channel-forming TM domain in SARS-CoV E (32,34,36). No structural data is available for the predicted N– or C–terminal extramembrane domains despite the latter being critical for viral assembly (37,38), although some results have been obtained using shorter synthetic peptides encompassing this domain (39). The C-terminal domain of E proteins contains a totally conserved proline residue, which in β- and γ-coronaviruses is at the centre of a predicted β-strand-coil-β-strand motif (Fig. 1A), reminiscent of viral internal fusion peptides (40-42). In SARS-CoV E, this motif was found to be responsible for redirecting a plasma membrane protein to the Golgi region. Conversely, mutations designed to increase its α-helical propensity disrupted localization to membranes (43). The C-terminal tail of E proteins is also important for its interaction with the C-terminal domain of M protein (18,44,45) at the cytoplasmic side of the ER-Golgi intermediate compartment (ERGIC), the budding compartment of the host cell. These interactions are the major drivers for envelope formation (46). The C-terminal tail of SARS-CoV E protein also interacts with the cellular protein PALS1 (47), which is thus depleted from the tight junctions in epithelial cells. Finally, SARS-CoV E interacts with the 7-domain (48) SARS non-structural protein 3 (nsp3) (49).

Thus, it is crucial to determine the structure of C-terminal predicted extramembrane domain of E proteins, especially in the context of a large construct that includes the TM domain. Until now, these structural studies have been hampered by problems in expression, purification and stabilization of E proteins. We report herein the first detailed structure of a truncated form of the SARS-CoV E monomer that includes both its TM domain and its predicted β-coil-β motif at the C-terminal tail.

MATERIALS AND METHODS

SARS CoV E protein constructs. A full length and a truncated version of SARS CoV E were used in this work. Full-length SARS-CoV E protein (EFL) was cloned into pTBMalE plasmid downstream of the MBP fusion tag and TEV cleavage sequence. The plasmid was transformed into E. coli strain BL21(DE3) codon plus for protein expression. The truncated form, ETR, was cloned into pNIC28-Bsa4 with an N-terminus 6-His tag followed by a TEV cleavage sequence. The plasmid was transformed into E. coli strain BL21(DE3) Rosetta T1R for protein expression. In both constructs, all three native cysteines (C40, C43 and C44) were mutated into alanines. In addition, two EFL mutants previously described by Cohen et al. (43) were prepared by site-directed mutagenesis, (i) P54A (E P54A) and (ii) V56A-Y57A-V58A-Y59A (E4ALA).

Protein expression and purification. Non-labeled E protein was produced by growing the culture in terrific broth (TB) media at 37°C until the culture density reached an OD600 of 2. Protein expression was induced by adding 0.5 mM IPTG and growing the culture overnight at 18°C. The cells were harvested by centrifugation at 7500×g and stored at -80°C. Stable isotope-labeled E protein was produced by growing the culture in LB media at 37°C until the culture density reached an OD600 of 2. Protein expression was induced by adding 0.5 mM IPTG and growing the culture overnight at 18°C. The cells were harvested by centrifugation at 7500×g and stored at -80°C. Stable isotope-labeled E protein was produced by growing the culture in LB media at 37°C. When the culture density reached an OD600 of 0.7, the media was exchanged to M9 minimal media at 25% of the initial volume to achieve a high-density culture as described previously (50). The M9 media was appropriately supplemented with 15N-NH4Cl and 13C-glucose (Cambridge Isotope Laboratories) to produce 15N-labeled and 15N/13C-labeled protein. Cultures were further grown for 1 h before inducing protein expression with 0.5 mM IPTG at 18°C. After 6 h, cells were harvested as described above and stored at -80°C. Frozen cell pellets were resuspended in lysis buffer (20 mM Tris pH 8.0, 300 mM NaCl, 5 mM imidazole, 2 mM β-mercaptoethanol and 10% glycerol) supplemented with 1 mM PMSF and 1.5% Triton X-100. The cells were completely lysed by sonication and microfluidization. Insoluble particles were
removed by centrifugation at 40,000×g and the supernatant was applied onto a pre-equilibrated Ni-NTA resin (Bio-Rad Profinity IMAC Ni²⁺-charged). The resin was washed with 20 mM Tris pH 8.0, 300 mM NaCl, 20 mM imidazole, 2 mM β-mercaptoethanol and 10% glycerol. Bound peptide was eluted in 20 mM Tris pH 8.0, 300 mM NaCl, 250 mM imidazole, 2 mM β-mercaptoethanol, 10% glycerol and 5 mM myristyl sulfobetaine (C14SB). Eₜ₅ protein was directly TCA-precipitated and lyophilized, while EFL was subjected to TEV protease cleavage for 3 hrs at 30°C prior to precipitation and lyophilization. Further purification was achieved by using reversed-phase HPLC on a Phenomenex Jupiter C4 semipreparative column (250 × 10 mm, 300 Å pore size, 5 μm particle size). Lyophilized peptide was dissolved with 1% TFA in acetonitrile and separated under isopropanol-acetonitrile linear gradient (4/1 (v/v) with 0.1% TFA). The identity and purity of peptide fractions were confirmed by SDS-PAGE and MALDI-TOF MS.

**Gel electrophoresis.** Standard SDS-PAGE was performed in 13.5% Tris-glycine gel with TGS running buffer, and stained with Coomassie Blue G-250. SDS-NuPAGE was performed in 4-12% NuPAGE® Bis-Tris gel (Invitrogen) with NuPAGE® MES SDS running buffer and stained with SimplyBlue™ SafeStain (Invitrogen) according to the manufacturer’s protocol. To perform electrophoresis in the presence of perfluorooctanoic acid (PFO) detergent, we modified Invitrogen’s SDS-NuPAGE protocol by replacing SDS with PFO. Lyophilized peptide was dissolved in sample buffer containing 4% PFO and heated at 65°C for 5 min prior to loading. The gel was run at 80 V for 2-3 hours with MES running buffer containing 0.5% PFO. Blue-native PAGE (BNPAGE) was performed as described previously (51). Lyophilized peptide was solubilized (0.1 mM) in sample buffer containing 25 mM SDS and either 25, 50, or 100 mM DPC. Aquaporin Z (AQPZ) in 20 mM SDS (heated at 65°C for 10 min) was included as an additional molecular weight marker.

**Fourier-Transform Infrared spectroscopy.** Sample preparation, data collection and H/D exchange were performed essentially as described (32) on a Nicolet Nexus spectrometer (Madison, USA). The NMR Sample Preparation. Approximately 1.2 mg of lyophilized Eₜ₅ protein was solubilized in sample of the SDS-DPC mixture, D₂O was added at 61.6%, 65%, and 72.4% respectively, to each SDS concentration. The samples were centrifuged in six-channel charcoal-filled Epon centerpieces using quartz windows. A radial distribution profile was acquired after sufficient time to reach equilibrium, as tested by HeteroAnalysis. The data were processed and fitted to several monomer/n-mer models in SEDFIT and SEDPHAT (54).

**Analytical Ultracentrifugation.** Sedimentation equilibrium experiments were performed using a Beckman XL-I analytical ultracentrifuge at 20°C (53), and monitored by measuring the absorbance at 280 nm. Lyophilized Eₜ₅ peptides were dissolved at OD₂₈₀ of 0.3, 0.5, and 0.8 (12 mm pathlength cell) in 20 mM sodium phosphate pH 5.5, 50 mM NaCl, and either 5 mM C14SB, or 100 mM DPC, and 12.5, 25, or 50 mM SDS. To match the density of the SDS-DPC mixture, D₂O was added at 61.6%, 65%, and 72.4% respectively, to each SDS concentration. The samples were centrifuged in six-channel charcoal-filled Epon centerpieces using quartz windows. A radial distribution profile was acquired after sufficient time to reach equilibrium, as tested by HeteroAnalysis. The data were processed and fitted to several monomer/n-mer models in SEDFIT and SEDPHAT (54).

**Circular Dichroism.** CD data was acquired on Chirascan CD Spectrometer (Applied Photophysics) using a 0.2 mm quartz cuvette (Hellma). Eₜ₅ peptide samples were dissolved at 1 mg/ml in 20 mM sodium phosphate, 50 mM NaCl, pH 5.5, 100 mM DPC, with or without 50 mM SDS. CD spectra were acquired from 180 to 260 nm with 1 nm spectral bandwidth and 3 replicates per spectra. Data points with excessive absorbance were excluded. After baseline subtraction, the data were analyzed in Dichroweb (55) by using the CDSSTR method (56) and the SMP180 reference set (57).
100 μl of methanol and dried under a dry stream of N₂ gas, resulting in a thin protein film deposit. The tube was placed in a vacuum lyophilizer overnight to remove any residual methanol. The thin protein film was then solubilized with sample buffer containing 20 mM sodium phosphate pH 5.5, 50 mM NaCl and 50 mM SDS. The sample was vortexed and sonicated for several times until a clear solution was obtained, indicating protein reconstitution into detergent micelles.

For Paramagnetic Relaxation Enhancement (PRE) experiments, a single point mutation (S60C) was introduced in ETR by site-directed mutagenesis using appropriate sets of primers. Expression and purification protocol of ETR-S60C mutant was the same as that of ETR protein. For labeling, 0.3 mM ¹⁵N-labeled ETR-S60C was dissolved in 20 mM sodium phosphate, 50 mM NaCl, 200 mM SDS, and 0.8 mM DTT at pH 5.5 and split into two equal portions for parallel labeling with (1-oxyl-2,2,5,5-tetramethyl-Δ³-pyrroline-3-methyl) methanethiosulfonate (MTSSL) (Toronto Research Chemicals Inc.) and a diamagnetic analog of MTSSL: (1-acetyl-2,2,5,5-tetramethyl-Δ³-pyrroline-3-methyl) methanethiosulfonate (dMTSSL, Toronto Research Chemicals Inc.). Ten-fold molar excess of both reagents were added from 75 mM stocks in methanol. The sample was vortexed for 30 minutes at high speed and incubated overnight at room temperature. Centrifugal filter unit (10 kDa MWCO, Milipore Corp.) was used to remove excess of both reagents. Labeled samples were washed for 4 times by concentrating to 100 μL. After a fourth wash, the sample was concentrated to 180 μL for NMR measurements.

Partial alignment of the ETR protein/micelle complexes relative to magnetic field was obtained by using stretched polyacrylamide hydrogels (58,59). A 7% polyacrylamide gel was then compressed into a 4.2 mm inner diameter open-ended tube using the gel press assembly (New Era Enterprise, Inc.).

NMR Spectroscopy. NMR experiments were performed at 308K using an Avance-II 700 NMR spectrometer with cryogenic probe. Sodium 2,2-dimethyl-2-silapentane-5-sulfonate (DSS) was used as the internal reference for ¹H nuclei. The chemical shifts of ¹³C and ¹⁵N nuclei were calculated from the ¹H chemical shifts. The NMR data were processed using TopSpin 3.1 (www.bruker-biospin.com) and analyzed using CARA (www.nmr.ch). Sequence-specific assignment of backbone ¹H, ¹⁵N, ¹³C and ¹³Ca was achieved by using 2D [¹H-¹⁵N]-TROSY-HSQC, 3D HNCO, HN(CA)CO, HNCA, HN(CO)CA, and HNCACB experiments on a ¹⁵N/¹³C-labeled ETR protein. Side-chain resonances were assigned using 3D ¹⁵N-resolved NOESY-HSQC (80, 100 and 150 ms mixing time), (H)CCH-TOCSY and ¹³C-resolved NOESY-HSQC (120 ms mixing time). To identify membrane-embedded residues, the NMR sample was lyophilized overnight and reconstituted in 99% D₂O. Immediately after reconstitution, 2D [¹H-¹⁵N]-TROSY-HSQC was collected. For paramagnetic probe measurements, ¹⁵N-HSQC spectra were recorded in the presence and absence of 1 mM dry 5-doxyl stearic acid (5-DSA) or 16-doxyl stearic acid (16-DSA). Axially symmetric alignment tensor coefficients (axiality and rhombicity) were calculated using MODULE (60). The PRE effect was measured using ¹⁵N-HSQC spectra of the S60C mutant before spin labeling and after MTSSL and dMTSSL labeling. The titration experiments with HMA, nsp3a or SH45-65 were performed with ¹⁵N-labeled ETR in 1:4 molar ratio SDS/DPC micelles at 318K. Chemical shift perturbation (CSP) values were calculated using the formula: CSP = \sqrt{\Delta \delta H^2 + (0.23 \times \Delta \delta N)^2}.

Structure Calculation. NOE distance restraints were obtained from ¹⁵N-NOESY-HSQC (mixing time 80, 100 and 150 ms) and ¹³C-NOESY-HSQC (mixing time 120 ms) spectra, respectively. Backbone dihedral angle restraints (φ and ψ) were derived from ¹³Cα, ¹³Cβ, ¹Hα.
and $^1$H$^\beta$ chemical shift values using TALOS+ (61). The short-range and medium range NOE connectivities were used to establish the sequence-specific $^1$H NMR assignment and to identify elements of the regular secondary structure. Hydrogen bonds were derived from the H/D exchange experiment and NOE connectivity. Distance restraints were obtained from the measured PRE effect using the procedures described previously (62-64). The intensities of cross-peaks in the MTSSL (Ip) and dMTSSL (Id) were calculated in CARA. The correlation time was set to 10 ns. The ratios of intensities (Ip/Id) were normalized against a set of 8 highest Ip/Id ratios, which were assumed to belong to peaks unaffected by PRE. For peaks with ratios below 0.15, no lower distance restraints were used, whereas upper restraints were set to 15 Å. For peaks with ratios above 0.9, only upper restraints of 25 Å were utilized. For peaks with ratios between 0.15 and 0.9, upper and lower distance restraints were generated using ± 3 Å margins. Structure calculations were performed using CYANA 3.0 (65,66) and visualized using PyMOL (Delano Scientific). CNS 1.3 (67,68) was used to refine the structure using the standard simulated annealing protocol. All the restraints used in the calculations to obtain a total of 15 structures, and the structure statistics, are summarized in Table 1.

**Single channel activity measurement.** Ion channel activity of $E_{TR}$ was measured by using Nanion Port-a-Patch®. Briefly, giant unilamellar vesicles (GUV) of 1,2-diphytanoyl-sn-glycero-3-phosphocholine (DPhPC) containing 10% cholesterol were prepared in 1 M sorbitol using Nanion Vesicle Prep Pro®, following the manufacturer’s protocol. Into 100 µL of a GUV solution, 0.5 µL of $E_{TR}$ peptide solution in ethanol (0.1 mg/ml) was added and incubated for 1 h at room temperature. $E_{TR}$-containing GUVs were subsequently deposited onto 6-8 MΩ NPC©-1 chips (Nanion). Conductance was measured under symmetrical buffer conditions (10 mM HEPES, 500 mM NaCl, pH 5.5).

**SPR.** The nsp3a sequence was subcloned from pcDNA3(+) into pET28b upstream of a C-terminus 6-His tag for expression in *E. coli*. The protein was expressed and purified as previously described by Serrano *et al.* (69). A negative control, consisting of C-terminal peptide from the small hydrophobic (SH) protein of human respiratory syncytial virus (RSV SH$_{45-65}$), was synthesized by standard solid phase and purified by reverse-phase HPLC. SPR measurements were performed on a Biacore 3000 system (GE Healthcare) using 10 mM phosphate buffer at pH 6.5, 100 mM NaCl, 3 mM EDTA, 0.05% n-octyl-β-D-glucopyranoside (OG) and 0.27% C14-betaine, at 25°C. $E_{TR}$ was immobilized to 15,000 RU onto a research-grade CM5 sensor chip (GE Healthcare) using standard amine-coupling chemistry. Briefly, a buffer-equilibrated carboxymethyl dextran surface was activated with a 10 min injection of a 1:1 mixture of 0.05 M N-hydroxysuccinimide (NHS) and 0.2 M N-ethyl-N-[3-(diethylamino)propyl] carbodiimide (EDC). $E_{TR}$ peptides dissolved in 10 mM sodium acetate, 15 mM DPC (pH 5.0) were passed over the activated surface to achieve the desired response level. Another 10 min injection of 1 M ethanolamine-HCl (pH 8.5) was used to deactivate the surface and remove any non-covalently bound protein. Kinetic measurements of immobilized $E_{TR}$ association with nsp3a and SH$_{45-65}$, (49 nM – 25 µM; in 10 two-fold serial dilutions) were performed with a 1 min association phase and 5 min dissociation phase at 30 µl/min flow rate. Each concentration was tested in duplicate. No regeneration was necessary, as all complexes dissociated within the monitored time. Sensorgrams were double-referenced (70) and globally fit to a steady-state model to obtain affinity values.
RESULTS and DISCUSSION

Expression and purification of SARS-CoV E protein. Initially, we successfully expressed and purified full-length SARS-CoV E protein (E_{FL}, Fig. 1B) by using either a β-barrel (71) and MBP as fusion tags. However, the yield of pure protein was low due to the presence of truncations (not shown). Nevertheless, the pure sample obtained (Fig. 1C-D) was sufficient for backbone assignment in SDS micelles, although not to produce a three-dimensional model. Therefore, a series of hexahistidine-tagged SARS-CoV E constructs were screened to obtain an expressing and well behaved sample. The best construct, encompassing residues 8–65, was successfully expressed in

\[ \text{E. coli} \]

and purified by affinity chromatography in milligram amounts without any enzymatic cleavage steps. This truncated construct (E_{TR}, see Fig. 1B) has an N-terminal His tag and a 16-residue linker that connects it to residues 8–65. The purified peptide appeared in MALDI-TOF MS as a single-charged peak at 8,997 Da and a double-charged peak at 4,512 Da, consistent with the calculated molecular weight of E_{TR}, 8,995 Da, and a small proportion of larger (dimer to pentamer) oligomers (Fig. 1E). After HPLC, the E_{TR} monomer (9 kDa) showed anomalous migration in standard SDS-PAGE (Fig. 1F), as reported previously for the full length protein (31).

Identification of the membrane-embedded region of E_{TR} by NMR. Screening of reconstitution conditions identified SDS as the best environment to achieve good peak dispersion in both dimensions (linker that connects it to residues 8–65. The purified peptide appeared in MALDI-TOF MS as a single-charged peak at 8,997 Da and a double-charged peak at 4,512 Da, consistent with the calculated molecular weight of E_{TR}, 8,995 Da, and a small proportion of larger (dimer to pentamer) oligomers (Fig. 1E). After HPLC, the E_{TR} monomer (9 kDa) showed anomalous migration in standard SDS-PAGE (Fig. 1F), as reported previously for the full length protein (31).

Effect of truncation and environment on E_{TR} secondary structure. To assess the effect of the
truncation and the presence of a His-tag on the ETR secondary structure, ETR and EFL were compared. The possible effect of the reconstitution environment was also determined.

**CD and IR spectra of ETR and EFL in detergent and lipid membranes** – The CD spectra of ETR in DPC, SDS and mixed (1:2 molar ratio) SDS/DPC micelles are almost superimposable (Fig. 4A) with minima at 209 and 222 nm. Also, these data are entirely consistent with the CD spectra of EFL, which was predominantly α-helical in both SDS and DPC micelles (71). When reconstituted in DMPC membranes, ETR and EFL produced an almost identical spectrum, with an amide I band centered at 1655 cm⁻¹ (Fig. 4B) characteristic of a predominantly α-helical conformation. Overall, these data show that both ETR and EFL (i) are predominantly α–helical and (ii) have a secondary structure that is not significantly affected by the reconstitution environment, supporting the relevance of the ETR structure (Fig. 3).

**13Cα chemical shifts** – 13Cα chemical shifts are highly correlated with secondary structure (72,73). Comparison of 13Cα chemical shifts of ETR and EFL in SDS (Fig. 4C) shows that ETR shifts (red) are almost identical to their counterpart residues (8-65) in EFL (blue). In EFL, the shifts for the last 10 residues (residues 66-76) are predicted to correspond to random coil, which is consistent with results obtained previously with a synthetic peptide spanning residues 59–76 (39). This peptide produced a broad amide I band in the IR spectrum centered around 1650 cm⁻¹, and immediately experienced complete H/D exchange.

When the 13C shifts for ETR were compared for SDS and an SDS/DPC mixture (1:4 molar ratio), only the N-terminal region (residues 7–15) showed minor differences. Finally, a comparison between EFL and ETR in an SDS/DPC mixture (1:4 molar ratio) also revealed an almost identical pattern in the C-terminal tail (not shown), although the spectral resolution for EFL was reduced in other regions. Overall, these results are consistent with those shown above (Fig. 4A-B), indicating that the detergent used, the tag, and the truncation has a minimal on ETR secondary structure.

**Ion channel activity of ETR and its inhibition by HMA.** Purified EFL has channel activity (71), which is inhibited by the drug HMA (30). An I/V plot obtained in a symmetrical 0.5 M NaCl experiment for purified ETR in DPhPC (Fig. 5A) was used to determine a conductance of 0.39 ± 0.02 nS. For comparison, synthetic full length SARS-CoV E and E_TM (residues 7-38) produced single channel conductances of 0.19 ± 0.06 pS and 0.18 ± 0.12 nS in 1M NaCl (35), although the values were higher in 1M KCl, with 0.37 ± 0.16 and 0.31 ± 0.12 for full length and TM, respectively. The lower conductance observed in synthetic samples may be due to extraneous modifications or impurities resulting from exposure to harsh chemicals. Representative traces of ETR channel activity (Fig. 5B), and complete inhibition after addition of 10 μM HMA (Fig. 5C), suggest that ETR is entirely functional.

To determine the binding site for HMA, we measured the differences in chemical shift perturbation (CSP) values before and after addition of the drug (Fig. 5D-E). In SDS, the average CSP value was low, 0.006 ppm (Fig. 5F), even at an HMA/ETR 10:1 molar ratio, suggesting no significant binding. However, the same panel shows that addition of HMA to ETR in mixed SDS/DPC (1:4 molar ratio) produced an average CSP value of 0.013 ppm, even at an HMA/ETR 2:1 molar ratio. Except Asn-64 and Leu-65, the residues that showed significant CSP (CSP ≥ 0.025 ppm) clustered near the membrane interface regions of the TM domain (Fig. 5F, see arrows). Near the N-terminal side of the TM domain the most affected were Glu-8, Gly-10, Thr-11, Val-14, Asn-15 and Ser-16, the latter two consistent with observations made on the TM channel in DPC (34). At the C-terminal end of the TM, Leu-37 was the most affected, suggesting that the interaction of HMA at Hε of Arg-38 reported previously (34) may have been an artifact due to the use of a TM peptide.

**Oligomeric state of ETR.** Gel electrophoresis – The localized changes in chemical shifts observed after HMA addition to ETR in mixed DPC/SDS micelles (Fig. 5E), but not in SDS micelles (Fig. 5D), suggest that binding sites for HMA may have been induced after DPC
addition to SDS, possibly through ETR increased oligomerization and population of the pentameric form. Thus, we examined the oligomeric state of ETR using gel electrophoresis performed in the presence of SDS or a mixture SDS/DPC. In contrast with the results obtained in normal SDS-PAGE (Fig. 1), we used NuPAGE, where ETR migrated with its expected monomer molecular weight, as a single band at 9 kDa (Fig. 6A). No oligomerization is consistent with the lack of interaction between HMA and ETR observed in SDS micelles.

To test the effect of DPC, in a Blue Native-PAGE gel, a constant SDS concentration (25 mM) was titrated with increasing DPC (Fig. 6B), from 1:1 to a 1:4 molar ratio. Bands consistent with dimers and trimers were observed at 1:1 and 1:2 molar ratios, whereas tetramers and pentamers were observed at a 1:4 molar ratio. This increasing oligomerization is again consistent with the binding of HMA observed in presence of DPC.

**Sedimentation equilibrium, SDS/DPC micelles** – For a sample corresponding to 1:4 molar ratio SDS/DPC, the equilibrium sedimentation data (Fig. 7A) could be fitted to several oligomeric models, from trimers to heptamers (Fig. 7B). Similar ambiguous results were obtained for a 1:2 molar ratio SDS/DPC mixture (Fig. 7C), suggesting that ETR forms a mixture of oligomers in SDS/DPC micelles, consistent with electrophoresis results (Fig. 6). The fit of the traces (Fig. 7A) to a monomer-pentamer model produced an apparent $K_s$ of $1.25 \cdot 10^{15}$ M$^{-4}$, i.e., $6 \cdot 10^3$ M$^{-1}$, or a $K_d$ for monomer-monomer interaction of 0.17 mM, and a mole fraction standard free energy change ($\Delta G_x^0$), of approximately -4 Kcal-mol$^{-1}$. Overall, the above results show that higher oligomeric states, including pentamers, are observed when increasing the DPC concentration, consistent with the larger shifts observed after adding HMA to mixed DPC/SDS micelles.

**ETR forms pentamers in C14 betaine and PFO.** We have shown previously that SARS-CoV E TM domain and EFL form pentamers (36,71). When ETR was solubilized in C14 betaine detergent (Fig. 7D), data could be optimally fitted to a monomer-pentamer model ($K_s=10^{16}$ M$^{-4}$) with a significantly lower $\chi^2$ than that obtained in SDS/DPC. Also, the mobility of ETR in PFO-NuPAGE (Fig. 7E) corresponds to a molecular weight of 45 kDa, i.e., a pentameric form. Thus, E-TM, ETR and EFL show a similar oligomerization behavior, being able to form pentameric channels that are inhibited by HMA.

**Interaction of ETR with nsp3a.** The interaction of nsp3 with SARS-CoV E protein, thought to be related to E ubiquitination, was previously mapped to an N-terminal acidic domain, nsp3a (49). Although the site of interaction with E is not known, it is likely to involve the C-terminal tail, as it is the largest extramembrane domain. Thus, to test that the conformation of ETR is structurally equivalent to the corresponding sequence in EFL, we tested its ability to interact with nsp3a (49) using surface plasmon resonance (SPR).

As a negative control, the mock peptide RSV SH45-65 (see Materials and Methods) was used to bind to immobilised ETR in detergent (see Materials and Methods) (Fig. 8A). Binding and unbinding of RSV SH45-65 was very slow (over > 1 min) and the data could not be fitted to any model, suggesting a non-specific interaction. In contrast, rapid and reversible interaction was observed using purified nsp3a (Fig. 8B). These data could be fitted with a stoichiometry of 1:1, although the end point could only be estimated. From that model, affinity was determined as 1.6 mM, although we note that binding is already evident at concentrations as small as 1-10 µM (see inset in Fig. 8B).

The residues involved in the interaction between nsp3a and ETR were identified by the differences in CSP values before and after addition of nsp3a to ETR in SDS/DPC micelles (1:4 molar ratio). Large chemical shift changes at Leu-39, Val-49 and Leu-65 (Fig. 8C) indicated a potential binding site at the C-terminal tail, whereas the central TM region did not experience any change. Overall, these data suggest that ETR has a structure in SDS/DPC micelles that is similar to the native fold of EFL in biological membranes.

**Structural discordance at the C-terminal tail.** Having determined the suitability of the construct used by comparison with the full length protein, and shown that the nature of the detergent did not affect the secondary structure
significantly, we investigated the structure of the C-terminal extramembrane domain of E\textsubscript{TR} and its apparent contrast with secondary structure predictions. The latter clearly suggest the presence of a $\beta$-turn-$\beta$ motif in the C-terminal region, not only in SARS-CoV E, but also in other representative sequences of other coronaviruses (see Fig. 1A). Indeed, we have reported previously that the synthetic peptide E\textsubscript{46-60}, which encompasses this putative $\beta$-hairpin, folds as $\beta$-strands, was completely resistant to H/D exchange and had a very high tendency to aggregate in solution (39). However, we have shown here that in the context of E\textsubscript{FL} or E\textsubscript{TR}, this domain does not adopt a $\beta$-structure. To test if at least some part of the population adopts this structure in lipid membranes, we mutated E\textsubscript{FL} residues 56-59 to increase the predicted helicity of this part of the molecule (43), from $\beta$-branched Val or bulky Tyr to small-side-chain Ala, to obtain the construct E\textsubscript{4ALA}. Lastly, to test the effect of the conserved Pro-54 at the center of this putative motif, we generated the mutant E\textsubscript{P54A}. The IR spectrum of these mutants when reconstituted in DMPC membranes (Fig. 9A) shows a similar amide I band. However, for the mutant E\textsubscript{4ALA}, two shoulders are eliminated (see arrows). This indicates that these shoulders in E\textsubscript{FL} do not represent misfolded protein, but they may correspond to bona fide $\beta$-structure conformation present in a small part of the population. Yet, the significance of these mutations is not completely clear since although the mutations introduced in E\textsubscript{4ALA} prevented Golgi complex accumulation when the C-terminus tail of SARS-E was coupled to VSV-G (43), a similar effect was also observed for the P54A mutant, which in our IR spectrum showed no obvious differences respect to E\textsubscript{FL}. As stated above, the synthetic peptide E\textsubscript{46-60} (39), which encompasses the predicted $\beta$-hairpin in SARS-CoV E was found to produce $\sim$100% $\beta$-structure and was completely resistant to H/D exchange (Fig. 9C). Combined with the effect observed for the E\textsubscript{4ALA} mutant, we propose that this $\beta$ structure may be in dynamic equilibrium with the much more abundant $\alpha$-helical form (Fig. 9D). A delicate balance between these two forms may alter processes in the infected cell, e.g., membrane scission, binding to protein partners, or E protein localization. Finally, the HMA titration results showed large CSP values for residues Val-49 and Leu-65, which are far in the sequence. Using the SymmDock server (75,76), a reconstructed E\textsubscript{TR} pentameric model was obtained based on the published structure of the E-TM pentamer (34) and the current E\textsubscript{TR} structure. The model suggests that these two residues may be in fact spatially close (Fig. 10), and belong to different monomers, providing a rationale for the observation above. CoV E proteins have been proposed to have at least two roles. One is related to their TM channel domain. This would be active in the secretory pathway, altering lumenal environments and rearranging secretory organelles, and leading to efficient trafficking of virions (38,77). The other would be related to their extramembrane domains, particularly the C-terminal domain. This is involved among others in protein-protein interactions and targeting. E protein participates in M-M, and E-M interactions (17,44) which are interesting targets for drug discovery. Also, formation of viral particles appears to be facilitated by a broad range of E sequences (78), which suggests that a common topology is more important than sequence requirements. E proteins have been suggested to act as chaperones during packaging (79), but the precise mechanism by which this takes place is not known. In this context, the structure determined here sheds light on a critical domain present in most CoV E proteins.
ACKNOWLEDGEMENTS

We thank T. Cornvik, A. Oi and E.K.M. Nilsson from the NTU/SBS Protein Production Platform for the expression and screening of truncated SARS-CoV E constructs, Luis Enjuanes for the kind gift of the plasmid carrying nsp3a cDNA, and Yin Hoe Yau and Susana Geifman Shochat for their guidance in conducting surface plasmon resonance experiments.
REFERENCES


**FIGURE LEGENDS**

**Figure 1. Sequences, expression and purification of SARS-CoV E<sub>TR</sub>.** (A) Alignment of representative sequences of E proteins in α-, β- and γ-coronaviruses. The cysteine residues are underlined, the conserved proline is highlighted (grey), and the four residues mutated to alanine in the E<sub>HALA</sub> mutant (see text) are shown in red. For these four proteins, the prediction of secondary structure is shown below in a color code, with the TM domain indicated as a black line; (B) proteins used in the present work: a His-tagged construct (E<sub>TR</sub>) encompassing residues 8-65 (bold–underlined), and full length SARS-CoV E (E<sub>FL</sub>). In E<sub>FL</sub>, the fragment SNA results from the cleavage of the tag. In both proteins, the native cysteines were mutated to alanine (C40A, C43A and C44A, see asterisks); (C-D) MALDI-TOF MS spectra (C) and standard SDS-PAGE (D) of pure E<sub>FL</sub> with the species labeled; (E-F) same for purified E<sub>TR</sub>; the identity of various single- and double-charged species is indicated. The calculated mass of E<sub>TR</sub> is 8995 Da.

**Figure 2. Topology and secondary structure of E<sub>TR</sub>.** [1H-15N]-TROSY-HSQC spectra of 0.2 mM E<sub>TR</sub> in 50 mM SDS in H<sub>2</sub>O (A) and in 99% D<sub>2</sub>O (B). The cross peaks are labeled by one-letter code and residue number; (C–D) peak intensity reduction upon addition of 5-DSA (C) and 16-DSA (D), calculated as the ratio of peak intensity before and after addition of the paramagnetic reagents; (E) [1H-15N] steady-state heteronuclear NOE experiment; (F) sequential and medium-ranged NOE connectivity between residues, displayed as bands under the respective residues.

**Figure 3. Structural model of E<sub>TR</sub>.** (A) Superposition of an ensemble of 15 calculated simulated annealing structures of E<sub>TR</sub> (only the sequence corresponding to E protein, 8-65, is shown). Side-chains are shown as line representation; the residues at the ends of the two helical segments are indicated; (B) residues of the C-terminal extramembrane α-helix oriented towards the micelle surface (blue); (C) ribbon representation of the TM central region, with the carbonyl oxygen of Phe-26 forming a H-bond to the side-chain of Thr-30.

**Figure 4. Equivalence in secondary structure of E<sub>TR</sub> and E<sub>FL</sub>.** (A) CD spectra of E<sub>TR</sub> in DPC (black), 1:2 molar ratio SDS/DPC mixture (blue), and SDS (red); (B) infrared amide I band of E<sub>TR</sub> (red) and E<sub>FL</sub> (blue) in DMPC lipid bilayers, and their respective Fourier self-deconvolved spectra (dotted lines); (C) comparison of secondary 13Cα chemical shifts (deviation from tabulated random-coil 13Cα chemical shift values) for E<sub>TR</sub> (red dots) and E<sub>FL</sub> (blue dots) in SDS micelles, and for E<sub>TR</sub> in (1:4 molar ratio) mixed SDS/DPC micelles (white dots). For the latter, Pro-54 and Thr-55 were excluded from the analysis due to significant line broadening; Arg-38 was excluded from the analysis due to the peak overlapping.

**Figure 5. Channel activity of E<sub>TR</sub> and interaction with HMA.** (A) I/V plot for E<sub>TR</sub> in DPhPC bilayers in a symmetrical experiment where both CIS and TRANS compartments contained 10 mM HEPES and 500 mM NaCl, at pH 5.5. Each point represents the mean of at least three current readings. The line is a linear regression fit of data points, which produced a slope of 0.39 ± 0.02 nS; (B) selected traces of 12 s each, recorded at various holding potentials of E<sub>TR</sub>; (C) channel activity recorded at 60 mV holding potential and after addition of 10 µM HMA (arrow); (D) [1H-15N]-TROSY-HSQC spectra of 0.1 mM 15N-labeled E<sub>TR</sub> in SDS micelles (D) before (blue) and after (red) addition of 1 mM HMA; (E) same for 0.2 mM 15N-labeled E<sub>TR</sub> in (1:4 molar ratio) SDS/DPC micelles upon titration with 0.4 mM HMA. Some shifts are indicated with arrows; (F) chemical shift perturbation (CSP) of the backbone amide resonances of E<sub>TR</sub> before and after addition of HMA in SDS (red) and (1:4 molar ratio) SDS/DPC micelles (blue). Not that the HMA/E<sub>TR</sub> molar ratio was 10 in SDS, and only 2 in SDS/DPC micelles. The arrows show
residues with significant change in chemical shifts after addition of HMA. The TM domain is indicated only to guide the eye.

**Figure 6. Gel electrophoresis of E
TR in SDS and in a mixture SDS/DPC.** (A) E
TR in 4-12% SDS-NuPAGE and (B) 4-16% Blue Native-PAGE of E
TR in 25 mM SDS with increasing concentration of DPC, as indicated. *E. coli* aquaporin Z (AQPZ) was included as an additional molecular weight marker. Bands and oligomeric states are indicated by arrows and black dots, respectively.

**Figure 7. Sedimentation equilibrium of E
TR in SDS/DPC micelles.** (A) Radial distribution profiles (open circles) of E
TR in a (25:100 mM) SDS/DPC mixture at 16600 rpm (red), 20300 rpm (green), and 24900 rpm (blue). The profile was fitted to a monomer-pentamer self-association model (black line) and the fitting residuals are shown below each plot; (B-D) global reduced chi-square values obtained after data fits to different monomer : n-mer models of E
TR association in SDS/DPC micelles (B), a 50:100 mM SDS/DPC mixture (C), and 5 mM C14-betaine (D); (E) 4-12% PFO-NuPAGE of E
TR.

**Figure 8. Interaction between E
TR and nsp3a.** (A) Sensorgrams corresponding to the interaction between purified RSV SH45-65 and immobilised E
TR (red). The steady-state model (dark red) did not produce a good fit to a 1:1 model of interaction. The association phase extends from 0-60 s, whereas the dissociation phase extends for minutes; (B) same for nsp3a and immobilised E
TR (blue) and fit to a steady-state model (red). Inset: dose-response plot, where the equilibrium responses of nsp3a were plotted against the log10[concentration] of nsp3a. Although the fit (blue line) yields an affinity of 1.6 mM, binding is already evident even in the interval 1-10 µM nsp3a concentration; (C) chemical shift perturbation (CSP) of the backbone amide resonances of 0.2 mM 15N-labeled E
TR in (1:4 molar ratio) SDS/DPC micelles upon titration with 0.4 mM of nsp3a (blue dots) or the negative control RSV SH45-65 (red dots).

**Figure 9. Structural flexibility at the putative β-coil-β motif.** (A) Amide I band corresponding to E
FL, E
P54A and E
4ALA in DMPC bilayers. Regions that change after mutation of the four residues indicated in Fig. 1A (E
4ALA) are shown as arrows; (B) Fourier self-deconvolved spectra corresponding to the amide I bands shown in (A); (C) infrared amide I and II bands corresponding to SARS-CoV E peptide E46-60 in DMPC lipid bilayers before (blue) and after (red) being exposed to D2O (39); (D) Possible equilibrium between two conformations at the putative β-coil-β motif, shifted towards an α-helical form, between the model determined experimentally for E
TR and E
FL and one built with prediction tools (PEP FOLD) (80), and consistent with data obtained with synthetic fragment E
46-60 shown in (C).

**Figure 10. Pentameric model formed by E
TR.** Side (A) and top (B) views of a proposed E
TR pentamer structure shown in ribbon representation. The side chains of Val-49 and Leu-65, which have been shown to interact with HMA, are shown as line representation.
Table 1. Restraints and structure statistics for the selected 15 structures of E_{TR}.

<table>
<thead>
<tr>
<th>NMR restraints</th>
<th>Total unambiguous distance restraints</th>
<th>2258</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Intra residual</td>
<td>1334</td>
</tr>
<tr>
<td></td>
<td>Sequential (</td>
<td>i – j</td>
</tr>
<tr>
<td></td>
<td>Short-range (</td>
<td>i – j</td>
</tr>
<tr>
<td></td>
<td>Medium (2 ≤</td>
<td>i – j</td>
</tr>
<tr>
<td></td>
<td>Long range (</td>
<td>i – j</td>
</tr>
<tr>
<td>Dihedral angle restraints</td>
<td></td>
<td>85</td>
</tr>
<tr>
<td>Hydrogen bond restraints(^a)</td>
<td></td>
<td>20</td>
</tr>
<tr>
<td>RDC restraints</td>
<td></td>
<td>44</td>
</tr>
<tr>
<td>PRE restraints</td>
<td></td>
<td>38</td>
</tr>
<tr>
<td>RMSD from the experimental residual dipolar couplings (Hz)</td>
<td>(^1\text{D}_{\text{NH}})</td>
<td>0.71±0.03</td>
</tr>
<tr>
<td>RMSD from the average atomic coordinates (residues 12-63, Å)(^b)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Backbone atoms</td>
<td></td>
<td>0.27 ± 0.11</td>
</tr>
<tr>
<td>All heavy atoms</td>
<td></td>
<td>0.70 ± 0.13</td>
</tr>
<tr>
<td>Ramachandran analysis (%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Residues in most favored regions</td>
<td></td>
<td>87.5</td>
</tr>
<tr>
<td>Residues in additional allowed regions</td>
<td></td>
<td>12.5</td>
</tr>
<tr>
<td>Residues in generously allowed regions</td>
<td></td>
<td>0.0</td>
</tr>
<tr>
<td>Residues in disallowed regions</td>
<td></td>
<td>0.0</td>
</tr>
</tbody>
</table>

\(^a\)Backbone hydrogen bonds of \(\alpha\)-helix were applied to regions confirmed to be \(\alpha\)-helical, according to the local NOE pattern and \(^{1}\text{H}\text{-H}_2\text{O}\) chemical exchange experiments.

\(^b\)Statistics were calculated and averaged over an ensemble of 15 structures with lowest target function according to CYANA.
Figure 1
Figure 2
Figure 3
Figure 4
Figure 5
Figure 7
Figure 8
Figure 9