Authors for correspondence:
Luis Martínez-Gil
e-mail: luis.martinez-gil@uv.es
Ismael Mingarro
e-mail: ismael.mingarro@uv.es

†Equal contribution (alphabetical order).

THE ROYAL SOCIETY
PUBLISHING

SARS-CoV-2 envelope protein topology in eukaryotic membranes

Gerard Duart†, Mª Jesús García-Murria†, Brayan Grau†, José M. Acosta-Cáceres, Luis Martínez-Gil and Ismael Mingarro

Departament de Bioquímica i Biologia Molecular, Estructura de Recerca Interdisciplinar en Biotecnologia i Biomedicina (ERI BioTecMed), Universitat de València E-46100 Burjassot, Spain
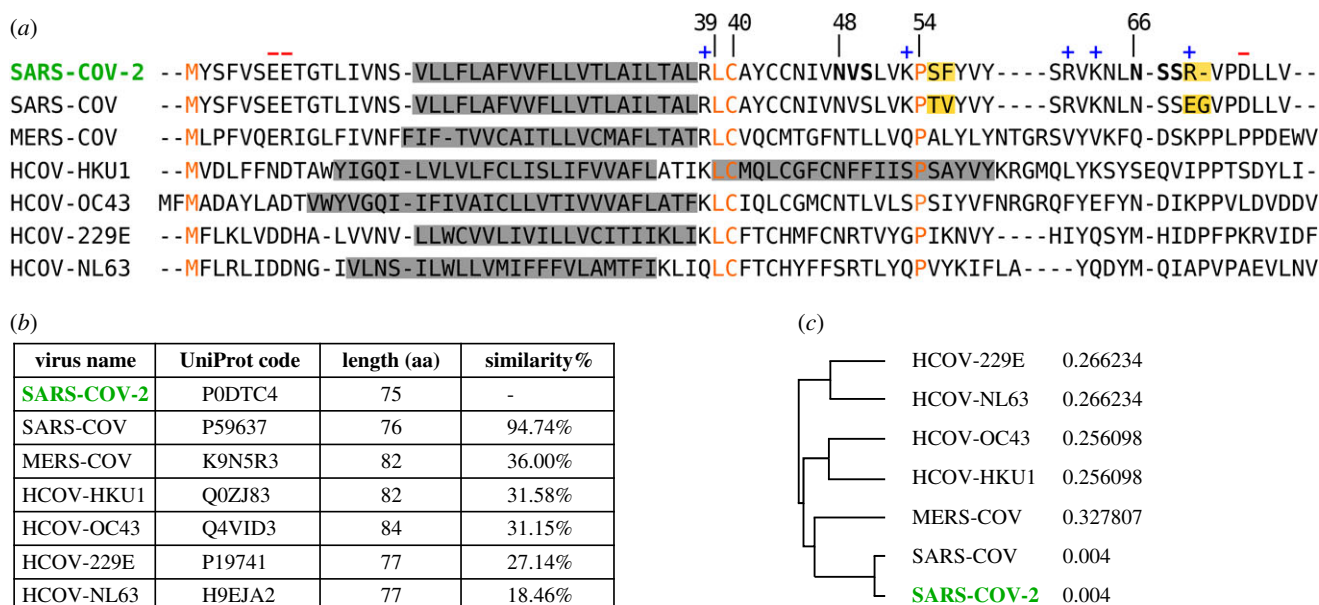
MJG-M, 0000-0002-8396-0335; BG, 0000-0002-8582-023X; LM-G, 0000-0002-9076-7760; IM, 0000-0002-1910-1229

Coronavirus E protein is a small membrane protein found in the virus envelope. Different coronavirus E proteins share striking biochemical and functional similarities, but sequence conservation is limited. In this report, we studied the E protein topology from the new SARS-CoV-2 virus both in microsomal membranes and in mammalian cells. Experimental data reveal that E protein is a single-spanning membrane protein with the N-terminus being translocated across the membrane, while the C-terminus is exposed to the cytoplasmic side (Nt_{lum}/Ct_{cyt}). The defined membrane protein topology of SARS-CoV-2 E protein may provide a useful framework to understand its interaction with other viral and host components and contribute to establish the basis to tackle the pathogenesis of SARS-CoV-2.

1. Introduction

Coronavirus disease 2019 (COVID-19), an extremely infectious human disease caused by coronavirus SARS-CoV-2, has spread around the world at an unprecedented rate, causing a worldwide pandemic. While the number of confirmed cases continues to grow rapidly, the molecular mechanisms behind the biogenesis of viral proteins are not fully unravelled. The SARS-CoV-2 genome encodes up to 29 proteins, although some may not get expressed [1]. The viral RNA is packaged by the structural proteins to assemble viral particles at the ERGIC (ER-Golgi intermediate compartment). The four major structural proteins are the spike (S) surface glycoprotein, the membrane (M) matrix protein, the nucleocapsid (N) protein, and the envelope (E) protein. These conserved structural proteins are synthesized from sub-genomic RNAs (sgRNA) encoded close to the 3′ end of the viral genome [2].

Among the four major structural proteins, the E protein is the smallest and has the lowest copy number of the membrane proteins found in the lipid envelope of mature virus particles (reviewed [3,4]). However, it is critical for pathogenesis of other human coronaviruses [5,6]. Interestingly, the sgRNA encoding E protein is one of the most abundantly expressed transcripts despite the protein having a low copy number in mature viruses [1]. It encodes a 75 residues long polypeptide with a predicted molecular weight of approximately 8 kDa. Two aliphatic amino acids (Leu and Val) constitute a substantial portion (36%, 27/75) of the E protein, which accounts for the high grand average of hydropathicity (GRAVY) index of the protein (1.128), as calculated using the ExPASy ProtParam tool (https://web.expasy.org/protparam/). Comparative sequence analysis of the E protein of SARS-CoV-2 and the other six known human coronaviruses do not reveal any large homologous/identical regions (figure 1), with only the initial methionine, Leu39, Cys40 and Pro54 being ubiquitously conserved. With regard to overall sequence similarity SARS-CoV-2 E protein has the highest similarity to SARS-CoV (94.74%) with only minor differences (figure 1*b*), followed by MERS-CoV (36.00%). Interestingly, sequence similarities are significantly lower for the other four human coronaviruses, which usually cause mild to moderate

(a)

```
                                          39 40      48      54           66
                                          |  |       |       |            |
SARS-COV-2  --MYSFVSEETGTLIVNS-VLLFLAFVVFLLVTLAILTALRLCAYCCNIVNVSLVKPSFYVY----SRVKNLN-SSRVPDLLV--
SARS-COV    --MYSFVSEETGTLIVNS-VLLFLAFVVFLLVTLAILTALRLCAYCCNIVNVSLVKTVYVY----SRVKNLN-SSEGVPDLLV--
MERS-COV    --MLPFVQERIGLFIVNFFIF-TVVCAITLLVCMAFLTATRLCVQCMTGFNTLLVQPALYLYNTGRSVYVKFQ-DSKPPLPPDEWV
HCOV-HKU1   --MVDLFFNDTAWYIGQI-LVLVLFCLISLIFVVAFLATIKLCMQLCGFCNFFIISPSAYVYKRGMQLYKSYSEQVIPPTSDYLI-
HCOV-OC43   MFMADAYLADTVWYVGQI-IFIVAICLLVTIVVVAFLATFKLCIQLCGMCNTLVLSPSIYVFNRGRQFYEFYN-DIKPPVLDVDDV
HCOV-229E   --MFLKLVDDHA-LVVNV-LLWCVVLIVILLVCITIIKLIKLCFTCHMFCNRTVYGPIKNVY----HIYQSYM-HIDPFPKRVIDF
HCOV-NL63   --MFLRLIDDNG-IVLNS-ILWLLVMIFFFVLAMTFIKLIQLCFTCHYFFSRTLYQPVYKIFLA----YQDYM-QIAPVPAEVLNV
```

(b)

virus name	UniProt code	length (aa)	similarity %
SARS-COV-2	P0DTC4	75	-
SARS-COV	P59637	76	94.74%
MERS-COV	K9N5R3	82	36.00%
HCOV-HKU1	Q0ZJ83	82	31.58%
HCOV-OC43	Q4VID3	84	31.15%
HCOV-229E	P19741	77	27.14%
HCOV-NL63	H9EJA2	77	18.46%

(c)

HCOV-229E	0.266234
HCOV-NL63	0.266234
HCOV-OC43	0.256098
HCOV-HKU1	0.256098
MERS-COV	0.327807
SARS-COV	0.004
SARS-COV-2	0.004

Figure 1. (a) Multi-alignment of amino acid sequences of the E protein of SARS-CoV-2 and the other six human coronavirus. SARS-CoV severe acute respiratory syndrome coronavirus (UniProt P59637), MERS-CoV Middle East respiratory syndrome coronavirus (UniProt K9N5R3), HCoV-HKU1 (UniProt Q0ZJ83), HCoC-OC43 (UniProt Q4VID3), HCoC-229E (UniProt P19741) and HCoV-NL63 (UniProt Q5SBN7). Predicted TM segments at UniProt are highlighted in a grey box. Native predicted glycosylation acceptor sites in SARS-CoV-2 are shown in bold and charged residues highlighted with + or − symbols on top. Conserved residues are shown in orange. Differences between SARS-CoV-2 and SARS-CoV are highlighted as yellow boxes. (b) Phylogenetic data and (c) tree obtained with Clustal Omega (EMBL-EBI) using the default parameters.

upper-respiratory tract illness typical for common cold, with the lowest similarity found for HCOV-NL63 (18.46%). These findings are consistent with the phylogenetic tree proposed based on the amino acid sequences of the human coronavirus E proteins using ClustalW (figure 1c).

2. Results and discussion

2.1. E protein topology prediction

Computer-assisted analysis of the SARS-CoV-2 E protein amino acid sequence using seven popular prediction methods showed that all membrane protein prediction algorithms except MEMSAT-SVM suggested the presence of one transmembrane (TM) segment located roughly around amino acids 12 to 39 (table 1), which is not predicted as a cleavable signal sequence according to SignalP-5.0 [7]. Regarding E protein topology, TMHMM and Phobius predicted an N-terminus cytosolic orientation, while MEMSAT-SVM, TMpred, HMMTop and TOPCONS predicted an N-terminus luminal orientation. These discrepancies found among the predictions from different algorithms motivated experimental approaches.
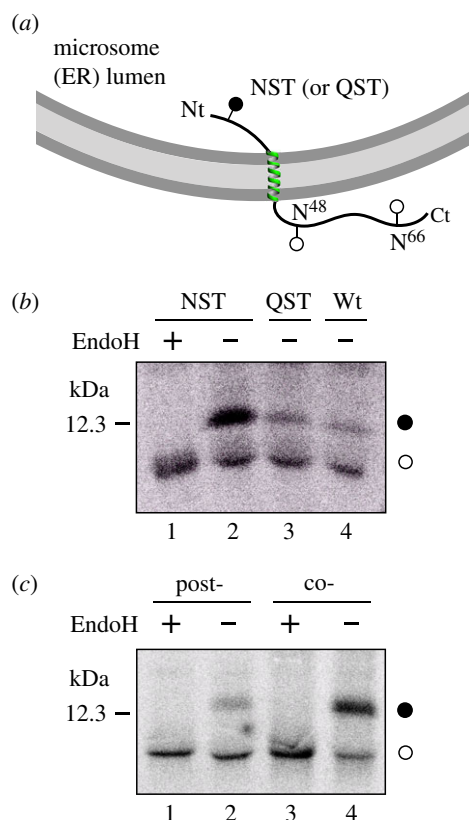
2.2. Insertion into microsomal membranes

First, we performed *in vitro* E protein transcription/translation experiments in the presence of ER-derived microsomes and [^{35}S]-labelled amino acids. The membrane insertion orientation of the predicted TM segment into microsomal membranes was based on N-linked glycosylation and summarized in figure 2a. N-linked glycosylation has been extensively used as topological reporter for more than two decades [8]. In eukaryotic cells, proteins can only be glycosylated in the lumen of the ER because the active site of oligosaccharyl transferase (OST), a translocon-associated protein responsible for N-glycosylation

[9], is located there [10]; no N-linked glycosylation occurs within the membrane or in the cytosol. It is important to note that two possible N-linked glycosylation sites are located C-terminally of the predicted TM segment in E protein wild-type sequence at positions N48 and N66 (figure 1). However, N48 is not expected to be modified even if situated lumenally due to the close proximity of this glycosylation acceptor site to the membrane if the hydrophobic region is recognized as TM by the translocon [11,12]. Thus, mono-glycosylation (at N66) would serve as a C-terminal translocation reporter. To test N-terminal translocation, a construct was engineered where a predicted highly efficient glycosylation acceptor site (i.e. Asn-Ser-Thr, NST) was designed at the N-terminus. When E protein constructs were translated *in vitro* in the presence of microsomes, the protein was significantly glycosylated when the N-terminal designed glycosylation site was present, as shown by the increase in the electrophoretic mobility of the slower radioactive band after an endoglycosidase H (Endo H) treatment (figure 2b, lanes 1 and 2). However, when a control (Gln-Ser-Thr, QST) that is not a glycosylation acceptor site (lane 3) or the wild-type (lane 4) sequences were translated, E protein molecules were minimally glycosylated. Since multiple topologies have been reported for previous coronavirus E proteins [13–17], SARS-CoV-2 E protein insertion into the microsomal membranes in two opposite orientations cannot be discounted, but according to our data being dominant an Nt$_{lum}$/Ct$_{cyt}$ orientation.

2.3. E protein integrates cotranslationally into microsomal membranes

We have previously reported that several viral membrane proteins are cotranslationally inserted into ER-derived microsomal membranes [18–20]. Since membrane protein insertion and N-glycosylation are coupled at the ER by complex

Figure 2. Translocon-mediated insertion of E protein variants into microsomal membranes. (*a*) Schematic representation of E protein constructs. Glycosylation acceptor Asn residues are indicated. (*b*) *In vitro* translation in the presence of microsomes of the different E protein constructs. Construct containing inserted asparagine and threonine residues at positions 3 and 5 (NST; lanes 1–2) or glutamine and threonine at positions 3 and 5 (lane 3), and wild-type variants (lane 4) were translated in the presence of microsomes. NST variant was split and half of the sample was Endo H treated (lane 1). Bands of non-glycosylated and glycosylated proteins are indicated by white and black dots, respectively. (*c*) E protein (harbouring an engineered glycosylation site at the N-terminus, positions 3–5) was translated in either the absence (lanes 1 and 2) or the presence (lanes 3 and 4) of microsomal membranes. In lanes 1 and 2, microsomal membranes were added posttranslationally (after 1 h, post-) and incubation was continued for another 1 h. Samples in lanes 1 and 3 were treated later with EndoH. The gels are representative of at least three independent experiments.

Table 1. Computer analysis of the SARS-CoV-2 E protein amino acid sequence topology. n.p., non-predicted.

algorithm	Nt	Ct	TMDs (start-end)
ΔG predictor	n.p.	n.p.	1 (17–39)
TMHMM	cytosol	lumen	1 (12–34)
MEMSAT-SVM	lumen	lumen	2 (10–39) (43–58)
TMpred	lumen	cytosol	1 (17–34)
HMMTop	lumen	cytosol	1 (11–35)
Phobius	cytosol	lumen	1 (12–37)
TOPCONS	lumen	cytosol	1 (16–36)

formation of a ribosome, the translocon and the OST [10], we sought to investigate whether or not SARS-CoV-2 E protein is cotranslationally inserted into the ER membrane by blocking protein synthesis after E protein (harbouring N-terminal NST
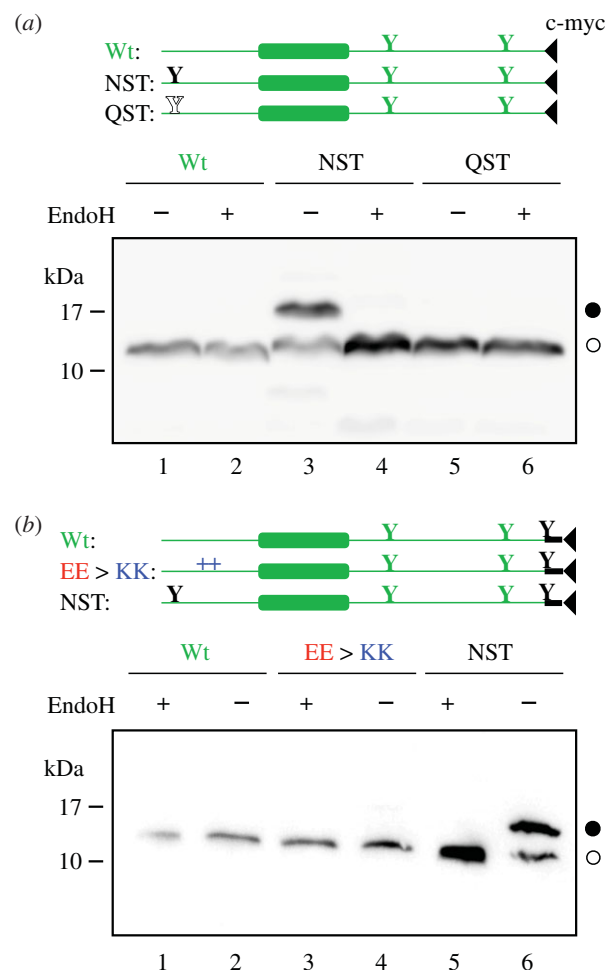
Figure 3. E protein topology in mammalian cells. To determine the topology *in vivo* HEK-293T cells were transfected with C-terminal tagged (c-myc) E protein variants. (*a*) Constructs encoding wild-type (Wt; lanes 1 and 2), inserted asparagine and threonine at positions 3 and 5 (NST; lanes 3 and 4) or glutamine and threonine at positions 3 and 5 (QST; lanes 5 and 6) were Endo H (+) or mock (−) treated. Filled and empty Y-shaped symbols denoted acceptor (NST) and non-acceptor (QST) glycosylation sites, respectively. (*b*) Additionally, we included constructs containing similar Wt (lanes 1 and 2), replaced glutamic acids at positions 7 and 8 by lysine residues (EE > KK; lanes 3 and 4) or NST (lanes 5 and 6) variants with an extra glycosylation site inserted at the Ct end of the protein. Once again, to confirm the glycosylated nature of the higher molecular weight bands, samples were either Endo H (+) or mock (−) treated. Designed glycosylation sites and tags are shown in black, while native E protein features are shown in grey.

glycosylation site) has been translated in the absence of membranes. As shown in figure 2*c*, E protein (NST) was efficiently glycosylated when microsomal membranes were added to the translation mixture cotranslationally (lane 4). But when microsomal membranes were included posttranslationally after the translation was inhibited by cycloheximide, the protein was only residually glycosylated (lane 2), suggesting that E protein is mainly integrated cotranslationally through the ER translocon. This means that the microsomal insertion machinery recognizes, orients and provides a path into the membrane for this viral protein.

2.4. Membrane topology in mammalian cells

To analyse protein topology in mammalian cells, a series of E protein variants tagged with c-myc epitope at the C-terminus

were transfected into HEK-293T cells. As shown in figure 3a, only an E protein construct harbouring the N-terminal engineered acceptor site was efficiently modified (lanes 1–4), denoting an N-terminal ER luminal localization (Nt_{lum}). Several topological parameters have been proposed to govern membrane protein topology, among which the preferential distribution of positively charged residues in the cytosol (positive-inside rule) has been established as the primary topology determinant both experimentally [21] and statistically [22]. E protein is a single-spanning membrane protein with an even net charge distribution on both sides of the membrane. There are only eight charged residues along the protein sequence (two negatively charged residues preceding the TM segment, and five positively and one negatively charged residues at the C-terminal domain; figure 1a), which correlates well the observed topology with the 'positive-inside rule'. However, negatively charged residues have also been proved to significantly affect the topology [23]. To test the robustness of the observed topology, we added an optimized Ct glycosylation tag [24] and replaced the two negatively charged residues located in the translocated N-terminal domain (E7 and E8) by two lysine residues (figure 3b). In cells expressing this mutant E protein (EE > KK), the protein retained its C-terminal tail at the cytosolic side of the membrane as indicated by the absence of glycosylated forms (figure 3b, lanes 3 and 4). These data reveal that topological determinants have only a minor effect on viral membrane protein topology as previously demonstrated for other viruses [25] and suggest that viral membrane protein topology could have co-evolved with the protein environment of its natural host, ensuring proper membrane protein orientation. Altogether, the present *in vivo* results demonstrate that SARS-CoV-2 E protein is a single-spanning membrane protein with an Nt_{lum}/Ct_{cyt} orientation in mammalian cell membranes. Similarly, SARS-CoV E protein was shown to mainly adopt an Nt_{lum}/Ct_{cyt} topology in the infected cell and mammalian cells expressing the recombinant protein [26]. This topology is compatible with the ion channel capacity described previously [27], and with the recently published pentameric structural model of SARS-CoV E protein in micelles [28], in which the C-terminal tail of the protein is α-helical and extramembrane.

3. Concluding remarks

The membrane topology described here would allow the cytoplasmic C-terminal tail of the E protein to interact with the C-termini of M and/or S SARS-CoV-2 membrane-embedded proteins [3], and/or with Golgi scaffold proteins as previously described for other coronaviruses [29], to induce virus budding or influence vesicular traffic through the Golgi complex by collecting viral membrane proteins for assembly at Golgi membranes. Future experiments will have to unravel whether these functions involve the SARS-CoV-2 E protein.

4. Material and methods

4.1. Enzymes and chemicals

TNT T7 Quick for PCR DNA was from Promega (Madison, WI, USA). Dog pancreas ER rough microsomes were from tRNA Probes (College Station, TX, USA). EasyTag EXPRESS[35]S

Protein Labeling Mix, ^{35}S-L-methionine and ^{35}S-L-cysteine, for *in vitro* labelling was purchased from Perkin Elmer (Waltham, MA, USA). Restriction enzymes were from New England Biolabs (Massachusetts, USA) and endoglycosidase H was from Roche Molecular Biochemicals (Basel, Switzerland). PCR and plasmid purification kits were from Thermo Fisher Scientific (Ulm, Germany). All oligonucleotides were purchased from Macrogen (Seoul, South Korea).

4.2. Computer-assisted analysis of E protein sequence

Prediction of transmembrane segments was done using up to 7 of the most common methods available on the Internet: ΔG Predictor [30,31] (http://dgpred.cbr.su.se/), TMHMM [32] (http://www.cbs.dtu.dk/services/TMHMM/), MEMSAT-SVM [33] (http://bioinf.cs.ucl.ac.uk/psipred/), TMpred (https://embnet.vital-it.ch/software/TMPRED_form.html), HMMTop [34] (http://www.enzim.hu/hmmtop/), Phobius [35] (http://phobius.sbc.su.se/) and TOPCONS [36] (http://topcons.net/). All user-adjustable parameters were left at their default values.

4.3. DNA manipulation

Full-length E protein was synthesized by Invitrogen (*GeneArt* gene synthesis) and subcloned into *Kpn*I linearized pCAGGS in-house version [37] using In-Fusion HD cloning Kit (Takara) according to the manufacturer's instructions. For *in vitro* assays, DNA was amplified by PCR adding the T7 promoter and the relevant glycosylation sites during the process. N-terminal NST glycosylation site was designed by inserting an asparagine and a threonine before and after Ser3, respectively. Control no-glycosylable QST site was introduced in similarly inserting a glutamine residue instead of an asparagine. All E protein variants were obtained by site-directed mutagenesis using QuikChange kit (Stratagene, La Jolla, California) and were confirmed by sequencing the plasmid DNA at Macrogen Company (Seoul, South Korea).

4.4. Translocon-mediated insertion into microsomal membranes

E protein variants, PCR amplified from pCAGGS, were transcribed and translated using the TNT T7 Quick for PCR DNA coupled transcription/translation system (Promega, USA). The reactions contained 10 µl of TNT, 2 µl of PCR product, 1 µl of EasyTag (5 µC$_i$) and 0.6 µl of column-washed microsomes (tRNA Probes, USA) and were incubated for 60 min at 30°C. Translation products were ultracentrifuged (100 000g for 15 min) on a 0.5 M sucrose cushion and analysed by SDS-PAGE. For the endoglycosidase H (Endo H), the treatment was done as previously described [20]. Briefly, the translation mixture was diluted in 120 µl of PBS and centrifuged on a 0.5 M sucrose cushion (100 000g 15 min 4°C). The pellet was then suspended in 50 µl of sodium citrate buffer with 0.5% SDS and 1% β-mercaptoethanol, boiled 5 min, and incubated 1 h at 37°C with 1 unit of Endo H. Then, the samples were analysed by SDS-PAGE and gels were visualized on a Fuji FLA3000 phosphorimager using Image Reader 8.1j software.

4.5. Postranslational and cotranslational insertion assay

E protein DNAs were transcribed/translated (30°C 1 h) either in the absence (figure 2c, post- samples) or in the presence (co- samples) of microsomal membranes. The translation was inhibited with cycloheximide (10 min, 26°C, 2 mg ml^{-1} final concentration), after which microsomes were added to those samples labelled as posttranslational and incubated for an additional hour at 30°C. Subsequently, membranes were collected by ultracentrifugation; half of the samples were EndoH treated and analysed by SDS-PAGE (double volume was loaded for the post-samples due to the lower translation levels observed). Protein molecules were visualized on a Fuji FLA3000 phosphorimager.

4.6. E protein expression in mammalian cells

E protein sequence variants were tagged with a c-myc epitope at their C-terminus (Glu-Gln-Lys-Leu-Ile-Ser-Glu-Glu-Asp-Leu, EQKLISEEDL) and inserted in a pCAGGS-ampicillin plasmid. When appropriate (figure 3b), an optimized glycosylation site followed by a flexible dipeptide (Asn-Ser-Thr-Gly-Ser, NSTGS) [24,38] preceded the c-myc epitope. Once the sequence was verified, plasmids were transfected into HEK293-T cells using Lipofectamine 2000 (Life Technologies) according to the manufacturer's protocol. Approximately 24 h post-transfection cells were harvested and washed with PBS buffer. After short centrifugation (1000 r.p.m. for 5 min on a table-top centrifuge), cells were lysed by adding 100 μl of lysis buffer (30 mM Tris-HCl, 150 mM NaCl, 0.5% Nonidet P-40), sonicated in an ice bath in a bioruptor (Diagenode) during 10 min and were centrifugated. Total protein was quantified and equal amounts of protein submitted to Endo H treatment or mock-treated, followed by SDS-PAGE analysis and transferred into a PVDF transfer membrane (ThermoFisher Scientific). Protein glycosylation status was analysed by Western Blot using an anti-c-myc antibody (Sigma), anti-rabbit IgG-peroxidase conjugated (Sigma) and with ECL developing reagent (GE Healthcare). Chemiluminescence was visualized using an ImageQuantTM LAS 4000mini Biomolecular Imager (GE Healthcare).

References

1. Kim D, Lee J-Y, Yang J-S, Kim JW, Kim VN, Chang H. 2020 Embracing the landscape of therapeutics. Cell 181, 1–19. (doi:10.1016/j.cell.2020.03.025)

2. Wu A et al. 2020 Genome composition and divergence of the novel coronavirus (2019-nCoV) originating in China. Cell Host Microbe 27, 325–328. (doi:10.1016/j.chom.2020.02.001)

3. Schoeman D, Fielding BC. 2019 Coronavirus envelope protein: current knowledge. Virol. J. 16, 1–22. (doi:10.1186/s12985-019-1182-0)

4. Ruch TR, Machamer CE. 2012 The coronavirus E protein: assembly and beyond. Viruses 4, 363–382. (doi:10.3390/v4030363)

5. Almazán F, DeDiego ML, Sola I, Zuñiga S, Nieto-Torres JL, Marquez-Jurado S, Andrés G, Enjuanes L. 2013 Engineering a replication-competent, propagation-defective middle east respiratory syndrome coronavirus as a vaccine candidate. mBio 4, 20282-11. (doi:10.1128/mBio.00650-13)

6. Netland J, DeDiego ML, Zhao J, Fett C, Álvarez E, Nieto-Torres JL, Enjuanes L, Perlman S. 2010 Immunization with an attenuated severe acute respiratory syndrome coronavirus deleted in E protein protects against lethal respiratory disease. Virology 399, 120–128. (doi:10.1016/j.virol.2010.01.004)

7. Armenteros JJA, Tsirigos KD, Sønderby CK, Petersen TN, Winther O, Brunak S, Heijne G, Nielsen H. 2019 The blind babymaker. Nat. Biotechnol. 37, 1–8. (doi:10.1038/nbt.4341)

8. Mingarro I, von Heijne G, Whitley P. 1997 Membrane-protein engineering. Trends Biotechnol. 15, 432–437. (doi:10.1016/S0167-7799(97)01101-3)

9. Martínez-Gil L, Saurí A, Marti-Renom MA, Mingarro I. 2011 Membrane protein integration into the endoplasmic reticulum. FEBS J 278, 3846–3858. (doi:10.1111/j.1742-4658.2011.08185.x)

10. Braunger K, Pfeffer S, Shrimal S, Gilmore R, Berninghausen O, Mandon EC, Becker T, Förster F, Beckmann R. 2018 Structural basis for coupling protein transport and N-glycosylation at the mammalian endoplasmic reticulum. Science 360, 215–219. (doi:10.1126/science.aar7899)

11. Nilsson IM, von Heijne G. 1993 Determination of the distance between the oligosaccharyltransferase active site and the endoplasmic reticulum membrane. J. Biol. Chem. 268, 5798–5801.

12. Orzáez M, Salgado J, Giménez-Giner A, Pérez-Payá E, Mingarro I. 2004 Influence of proline residues in transmembrane helix packing. J. Mol. Biol. 335, 631–640. (doi:10.1016/j.jmb.2003.10.062)

13. Corse E, Machamer CE. 2000 Infectious bronchitis virus E protein is targeted to the Golgi complex and directs release of virus-like particles. J. Virol. 74, 4319–4326. (doi:10.1128/JVI.74.9.4319-4326.2000)

14. Raamsman MJ, Locker JK, de Hooge A, de Vries AA, Griffiths G, Vennema H, Rottier PJ. 2000 Characterization of the coronavirus mouse hepatitis virus strain A59 small membrane protein E. J. Virol. 74, 2333–2342. (doi:10.1128/JVI.74.5.2333-2342.2000)

15. Godet M, L'Haridon R, Vautherot JF, Laude H. 1992 TGEV corona virus ORF4 encodes a membrane protein that is incorporated into virions. Virology 188, 666–675. (doi:10.1016/0042-6822(92)90521-P)

16. Maeda J, Repass JF, Maeda A, Makino S. 2001 Membrane topology of coronavirus E protein. Virology 281, 163–169. (doi:10.1006/viro.2001.0818)

17. Yuan Q, Liao Y, Torres J, Tam JP, Liu DX. 2006 Biochemical evidence for the presence of mixed membrane topologies of the severe acute respiratory syndrome coronavirus envelope protein expressed in mammalian cells. FEBS Lett 580, 3192–3200. (doi:10.1016/j.febslet.2006.04.076)

18. Saurí A, Saksena S, Salgado J, Johnson AE, Mingarro I. 2005 Double-spanning plant viral movement protein integration into the endoplasmic reticulum membrane is signal recognition particle-dependent, translocon-mediated, and concerted. J. Biol. Chem. 280, 25 907–25 912. (doi:10.1074/jbc.M412476200)

19. Martínez-Gil L, Saurí A, Vilar M, Pallás V, Mingarro I. 2007 Membrane insertion and topology of the p7B movement protein of melon necrotic spot virus

(MNSV). *Virology* **367**, 348–357. (doi:10.1016/j. virol.2007.06.006)

20. Martínez-Gil L, Johnson AE, Mingarro I. 2010 Membrane insertion and biogenesis of the *Turnip Crinkle Virus* p9 movement protein *J. Virol.* **11**, 5520–5527. (doi:10.1128/JVI.00125-10)

21. von Heijne G. 1989 Control of topology and mode of assembly of a polytopic membrane protein by positively charged residues. *Nature* **341**, 456–458. (doi:10.1038/341456a0)

22. Baeza-Delgado C, Marti-Renom MA, Mingarro I. 2013 Structure-based statistical analysis of transmembrane helices. *Eur. Biophys. J.* **42**, 199–207. (doi:10.1007/s00249-012-0813-9)

23. Nilsson I, von Heijne G. 1990 Fine-tuning the topology of a polytopic membrane protein: role of positively and negatively charged amino acids. *Cell* **62**, 1135–1141. (doi:10.1016/0092-8674(90) 90390-Z)

24. Bañó-Polo M, Baldin F, Tamborero S, Marti-Renom MA, Mingarro I. 2011 *N*-glycosylation efficiency is determined by the distance to the C-terminus and the amino acid preceding an Asn-Ser-Thr sequon. *Protein Sci* **20**, 179–186. (doi:10.1002/pro.551)

25. Saurí A, Tamborero S, Martínez-Gil L, Johnson AE, Mingarro I. 2009 Viral membrane protein topology is dictated by multiple determinants in its sequence. *J. Mol. Biol.* **387**, 113–128. (doi:10.1016/ j.jmb.2009.01.063)

26. Nieto-Torres JL, DeDiego ML, Álvarez E, Jimenez-Guardeño JM, Regla-Nava JA, Llorente M, Kremer L, Shuo S, Enjuanes L. 2011 Subcellular location and topology of severe acute respiratory syndrome coronavirus envelope protein. *Virology* **415**, 69–82. (doi:10.1016/j.virol.2011.03.029)

27. Verdiá-Báguena C, Nieto-Torres JL, Alcaraz A, DeDiego ML, Torres J, Aguilella VM, Enjuanes L. 2012 Coronavirus E protein forms ion channels with functionally and structurally-involved membrane lipids. *Virology* **432**, 485–494. (doi:10.1016/j.virol. 2012.07.005)

28. Surya W, Li Y, Torres J. 2018 Structural model of the SARS coronavirus E channel in LMPG micelles. *Biochimica et Biophysica Acta (BBA) - Biomembranes* **1860**, 1309–1317. (doi:10.1016/j.bbamem.2018. 02.017)

29. Corse E, Machamer CE. 2003 The cytoplasmic tails of infectious bronchitis virus E and M proteins mediate their interaction. *Virology* **312**, 25–34. (doi:10. 1016/S0042-6822(03)00175-2)

30. Hessa T, Meindl-Beinker NM, Bernsel A, Kim H, Sato Y, Lerch-Bader M, Nilsson I, White SH, von Heijne G. 2007 Molecular code for transmembrane-helix recognition by the Sec61 translocon. *Nature* **450**, 1026–1030. (doi:10.1038/nature06387)

31. Hessa T, Kim H, Bihlmaier K, Lundin C, Boekel J, Andersson H, Nilsson I, White SH, von Heijne G. 2005 Recognition of transmembrane helices by the endoplasmic reticulum translocon. *Nature* **433**, 377–381. (doi:10.1038/nature03216)

32. Krogh A, Larsson B, von Heijne G, Sonnhammer ELL. 2001 Predicting transmembrane protein topology with a hidden markov model: application to complete genomes. *J. Mol. Biol.* **305**, 567–580. (doi:10.1006/jmbi.2000.4315)

33. Nugent T, Jones DT. 2009 Transmembrane protein topology prediction using support vector machines. *BMC Bioinformatics* **10**, 980-11. (doi:10.1186/1471-2105-10-159)

34. Tusnády GE, Simon I. 2001 The HMMTOP transmembrane topology prediction server. *Bioinformatics* **17**, 849–850. (doi:10.1093/ bioinformatics/17.9.849)

35. Käll L, Krogh A, Sonnhammer ELL. 2004 A combined transmembrane topology and signal peptide prediction method. *J. Mol. Biol.* **338**, 1027–1036. (doi:10.1016/j.jmb.2004. 03.016)

36. Tsirigos KD, Peters C, Shu N, Käll L, Elofsson A. 2015 The TOPCONS web server for consensus prediction of membrane protein topology and signal peptides. *Nucleic Acids Res.* **43**, W401–W407. (doi:10.1093/ nar/gkv485)

37. Vera-Velasco NM, García-Murria MJ, del Pino MMS, Mingarro I, Martínez-Gil L. 2018 Proteomic composition of Nipah virus-like particles. *J. Proteomics* **172**, 190–200. (doi:10.1016/j.jprot. 2017.10.012)

38. Tamborero S, Vilar M, Martínez-Gil L, Johnson AE, Mingarro I. 2011 Membrane insertion and topology of the translocating chain-associating membrane protein (TRAM). *J. Mol. Biol.* **406**, 571–582. (doi:10. 1016/j.jmb.2011.01.009)