May 27, 2020

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.

# Equal contribution (alphabetical order)

* Corresponding authors

ABSTRACT

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 establish the basis to tackle the pathogenesis of SARS-CoV-2.

KEYWORDS

Coronavirus; Envelope protein; membrane insertion; SARS-CoV-2; topology;

INTRODUCTION

The coronavirus disease 19 (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 unraveled. The SARS-CoV-2 genome encodes for 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 being low copy number in mature viruses [1]. It encodes a 75 residues long polypeptide with a predicted molecular weight of ~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 1B), and the lowest with 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).

RESULTS AND DISCUSSION

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, whilst MEMSAT-SVM, TMpred, HMMTop and TOPCONS predicted an N-terminus luminal orientation. Firstly, we performed *in vitro* E protein transcription/translation experiments in the presence of ER-derived microsomes and [^{35}S]-labeled amino acids. The membrane insertion orientation of the predicted TM segment into microsomal membranes was based on N-linked glycosylation and summarized in Figure 2 (top).

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, 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 serves as a C-terminal translocation reporter. To test N-terminal translocation a construct was engineered where a predicted highly efficient glycosylation acceptor site (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 2, lanes 1 and 2). However, when a control (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 discarded, but being dominant an Nt_{lum}/Ct_{cyt} orientation.

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 localisation (Nt_{lum}). Several topological parameters have been proposed to govern membrane protein topology, among them the preferential distribution of positively charged residues in the cytosol ('positive-inside rule') has been

established as the primary topology determinant both experimentally [18] and statistically [19]. 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 (Fig. 1A), that nicely correlates the observed topology with the 'positive-inside rule'. However, negatively charged residues have also been proved to significantly affect the topology [20]. To test the robustness of the observed topology, we added an optimized Ct glycosylation tag [21] and replaced the two negatively charged residues located in the translocated N-terminal domain (E7 and E8) by two lysine residues (Fig. 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 (Fig. 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 [22], 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 demonstrated 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 infected and transiently expressing mammalian cells [23]. This topology is compatible with the ion channel capacity described previously [24], and with the recently published pentameric structural model of SARS-CoV E protein in micelles [25], in which the C-terminal tail of the protein is α-helical and extramembrane.

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 [26], 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.

EXPERIMENTAL METHODS

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* labeling 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).

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 [27,28] (http://dgpred.cbr.su.se/), TMHMM [29] (http://www.cbs.dtu.dk/services/TMHMM/), MEMSAT-SVM [30] (http://bioinf.cs.ucl.ac.uk/psipred/), TMpred (https://embnet.vital-it.ch/software/TMPRED_form.html), HMMTop [31] (http://www.enzim.hu/hmmtop/),

Phobius [32] (http://phobius.sbc.su.se/) and TOPCONS [33] (http://topcons.net/). All user-adjustable parameters were left at their default values.

DNA Manipulation. Full-length E protein was synthesized by Invitrogen (*GeneArt* gene synthesis) and subcloned into *Kpn*I linearized pCAGGS 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)

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 μCi), and 0.6 μL of column-washed microsomes (tRNA Probes, USA) and were incubated for 60 min at 30 °C. Translation products were ultracentrifuged (100,000 *g* for 15 min) on a 0.5 M sucrose cushion, and analyzed by SDS-PAGE. For the endoglycosidase H (Endo H) the treatment was done as previously described [34]. Briefly, the translation mixture was diluted in 120 μL of PBS and centrifuged on a 0.5 M sucrose cushion (100 000 × g 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 analyzed by SDS-PAGE.

E protein expression in mammalian cells. E protein sequence variants were tagged with an optimized C-terminal glycosylation site [21,35] plus a c-myc epitope at their C-terminus and inserted in a pCAGGS-ampicillin plasmid. 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 a short centrifugation (1000 rpm 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) were sonicated in an ice bath in a bioruptor (Diagenode) during 10min and 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).

Acknowledgments

We thank Prof. Paul Whitley (University of Bath) for critical reading of the manuscript. This work was supported by grants PROMETEU/2019/065 from Generalitat Valenciana and COV20/01265 from ISCIII (to I.M.). G.D. was recipient of a predoctoral contract from the Spanish Ministry of Education (FPU). J.A.-C. was recipient of a predoctoral fellowship from the Ministerio de Educación (Perú). B.G.

was recipient of a predoctoral contract from the University of Valencia (Atracció de Talent).

Table 1. Computer analysis of the SARS-CoV-2 E protein amino acid sequence topology

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)

n.p., non-predicted

Figure legend

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), 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.

Figure 2. Translocon-mediated insertion of E protein variants into microsomal membranes. (Top) Schematic representation of E protein constructs. (Bottom) *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. The gel is representative of at least four independent experiments.

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 gray.

References

[1] D. Kim, J.-Y. Lee, J.-S. Yang, J.W. Kim, V.N. Kim, H. Chang, The Architecture of SARS-CoV-2 Transcriptome, Cell. (2020) 1–19. doi:10.1016/j.cell.2020.04.011.

[2] A. Wu, Y. Peng, B. Huang, X. Ding, X. Wang, P. Niu, et al., Genome Composition and Divergence of the Novel Coronavirus (2019-nCoV) Originating in China, Cell Host and Microbe. 27 (2020) 325–328. doi:10.1016/j.chom.2020.02.001.

[3] D. Schoeman, B.C. Fielding, Coronavirus envelope protein: current knowledge, (2019) 1–22. doi:10.1186/s12985-019-1182-0.

[4] T.R. Ruch, C.E. Machamer, The Coronavirus E Protein: Assembly and Beyond, Viruses. 4 (2012) 363–382. doi:10.3390/v4030363.

[5] F. Almazán, M.L. DeDiego, I. Sola, S. Zuñiga, J.L. Nieto-Torres, S. Marquez-Jurado, et al., Engineering a Replication-Competent, Propagation-Defective Middle East Respiratory Syndrome Coronavirus as a Vaccine Candidate, mBio. 4 (2013) 20282–11. doi:10.1128/mBio.00650-13.

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

[7] J.J.A. Armenteros, K.D. Tsirigos, C.K. Sønderby, T.N. Petersen, O. Winther, S. Brunak, et al., SignalP 5.0 improves signal peptide predictions using deep neural networks, Nat. Biotechnol. (2019) 1–8. doi:10.1038/s41587-019-0036-z.

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

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

[10] K. Braunger, S. Pfeffer, S. Shrimal, R. Gilmore, O. Berninghausen, E.C. Mandon, et al., Structural basis for coupling protein transport and N-glycosylation at the mammalian endoplasmic reticulum, Science. 360 (2018) 215–219. doi:10.1126/science.aar7899.

[11] I.M. Nilsson, G. von Heijne, Determination of the distance between the oligosaccharyltransferase active site and the endoplasmic reticulum membrane, J Biol Chem. 268 (1993) 5798–5801.

[12] M. Orzáez, J. Salgado, A. Giménez-Giner, E. Pérez-Payá, I. Mingarro, Influence of proline residues in transmembrane helix packing, J Mol Biol. 335 (2004) 631–640.

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

[14] M.J. Raamsman, J.K. Locker, A. de Hooge, A.A. de Vries, G. Griffiths, H. Vennema, et al., Characterization of the coronavirus mouse hepatitis virus strain A59 small membrane protein E, J Virol. 74 (2000) 2333–2342. doi:10.1128/jvi.74.5.2333-2342.2000.

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

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

[17] Q. Yuan, Y. Liao, J. Torres, J.P. Tam, D.X. Liu, 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 (2006) 3192–3200. doi:10.1016/j.febslet.2006.04.076.

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

[19] C. Baeza-Delgado, M.A. Marti-Renom, I. Mingarro, Structure-based statistical analysis of transmembrane helices, Eur Biophys J. 42 (2012) 199–207. doi:10.1007/s00249-012-0813-9.

[20] I. Nilsson, G. von Heijne, Fine-tuning the topology of a polytopic membrane protein: role of positively and negatively charged amino acids, Cell. 62 (1990) 1135–1141.

[21] M. Bañó-Polo, F. Baldin, S. Tamborero, M.A. Marti-Renom, I. Mingarro, 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 (2011) 179–186. doi:10.1002/pro.551.

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

[23] J.L. Nieto-Torres, M.L. DeDiego, E. Álvarez, J.M. Jimenez-Guardeño, J.A. Regla-Nava, M. Llorente, et al., Subcellular location and topology of severe acute respiratory syndrome coronavirus envelope protein, Virology. 415 (2011) 69–82. doi:10.1016/j.virol.2011.03.029.

[24] C. Verdiá-Báguena, J.L. Nieto-Torres, A. Alcaraz, M.L. DeDiego, J. Torres, V.M. Aguilella, et al., Coronavirus E protein forms ion channels with functionally and structurally-involved membrane lipids, Virology. 432 (2012) 485–494. doi:10.1016/j.virol.2012.07.005.

[25] W. Surya, Y. Li, J. Torres, Structural model of the SARS coronavirus E channel in LMPG micelles, Biochimica Et Biophysica Acta (BBA) - Biomembranes. 1860 (2018) 1309–1317. doi:10.1016/j.bbamem.2018.02.017.

[26] E. Corse, C.E. Machamer, The cytoplasmic tails of infectious bronchitis virus E and M proteins mediate their interaction, Virology. 312 (2003) 25–34. doi:10.1016/s0042-6822(03)00175-2.

[27] T. Hessa, N.M. Meindl-Beinker, A. Bernsel, H. Kim, Y. Sato, M. Lerch-Bader, et al., Molecular code for transmembrane-helix recognition by the Sec61 translocon, Nature. 450 (2007) 1026–1030. doi:10.1038/nature06387.

[28] T. Hessa, H. Kim, K. Bihlmaier, C. Lundin, J. Boekel, H. Andersson, et al., Recognition of transmembrane helices by the endoplasmic reticulum translocon, Nature. 433 (2005) 377–381. doi:10.1038/nature03216.

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

[30] T. Nugent, D.T. Jones, Transmembrane protein topology prediction using support vector machines, BMC Bioinformatics. 10 (2009) 980–11. doi:10.1186/1471-2105-10-159.

[31] G.E. Tusnády, I. Simon, The HMMTOP transmembrane topology prediction server, Bioinformatics. 17 (2001) 849–850. doi:10.1093/bioinformatics/17.9.849.

[32] L. Käll, A. Krogh, E.L.L. Sonnhammer, A Combined Transmembrane Topology and Signal Peptide Prediction Method, J Mol Biol. 338 (2004) 1027–1036. doi:10.1016/j.jmb.2004.03.016.

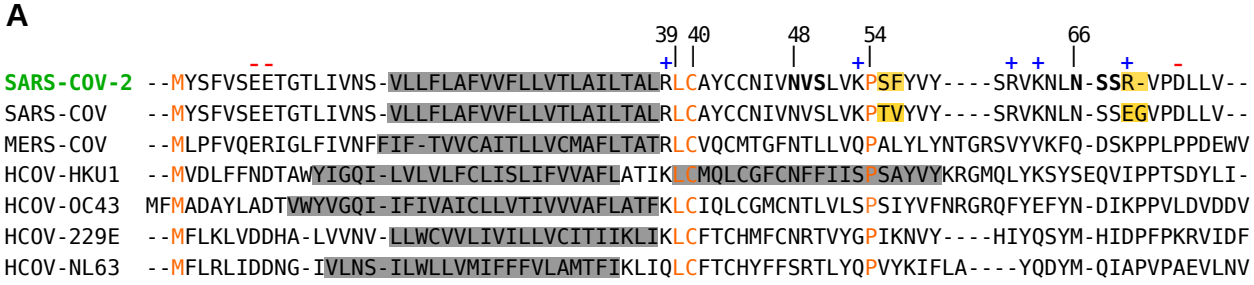
[33] K.D. Tsirigos, C. Peters, N. Shu, L. Käll, A. Elofsson, The TOPCONS web server for consensus prediction of membrane protein topology and signal peptides, Nucleic Acids Res. 43 (2015) W401–W407. doi:10.1093/nar/gkv485.

[34] L. Martínez-Gil, A.E. Johnson, I. Mingarro, Membrane insertion and biogenesis of the Turnip Crinkle Virus p9 movement protein, J Virol. (2010). doi:10.1128/JVI.00125-10.

[35] S. Tamborero, M. Vilar, L. Martínez-Gil, A.E. Johnson, I. Mingarro, Membrane insertion and topology of the translocating chain-associating membrane protein (TRAM), J Mol Biol. 406 (2011) 571–582. doi:10.1016/j.jmb.2011.01.009.

Figure 1

A



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

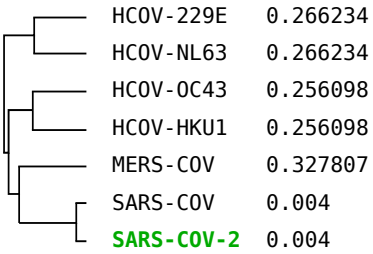


Figure 2

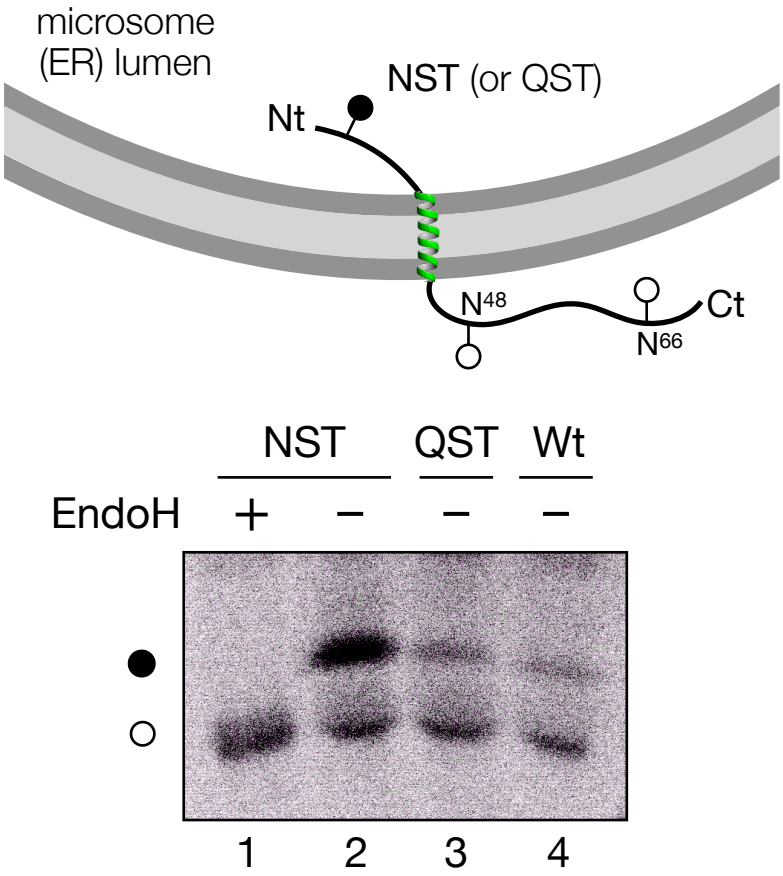


Figure 3