



Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.



BBA Research Letter

## The SARS-CoV-2 envelope (E) protein has evolved towards membrane topology robustness

Gerard Duart, Maria J. García-Murria, Ismael Mingarro \*

Departament de Bioquímica i Biologia Molecular, Institut Universitari de Biotecnologia i Biomedicina (BIOTECMED), Universitat de València, E-46100 Burjassot, Spain



## ARTICLE INFO

## Keywords:

Coronavirus  
Envelope protein  
Membrane topology  
SARS-CoV-2  
Evolution

## ABSTRACT

- Single-spanning SARS-CoV-2 envelope (E) protein topology is a major determinant of protein quaternary structure and function.
- Charged residues distribution in E protein sequences from highly pathogenic human coronaviruses (*i.e.*, SARS-CoV, MERS-CoV and SARS-CoV-2) stabilize N<sub>t<sub>out</sub></sub>-C<sub>t<sub>in</sub></sub> membrane topology.
- E protein sequence could have evolved to ensure a more robust membrane topology from MERS-CoV to SARS-CoV and SARS-CoV-2.

In the past 20 years, the world has seen three human coronaviruses responsible for severe disease outbreaks: the Severe Acute Respiratory Syndrome (SARS-CoV) that emerged in 2002, the Middle East Respiratory Syndrome (MERS-CoV) in 2012 and recently the emergence of SARS-CoV-2, which has spread around the world at an unprecedented rate, causing a worldwide pandemic.

Coronaviruses' genome includes four major structural proteins: membrane (M), spike (S), nucleocapsid (N) and envelope (E). The multifunctional E protein is the smallest of the structural proteins (between 8 and 12 kDa) and has the lowest copy number in the lipid envelope of mature virus particles [1]. The majority of the E protein pool localizes to the endoplasmic reticulum Golgi intermediate compartment (ERGIC) in the host cell where it participates in virus budding, assembly and trafficking [2]. In addition to this structural role the E protein oligomerizes to form pentameric ion channels similar to viroporins [3–5] and possesses a C-terminal PDZ-binding motif that induce immunopathology by overexpression of inflammatory cytokines [6]. These features of E protein play a major role in the exacerbated immune response causing the acute respiratory syndrome, the leading cause of death in SARS-CoV and SARS-CoV-2 [7], and have been shown to be critical for propagation of other human coronaviruses. The assembly of E protein into the ER membrane in the correct orientation (topology) is critical for its functions. In the evolution of membrane proteins it is not rare to observe mutations leading to a more fixed orientation relative to the

membrane.

SARS-CoV-2 E protein is a single-spanning membrane protein with a skewed distribution of charged residues on both sides of the membrane. There are only eight charged residues in the protein sequence, two negatively charged residues N-terminal to the transmembrane (TM) domain, and five positively plus one negatively charged residues in the C-terminal domain (Fig. 1A). The observed N<sub>t<sub>out</sub></sub>-C<sub>t<sub>in</sub></sub> topology [8] is in good agreement with the 'positive-inside' rule [9].

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 [8]. Interestingly, sequence similarities are significantly higher for the coronaviruses that usually cause severe illness than for those that cause mild to moderate upper-respiratory tract symptoms typical for common cold. SARS-CoV-2 E protein has the highest similarity to SARS-CoV (94.74%) with only minor differences (Fig. 1A), followed by MERS-CoV (36.00%) [8]. Nevertheless, regarding topology determination, there is a common feature in all of them. There is a positively charged residue strongly conserved and strategically located proximal to the hydrophobic region C-terminal end (Supplementary Fig. 1) in all seven human coronaviruses. It is worth mentioning that this positively charged residue is an Arg (Arg38) in MERS-CoV, SARS-CoV and SARS-CoV-2 (Fig. 1A), while in the other human coronaviruses is a Lys [8]. Interestingly, in the analysis of 81,818 E protein sequences of SARS-CoV-2 globally available, no change was detected in

\* Corresponding author.

E-mail address: [mingarro@uv.es](mailto:mingarro@uv.es) (I. Mingarro).<https://doi.org/10.1016/j.bbamem.2021.183608>

Received 18 February 2021; Received in revised form 16 March 2021; Accepted 17 March 2021

Available online 24 March 2021

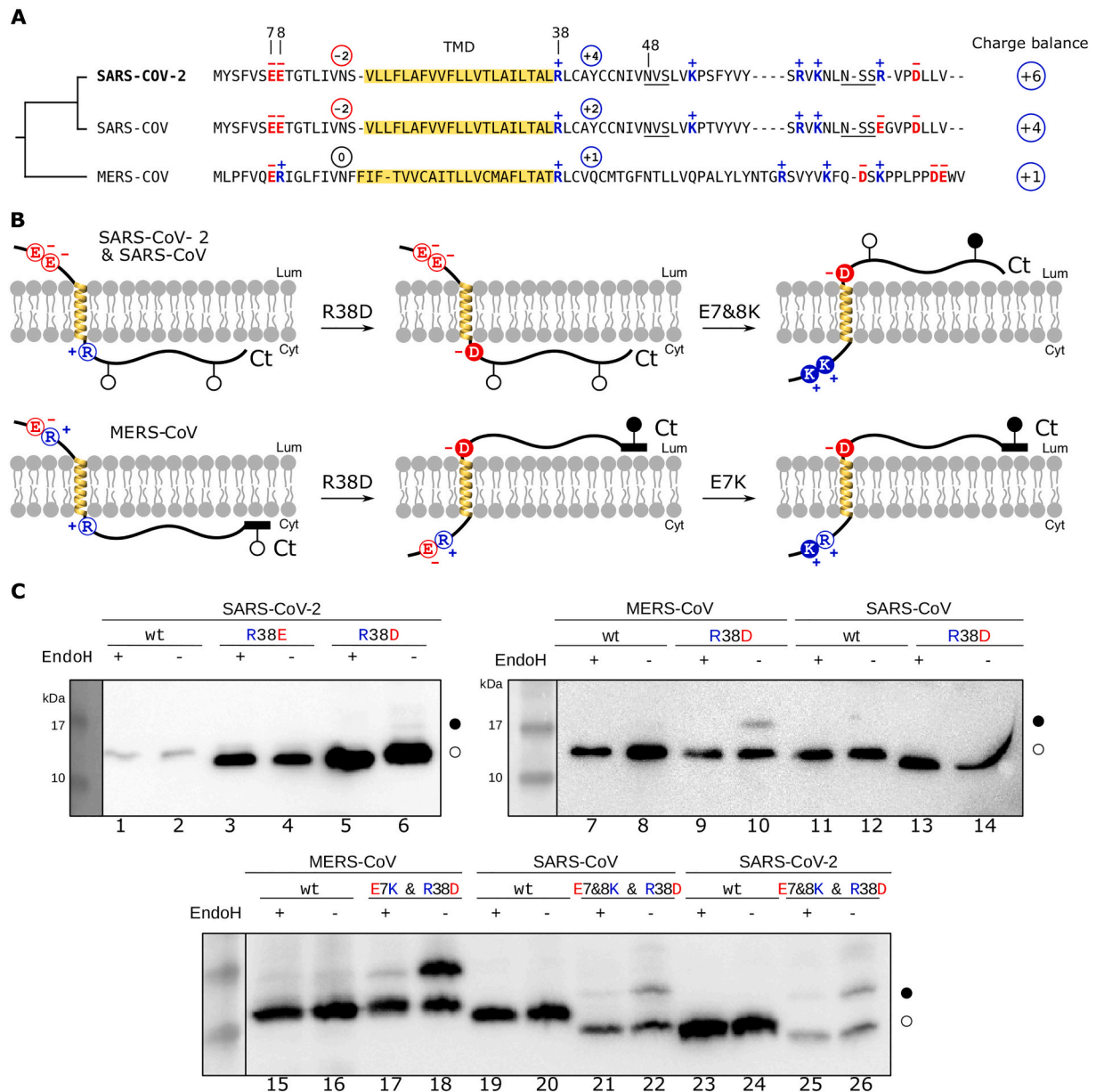
0005-2736/© 2021 The Authors.

Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license

[\(http://creativecommons.org/licenses/by-nc-nd/4.0/\)](http://creativecommons.org/licenses/by-nc-nd/4.0/).

this position [10]. Positively charged residues located near the cytoplasmic end of hydrophobic segments in membrane proteins promote correct membrane insertion of TM helices. It has been determined that a single Arg or Lys residue typically contributes approximately  $-0.5$  kcal/mol to the apparent free energy of membrane insertion when placed at this location [11].

In comparison to globular (water-soluble) proteins, topology provides an extra dimension that membrane proteins can evolve. Topology can evolve, for example, by redistribution of charged residues on both sides of the membrane. The alignment of MERS-CoV, SARS-CoV and SARS-CoV-2 E proteins unveils a tendency to accumulate a net positive charge balance C-terminally to the TM domain (Fig. 1A), which



**Fig. 1.** A. Multi-alignment of amino acid sequences of the E protein from MERS-CoV (UniProt K9N5R3), SARS-CoV (UniProt P59637) and SARS-CoV-2 (UniProt P0DTC4). Predicted TM segments are highlighted in a yellow box. Negatively charged amino acids are shown in red with - symbol on top while the positive ones are shown in blue with + symbol on top. Native predicted glycosylation acceptor sites are underlined. Conserved and relevant residues are marked with the number on top (7, 8, 38 and 48). The net charge summation before and after the TM segment is shown encircled. The charge balance (charge balance at the region following the TM segment minus charge balance at the region preceding the TM segment) is shown at the right side. Tree obtained with Clustal Omega (EMBL-EBI) using the default parameters.

B. Schematic representations of E protein topology in the presence of the different mutations. Wild type residues 7, 8 and 38 are shown in an empty colored circle (red for glutamic acids and blue for arginines) accompanied with - or + symbol depending on the charge of the residue. Point mutations are shown in red (negative) or blue (positive) solid circles emphasizing the charge change. Glycosylation acceptor sites are indicated with white (non-glycosylated) or black (glycosylated) dots. In MERS-CoV, Ct-tail containing the glycosylation site is represented with a black rectangle.

C. To determine the topology *in vivo*, HEK-293T cells were transfected with Ct tagged (c-myc) E protein variants. The E protein virus and the proper mutations are indicated on top of each gel. Lanes with odd numbers are Endo H treated (+) and even numbers are mock treated (-). Samples were separated on SDS-PAGE (14% polyacrylamide) and analyzed by Western blot using an anti-c-myc antibody (Sigma). Bands of non-glycosylated and glycosylated proteins are indicated by white and black dots, respectively. The gels are representative of at least three independent experiments. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

correlates with the ‘positive-inside’ rule, but also suggests an increasing robustness in the topology determination from MERS-CoV to SARS-CoV-2. MERS-CoV E protein sequence contains one positively and one negatively charged residues in the translocated N-terminus and four positively charged residues plus three negatively charged residues in the C-terminal cytosolic domain, giving a net charge balance of +1. In the case of SARS-CoV, charge balance increases substantially with a net charge of  $-2$  in the N-terminal extra-membranous domain and  $+2$  (4 positively plus 2 negatively charged residues) in the C-terminus, giving a net charge balance of  $+4$ . In the case of SARS-CoV-2, this balance is higher due to E69R substitution, giving a net charge balance of  $+6$  (Fig. 1A).

The topology of the SARS-CoV-2 envelope protein was recently proved to be  $N_{\text{out}}-C_{\text{in}}$  in eukaryotic membranes [8]. To test the topological relevance of the conserved Arg38 residue we designed two replacement mutants in which the positively charged residue was mutated to aspartic or glutamic acid residues (R38D or R38E, respectively). The topology was determined by monitoring glycosylation of the consensus acceptor sites that the E protein has downstream of the TM segment (Fig. 1A). Glycosylation at a single site increases the molecular weight of the protein by  $\sim 2.5$  kDa. In eukaryotic cells, proteins can only be glycosylated in the lumen of the ER because the active site of oligosaccharyl transferase, the enzyme responsible for co-translational glycosylation, is located there. To analyze protein topology in mammalian cells, E protein variants tagged with c-myc epitope at the C-terminus were transfected into HEK-293T cells. As shown in Fig. 1C, neither R38E mutant (lanes 3 and 4) nor R38D mutant (lanes 5 and 6) resulted on alteration of the original E protein topology (lanes 1 and 2). The N-terminal translocation of these mutants was demonstrated by engineering two highly efficient glycosylation sites, one at the N-terminus and another one in a C-terminal tag (Supplementary Fig. 2). Similarly, R38D mutation in SARS-CoV E protein displayed the same glycosylation pattern as the wild-type equivalent (Fig. 1C, lanes 11–14). The MERS-CoV E protein sequence does not contain natural glycosylation consensus acceptor sites (Fig. 1A). Therefore an optimized C-terminal glycosylation tag was added to the C-terminal domain (Fig. 1B) [12]. In this case, no glycosylation band was observed when the wild-type protein was expressed (Fig. 1C, lanes 7 and 8). However, a higher molecular weight band was detected when the R38D mutant was expressed (lane 10). The nature of the higher molecular weight protein species was confirmed by endoglycosidase H (EndoH) treatment (lane 9), a highly specific enzyme that cleaves oligosaccharides regardless of their location. Thus, in the case of MERS-CoV E protein some inverted molecules were observed when R38D mutant was expressed. This replacement eliminates the positive charge balance at the C-terminal domain. These data reveal that topological determinants have only a limited effect on viral membrane protein topology as previously observed for other viruses [13] and suggests that E protein in coronaviruses could have evolved to ensure a more robust membrane topology.

Recent statistical studies have suggested that negatively charged residue enrichment in the non-cytoplasmic regions can modulate membrane protein topology [14]. To challenge the robustness of the topology observed for E proteins, we decided to replace the negatively charged residues found in the translocated N-termini in combination with the designed R38D mutations. In the case of MERS-CoV E protein there is only one negatively charged residue (Glu7) in the N-terminal domain, while both SARS-CoV and SARS-CoV-2 E proteins have two (Glu7 and Glu8, Fig. 1A). The combination mutant (E7K & R38D) showed a stronger topology effect on MERS-CoV E protein, since this protein was strongly glycosylated when expressed (Fig. 1C, lanes 15–18). SARS-CoV and SARS-CoV-2 E proteins with the combined mutations (E7&8K & R38D) had only a small proportion of molecules with the reversed topology, suggesting stronger topology determination (Fig. 1C, lanes 19–26), especially if we take into account that the observed effect is generated by a triple mutation. It should be mentioned that the consensus glycosylation acceptor site at Asn48 is not expected to

be modified even if situated luminally due to its close proximity to the membrane [8].

In all three cases, the conserved Arg38 residue plays a limited role in the topology determination. However, its relevance is likely ameliorated with other topological determinants in human coronavirus E protein sequences. Our results suggest that viral evolution has played an important role in strengthening the E protein ( $N_{\text{out}}-C_{\text{in}}$ ) topology from MERS-CoV to SARS coronaviruses. Probably, the R8E mutation present in both SARS-CoVs compared with MERS-CoV is one of the factors contributing to topology robustness, by converting the net charge of 0 at N-terminal region of MERS-CoV into a  $-2$  in both SARS-CoVs, in good agreement with the “negative outside enrichment” rule suggested from statistics derived from a large body of membrane protein sequences [14] and observed in membrane protein structures [15]. At the same time, an evolutionary tendency to accumulate positively charged residues in the cytoplasmic C-terminal domain of these E proteins could be observed (Fig. 1A), contributing to a multifactorial effect on membrane topology, which allows quaternary protein structure formation [4,5] and plays an essential role in viral infection and pathogenesis.

### Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

### Acknowledgements

We thank Prof. Paul Whitley (University of Bath) for critical reading of the manuscript and Pilar Selvi for excellent technical assistance. 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, Culture and Sports (FPU18/05771).

### Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.bbmem.2021.183608>.

### References

- [1] Y.M. Bar-On, A. Flamholz, R. Phillips, R. Milo, SARS-CoV-2 (COVID-19) by the numbers, *eLife* 9 (2020), 1787–15.
- [2] J.L. Nieto-Torres, et al., Subcellular location and topology of severe acute respiratory syndrome coronavirus envelope protein, *Virology* 415 (2011) 69–82.
- [3] C. Verdiá-Báguena, et al., Coronavirus E protein forms ion channels with functionally and structurally-involved membrane lipids, *Virology* 432 (2012) 485–494.
- [4] W. Surya, Y. Li, J. Torres, Structural model of the SARS coronavirus E channel in LMPG micelles, *Biochim. Biophys. Acta Biomembr.* 1860 (2018) 1309–1317.
- [5] V.S. Mandala, et al., Structure and drug binding of the SARS-CoV-2 envelope protein transmembrane domain in lipid bilayers, *Nat. Struct. Mol. Biol.* 27 (2020) 1–24.
- [6] J.M. Jimenez-Guardaño, et al., The PDZ-binding motif of severe acute respiratory syndrome coronavirus envelope protein is a determinant of viral pathogenesis, *PLoS Pathog.* 10 (2014) e1004320.
- [7] I. Alam, et al., Functional pangenome analysis shows key features of E protein are preserved in SARS and SARS-CoV-2, *Front. Cell. Infect. Microbiol.* 10 (2020) e82210–e82219.
- [8] G. Duarte, et al., SARS-CoV-2 envelope protein topology in eukaryotic membranes, *Open Biol.* 10 (2020), 200209–6.
- [9] G. von Heijne, Control of topology and mode of assembly of a polytopic membrane protein by positively charged residues, *Nature* 341 (1989) 456–458.
- [10] M.S. Rahman, et al., Mutational insights into the envelope protein of SARS-CoV-2, *Gene Rep.* 22 (2021) 100997.
- [11] M. Lerch-Bader, C. Lundin, H. Kim, I. Nilsson, G. von Heijne, Contribution of positively charged flanking residues to the insertion of transmembrane helices into the endoplasmic reticulum 105 (2008) 4127–4132.
- [12] 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.

- [13] 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.
- [14] J.A. Baker, W.-C. Wong, B. Eisenhaber, J. Warwicker, F. Eisenhaber, Charged residues next to transmembrane regions revisited: “positive-inside rule” is complemented by the “negative inside depletion/outside enrichment rule”, *BMC Biol.* 15 (2017) 1–29.
- [15] C. Baeza-Delgado, M.A. Martí-Renom, I. Mingarro, Structure-based statistical analysis of transmembrane helices, *Eur. Biophys. J.* 42 (2012) 199–207.