bioRxiv preprint doi: https://doi.org/10.1101/2021.03.08.434384; this version posted March 8, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Comparative studies of the seven human coronavirus envelope proteins using topology prediction and molecular modelling to understand their pathogenicity

Dewald Schoeman[1ϕ], Ruben Cloete[2ϕ], and Burtram C. Fielding[1]*

[1]Molecular Biology and Virology Research Laboratory, Department of Medical Biosciences, University of the Western Cape, Cape Town, South Africa,

[2]South African Medical Research Council Bioinformatics Unit, South African National Bioinformatics Institute, University of the Western Cape, Cape Town, South Africa.

*To whom correspondence should be addressed. Contact: bfielding@uwc.ac.za

[ϕ]Made an equal contribution.

ABSTRACT

Human (h) coronaviruses (CoVs) 229E, NL63, OC43, and HKU1 are less virulent and cause mild, self-limiting respiratory tract infections, while SARS-CoV, MERS-CoV, and SARS-CoV-2, are more virulent and have caused severe outbreaks. The CoV envelope (E) protein, an important contributor to the pathogenesis of severe hCoVs infections, may provide insight into this disparate severity of the disease. Topology prediction programs and 3D modelling software was used to predict and visualize structural aspects of the hCoV E protein related to its functions. All seven hCoV E proteins largely adopted different topologies, with some distinction between the more virulent and less virulent ones. The 3D models refined this distinction, showing the PDZ-binding motif (PBM) of SARS-CoV, MERS-CoV, and SARS-CoV-2 to be more flexible than the PBM of hCoVs 229E, NL63, OC43, and HKU1. We speculate that the increased flexibility of the PBM may provide the more virulent hCoVs with a greater degree of freedom, which can allow them to bind to different host proteins and can contribute to a more severe form of the disease. This is the first paper to predict the topologies and model 3D structures of all seven hCoVs E proteins, providing novel insights for possible drug and/or vaccine development.

Keywords: Human coronaviruses, Envelope protein, Topology, PDZ-binding motif, Molecular modelling, Pathogenicity

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.08.434384; this version posted March 8, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

INTRODUCTION

Of the seven human (h) coronaviruses (CoVs) identified, the recent three, severe acute respiratory syndrome (SARS)-CoV, Middle East respiratory syndrome (MERS)-CoV, and SARS-CoV-2, are the most virulent and have caused severe outbreaks in the last two decades (Broadbent, 2020; Hewings-Martin, 2020). The other four hCoVs, hCoV-229E, -NL63, -OC43, and -HKU1, are less virulent and circulate continuously within the human population, peaking seasonally in different countries all over the world (Aldridge, et al., 2020; Cui, et al., 2011; Edridge, et al., 2020; Gaunt, et al., 2010; Killerby, et al., 2018; Lau, et al., 2006; Su, et al., 2016; Zeng, et al., 2018). They generally cause less severe acute respiratory tract infections, such as the common cold, in immunocompetent persons, but can become more severe in immunocompromised persons, the elderly, and those with chronic, underlying medical conditions (Killerby, et al., 2018; Liu, et al., 2020; Trombetta, et al., 2016). It is unclear, though, why these four hCoVs generally cause less severe disease in immunocompetent persons than the three more virulent hCoVs.

The disease severity and immunopathology of SARS-CoV and SARS-CoV-2 infections have largely been attributed to the envelope (E) protein (DeDiego, et al., 2007; Fett, et al., 2013; Jimenez-Guardeño, et al., 2014; Lamirande, et al., 2008; Nieto-Torres, et al., 2014; Nieto-Torres, et al., 2015; Regla-Nava, et al., 2015; Schoeman and Fielding, 2020; Xia, et al., 2020). No research, however, has demonstrated whether this is also the case for MERS-CoV infections. Despite the poor sequence identity of the E protein between the seven hCoVs, certain functional aspects, such as its postsynaptic density protein 95 (PSD95)/Drosophila disc large tumour suppressor (Dlg1)/zonula occludens-1 protein (zo-1) (PDZ)-binding motif (PBM) and ion-channel (IC) activity, appear to remain conserved and have both been implicated in the pathogenesis and clinical presentation of SARS-CoV and SARS-CoV-2 infections (Huang, et al., 2020; Nieto-Torres, et al., 2014; Nieto-Torres, et al., 2015; Schoeman and Fielding, 2020; Teoh, et al., 2010; Wang, et al., 2020; Wolff, et al., 2020; Xia, et al., 2020). The SARS-CoV E protein can interact with, among others, host cell proteins syntenin and the protein associated with *Caenorhabditis elegans* lin-7 protein 1 (PALS1) in the cytoplasm of host cells through its C-terminal PBM, highlighting the importance of its topology (Jimenez-Guardeño, et al., 2014; Teoh, et al., 2010). Recently, a peptide mimicking the PBM of SARS-CoV-2 E was also shown to be capable of interacting with PALS1 (Toto, et al., 2020). Despite some experimental evidence, the topology of the hCoV E protein remains largely under debate and data is limited to the prototypic SARS-CoV and to SARS-CoV-2 (Arbely, et al., 2004; Duart, et al., 2020; Khattari, et al., 2006; Nieto-Torres, et al., 2011; Yuan, et al., 2006). The topology of the E protein for the less virulent hCoVs, however, is unknown with no evidence to suggest whether or not it adopts a topology similar to the more virulent hCoVs. Similarly, the three-dimensional (3D) structure of the less virulent hCoVs has not been experimentally resolved to date. To date, the only hCoV E protein structure that has been resolved experimentally is that of SARS-CoV E for which two templates, viz. 5x29 and 2mm4 are available, both spanning residues 8-65 of the protein (Li, et al., 2014; Surya, et al., 2018). The SARS-CoV E protein also has other functions and is involved in other host cell process based on its interaction with different host cell proteins (Schoeman and Fielding, 2019). Given the various functions of E protein in the CoV life

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.08.434384; this version posted March 8, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

cycle, it is suggested to be capable of adopting multiple topologies, depending on the respective function (Kuo, et al., 2007).

The complexity and largely hydrophobic nature of membrane proteins, such as the CoV E protein, presents a challenge in studying their structure and dynamics experimentally (Arbely, et al., 2004; Latek, et al., 2019; Wu, et al., 2003). In the absence of experimental data, bioinformatics software tools may provide useful answers in dissecting the topologies, which can then be explored experimentally. Recently, Duart, et al. (2020) used the prediction programs ΔG predictor, TMHMM, MEMSAT-SVM, TMpred, HMMTop, Phobius, TOPCONS to predict the topology of SARS-CoV-2 E protein, and validated the topology experimentally. The N-terminus of the SARS-CoV-2 E protein was oriented to the lumen of the endoplasmic reticulum-Golgi intermediate compartment (ERGIC) while the C-terminus was located cytoplasmically – a topology that would enable the C-terminal PBM to interact with host cell proteins in the cytoplasm. Since there is very little consensus on the topology of the E protein of more virulent hCoVs and none for the less virulent hCoVs, this study aims to employ the prediction programs used by Duart, et al. (2020) to predict the topology of the E protein for all seven hCoVs. Defining the membrane topology of the less virulent hCoVs can allude to whether they can interact with host cell proteins in a way similar to what the more virulent hCoVs do. This may provide insight into understanding the mechanism behind why hCoVs-229E, -NL63, -OC43, and -HKU1 are less virulent than their more virulent and recently emerging counterparts. It can also provide more insight into whether all hCoVs share a similar mechanism of assembly and release (Ruch and Machamer, 2012). The 3D structural models will provide insight into the pathogenicity of the virulent hCoVs by predicting the secondary and tertiary structural fold of the E protein's PBM and thereby, establishing a structure-function relationship between the pathogenicity of the more virulent hCoVs vs the less virulent forms of hCoVs.

METHODS

Topology prediction of hCoV E protein

The amino acid sequences for the E protein of all seven hCoVs were obtained from the UniProt KB database (https://www.uniprot.org/) and only complete and "Reviewed" sequences were selected. The topology of the E proteins was predicted using the prediction programs ΔG Predictor (Hessa, et al., 2007) (http://dgpred.cbr.su.se/), TMHMM (Sonnhammer, et al., 1998) (http://www.cbs.dtu.dk/services/TMHMM/), MEMSAT-SVM (Nugent and Jones, 2009) (http://bioinf.cs.ucl.ac.uk/psipred/), TMpred (Hofmann, 1993) (https://embnet.vital-it.ch/software/TMPRED_form.html), HMMTop (Tusnady and Simon, 1998; Tusnady and Simon, 2001) (http://www.enzim.hu/hmmtop/), Phobius (Käll, et al., 2004) (http://phobius.sbc.su.se/), and TOPCONS (Tsirigos, et al., 2015) (http://topcons.net/). For each program, the user-adjustable parameters were kept at their default settings.

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.08.434384; this version posted March 8, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Molecular modelling

Three-dimensional protein structures were constructed for all seven hCoV E proteins using MODELLER software to compare the structural features between the more virulent and less virulent forms (Eswar, et al., 2008; Sali and Blundell, 1993).

Template selection and model construction

Two nuclear magnetic resonance (NMR)-resolved structures for SARS-CoV E (PDBID: 2mm4 and PDBID: 5x29) were obtained from the protein data bank (PDB) (Li, et al., 2014; Surya, et al., 2018). These structures were used to generate full-length 3D models for SARS-CoV (template: 5x29), SARS-CoV-2 (template: 5x29), MERS-CoV (template: 2mm4), and hCoVs 229E (template: 5x29), NL63 (template: 229E protein model), OC43 (template: 5x29), and HKU1 (template: 229E protein model). The routinely used python script, align2d.py, in MODELLER was used to perform an alignment prediction between each of the hCoV E protein sequences and the respective template sequence (Eswar, et al., 2008; Sali and Blundell, 1993). Thereafter, a 3D model was built for each hCoV E protein using the model-ligand.py script (Eswar, et al., 2008; Sali and Blundell, 1993). As a validation step, we also used the protein structure prediction server, ITASSER, to generate 3D structures for the seven hCoV E proteins and compared the structures to the MODELLER predicted structures (Zhang, 2008).

Quality assessment

The phi and psi dihedral angle parameters for the Ramachandran plot was calculated for each of the predicted protein models using PROCHECK webserver (Laskowski, et al., 1993). The root mean square deviation analysis was done by performing structural alignment between the predicted model structure and the homologous template protein. This was done to assess if any structural deviation exists within the main chain atoms of the two protein structures. The predicted structures were visualized using PyMol molecular graphics software (DeLano, 2002).

RESULTS

Predicted topology of hCoV E protein

The output from each prediction program for the E protein of each hCoV can be found in the supplementary information (Tables S1-S7). The predicted topology of the E protein for all seven hCoVs exhibits little consensus, with only a marginal similarity within the more virulent hCoVs and the less virulent hCoVs, respectively (Table 1). While some of the more virulent and the less virulent hCoV E proteins have been predominantly predicted to possess two transmembrane domains (TMDs), the majority of the E proteins were predicted to have only one TMD.

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.08.434384; this version posted March 8, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Topology of virulent hCoV E proteins

Both SARS-CoV E and SARS-CoV-2 E were predicted to have one TMD with a lumenal N-terminus (N_{lumen}) and cytoplasmic C-terminus (C_{cyto}), while MERS-CoV E was predicted to adopt a contrasting topology of either a N_{cyto}/C_{lumen} with a single TMD, or a N_{lumen}/C_{lumen} with two TMDs, with an equal number of predictions for each topology (Table S3).

Topology of less virulent hCoV E proteins

The predicted topologies of the less virulent hCoVs were almost equally as inconsistent as the predicted topologies of the more virulent hCoVs. While hCoVs OC43 and HKU1 predominantly exhibited the same predicted topologies in having two TMDs and both N_{lumen}/C_{lumen} orientations, hCoVs 229E and NL63 did not share this topology, nor were the respective topologies of the latter two similar. The hCoV-229E E protein was predominantly predicted to adopt a N_{lumen}/C_{cyto} topology with a single TMD, while the hCoV-NL63 E protein was predicted to exhibit a N_{cyto}/C_{lumen} topology also with a single TMD (Table 1).

Table 1. The predominant predicted topologies and number of transmembrane domains (TMDs) for all seven human (h) coronavirus (CoV) envelope (E) proteins.

hCoV	N-terminus	C-Terminus	Number of TMDs
SARS-CoV	Lumen	Cytoplasm	1
SARS-CoV-2	Lumen	Cytoplasm	1
MERS-CoV	Cytoplasm	Lumen	1
	Lumen	Lumen	2
hCoV-229E	Lumen	Cytoplasm	1
hCoV-NL63	Cytoplasm	Lumen	1
hCoV-OC43	Lumen	Lumen	2
hCoV-HKU1	Lumen	Lumen	2

Accession numbers for the E proteins of SARS-CoV (P59637), SARS-CoV-2 (P0DTC4), MERS-CoV (K9N5R3), hCoV-229E (P19741), hCoV-NL63 (Q6Q1S0), hCoV-OC43 (Q04854), and hCoV-HKU1 (Q5MQC8) were obtained from the UniProt KB database. Predictions were made using prediction programs ΔG Predictor, TMHMM, MEMSAT-SVM, TMpred, HMMTop, Phobius, TOPCONS. For each hCoV, the predicted orientation of the amino (N) and carboxy (C) terminus of the respective E protein is indicated as being oriented towards the lumen of the endoplasmic reticulum-Golgi intermediate compartment (ERGIC) or the cytoplasm of infected host cells.

3D models of hCoV E protein

Virulent hCoVs: SARS-CoV, SARS-CoV-2, and MERS-CoV

The SARS-CoV E protein shared 91% sequence identity with the amino acid sequence of template 5x29 (Figure S1A). The 3D model showed three α-helices and four coil regions, with the C-terminal PBM adopting a flexible coil region (Figure 1A). Quality assessment of

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.08.434384; this version posted March 8, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

the model revealed that 90% of the residues were in the most favoured regions of the Ramachandran plot and 1.4% were in the disallowed regions. The RMSD analysis showed a 0.698Å difference between SARS-CoV E and the template 5x29.

Similarly, SARS-CoV-2 E shared 91% sequence identity with template 5x29 (Figure S1B). The 3D model also showed three α-helices and four coil regions, with the PBM adopting a flexible coil region (Figure 1B). Quality assessment revealed that 88.6% of the residues were in the most favoured regions of the Ramachandran plot and 0% were in the disallowed regions. The RMSD analysis showed a 2.155Å difference between SARS-CoV-2 E and template 5x29.

The MERS-CoV E protein shared a 35% sequence identity to template 2mm4 (Figure S1C). While the 3D model showed only two α-helices and three coil regions, the PBM also adopted a flexible coil region (Figure 2C). Quality assessment indicated that 95.8% of the residues were in the most favoured regions of the Ramachandran plot and 1.4% were in the disallowed regions. The RMSD analysis between MERS-CoV E and the homologous template structure 2mm4 indicated a difference of 1.458Å.

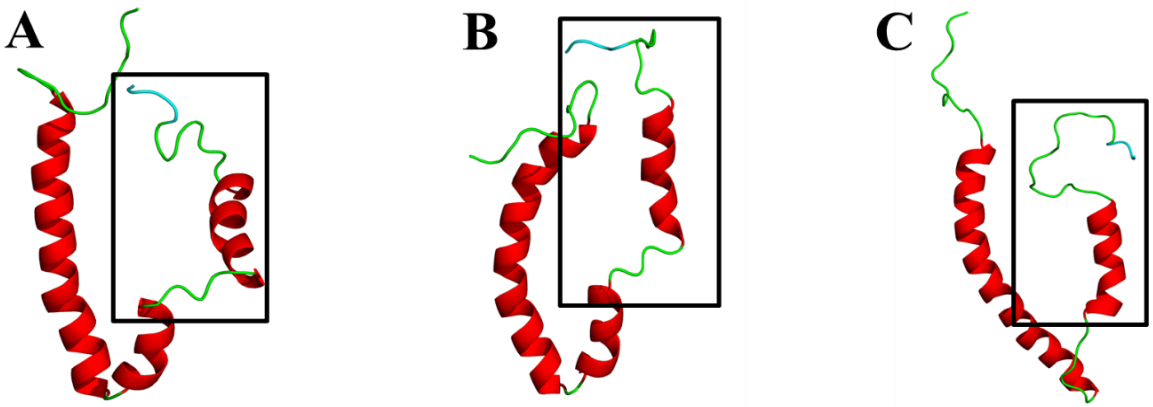

Fig. 1 Cartoon representation of the three-dimensional (3D) models of the envelope (E) protein of the more virulent hCoVs. A. SARS-CoV E. **B.** SARS-CoV-2 E. **C.** MERS-CoV E. Models were generated using MODELLER software and based on two nuclear magnetic resonance (NMR)-resolved structures for SARS-CoV E (PDBID: 5x29 and PDBID: 2mm4) obtained from the protein data bank (PDB) (Eswar, et al., 2008; Li, et al., 2014; Sali and Blundell, 1993; Surya, et al., 2018). Models of the SARS-CoV and SARS-CoV-2 E proteins were generated from the 5x29 template, whereas the MERS-CoV E protein was generated from 2mm4 template. The PDZ-binding motif (PBM) is coloured in cyan and the squared region shows the variable C-terminal domain.

Non-virulent hCoVs: hCoV-229E, -NL63, -OC43, and -HKU1

The hCoV-229E E protein shared a 29% sequence identity with 5x29 amino acid sequence (Figure S2A). While the 3D model showed four α-helices and four coil regions, the C-terminal PBM adopted one α-helical turn conformation with reduced flexibility (Figure 2A) and Val74 shared conservation with Val78 from 5x29. Quality assessment of the model indicated that 91.5% of the residues were in most favoured regions of the Ramachandran plot

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.08.434384; this version posted March 8, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

and 0% were in the disallowed regions. The RMSD analysis between the hCoV-229E E protein and the homologous template structure 5x29 indicated a 1.776Å difference.

The hCoV-NL63 E protein shared 47% sequence identity with the amino acid sequence of the hCoV-229E E protein. (Figure S2B). Similar to E of hCoV-229E, the 3D model of hCoV-NL63 showed four α-helices and four coil regions, where the PBM adopted one α-helical turn conformation with reduced flexibility (Figure 2B), and conservation of Val74 was shared with Val74 from the E protein of hCoV-229E. Quality assessment indicated that 91% of the residues were in most favoured regions of the Ramachandran plot and 0% were in the disallowed regions. The RMSD analysis between hCoV-NL63 E and the homologous template structure (229E protein model) indicated a 1.217Å difference.

The hCoV-OC43 shared 22% sequence identity with the 5x29 amino acid sequence (Figure S2C). This structure also showed four α-helices and four coil regions, the PBM also adopted one α-helical turn conformation with reduced flexibility (Figure 2C), and Val81 shared conservation with Val78 from template 5x29. Quality assessment indicated that 90.8% of the residues were in the most favoured regions of the Ramachandran plot and 1.3% were in the disallowed regions. The RMSD analysis between hCoV-OC43 E and the homologous template structure 5x29 indicated a 3.328Å difference.

The hCoV-HKU1 E protein shared 27% sequence identity with the E protein amino acid sequence of hCoV-229E (Figure S2D). This model showed only three α-helices and three coil regions, while the PBM also adopted one α-helical turn conformation with reduced flexibility (Figure 2D), and Ile82 shared conservation with Ile75 from hCoV-229E E protein. Quality assessment indicated that 89.2% of the residues were in most favoured regions of the Ramachandran plot and 1.4% were in the disallowed regions. The RMSD analysis between hCoV-HKU1 E and the homologous template structure 229E indicated a 5.343Å difference.

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.08.434384; this version posted March 8, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

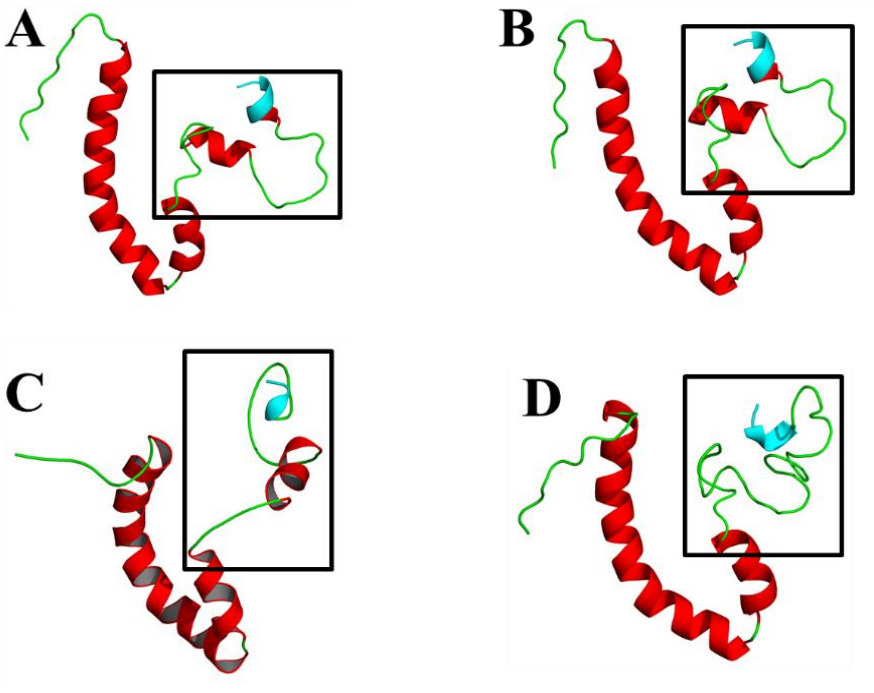

Fig. 2 Cartoon representation of the three-dimensional (3D) models of the envelope (E) protein of the less virulent hCoVs. A. hCoV-229E E. **B.** hCoV-NL63 E. **C.** hCoV-OC43 E. **D.** hCoV-HKU1 E. Models were generated using MODELLER software and based on two nuclear magnetic resonance (NMR)-resolved structures for SARS-CoV E (PDBID: 2mm4 and PDBID: 5x29) obtained from the protein data bank (PDB) (Eswar, et al., 2008; Li, et al., 2014; Sali and Blundell, 1993; Surya, et al., 2018). Models of the hCoV-NL63 and -OC43 E proteins were generated from the 5x29 template, whereas models of the hCoV-229E and -HKU1 E proteins were generated from a 229E template. The PDZ-binding motif (PBM) is coloured in cyan and the squared region shows the variable C-terminal domain.

DISCUSSION

Despite the differences in the predicted number of TMDs for all seven hCoVs, the majority of hCoVs were predicted to have only one TMD, the length and number of which is largely consistent with reports for E in other CoVs (Lim and Liu, 2001; Liu, et al., 2007; Nieto-Torres, et al., 2011; Torres, et al., 2007; Torres, et al., 2005; Ye and Hogue, 2007; Yuan, et al., 2006). The predicted N_{lumen} and C_{cyto} topology of both the SARS-CoV E and SARS-CoV-2 E is consistent with what has been experimentally determined for the topology of untagged SARS-CoV E and c-myc-tagged SARS-CoV-2 E (Duart, et al., 2020; Nieto-Torres, et al., 2011). This similarity in topology is to be expected given the approximate 94% similarity in E protein amino acid sequence (Grifoni, et al., 2020; Schoeman and Fielding, 2020).

Given the lack of experimental evidence for the topology of MERS-CoV E, the reason(s) for and/or consequence(s) of these predicted topologies is not known. It may, however, provide some insight into why the human-human transmission of MERS-CoV is less effective than that of SARS-CoV and SARS-CoV-2 (Drosten, et al., 2014; Grijalva, et al., 2020; Wang, et al., 2020; Wilson-Clark, et al., 2006). The topology of membrane proteins, such as the E protein, plays an important part in their function, with certain functions, such as viral

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.08.434384; this version posted March 8, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

assembly and release of virions, being dependent on protein topology (Ruch and Machamer, 2012; Seppälä, et al., 2010; White, et al., 2015). The exact role of the CoV E protein in the assembly and release of newly formed virions was recently made clearer from the study of the infectious bronchitis virus (IBV) E protein (Westerbeck and Machamer, 2019). A low-molecular-weight (LMW), likely monomeric, form of IBV E increased the lumenal pH of the Golgi network employed by CoVs to assemble and egress from infected cells. This increased Golgi pH likely protects the spike (S) protein from premature cleavage, preventing the release of the trimer subunits prior to receptor binding. Interestingly, this disruption of the host secretory pathway was mediated by the E protein independent of its ion-channel activity, likely requiring interaction between the cytoplasmic tail of IBV E and a host protein capable of altering the pH of the Golgi network. Without the C-terminus of IBV E exposed to the cytoplasm, interaction with the necessary host protein would be unlikely and could result in the production of impaired, possibly non-infectious virions (Corse and Machamer, 2000; Westerbeck and Machamer, 2019). Neither of the predicted MERS-CoV E topologies reflects a cytoplasmic C-terminus, suggesting that MERS-CoV E might likely not be capable of interacting with a host protein in a way similar to IBV E to alter the host secretory pathway for the effective release of infectious virions. If MERS-CoV E does, in fact, adopt either of these topologies, it could explain why the virus is less effective at human-human transmission. Interestingly, an early X-ray scattering study suggested that E adopts a helical hairpin topology, whereas subsequent studies of solution NMR consistently demonstrated a single-span topology in which E was bound to several detergent micelles (Arbely, et al., 2004; Li, et al., 2014; Pervushin, et al., 2009; Surya, et al., 2018). This demonstrates the need to be cautious about which technique is used to determine E topology and take into account the environment to which E is exposed during the study. Nevertheless, the predicted topologies may merely reflect the ability of the E protein to adopt multiple topologies to perform specific functions (Kuo, et al., 2007; Westerbeck and Machamer, 2015; Westerbeck and Machamer, 2019). The lack of experimental evidence also impedes the reason(s) for and consequences of the predicted topologies.

The severe lack of experimental data on the E protein of the less virulent hCoVs makes it especially difficult to validate both the accuracy of the predicted topologies and their relevance. The requirement of the C-terminally located PBM of SARS-CoV E in the cytoplasm has been well-established as crucial to its pathogenesis, while no distinct function has been ascribed to the N-terminus (Castaño-Rodriguez, et al., 2018; Jimenez-Guardeño, et al., 2014; Teoh, et al., 2010). Thus far, only one study has investigated a possible function of the protein's N-terminus whereby a chimeric SARS-CoV E protein was shown to contain possible Golgi-targeting information in its N-terminus (Cohen, et al., 2011). It should, however, be confirmed using the full-length protein as well.

A N_{lumen}/C_{cyto} topology enables the SARS-CoV E PBM to interact with host cell proteins such as syntenin and PALS1. More recently, SARS-CoV-2 E has also been shown to interact with PALS1 in a similar manner, very likely since it adopts the same topology as SARS-CoV E (Duart, et al., 2020; Nieto-Torres, et al., 2011). If the C-terminus of SARS-CoV E was not oriented towards the cytoplasm, it would not be able to interact with these host proteins. Therefore, since the E proteins of the less virulent hCoVs NL63, OC43, and HKU1 were

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.08.434384; this version posted March 8, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

predicted to have a lumenal C-terminus, it might not allow for interaction with the same host proteins as SARS-CoV E; a feature that could likely explain why these hCoVs cause less severe infections compared to SARS-CoV. The same, however, cannot be said for the predicted topology of the less virulent hCoV-229E E protein. Despite sharing a topology similar to SARS-CoV E and SARS-CoV-2 E, hCoV-229E is self-limiting and causes less severe symptoms than the former two hCoVs (Lew, et al., 2003; Pappas, et al., 2008; Poutanen, 2018). Nevertheless, given the various functions of E in the CoV life cycle, it has also been proposed to assume multiple topologies, depending on the function (Nieto-Torres, et al., 2011; Ruch and Machamer, 2012; Yuan, et al., 2006). Since only a small proportion of the E protein produced during an infection is incorporated into CoV virions, it stands to reason that one topology might be adopted to facilitate the assembly of new viral progeny, while another is involved in the release through the secretory pathway (Westerbeck and Machamer, 2019). This was attempted where the IBV E protein was produced to adopt either a transmembrane topology, with one TMD or a membrane hairpin topology, containing two TMDs (Ruch and Machamer, 2012).

Both SARS-CoV E and SARS-CoV-2 E shared a 91% sequence identity with the amino acid sequence of template 5x29, suggesting high sequence homology (Figures S1A and S1B). This is to be expected given the 94% sequence similarity between SARS-CoV E and SARS-CoV-2 E, and since 5x29 spans residues 8-65 of the SARS-CoV E UniProt reference sequence (P59637), which excludes the PBM (Grifoni, et al., 2020; Schoeman and Fielding, 2020). The RMSD analysis showed a 0.698Å and 2.155Å difference between SARS-CoV E and SARS-CoV-2 E and the template 5x29, respectively, suggesting very little structural deviation between the respective two structures and the template. The RMSD analysis between MERS-CoV E and the homologous template structure 2mm4 indicated 1.458Å difference also suggesting a slight amount of structural deviation between the two structures. The 3D structures predicted for the E proteins of SARS-CoV, SARS-CoV-2, and MERS-CoV, therefore, successfully satisfied the Ramachandran plot dihedral angle distributions and, based on the low RMSD values, the correct protein folds were assigned to the respective protein sequences.

Despite the moderate sequence identity between the target sequence of the hCoV-229E E protein and the 5x29 template, the protein model from this pairwise sequence alignment will provide a medium accuracy protein model which is still useful for domain comparisons. The RMSD analysis between hCoV-229E E and the homologous template structure 5x29 indicated a 1.776Å difference, suggesting very little deviation between the two structures. The RMSD analysis between hCoV-NL63 E and the homologous template structure (229E protein model) similarly indicated a 1.217Å difference, also suggesting very little deviation between the generated model and the homologous template structure. Thus, the 3D structures predicted for the E proteins of hCoVs 229E and NL63 successfully passed quality assessment and based on the low RSMD values, the correct protein folds were assigned to the respective protein sequences.

The RMSD analysis between hCoV-OC43 E and the homologous template structure 5x29 indicated 3.328Å difference, whereas the RMSD analysis between hCoV-HKU1 E and the

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.08.434384; this version posted March 8, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

354 homologous template structure 229E indicated a 5.343Å difference, both of which suggest a
355 large deviation between the generated models and their respective homologous template
356 structures. Although the 3D structure predicted for the hCoV-HKU1 E protein satisfied the
357 Ramachandran quality check, a high RSMD value suggests the incorrect fold was assigned to
358 the protein sequence.

359 The E protein structures of the less virulent hCoVs OC43 and HKU1 predicted by
360 MODELLER had the lowest sequence identities and the highest RMSD values compared to
361 their templates. The high RMSD values are due to misorientations of the C-terminal domain
362 regions. This is a common limitation of homology modelling if sequence identity and
363 coverage are low in a specific region resulting in low accuracy protein models. However, the
364 ITASSER predicted protein models for OC43 and HKU1 displayed lower RMSD values
365 when aligned to template 2mm4, although the PBM was not conserved and displayed a
366 flexible coil region (Figures S3A and S3B). We are confident in the E protein models
367 predicted using MODELLER that display a less flexible PBM for the less virulent hCoVs.

368 The flexibility of the region in which the E protein PBM is located appears to distinguish the
369 more virulent hCoVs quite clearly from the less virulent hCoVs; the former containing a
370 PBM with a more flexible coil region, while the latter is characterised by a PBM that contains
371 a less flexible α-helical turn. While so far only five interaction partners have been identified
372 for SARS-CoV E, more may yet be identified and those that have already been identified
373 have very different structures (Jimenez-Guardeño, et al., 2014; Nieto-Torres, et al., 2011;
374 Teoh, et al., 2010; Yang, et al., 2005). This could quite possibly be a pathogenic feature that
375 allows the more virulent hCoVs to interact with a wider range of host proteins which could
376 explain their increased pathogenicity and severity of the disease they cause. By comparison,
377 the less flexible α-helical turn of the less virulent hCoV E protein PBM could very likely be
378 an impeding characteristic, preventing or limiting interaction with host proteins, which could
379 possibly explain the limited pathogenic capabilities of the less virulent hCoVs. Without
380 experimental data, however, this remains to be determined.

381 Clearly, much is still unknown about the topology of the hCoV E protein and how different
382 hCoVs can exhibit different topologies while still performing the same functions. However,
383 this is the first study, to predict and generate 3D models of the E protein for all seven hCoVs.
384 This study also attempted to determine a distinguishing feature between the E proteins of the
385 more virulent and the less virulent hCoVs to provide a possible explanation for the difference
386 in pathogenicity and disease severity. SARS-CoV is the only hCoV for which E protein
387 interaction partners have been identified. It, therefore, remains to be determined whether the
388 E protein of any of the less virulent hCoVs can also interact with host cell proteins similar to
389 SARS-CoV E, and, if so, which host cell proteins the less virulent hCoV's E proteins can
390 interact with. As it stands, there is no experimental evidence on the topology of the E protein
391 for any of the less virulent hCoVs, leaving questions to be answered: can one protein adopt
392 multiple topologies, each one for a different function? If the E protein of less virulent hCoVs
393 exhibits a topology opposite to that of the more virulent hCoVs, can the less virulent hCoVs
394 still perform similar pathogenic functions like ion-channel activity and viral-host protein-
395 protein interactions with a PBM? Answers to such questions could further help to explain the

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.08.434384; this version posted March 8, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

396 difference in severity between the different hCoVs and could provide insight into membrane
397 protein topology- and structure-function relationships as well.

398

399 **Acknowledgements**

400 We apologize to any author whose work has been inadvertently omitted from this article.

401

402 **Funding**

403 This work has been supported by the National Research Foundation of South Africa; the
404 Poliomyelitis Research Foundation (PRF) of South Africa [17/53 and 19/06]; and the Higher
405 Education Department, next Generation of Academic Programme (nGAP) in the form of full-
406 time academic positions and salaries.

407

408 **Conflict of Interest**: none declared.

409

410 **REFERENCES**

411 1. Aldridge, R.W., *et al.* Seasonality and immunity to laboratory-confirmed seasonal
412 coronaviruses (HCoV-NL63, HCoV-OC43, and HCoV-229E): results from the Flu Watch
413 cohort study. *Wellcome Open Research* 2020;5(52):52.

414 2. Arbely, E., *et al.* A highly unusual palindromic transmembrane helical hairpin formed by
415 SARS coronavirus E protein. *J. Mol. Biol.* 2004;341(3):769-779.

416 3. Broadbent, L. Coronaviruses – a brief history. In.; 2020.

417 4. Castaño-Rodriguez, C., *et al.* Role of severe acute respiratory syndrome Coronavirus
418 Viroporins E, 3a, and 8a in replication and pathogenesis. *mBio* 2018;9(3):e02325-02317.

419 5. Cohen, J.R., Lin, L.D. and Machamer, C.E. Identification of a Golgi complex-targeting
420 signal in the cytoplasmic tail of the severe acute respiratory syndrome coronavirus
421 envelope protein. *J. Virol.* 2011;85(12):5794-5803.

422 6. Corse, E. and Machamer, C.E. Infectious bronchitis virus E protein is targeted to the
423 Golgi complex and directs release of virus-like particles. *J. Virol.* 2000;74(9):4319-4326.

424 7. Cui, L.-J., *et al.* Human coronaviruses HCoV-NL63 and HCoV-HKU1 in hospitalized
425 children with acute respiratory infections in Beijing, China. *Adv. Virol.* 2011;2011.

426 8. DeDiego, M.L., *et al.* A severe acute respiratory syndrome coronavirus that lacks the E
427 gene is attenuated *in vitro* and *in vivo*. *J. Virol.* 2007;81(4):1701-1713.

428 9. DeLano, W.L. Pymol: An open-source molecular graphics tool. *CCP4 Newsletter on*
429 *protein crystallography* 2002;40(1):82-92.

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.08.434384; this version posted March 8, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

10. Drosten, C., *et al.* Transmission of MERS-coronavirus in household contacts. *N. Engl. J. Med.* 2014;371(9):828-835.

11. Duart, G., *et al.* SARS-CoV-2 envelope protein topology in eukaryotic membranes. *Open Biol* 2020;10(9):200209.

12. Edridge, A.W., *et al.* Coronavirus protective immunity is short-lasting. *medRxiv* 2020:2020.2005.2011.20086439.

13. Eswar, N., *et al.* Protein structure modeling with MODELLER. *Methods Mol. Biol.* 2008;426:145-159.

14. Fett, C., *et al.* Complete protection against severe acute respiratory syndrome coronavirus-mediated lethal respiratory disease in aged mice by immunization with a mouse-adapted virus lacking E protein. *J. Virol.* 2013;87(12):6551-6559.

15. Gaunt, E.R., *et al.* Epidemiology and clinical presentations of the four human coronaviruses 229E, HKU1, NL63, and OC43 detected over 3 years using a novel multiplex real-time PCR method. *J. Clin. Microbiol.* 2010;48(8):2940-2947.

16. Grifoni, A., *et al.* A Sequence Homology and Bioinformatic Approach Can Predict Candidate Targets for Immune Responses to SARS-CoV-2. *Cell Host & Microbe* 2020;27(4):671-680.e672.

17. Grijalva, C.G., *et al.* Transmission of SARS-COV-2 Infections in Households - Tennessee and Wisconsin, April-September 2020. *Morbidity and Mortality Weekly Report* 2020;69(44):1631-1634.

18. Hessa, T., *et al.* Molecular code for transmembrane-helix recognition by the Sec61 translocon. *Nature* 2007;450(7172):1026-1030.

19. Hewings-Martin, Y. How do SARS and MERS compare with COVID-19? In.; 2020.

20. Hofmann, K. TMbase-A database of membrane spanning proteins segments. *Biol. Chem. Hoppe Seyler* 1993;374:166.

21. Huang, C., *et al.* Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *The Lancet* 2020;395(10223):497-506.

22. Jimenez-Guardeño, J.M., *et al.* The PDZ-binding motif of severe acute respiratory syndrome coronavirus envelope protein is a determinant of viral pathogenesis. *PLoS Pathog.* 2014;10(8):e1004320.

23. Käll, L., Krogh, A. and Sonnhammer, E.L. A combined transmembrane topology and signal peptide prediction method. *J. Mol. Biol.* 2004;338(5):1027-1036.

24. Khattari, Z., *et al.* SARS coronavirus E protein in phospholipid bilayers: an x-ray study. *Biophys. J.* 2006;90(6):2038-2050.

25. Killerby, M.E., *et al.* Human coronavirus circulation in the United States 2014–2017. *J. Clin. Virol.* 2018;101:52-56.

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.08.434384; this version posted March 8, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

26. Kuo, L., Hurst, K.R. and Masters, P.S. Exceptional flexibility in the sequence requirements for coronavirus small envelope protein function. *J. Virol.* 2007;81(5):2249-2262.

27. Lamirande, E.W., *et al.* A live attenuated severe acute respiratory syndrome coronavirus is immunogenic and efficacious in golden Syrian hamsters. *J. Virol.* 2008;82(15):7721-7724.

28. Laskowski, R.A., *et al.* PROCHECK: a program to check the stereochemical quality of protein structures. *Journal of applied crystallography* 1993;26(2):283-291.

29. Latek, D., *et al.* Modeling of Membrane Proteins. In, *Computational Methods to Study the Structure and Dynamics of Biomolecules and Biomolecular Processes*. Springer; 2019. p. 371-451.

30. Lau, S.K., *et al.* Coronavirus HKU1 and other coronavirus infections in Hong Kong. *J. Clin. Microbiol.* 2006;44(6):2063-2071.

31. Lew, T.W.K., *et al.* Acute Respiratory Distress Syndrome in Critically Ill Patients With Severe Acute Respiratory Syndrome. *JAMA* 2003;290(3):374-380.

32. Li, Y., *et al.* Structure of a conserved Golgi complex-targeting signal in coronavirus envelope proteins. *J. Biol. Chem.* 2014;289(18):12535-12549.

33. Lim, K. and Liu, D. The missing link in coronavirus assembly: retention of the avian coronavirus infectious bronchitis virus envelope protein in the pre-golgi compartments and physical interaction between the envelope and membrane proteins. *J. Biol. Chem.* 2001;276(20):17515-17523.

34. Liu, D., Yuan, Q. and Liao, Y. Coronavirus envelope protein: a small membrane protein with multiple functions. *Cell. Mol. Life Sci.* 2007;64(16):2043-2048.

35. Liu, D.X., Liang, J.Q. and Fung, T.S. Human Coronavirus-229E, -OC43, -NL63, and -HKU1. *Reference Module in Life Sciences* 2020:B978-970-912-809633-809638.821501-X.

36. Nieto-Torres, J.L., *et al.* Subcellular location and topology of severe acute respiratory syndrome coronavirus envelope protein. *Virology* 2011;415(2):69-82.

37. Nieto-Torres, J.L., *et al.* Severe acute respiratory syndrome coronavirus envelope protein ion channel activity promotes virus fitness and pathogenesis. *PLoS Pathog.* 2014;10(5):e1004077.

38. Nieto-Torres, J.L., *et al.* Severe acute respiratory syndrome coronavirus E protein transports calcium ions and activates the NLRP3 inflammasome. *Virology* 2015;485:330-339.

39. Nugent, T. and Jones, D.T. Transmembrane protein topology prediction using support vector machines. *BMC Bioinformatics* 2009;10(1):159.

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.08.434384; this version posted March 8, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

40. Pappas, D.E., *et al.* Symptom profile of common colds in school-aged children. *The Pediatric Infectious Disease Journal* 2008;27(1):8-11.

41. Pervushin, K., *et al.* Structure and inhibition of the SARS coronavirus envelope protein ion channel. *PLoS Pathog.* 2009;5(7):e1000511.

42. Poutanen, S.M. 222 - Human Coronaviruses. In: Long, S.S., Prober, C.G. and Fischer, M., editors, *Principles and Practice of Pediatric Infectious Diseases (Fifth Edition)*. Elsevier; 2018. p. 1148-1152.e1143.

43. Regla-Nava, J.A., *et al.* SARS coronaviruses with mutations in E protein are attenuated and promising vaccine candidates. *J. Virol.* 2015:JVI. 03566-03514.

44. Ruch, T.R. and Machamer, C.E. The coronavirus E protein: Assembly and beyond. *Viruses* 2012;4(3):363-382.

45. Ruch, T.R. and Machamer, C.E. A single polar residue and distinct membrane topologies impact the function of the infectious bronchitis coronavirus E protein. *PLoS Pathog.* 2012;8(5):e1002674.

46. Sali, A. and Blundell, T.L. Comparative protein modelling by satisfaction of spatial restraints. *J. Mol. Biol.* 1993;234(3):779-815.

47. Schoeman, D. and Fielding, B.C. Coronavirus envelope protein: current knowledge. *Virol J.* 2019;16(1):69.

48. Schoeman, D. and Fielding, B.C. Is There a Link Between the Pathogenic Human Coronavirus Envelope Protein and Immunopathology? A Review of the Literature. *Front. Microbiol.* 2020;11:2086.

49. Seppälä, S., *et al.* Control of Membrane Protein Topology by a Single C-Terminal Residue. *Science* 2010;328(5986):1698-1700.

50. Sonnhammer, E.L., von Heijne, G. and Krogh, A.S. A hidden Markov model for prediction transmembrane helices in protein sequences. In, *Sixth International Conference on Intelligent Systems for Molecular Biology*. AAAI Press; 1998. p. 175-118.

51. Su, S., *et al.* Epidemiology, genetic recombination, and pathogenesis of coronaviruses. *Trends Microbiol.* 2016;24(6):490-502.

52. Surya, W., Li, Y. and Torres, J. Structural model of the SARS coronavirus E channel in LMPG micelles. *Biochim Biophys Acta Biomembranes* 2018;1860(6):1309-1317.

53. Teoh, K.-T., *et al.* The SARS coronavirus E protein interacts with PALS1 and alters tight junction formation and epithelial morphogenesis. *Mol. Biol. Cell* 2010;21(22):3838-3852.

54. Torres, J., *et al.* Conductance and amantadine binding of a pore formed by a lysine-flanked transmembrane domain of SARS coronavirus envelope protein. *Protein Sci.* 2007;16(9):2065-2071.

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.08.434384; this version posted March 8, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

55. Torres, J., *et al.* The transmembrane oligomers of coronavirus protein E. *Biophys. J.* 2005;88(2):1283-1290.

56. Toto, A., *et al.* Comparing the binding properties of peptides mimicking the Envelope protein of SARS-CoV and SARS-CoV-2 to the PDZ domain of the tight junction-associated PALS1 protein. *Protein Sci.* 2020;29(10):2038-2042.

57. Trombetta, H., *et al.* Human coronavirus and severe acute respiratory infection in Southern Brazil. *Pathogens and global health* 2016;110(3):113-118.

58. Tsirigos, K.D., *et al.* The TOPCONS web server for consensus prediction of membrane protein topology and signal peptides. *Nucleic Acids Res.* 2015;43(W1):W401-W407.

59. Tusnady, G.E. and Simon, I. Principles governing amino acid composition of integral membrane proteins: application to topology prediction. *J. Mol. Biol.* 1998;283(2):489-506.

60. Tusnady, G.E. and Simon, I. The HMMTOP transmembrane topology prediction server. *Bioinformatics* 2001;17(9):849-850.

61. Wang, D., *et al.* Clinical characteristics of 138 hospitalized patients with 2019 novel coronavirus–infected pneumonia in Wuhan, China. *JAMA* 2020.

62. Wang, Z., *et al.* Household transmission of SARS-CoV-2. *J. Infect.* 2020;81(1):179-182.

63. Westerbeck, J.W. and Machamer, C.E. A coronavirus E protein is present in two distinct pools with different effects on assembly and the secretory pathway. *J. Virol.* 2015:JVI. 01237-01215.

64. Westerbeck, J.W. and Machamer, C.E. The Infectious Bronchitis Coronavirus Envelope Protein Alters Golgi pH To Protect the Spike Protein and Promote the Release of Infectious Virus. *J. Virol.* 2019;93(11).

65. White, C., Nixon, A. and Bradbury, N.A. Determining Membrane Protein Topology Using Fluorescence Protease Protection (FPP). *JoVE* 2015(98):e52509.

66. Wilson-Clark, S.D., *et al.* Household transmission of SARS, 2003. *CMAJ* 2006;175(10):1219-1223.

67. Wolff, D., *et al.* Evaluation of the Cost of Survivorship Care After Allogeneic Hematopoeitic Stem Cell Transplantation-An Analysis of 2 German Transplantation Centers. *Front Public Health* 2020;8:572470.

68. Wu, Q., *et al.* The E protein is a multifunctional membrane protein of SARS-CoV. *Genomics, proteomics & bioinformatics* 2003;1(2):131-144.

69. Xia, B., *et al.* SARS-CoV-2 envelope protein causes acute respiratory distress syndrome (ARDS)-like pathological damage and constitutes an antiviral target. *bioRxiv* 2020:2020.2006.2027.174953.

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.08.434384; this version posted March 8, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

572 70. Yang, Y., *et al.* Bcl-xL inhibits T-cell apoptosis induced by expression of SARS
573 coronavirus E protein in the absence of growth factors. *Biochem. J.* 2005;392(1):135-143.

574 71. Ye, Y. and Hogue, B.G. Role of the coronavirus E viroporin protein transmembrane
575 domain in virus assembly. *J. Virol.* 2007;81(7):3597-3607.

576 72. Yuan, Q., *et al.* Biochemical evidence for the presence of mixed membrane topologies of
577 the severe acute respiratory syndrome coronavirus envelope protein expressed in
578 mammalian cells. *FEBS Lett.* 2006;580(13):3192-3200.

579 73. Zeng, Z.-Q., *et al.* Epidemiology and clinical characteristics of human coronaviruses
580 OC43, 229E, NL63, and HKU1: a study of hospitalized children with acute respiratory
581 tract infection in Guangzhou, China. *European journal of clinical microbiology &*
582 *infectious diseases: official publication of the European Society of Clinical Microbiology*
583 2018;37(2):363-369.

584 74. Zhang, Y. I-TASSER server for protein 3D structure prediction. *BMC Bioinformatics*
585 2008;9:40.

586

587

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.08.434384; this version posted March 8, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Supplementary data

Table S2. Predicted topologies of severe acute respiratory coronavirus (SARS-CoV) envelope (E) protein (Accession number: P59637) using prediction programs ΔG Predictor, TMHMM, MEMSAT-SVM, TMpred, HMMTop, Phobius, TOPCONS. For each program, the predicted location of the amino (N) and carboxy (C) terminus of the E protein is indicated as being in the lumen of the endoplasmic reticulum-Golgi intermediate compartment (ERGIC) or in the cytoplasm of infected host cells. The number of transmembrane domains (TMDs) and the corresponding length of residues that they span is also indicated as predicted by each program.

Prediction Program	N-terminus	C-terminus	Number and position of transmembrane domains (TMDs)
ΔG Predictor	N.P.	N.P.	1 (17-39)
TMHMM	Cytoplasm	Lumen	1 (12-34)
MEMSAT-SVM	Lumen	Lumen	2 (9-38); (42-57)
TMpred	Lumen	Cytoplasm	1 (17-34)
HMMTop	Lumen	Cytoplasm	1 (11-35)
Phobius	Cytoplasm	Lumen	1 (12-37)
TOPCONS	Lumen	Cytoplasm	1 (16-36)

N.P. = not predicted

Table S3. Predicted topologies of severe acute respiratory coronavirus 2 (SARS-CoV-2) envelope (E) protein (Accession number: P0DTC4) using prediction programs ΔG Predictor, TMHMM, MEMSAT-SVM, TMpred, HMMTop, Phobius, TOPCONS. For each program, the predicted location of the amino (N) and carboxy (C) terminus of the E protein is indicated as being in the lumen of the endoplasmic reticulum-Golgi intermediate compartment (ERGIC) or in the cytoplasm of infected host cells. The number of transmembrane domains (TMDs) and the corresponding length of residues that they span is also indicated as predicted by each program.

Prediction Program	N-terminus	C-terminus	Number and position of transmembrane domains (TMDs)
ΔG Predictor	N.P.	N.P.	1 (17-39)
TMHMM	Cytoplasm	Lumen	1 (12-34)
MEMSAT-SVM	Lumen	Lumen	2 (9-38); (42-57)
TMpred	Lumen	Cytoplasm	1 (17-34)
HMMTop	Lumen	Cytoplasm	1 (11-35)
Phobius	Cytoplasm	Lumen	1 (12-37)
TOPCONS	Lumen	Cytoplasm	1 (16-36)

N.P. = not predicted

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.08.434384; this version posted March 8, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Table S4. Predicted topologies of Middle East respiratory coronavirus (MERS-CoV) envelope (E) protein (Accession number: K9N5R3) using prediction programs ΔG Predictor, TMHMM, MEMSAT-SVM, TMpred, HMMTop, Phobius, TOPCONS. For each program, the predicted location of the amino (N) and carboxy (C) terminus of the E protein is indicated as being in the lumen of the endoplasmic reticulum-Golgi intermediate compartment (ERGIC) or in the cytoplasm of infected host cells. The number of transmembrane domains (TMDs) and the corresponding length of residues that they span is also indicated as predicted by each program.

Prediction Program	N-terminus	C-terminus	Number and position of transmembrane domains (TMDs)
ΔG Predictor	N.P.	N.P.	2 (16-38); (39-59)
TMHMM	Cytoplasm	Lumen	1 (13-35)
MEMSAT-SVM	Lumen	Lumen	2 (10-37); (41-58)
TMpred	Lumen	Lumen	2 (16-34); (32-59)
HMMTop	Cytoplasm	Lumen	1 (9-33)
Phobius	Cytoplasm	Lumen	1 (12-34)
TOPCONS	Lumen	Lumen	2 (10-30); (32-52)

N.P. = not predicted

Table S5. Predicted topologies of human coronavirus 229E (hCoV-229E) envelope (E) protein (Accession number: P19741) using prediction programs ΔG Predictor, TMHMM, MEMSAT-SVM, TMpred, HMMTop, Phobius, TOPCONS. For each program, the predicted location of the amino (N) and carboxy (C) terminus of the E protein is indicated as being in the lumen of the endoplasmic reticulum-Golgi intermediate compartment (ERGIC) or in the cytoplasm of infected host cells. The number of transmembrane domains (TMDs) and the corresponding length of residues that they span is also indicated as predicted by each program.

Prediction Program	N-terminus	C-terminus	Number and position of transmembrane domains (TMDs)
ΔG Predictor	N.P.	N.P.	1 (15-36)
TMHMM	Lumen	Cytoplasm	1 (10-32)
MEMSAT-SVM	Cytoplasm	Lumen	1 (12-42)
TMpred	Lumen	Cytoplasm	1 (15-33)
HMMTop	Cytoplasm	Lumen	1 (11-33)
Phobius	Lumen	Cytoplasm	1 (15-36)
TOPCONS	Lumen	Cytoplasm	1 (13-33)

N.P. = not predicted

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.08.434384; this version posted March 8, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Table S6. Predicted topologies of human coronavirus NL63 (hCoV-NL63) envelope (E) protein (Accession number: Q6Q1S0) using prediction programs ΔG Predictor, TMHMM, MEMSAT-SVM, TMpred, HMMTop, Phobius, TOPCONS. For each program, the predicted location of the amino (N) and carboxy (C) terminus of the E protein is indicated as being in the lumen of the endoplasmic reticulum-Golgi intermediate compartment (ERGIC) or in the cytoplasm of infected host cells. The number of transmembrane domains (TMDs) and the corresponding length of residues that they span is also indicated as predicted by each program.

Prediction Program	N-terminus	C-terminus	Number and position of transmembrane domains (TMDs)
ΔG Predictor	N.P.	N.P.	1 (16-36)
TMHMM	Lumen	Cytoplasm	1 (10-32)
MEMSAT-SVM	Cytoplasm	Lumen	1 (12-42)
TMpred	Cytoplasm	Lumen	1 (15-33)
HMMTop	Cytoplasm	Lumen	1 (11-33)
Phobius	Cytoplasm	Lumen	1 (12-33)
TOPCONS	Lumen	Cytoplasm	1 (23-43)

N.P. = not predicted

Table S7. Predicted topologies of human coronavirus OC43 (hCoV-OC43) envelope (E) protein (Accession number: Q04854) using prediction programs ΔG Predictor, TMHMM, MEMSAT-SVM, TMpred, HMMTop, Phobius, TOPCONS. For each program, the predicted location of the amino (N) and carboxy (C) terminus of the E protein is indicated as being in the lumen of the endoplasmic reticulum-Golgi intermediate compartment (ERGIC) or in the cytoplasm of infected host cells. The number of transmembrane domains (TMDs) and the corresponding length of residues that they span is also indicated as predicted by each program.

Prediction Program	N-terminus	C-terminus	Number and position of transmembrane domains (TMDs)
ΔG Predictor	N.P.	N.P.	2 (18-39); (41-61)
TMHMM	Lumen	Cytoplasm	1 (15-37)
MEMSAT-SVM	Lumen	Lumen	2 (12-39); (43-58)
TMpred	Cytoplasm	Cytoplasm	2 (15-33); (41-61)
HMMTop	Lumen	Lumen	2 (18-36); (43-61)
Phobius	Lumen	Cytoplasm	1 (12-39)
TOPCONS	Lumen	Lumen	2 (16-36); (38-58)

N.P. = not predicted

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.08.434384; this version posted March 8, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Table S8. Predicted topologies of human coronavirus HKU1 (hCoV-HKU1) envelope (E) protein (Accession number: Q5MQC8) using prediction programs ΔG Predictor, TMHMM, MEMSAT-SVM, TMpred, HMMTop, Phobius, TOPCONS. For each program, the predicted location of the amino (N) and carboxy (C) terminus of the E protein is indicated as being in the lumen of the endoplasmic reticulum-Golgi intermediate compartment (ERGIC) or in the cytoplasm of infected host cells. The number of transmembrane domains (TMDs) and the corresponding length of residues that they span is also indicated as predicted by each program.

Prediction Program	N-terminus	C-terminus	Number and position of transmembrane domains (TMDs)
ΔG Predictor	N.P.	N.P.	2 (17-39); (40-61)
TMHMM	Lumen	Lumen	2 (10-32); (39-58)
MEMSAT-SVM	Lumen	Lumen	2 (10-36); (40-55)
TMpred	Lumen	Lumen	2 (16-34); (39-59)
HMMTop	Lumen	Lumen	2 (17-36); (43-59)
Phobius	Lumen	Lumen	2 (12-34); (39-59)
TOPCONS	Lumen	Lumen	2 (10-30): (32-52)

N.P. = not predicted

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.08.434384; this version posted March 8, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

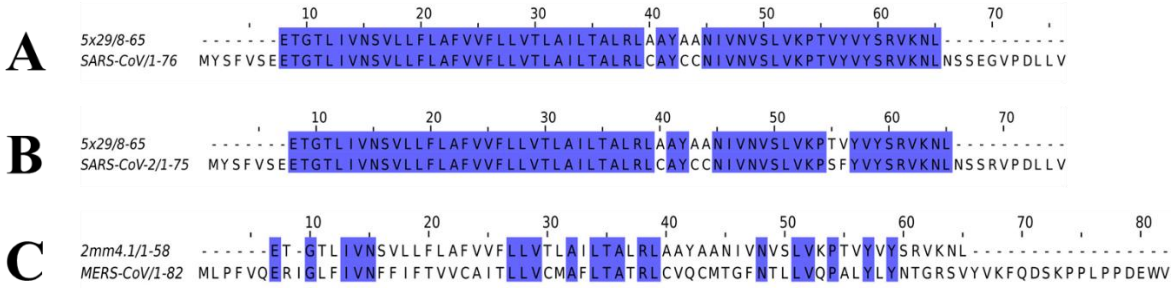

Figure S3. Pairwise sequence alignments between the more virulent human (h) coronavirus (CoV) envelope (E) proteins and the respective template used to generate the three-dimensional (3D) model. Sequence alignments were generated using Jalview (v2.11.1.3) and coloured by sequence identity (blue). **A.** Pairwise sequence alignment between the SARS-CoV E protein (P59637) and template 5x29. Sequences shared 91% identity with no conserved residues in the PDZ-binding motif (DLLV). **B.** Pairwise sequence alignment between the SARS-CoV-2 E protein (P0DTC4) and template 5x29. Sequences shared 91% identity with no conserved residues in the PDZ-binding motif (DLLV). **C.** Pairwise sequence alignment between the MERS-CoV E protein (K9N5R3) and template 2mm4. Sequences shared 35% identity with no conserved residues in the PDZ-binding motif (DEWV).

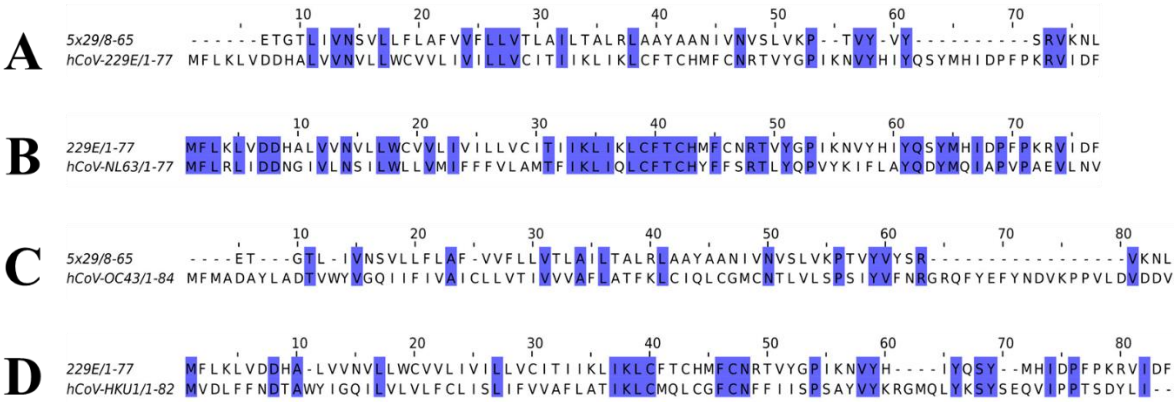

Figure S4. Pairwise sequence alignments between the less virulent human (h) coronaviruses (CoVs) and the respective template used to generate the three-dimensional (3D) model. Sequence alignments were generated using Jalview (v2.11.1.3) and coloured by sequence identity (blue). **A.** Pairwise sequence alignment between the hCoV-229E E protein (P19741) and template 5x29. Sequences shared 29% and PDZ-binding motif residues **V**IDF were conserved. **B.** Pairwise sequence alignment between the hCoV-NL63 E protein (Q6Q1S0) and the hCoV-229E E protein homologous structure. Sequences shared 47% identity and PDZ-binding motif residues **V**LNV were conserved. **C.** Pairwise sequence alignment between the hCoV-OC43 E protein (Q04854) and template 5x29. Sequences shared 22% identity and PDZ-binding motif residues **V**DDV were conserved. **D.** Pairwise sequence alignment between the hCoV-HKU1 E protein (Q5MQC8) and the hCoV-229E E protein homologous structure. Sequences shared 27% identity and PDZ-binding motif residues DYL**I** were conserved.

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.08.434384; this version posted March 8, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

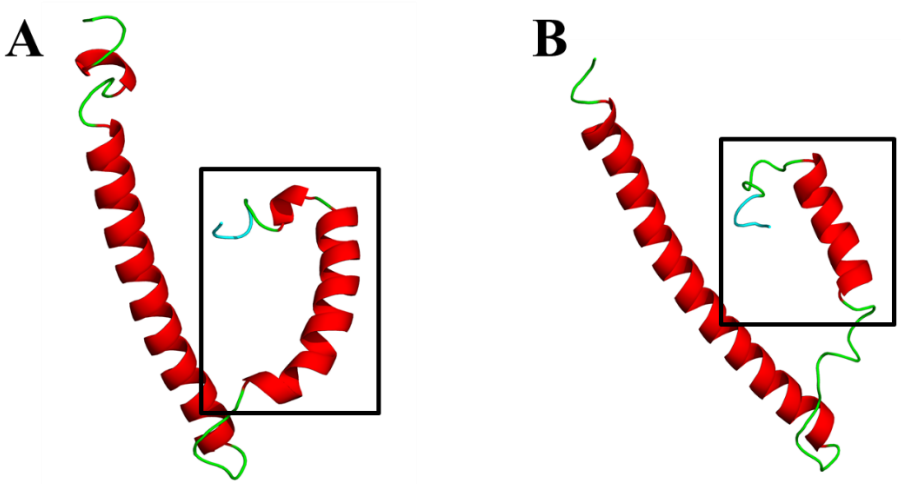

Figure S5. Cartoon representation of the three-dimensional (3D) models of the envelope (E) protein of hCoV-OC43 and hCoV-HKU1. **A.** hCoV-OC43 E. **B.** hCoV-HKU1 E. Models were generated using ITASSER Webserver and based on the nuclear magnetic resonance (NMR)-resolved structure for SARS-CoV E (PDBID: 2mm4) obtained from the protein data bank (PDB) (Li et al., 2014). The PDZ-binding motif (PBM) is coloured in cyan and the squared region shows the variable C-terminal domain.