



Sequence and structural analysis of COVID-19 E and M proteins with MERS virus E and M proteins—A comparative study

Ebtisam A. Aldaais^{a,1,*}, Subha Yegnaswamy^{b,1}, Fatimah Albahrani^{c,1}, Fatima Alsowaiket^{c,1}, Sarah Alramadan^{c,1}

^a Department of Radiological Sciences, Imam Abdulrahman Bin Faisal University, Dammam, P.O. Box 2435, 31441, Saudi Arabia

^b Aldaais Research Group, Imam Abdulrahman bin Faisal University, Dammam, P.O. Box 2435, 31451, Saudi Arabia

^c Department of Biomedical Engineering, Imam Abdulrahman Bin Faisal University, Dammam, P.O. Box 2435, 31451, Saudi Arabia

ARTICLE INFO

Dataset link: <https://www.ncbi.nlm.nih.gov/protein>, <https://zhanglab.ccmb.med.umich.edu/TASSER/>

Keywords:

COVID-19
Beta coronaviruses
E protein
M protein
sequence and structural analysis

ABSTRACT

The outbreak of SARS in 2003, MERS in 2012, and now COVID-19 in 2019 has demonstrated that Coronaviruses are capable of causing primary lethal infections in humans, and the pandemic is now a global concern. The COVID-19 belongs to the beta coronavirus family encoding 29 proteins, of which four are structural, the Spike, Membrane, Envelope, and Nucleocapsid proteins. Here we have analyzed and compared the Membrane (M) and Envelope (E) proteins of COVID-19 and MERS with SARS and Bat viruses. The sequence analysis of conserved regions of both E and M proteins revealed that many regions of COVID-19 are similar to Bat and SARS viruses while the MERS virus showed variations. The essential binding motifs found in SARS appeared in COVID-19. Besides, the M protein of COVID-19 showed a distinct serine phosphorylation site in the C-terminal domain, which looked like a catalytic triad seen in serine proteases. A Dileucine motif occurred many times in the sequence of the M protein of all the four viruses compared. Concerning the structural part, the COVID-19 E protein showed more similarity to Bat while MERS shared similarity with the SARS virus. The M protein of both COVID-19 and MERS displayed variations in the structure. The interaction between M and E proteins was also studied to know the additional binding regions. Our study highlights the critical motifs and structural regions to be considered for further research to design better inhibitors for the infection caused by these viruses.

1. Introduction

On December 31, 2019, viral pneumonia of undiscovered source in Wuhan city, the capital city of Hubei Province in the People's Republic of China, was reported to the World Health Organization (WHO) Country Office in China [1,2]. The declaration of a Public Health Emergency of International Concern was on January 30, 2020. While on February 11, 2020, the International Committee on Taxonomy of Viruses (ICTV) suggested the name of the novel virus to be "severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2)" since it has genetic similarity to Severe Acute Respiratory Syndrome (SARS), whereas WHO referred to the novel coronavirus disease as "COVID-19" [3,4]. Currently, COVID-19 has spread globally in more than 188 countries, with more than 14.60 million cases confirmed and more than 607,800 deaths, while the number of recovered people is more than 8.71 million as of July 20, 2020 [2]. The outbreak of COVID-19 covered several middle eastern countries, including Saudi Arabia. The Ministry of Health announced the first case in Saudi Arabia on March 2, 2020, with exponential

growth in the number of patients that reached 248,416 confirmed cases and a total of 2447 deaths as of July 19, 2020 [5,6].

Coronaviruses (CoVs) are a large family of enveloped viruses that cause respiratory tract infectious diseases with symptoms similar to the typical common cold [7]. CoVs are categorized into three genus groups, α -CoVs, β -CoVs, and γ -CoVs. COVID-19 belongs to β -CoVs similar to SARS and Middle East Respiratory Syndrome (MERS), which are identified as bat-origin that infect people through an intermediate host [8].

Betacoronavirus family is considered to be an enveloped, single, and positive-stranded RNA virus of zoonotic origin, which contains four lineages: A, B, C, and D. Beta coronaviruses include SARS-CoV, MERS-CoV, and SARS-CoV-2 (see Table 1) [9]. SARS-CoV is a virus that causes Severe Acute Respiratory Syndrome, which caused four outbreaks between 2002 and 2003. According to WHO, SARS started from Guangdong province in southern China [10]. Researches proved Chinese horseshoe bats to be the natural reservoir of SARS [11]. This

* Corresponding author.

E-mail address: ealdaais@iau.edu.sa (E.A. Aldaais).

¹ All authors have contributed equally to this research.

<https://doi.org/10.1016/j.bbrep.2021.101023>

Received 16 January 2021; Received in revised form 5 May 2021; Accepted 11 May 2021

Available online 14 May 2021

2405-5808/© 2021 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/>).

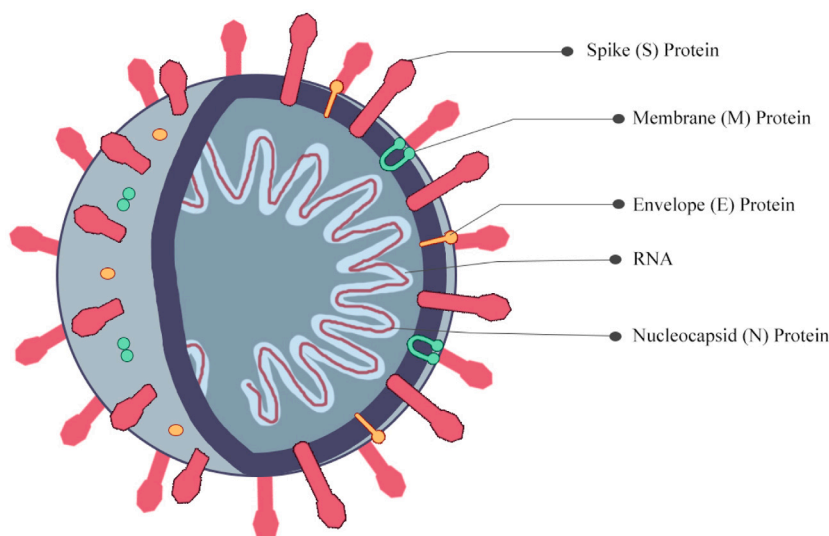


Fig. 1. 3D model visualizing coronavirus and its structural proteins.

Table 1

Statistics of SARS-CoV, MERS-CoV, and SARS-CoV2.

Beta coronaviruses	Region (Year)	Cases	Fatality	Origin	host	lineage	Incubation period
SARS-CoV	Guangdong, China (2002–2003)	8422	916 (9.6%)	Horseshoe bats	Palm civets	B	5 days
MERS-CoV	Saudi Arabia (2012)	2519	866 (34.3%)	bat	Dromedary camel	C	5 days
SARS-CoV-2	Wuhan, China (2019-)	Over 79,331	~2%	bat	Malayan pangolins	B	5-7 days

epidemic affected about 26 countries, including China, Hong Kong, Singapore, and Canada. This global outbreak was controlled under WHO regulations of quarantine and travel restrictions. The virus was finally brought to an end in early 2004, with a total number of 8422 cases and 916 confirmed deaths [10].

In 2012, a novel coronavirus called MERS, or Middle East Respiratory Syndrome, was declared as a new disease. The virus first identified in Saudi Arabia, considering the dromedary camels as the primary reservoir [12]. Since 2012, 27 countries have reported MERS cases, including Austria, Egypt, China, Malaysia, and Bahrain. The total number of cases was about 2519, where 80% of them were located in Saudi Arabia, and the total deaths were 866. During this epidemic, the WHO did not restrict travel and Trade. Instead, the Food and Agriculture Organization of the United Nations (FAO), the World Organization for Animal Health (OIE), and national governments worked with the WHO to develop public health prevention strategies to combat the virus [13].

SARS-CoV-2 is the virus behind the COVID-19 epidemic, which originated in 2019 at Wuhan city in China, as mentioned earlier. On the 27th of May, statistics confirmed more than 5.5 million cases, with about 350 thousand deaths (Table 1) [14].

Similar to other viruses, the structure of SARS-CoV-2 is composed of four structural proteins. These proteins are the spike (S) glycoprotein, small envelope (E) glycoprotein, membrane (M) protein, and nucleocapsid (N) protein that is bound to a single-stranded, positive-sense viral Ribonucleic acid (RNA) genome (see Fig. 1) [1,15,16]. In this study, we will focus on the E and M proteins.

The envelope (E) protein is considered the smallest of the main structural proteins of CoVs. As on date, many questions about it are not answered yet. It is a short protein that consists of 76–109 amino acids that occur with a short hydrophilic N-terminal (7–12 amino acids), and a bigger hydrophobic transmembrane domain (25 amino acids), ending in a long hydrophilic carboxyl C-terminal that cover most of the protein [7,17].

The importance of this protein comes from its vital role in SARS. It assists the release and insertion of the virus to host cell. Also, the protein alters some cellular processes, indicating its role in controlling the virus's pathogenicity. Thus, it should be considered as a significant virulence element in the novel SARS-CoV-2 [18].

With SARS, the envelope protein reveal its virulence through several mechanisms. Through a sequence motif at the last four amino acids of the C-terminus, where it binds to the postsynaptic density protein95 (PDS95), Drosophila disk large tumor suppressor protein (Glg1) and zonula occludens-1 protein (zo-1), all of these proteins ranges can be referred to as the PDZ domain [17,19].

On the other hand, the membrane (M) protein has a critical role in the virus, and especially during the budding and assemble processes. It mainly consists of a big carboxyl-terminal region at the interior of the virion, three hydrophobic transmembrane domains, and a short amino acid terminal on the virion exterior. The M protein is a dominant structural protein that can combine with other structural proteins such as the spike (S) and Envelope (E) proteins, making it crucial to understand the system of the new CoV [20]. As the M protein's assists to the S protein, which is involved in the cell attachment and entry to the host cell, any mutation occurs on the M protein is expected to have a remarkable impact on the interactions with the infected cell [21].

2. Methodology

For this study, we took the Envelope protein sequence from NCBI with the accession nos.: YP_009724392.1 envelope protein Wuhan-Hu-1 strain [Severe acute respiratory syndrome coronavirus 2], AVP78033.1 envelope protein Bat-SL-CoVZC45 strain [Bat SARS-like coronavirus], AAP51230.1 envelope protein E GDO1 strain [SARS coronavirus GD01], AVV62542.1 E protein NL140422 strain [Middle East respiratory syndrome-related coronavirus] [22]. Also, we took the membrane protein sequence from NCBI with the accession nos.: YP_009724393.1 membrane glycoprotein Wuhan-Hu-1 strain [Severe acute respiratory syndrome coronavirus 2], AVP78034.1 membrane protein Bat-SL-CoVZC45

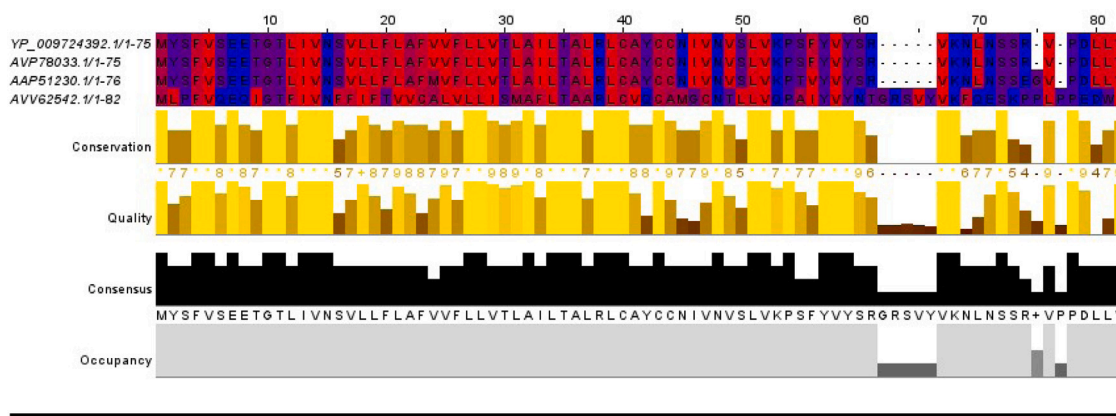


Fig. 2. A representation of a Multiple Sequence Alignment (MSA) of COVID-19 E protein. The MSA is showing the alignment of COVID-19 E protein with Bat, SARS, MERS Viruses. The alignment is colored based on the hydrophobicity. The blue-colored are hydrophilic, and the red ones are hydrophobic residues. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

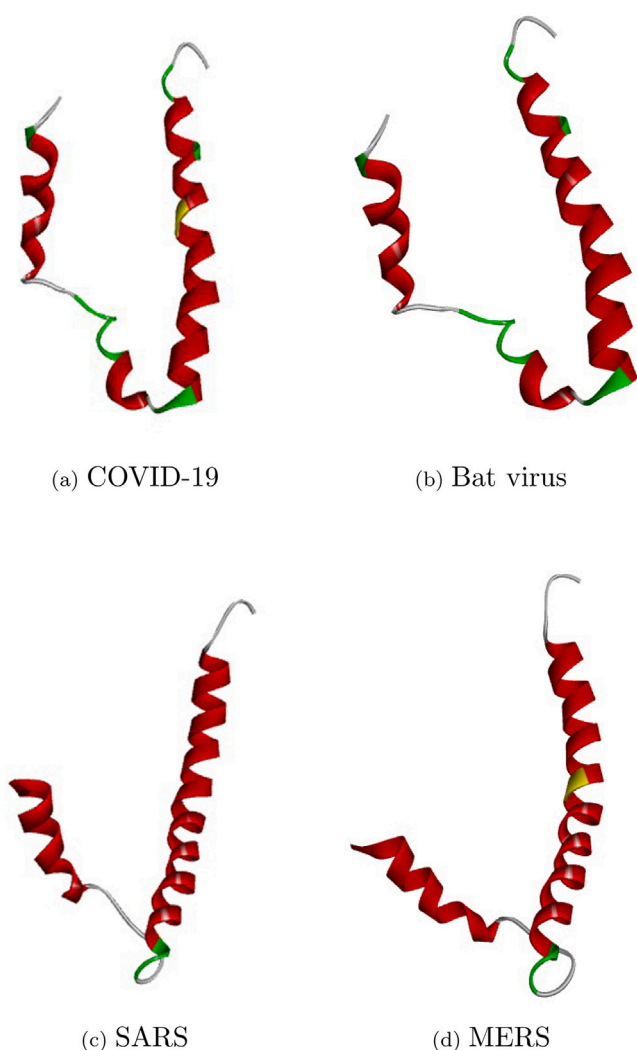


Fig. 3. Modeled structure of the E protein for COVID-19 (a), Bat (b), SARS (c), and MERS (d) viruses using Swiss-Model Server. The models were visualized using Discovery studio visualizer.

strain [Bat SARS-like coronavirus], AAP51231.1 membrane protein M GD01 strain [SARS coronavirus GD01], AVV62543.1 M protein

NL140422 strain [Middle East respiratory syndrome-related coronavirus] [22]. Protein Blast P search was done to determine the closely related hits of COVID-19 E and M proteins from Bat, SARS, and MERS virus [23]. The Multiple Sequence Alignment (MSA) was done using Multiple Alignment applying Fast Fourier Transform (MAFFT) server to find the conserved and consensus regions among the selected viruses [24,25]. The MSA results were viewed through Jalview viewer [26]. The probable motifs present in the virus were obtained from Motif-Scan [27]. The E proteins' structures were modeled using the Swiss-Model Homology Modeling server since the template for SARS virus E protein monomer 2mm4.1 A was available. This monomer was taken as a template to model all other viruses, COVID-19, BAT, SARS and MERS [28–33]. Since no template structure was available for the M protein, the I-Tasser server was used to model the M protein for all viruses except COVID-19 [34–36]. The model of COVID-19 (QHD43419) was already available in the I-Tasser server. The E protein and M protein models were evaluated using the procheck option of ProTSAV (Protein Structure Analysis and Validation server) [37]. The domains in the structures were analyzed using Discovery Studio Visualizer [38]. The docking studies between the M and E proteins were done using Cluspro [39–41] – an online server for protein–protein docking – and the interactions were analyzed using Discovery Studio Visualizer. The Dimplot [42] was further used to evaluate the interactions between M and E protein.

3. Results and discussion

Our studies were limited to the comparison of COVID-19 with Bat, SARS, and MERS because the Bat and SARS showed very close homology wherein MERS was as virulent as COVID-19 in the infection spread as well as morbidity.

3.1. Comparison of the COVID-19 E protein with SARS, Bat and MERS viruses

The E protein of COVID-19 was subjected to a blast search, and the results with the nearest hits of Bat, SARS, and MERS viruses were selected. The blast search was limited to exclude the hits from different strains of COVID-19. So the results with the top hits from the viruses mentioned above were taken for analysis. The Bat virus had almost 100% identity with COVID-19 E protein followed by SARS 93.42%, and the MERS virus had a decidedly less identity of 37.33% compared to the other two mentioned viruses. The COVID-19 and MERS E proteins are small hydrophobic transmembrane proteins with a short hydrophilic amino terminus followed by a transmembrane domain with hydrophobic residues and a long hydrophilic carboxy-terminal.

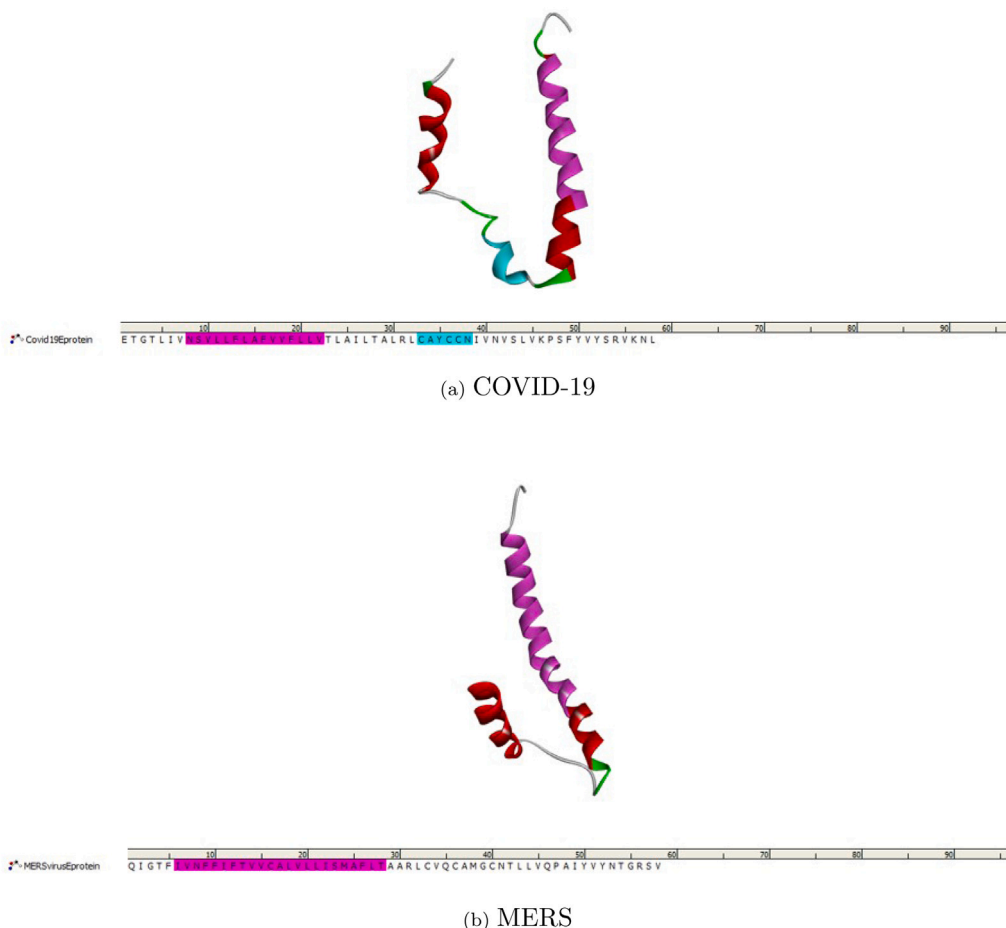


Fig. 4. A comparison between E protein regions for COVID-19 and MERS viruses. (a) A structure of COVID-19 E protein showing interaction regions. The region that could be involved in spike protein interaction is shown in cyan, while the one involved in forming a homopentameric channel is shown in magenta. The sequence also highlighted in the same manner. (b) A structure of the MERS virus showing interaction regions. The region which is involved in forming a homopentameric channel is shown in magenta. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

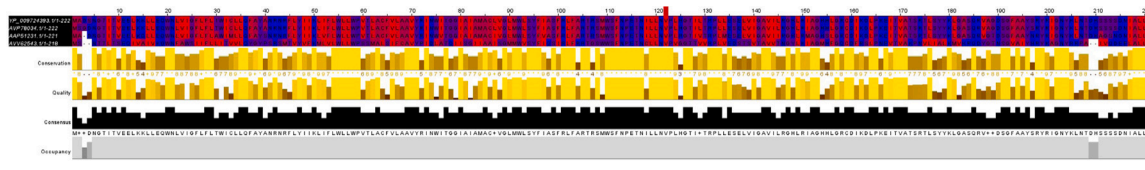


Fig. 5. A Multiple Sequence Alignment of COVID-19 M protein. The MSA is showing the alignment of COVID-19 M protein with Bat, SARS, MERS Virus. The alignment is colored based on the hydrophobicity. The blue-colored are hydrophilic, and the red ones are hydrophobic residues. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

The pattern is similar to the one expected for any transmembrane protein. Further, the E proteins of the four viruses were analyzed using MAFFT and compared for conserved regions (see Fig. 2). The IVNSVLLFLAFVVFLV hydrophobic region between 13 and 29 in the transmembrane domain is highly conserved among all the CoVs, which was found to be involved in forming a homopentameric ion channel called viroporins [43]. The cysteine motif CAYCCN was observed in the C-terminal domain between the 40th and 45th regions for all viruses, including the MERS virus, with a little diverse pattern. This motif could be involved in the interaction with the Spike protein [44]. These three cysteine residues undergo palmitoylation. They are also involved in oligomerization and alters membrane permeability of E-protein [45].

Another conserved region SRVKNLNSSR between positions 60 and 70 is highly conserved among COVID-19, Bat, and SARS, whereas in the MERS virus, a five amino acid insertion was observed between positions 61 and 67. This site is probably a potential phosphorylation site due to

the presence of SSR residues, which was also showed in motif scan as PKC-phospho site. The motif scan in COVID-19 revealed the presence of the ASN-Glycosylation site between residues 48–51, and 66–69. Previous Studies on SARS-CoV explained that N48 is not glycosylated, due to its proximity to the membrane, and only N66 undergoes glycosylation. The glycosylation of N66 might involve manipulating the membrane topology and thereby pathogenesis of the virus [46]. Alternatively, it was found to prevent oligomerization of E protein probably to promote or inhibit a specific function [7]. The C-terminal PDZ binding motif (PDM) DLLV motif, between 72 and 75, was also observed in line with SARS-CoV and Bat Virus. However, a hydrophilic substitution between D and V was observed in MERS Virus, contrary to the other virus's hydrophobic terminal. This DLLV motif is involved in the interaction with many host cell proteins. The C-terminal region is also involved in the interaction with the M protein, as demonstrated with the other beta coronaviruses [47]. The motif scan in the MERS virus showed

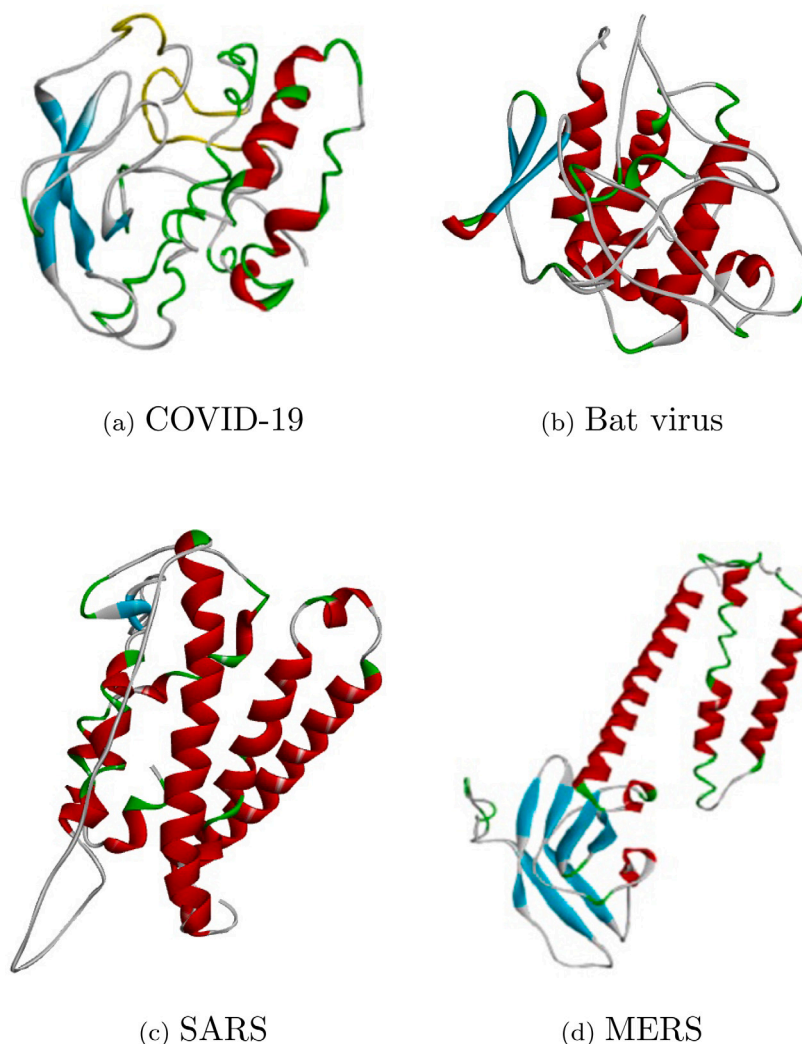


Fig. 6. Modeled structure of the M protein for COVID-19 (a), Bat (b), SARS (c), and MERS (d) viruses using I-Tasser Server. The models were visualized using Discovery studio visualizer.

different interesting regions. 1–55 regions were similar to the TonB dependent receptor seen in bacterial outer-membrane proteins. 41–66 regions were similar to calponin like repeat, 61–63 PKC-Phospho site, a phosphorylation site.

3.2. E protein structure

Only the structure of the Spike protein was available for all viruses in the protein data bank. Thus, the E proteins were modeled using the Swiss-Model Homology Modeling server. Furthermore, the structure of the SARS virus E protein monomer 2mm4.1. A was taken as a template for modeling COVID-19 (Fig. 3(a)), Bat (Fig. 3(b)), and MERS (Fig. 3(d)) E proteins, including the SARS E protein (Fig. 3(c)) since we had taken a different strain of SARS. The template showed an identity of 91.38% for COVID-19 and Bat, while 93.10% for SARS, with sequence coverage of 8 to 65 involving the transmembrane and part of the C-terminal domain. The MERS virus E protein with 36.21% identity and a sequence coverage of 8–65 was also constructed with the same approach. The validation of the E-protein models showed acceptable deviations for all viruses except MERS because of low template identity. The E-protein models were compared for conserved regions and domains for all viruses. The structure revealed that COVID-19 E protein shared a similar feature with the Bat virus protein, whereas MERS virus E protein shared similarity with SARS virus E protein.

The modeled structure of COVID-19 comprising mainly the transmembrane domain and part of the C-terminal portion had only helix and turns. A few portions of N- and C-terminal residues could not be modeled due to the lack of the template structure. The two critical regions in the structure are: (1) the triple cysteine motif found in SARS-CoV in the transmembrane domain, which is involved in spike protein interaction shown in cyan color. (2) The transmembrane leucine and valine rich region are involved in the formation of the homopentameric ion channel shown in magenta (Fig. 4(a)). The C-terminal PDZ binding motif (PDM) could not be shown as it was not modeled. Fig. 4(a) shows a structure of a double spanning membrane protein with the N-terminal and part of the C-terminal projecting on the same side of the cell.

Some early studies on SARS revealed that both the N- and C-termini of the SARS E protein might be located on the cytoplasmic side of the cell [48]. Another study provided a contradictory report suggesting the conformation spanning the intracellular membranes only once. The carboxy-terminal domain is towards the cell cytoplasm and the amino-terminal domain towards the lumen of intracellular membranes [49]. Earlier studies on MERS virus indicate that it has a single helical transmembrane domain and forms pentamers similar to SARS [50]. The motif 13–35 IVNFFIFTVVICALVLLISMAFLT in the transmembrane domain is highlighted in magenta (Fig. 4(b)).

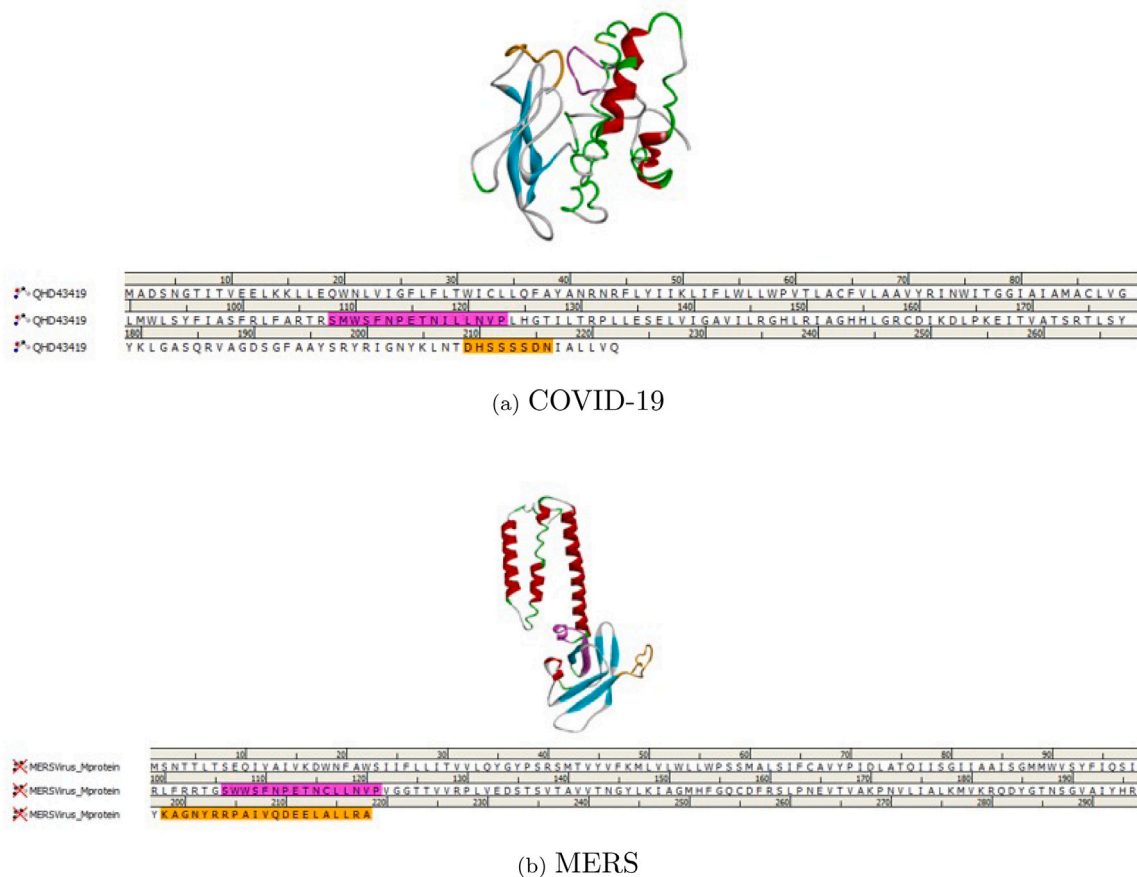


Fig. 7. A comparison between M protein regions for COVID-19 and MERS viruses. (a) A structure of COVID-19 M protein showing interaction regions. The region involved in M–M interaction is shown in magenta, and the one which is a phosphorylation site is shown in orange. (b) A structure of MERS M protein showing interaction regions. The region involved in M–M interaction is shown in magenta, and the carboxy-terminal motif involved in many interactions is shown in orange. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

3.3. Comparison of the COVID-19 M protein with Bat, SARS, and MERS viruses

The same procedure was followed for Membrane protein wherein the nearest hits of Bat, SARS, and MERS viruses were taken from Blast search considering the virus strain used for the analysis of Envelope proteins. All the virus sequences displayed a hydrophilic amino-terminal followed by intermittent hydrophilic and hydrophobic residues and a hydrophobic carboxy-terminal tail lined by I/LALLV motif. The MSA analysis revealed 98.6% identity to Bat Virus, 90% identity to SARS, and 42.3% identity to the MERS virus. The Dileucine motif is observed to occur at least five times in the sequence of all the viruses. A similar dileucine motif observed in the carboxy-terminal of SARS is required for efficient packaging of N protein in Virus-like particles (VLP) [51]. Dileucine motifs also observed in HIV-1 envelope protein play an important role in endocytosis [52].

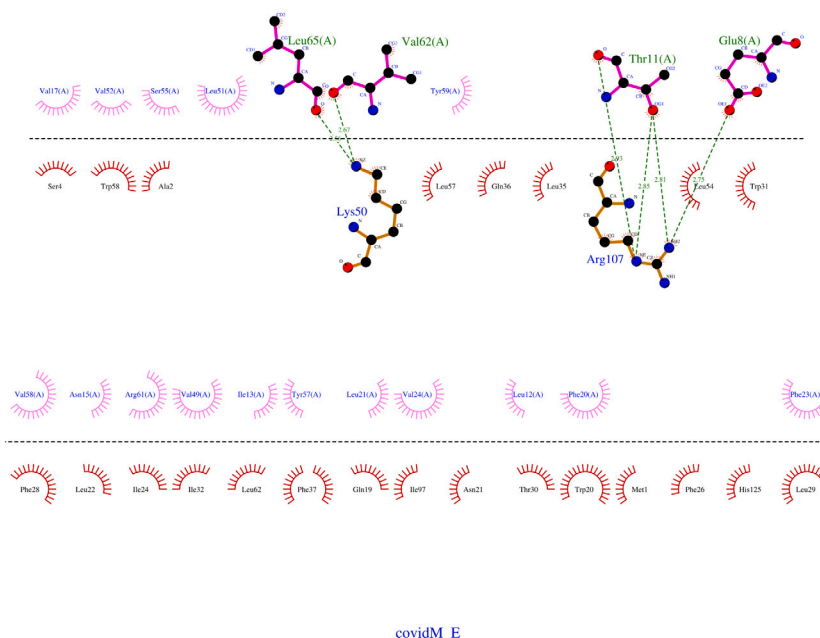
The conserved regions among the four viruses observed using MAFFT (Fig. 5) were F/ILWLLWP between 54 and 59, MWLSYF between 91 and 96, SM/ WWSFNPETNILLNVP between 108 and 123, RPLLE between 131 and 135, and GHRLRIAG 147 and 153. Although that the M protein of COVID-19, Bat, SARS seems to be similar in several sites (conserved and non-conserved), MERS virus showed similarity on conserved sites only. The importance of the SWWSFNPETNLL motif has already been studied in coronavirus, where it functions to facilitate M–M interactions, thereby helping in the formation of the viral envelope [53]. The region between 60 and 70 TLACFVLAAY and GLMWLSYFV transmembrane domain of M protein contains a T-cell epitope cluster that contributes significantly to the M protein-specific

cellular immunity in SARS [54]. Previous work on SARS revealed the N-terminal part of M protein containing the transmembrane domain induces the retrieval of S protein to the ERGIC, suggesting a similar function in COVID-19 [55].

The carboxy-terminal region is lined by I/LALLV motif except for the MERS virus, which differed in one position where “R” was present instead of “V” represented by the LALLR motif. Y195 in the cytoplasmic tail is necessary for the M–S interaction [56]. The other important motif between regions 209 and 216 DHSSSDN observed in COVID-19 and Bat Virus but not in SARS and MERS virus. The site is a serine phosphorylation site, and a catalytic triad “DHS” pattern seen in serine protease enzymes active site, which is essential for their activity, was also noticed. Moreover, the catalytic triad regions were noticed first in *Vibrio* bacteria and some viruses [57,58]. The diacidic motifs D211XE213 in the carboxy-terminal of the MERS virus were found to be functional ER export signals involved in the trafficking of the protein. Another C-terminal motif K199xG201xY203R204, is involved in specific localization in the TGN [59]. The motif scan in M protein of COVID-19 displayed ASN-Glycosylation site between 5 and 8, CK2-phospho site between 9–12 and 212–215, Myristyl site between 79 and 84, 126 and 131, PKC-phospho site between 99 and 101, 172 and 174, 184 and 186 suggesting the possible interactions it could encompass with the host proteins. There are many sites for phosphorylation which supports further the possible serine protease function. MERS virus protein also displayed a similar ASN-Glycosylation site between 3 and 6, Myristyl site between 105 and 110, 186 and 191, and PKC-phospho site between 97 and 99. The only glycosylation site seen in the M protein of SARS at N4 suggests a similar function in both COVID-19 and MERS virus [60].



(a)



(b)

Fig. 8. The interaction between M and E proteins of COVID-19. (a) The interacting regions are represented in the ball and stick model with the M protein highlighted in yellow and the E protein shown in purple. (b) A 2D-plot showing the interaction between M and E chains of COVID-19 and the interacting residues. The hydrogen bonds are shown in dotted lines. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

3.4. M protein structure

The M protein structure of COVID-19 (QHD43419) (Fig. 6(a)) was downloaded from the I-Tasser server while the SARS (Fig. 6(b)), BAT (Fig. 6(c)), and MERS (Fig. 6(d)) proteins were modeled using the I-Tasser server due to the lack of template structure. The COVID-19 model showed two helical regions, two beta-sheets, turns, and coils. Besides, there were many coils in the model; the ambiguity in the template for these regions might cause it. The SWWSFNPETNNL motif shown as coils in the structure is highlighted in magenta, and the carboxy-terminal also represented by coils is highlighted in orange (Fig. 7(a)). The carboxy-terminal is involved in M–M, M–N, M–S, and M–E interactions [47,61]. The studies on M protein in SARS revealed a triple-spanning transmembrane protein exhibits a Nexo–Cendo topology with a small ectodomain towards luminal N-terminus,

three transmembrane segments and a long cytosolic C-terminus also called endodomain [55]. MERS virus model shows helices and beta sheets with the SWWSFNPETNNL motif represented as beta sheets shown in magenta while the C-terminal I/LALLV motif (also a beta-sheet) is shown in orange color (Fig. 7(b)). The motifs are seen in the transmembrane domain and C-terminal end. The first part of the structure with amino-terminal and part of the transmembrane domain took up helix while the remaining took up beta-sheets. The structure of both COVID-19 and MERS displayed both helices and beta-sheet, while SARS had a major portion of helices in the model suggesting a difference in the structure could be possible between SARS, COVID-19, and MERS viruses.

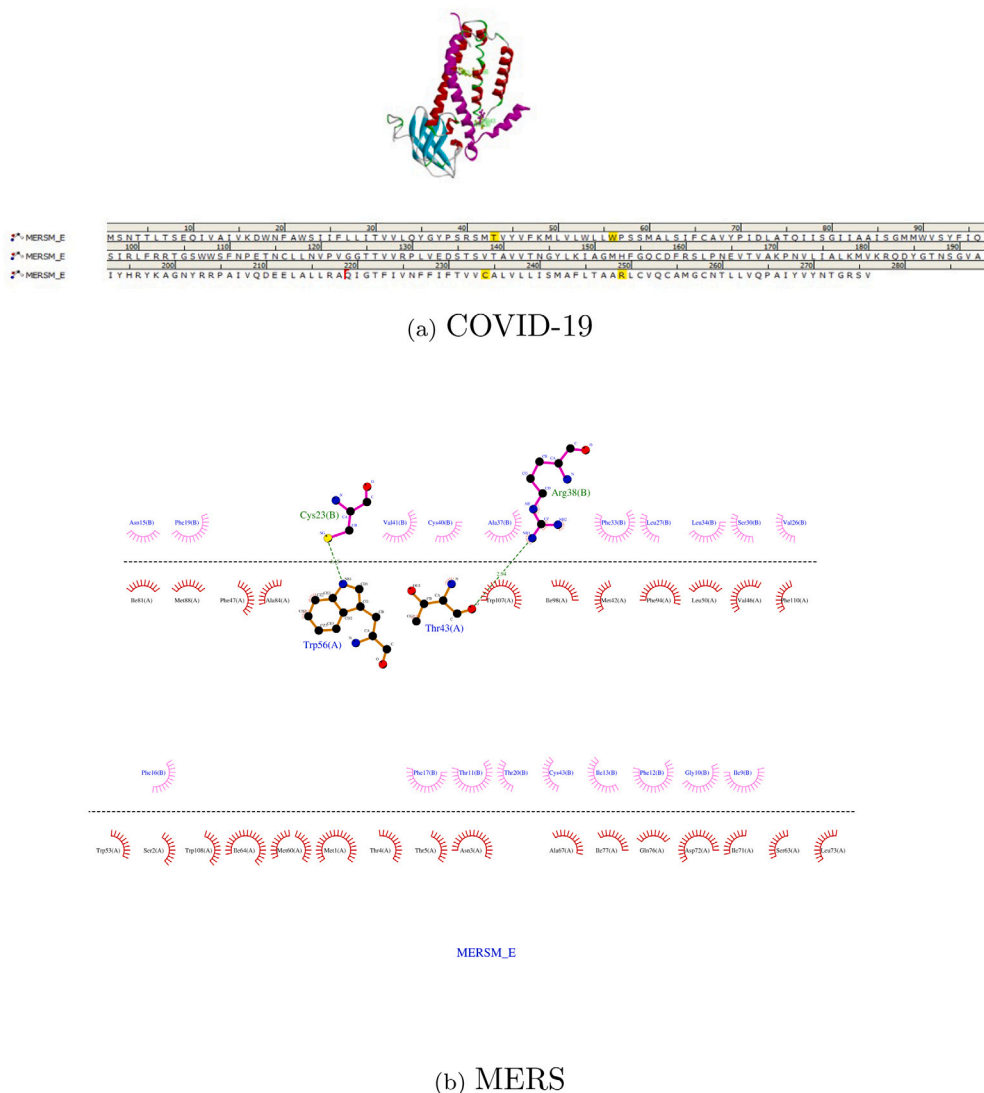


Fig. 9. The interaction between M and E proteins of MERS virus. (a) The interacting regions are represented in the ball and stick model. The E protein is shown in purple. The interacting residues are highlighted in yellow. (b) A 2D-plot of the interaction between M and E chains of MERS virus. The interacting residues are shown. The hydrogen bonds are shown in dotted lines. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

3.5. Interaction between M and E proteins of COVID-19 and MERS

We studied the interaction between E and M proteins for COVID-19 and MERS. Although only part of the model was available for E protein, the possible interaction with M protein was studied using the Cluspro web server, which does protein–protein docking. Previous studies on the interaction between M and E proteins revealed the carboxy-terminal of both M and E proteins are involved in the interaction [47,61]. Further, the interactions were visualized in Discovery Studio Visualizer, and the 2-D interactions were plotted using the Dimplot program of Ligplot+. The COVID-19 residues involved in the interaction were K50 from the transmembrane region of the M protein forms hydrogen (H_2) bond with the carboxy-terminal V62 and L65 of E protein. Also, R107 of the transmembrane region of M protein interacts with amino-terminal E8, T11 of E protein. The interactions are shown with the interacting residues represented by the ball and stick model (Fig. 8(a)). The E protein is shown in purple color. The 2-D plot of interactions was drawn using Dimplot (Fig. 8(b)), where the residues involved in hydrogen bonds between M and E proteins are shown.

Similarly, MERS virus's M and E proteins' interaction was also done, and the interacting residues are shown in (Fig. 9(a)). The M protein transmembrane residue W56 interacts with transmembrane region C23

of the E protein. The T43 transmembrane region of M protein interacts with the end of the transmembrane region R38 the 2D plot interaction is shown in (Fig. 9(b)). The docking suggests the possible interaction between M and E protein. Most of the interacting regions of the M-protein were seen in the C-terminal domain in SARS as per the earlier studies, whereas in the current docking studies, the interactions were seen in the transmembrane region. Due to the arbitrariness of the Insilco model, the regions involved could be considered for further studies.

4. Conclusion

Much research has focused on the Spike protein of CoVs, which is vital for viral attachment and transmission of the disease. The other critical structural proteins which deserve further attention are the Membrane and Envelope proteins, but not much information is available concerning the sequence and structural regions of these proteins. In this study, we analyzed the sequence and structural domains of M and E proteins in COVID-19 and MERS Viruses by comparing them with SARS and Bat viruses since they share a better homology with the viruses mentioned above. The sequence analysis of M and E proteins revealed that COVID-19 shared many similar regions with SARS and

Bat viruses, whereas the MERS virus showed variations in many regions except for few consensus regions. The structural analysis of M protein in COVID-19 and MERS showed an orientation towards helices and beta sheets, whereas the SARS virus showed more orientation towards helix. Similarly, the E protein structure of COVID-19 shared more similarity with the Bat virus, while MERS, which displayed a sequence variation, showed structural similarity with the SARS virus. Although both COVID-19 and MERS belong to the same beta CoV, they are far apart from the sequence and structure groups. Many significant motifs and interacting regions have been analyzed, which will provide useful information on the drug targeting regions. The conserved sequences and motifs will help in understanding the virus better and also the drug binding regions. The interaction between M and E proteins will throw light in the design of a novel inhibitor for COVID-19 and MERS. Aside from the spike protein, the M and E proteins inhibitors will also find an essential role in combating the infection soon.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

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.

Availability of data and materials

The protein sequences of M and E proteins were downloaded from NCBI

URL: <https://www.ncbi.nlm.nih.gov/protein>

The structure of Covid-19 M protein was downloaded from I-Tasser Server. I-Tasser server for protein structure and function prediction (2020)

URL: <https://zhanglab.ccmb.med.umich.edu/I-TASSER/>

Funding

Self funded.

References

- [1] M.T. ul Qamar, S.M. Alqahtani, M.A. Alamri, L.-L. Chen, Structural basis of SARS-CoV-2 3CLpro and anti-COVID-19 drug discovery from medicinal plants, *J. Pharm Anal* (2020).
- [2] Saudi Arabia coronavirus (COVID-19) tracker map: AccuWeather, 2020, URL <https://www.accuweather.com/en/sa/national/covid-19>.
- [3] Coronavirus disease (COVID-19) - events as they happen, 2020, URL <https://www.who.int/emergencies/diseases/novel-coronavirus-2019/events-as-they-happen>.
- [4] Naming the coronavirus disease (COVID-19) and the virus that causes it, 2020, URL [https://www.who.int/emergencies/diseases/novel-coronavirus-2019/technical-guidance/naming-the-coronavirus-disease-\(covid-2019\)-and-the-virus-that-causes-it](https://www.who.int/emergencies/diseases/novel-coronavirus-2019/technical-guidance/naming-the-coronavirus-disease-(covid-2019)-and-the-virus-that-causes-it).
- [5] Person, Saudi Arabia announces first case of coronavirus, 2020, URL <https://www.arabnews.com/node/1635781/saudi-arabia>,
- [6] WHO coronavirus disease (COVID-19) dashboard, 2020, URL https://covid19.who.int/?gclid=CjwKCAjw_LL2BRAKEiwAv2Y3SZ5qzQ1ZXULC-hBX6gPpO8vMQPs62ItUQNC9oHhblWbuEndKtMRwVRoCNJAQAvD_BwE.
- [7] D. Schoeman, B.C. Fielding, Coronavirus envelope protein: current knowledge, *Virol. J.* 16 (1) (2019) 1–22.
- [8] Z.-W. Ye, S. Yuan, K.-S. Yuen, S.-Y. Fung, C.-P. Chan, D.-Y. Jin, Zoonotic origins of human coronaviruses, *Int. J. Biol. Sci.* 16 (10) (2020) 1686.
- [9] I.C. on Taxonomy of Viruses (ICTV), Virus taxonomy: 2019 release, 2019, URL <https://talk.ictvonline.org/taxonomy/>.

- [10] W.H. Organization, SARS (Severe Acute Respiratory Syndrome), 2012, URL <https://www.who.int/ith/diseases/sars/en/>.
- [11] X.-Y. Ge, J.-L. Li, X.-L. Yang, A.A. Chmura, G. Zhu, J.H. Epstein, J.K. Mazet, B. Hu, W. Zhang, C. Peng, Y.-J. Zhang, C.-M. Luo, B. Tan, N. Wang, Y. Zhu, G. Cramer, S.-Y. Zhang, L.-F. Wang, P. Daszak, Z.-L. Shi, Isolation and characterization of a bat SARS-like coronavirus that uses the ACE2 receptor, *Nature* 503 (7477) (2013) 535–538.
- [12] Z. Memish, MERS, *Int. J. Infect. Dis.* 53 (2016) 23.
- [13] W.H. Organization, Middle east respiratory syndrome coronavirus (MERS-CoV), 2019, URL [https://www.who.int/news-room/fact-sheets/detail/middle-east-respiratory-syndrome-coronavirus-\(mers-cov\)](https://www.who.int/news-room/fact-sheets/detail/middle-east-respiratory-syndrome-coronavirus-(mers-cov)).
- [14] G. News, Coronavirus (COVID-19), 2020, URL <https://news.google.com/covid19/map?hl=en-US&gl=US&ceid=US%3Aen>.
- [15] I. Astuti, Ysrafil, Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2): An overview of viral structure and host response, *Diabetes Metab. Syndr. Clin. Res. Rev.* (2020).
- [16] S. Jiang, C. Hillyer, L. Du, Neutralizing antibodies against SARS-CoV-2 and other human coronaviruses, *Trends Immunol.* (2020).
- [17] D. Liu, Q. Yuan, Y. Liao, Coronavirus envelope protein: a small membrane protein with multiple functions, *Cell. Mol. Life Sci.* 64 (16) (2007) 2043–2048.
- [18] M. Cascella, M. Rajnik, A. Cuomo, S.C. Dulebohn, R. Di Napoli, Features, evaluation and treatment coronavirus (COVID-19), in: StatPearls [Internet], StatPearls Publishing, 2020.
- [19] J.M. Jimenez-Guardeño, J.L. Nieto-Torres, M.L. DeDiego, J.A. Regla-Nava, R. Fernandez-Delgado, C. Castaño Rodríguez, L. Enjuanes, The PDZ-binding motif of severe acute respiratory syndrome coronavirus envelope protein is a determinant of viral pathogenesis, *PLoS Pathog* 10 (8) (2014) e1004320.
- [20] J.Q. Liang, S. Fang, Q. Yuan, M. Huang, R.A. Chen, T.S. Fung, D.X. Liu, N-linked glycosylation of the membrane protein ectodomain regulates infectious bronchitis virus-induced ER stress response, apoptosis and pathogenesis, *Virology* 531 (2019) 48–56.
- [21] M. Bianchi, D. Benvenuto, M. Giovanetti, S. Angeletti, M. Ciccozzi, S. Pascarella, Sars-CoV-2 envelope and membrane proteins: Structural differences linked to virus characteristics? *BioMed Res. Int.* 2020 (2020).
- [22] Protein, 2020, URL <https://www.ncbi.nlm.nih.gov/protein>.
- [23] Protein BLAST: search protein databases using a protein query, 2020, URL https://blast.ncbi.nlm.nih.gov/Blast.cgi?PROGRAM=blastp&PAGE_TYPE=BlastSearch&LINK_LOC=blasthome.
- [24] Multiple sequence alignment, 2020, URL <https://www.ebi.ac.uk/Tools/msa/mafft/>.
- [25] F. Madeira, Y. Park, J. Lee, N. Buso, T. Gur, N. Madhusoodanan, P. Basutkar, A. Tivey, S. Potter, R. Finn, R. Lopez, The EMBL-EBI search and sequence analysis tools APIs in 2019, *Nucleic Acids Res.* 47 (W1) (2019) W636–W641.
- [26] A.M. Waterhouse, J.B. Procter, D.M. Martin, M. Clamp, G.J. Barton, Jalview version 2—a multiple sequence alignment editor and analysis workbook, *Bioinformatics* 25 (9) (2009) 1189–1191.
- [27] Motif scan, 2020, URL https://myhits.sib.swiss/cgi-bin/motif_scan.
- [28] SWISS-MODEL, 2020, URL <https://swissmodel.expasy.org/>.
- [29] A. Waterhouse, M. Bertoni, S. Bienert, G. Studer, G. Tauriello, R. Gumienny, F.T. Heer, T.A.P. de Beer, C. Rempfer, L. Bordoli, et al., SWISS-MODEL: homology modelling of protein structures and complexes, *Nucleic Acids Res.* 46 (W1) (2018) W296–W303.
- [30] M. Bertoni, F. Kiefer, M. Biasini, L. Bordoli, T. Schwede, Modeling protein quaternary structure of homo-and hetero-oligomers beyond binary interactions by homology, *Sci. Rep.* 7 (1) (2017) 1–15.
- [31] N. Guex, M.C. Peitsch, T. Schwede, Automated comparative protein structure modeling with SWISS-MODEL and Swiss-PdbViewer: A historical perspective, *Electrophoresis* 30 (S1) (2009) S162–S173.
- [32] S. Bienert, A. Waterhouse, T.A. de Beer, G. Tauriello, G. Studer, L. Bordoli, T. Schwede, The SWISS-MODEL repository—new features and functionality, *Nucleic Acids Res.* 45 (D1) (2017) D313–D319.
- [33] G. Studer, C. Rempfer, A.M. Waterhouse, R. Gumienny, J. Haas, T. Schwede, QMEANDisCo—distance constraints applied on model quality estimation, *Bioinformatics* 36 (6) (2020) 1765–1771.
- [34] I-TASSER server for protein structure and function prediction, 2020, URL <https://zhanglab.ccmb.med.umich.edu/I-TASSER/>.
- [35] A. Roy, A. Kucukural, Y. Zhang, I-TASSER: a unified platform for automated protein structure and function prediction, *Nat. Protoc.* 5 (4) (2010) 725–738.
- [36] J. Yang, R. Yan, A. Roy, D. Xu, J. Poisson, Y. Zhang, The I-TASSER Suite: protein structure and function prediction, *Nature Methods* 12 (1) (2015) 7–8.
- [37] Protein validation. URL <http://www.scfbio-iitd.res.in/software/proteomics/uploadaction.jsp>.
- [38] D.S. Biovia, et al., Discovery studio modeling environment, 2017.
- [39] S. Vajda, C. Yueh, D. Beglov, T. Bohnuud, S.E. Mottarella, B. Xia, D.R. Hall, D. Kozakov, New additions to the ClusPro server motivated by CAPRI, *Proteins Struct. Funct. Bioinform.* 85 (3) (2017) 435–444.
- [40] D. Kozakov, D.R. Hall, B. Xia, K.A. Porter, D. Padjhony, C. Yueh, D. Beglov, S. Vajda, The ClusPro web server for protein–protein docking, *Nat. Protoc.* 12 (2) (2017) 255.

- [41] D. Kozakov, D. Beglov, T. Bohnuud, S.E. Mottarella, B. Xia, D.R. Hall, S. Vajda, How good is automated protein docking?, *Proteins Struct. Funct. Bioinform.* 81 (12) (2013) 2159–2166.
- [42] A.C. Wallace, R.A. Laskowski, J.M. Thornton, LIGPLOT: a program to generate schematic diagrams of protein-ligand interactions, *Protein Eng. Des. Sel.* 8 (2) (1995) 127–134.
- [43] J. Torres, J. Wang, K. Parthasarathy, D.X. Liu, The transmembrane oligomers of coronavirus protein E, *Biophys. J.* 88 (2) (2005) 1283–1290.
- [44] Q. Wu, Y. Zhang, H. Lü, J. Wang, X. He, Y. Liu, C. Ye, W. Lin, J. Hu, J. Ji, J. Xu, J. Ye, Y. Hu, W. Chen, S. Li, J. Wang, J. Wang, S. Bi, H. Yang, The E protein is a multifunctional membrane protein of SARS-CoV, *Genom. Proteom. Bioinform.* 1 (2) (2003) 131–144, [http://dx.doi.org/10.1016/S1672-0229\(03\)01017-9](http://dx.doi.org/10.1016/S1672-0229(03)01017-9), URL <http://www.sciencedirect.com/science/article/pii/S1672022903010179>.
- [45] Y. Liao, Q. Yuan, J. Torres, J. Tam, D. Liu, Biochemical and functional characterization of the membrane association and membrane permeabilizing activity of the severe acute respiratory syndrome coronavirus envelope protein, *Virology* 349 (2) (2006) 264–275.
- [46] Q. Yuan, Y. Liao, J. Torres, J. Tam, D. 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 (13) (2006) 3192–3200.
- [47] E. Corse, C.E. Machamer, The cytoplasmic tails of infectious bronchitis virus E and M proteins mediate their interaction, *Virology* 312 (1) (2003) 25–34.
- [48] Y. Liao, J. Lescar, J. Tam, D. Liu, Expression of SARS-coronavirus envelope protein in *Escherichia coli* cells alters membrane permeability, *Biochem. Biophys. Res. Commun.* 325 (1) (2004) 374–380.
- [49] J.L. Nieto-Torres, M.L. Dediego, E. Alvarez, J.M. Jiménez-Guardaño, J.A. Regla-Nava, M. Llorente, L. Kremer, S. Shuo, L. Enjuanes, Subcellular location and topology of severe acute respiratory syndrome coronavirus envelope protein, *Virology* 415 (2) (2011) 69–82.
- [50] W. Surya, Y. Li, C. Verdà-Bàguena, V.M. Aguilera, J. Torres, MERS coronavirus envelope protein has a single transmembrane domain that forms pentameric ion channels, *Virus Res.* 201 (2015) 61–66.
- [51] Y.-T. Tseng, C.-H. Chang, S.-M. Wang, K.-J. Huang, C.-T. Wang, Identifying SARS-CoV membrane protein amino acid residues linked to virus-like particle assembly, *PLoS One* 8 (5) (2013) e64013.
- [52] R. Byland, P.J. Vance, J.A. Hoxie, M. Marsh, A conserved dileucine motif mediates clathrin and AP-2-dependent endocytosis of the HIV-1 envelope protein, *Mol. Biol. Cell* 18 (2) (2007) 414–425.
- [53] A.L. Arndt, B.J. Larson, B.G. Hogue, A conserved domain in the coronavirus membrane protein tail is important for virus assembly, *J. Virol.* 84 (21) (2010) 11418–11428.
- [54] J. Liu, Y. Sun, J. Qi, F. Chu, H. Wu, F. Gao, T. Li, J. Yan, G.F. Gao, The membrane protein of severe acute respiratory syndrome coronavirus acts as a dominant immunogen revealed by a clustering region of novel functionally and structurally defined cytotoxic T-lymphocyte epitopes, *J. Infect. Dis.* 202 (8) (2010) 1171–1180.
- [55] D. Voß, S. Pfefferle, C. Drosten, L. Stevermann, E. Traggiai, A. Lanzavecchia, S. Becker, Studies on membrane topology, N-glycosylation and functionality of SARS-CoV membrane protein, *Virol. J.* 6 (1) (2009) 1–13.
- [56] C.E. McBride, C.E. Machamer, A single tyrosine in the severe acute respiratory syndrome coronavirus membrane protein cytoplasmic tail is important for efficient interaction with spike protein, *J. Virol.* 84 (4) (2010) 1891–1901.
- [57] Catalytic triad - an overview, 2016, URL <https://www.sciencedirect.com/topics/biochemistry-genetics-and-molecular-biology/catalytic-triad>.
- [58] J.D. Cronk, The catalytic triad and its role in the serine protease mechanism, 2016, URL <http://guweb2.gonzaga.edu/faculty/cronk/CHEM440pub/catalytic-triad.html>.
- [59] A. Perrier, A. Bonnin, L. Desmarests, A. Danneels, A. Goffard, Y. Rouillé, J. Dubuisson, S. Belouzard, The C-terminal domain of the MERS coronavirus m protein contains a trans-Golgi network localization signal, *J. Biol. Chem.* 294 (39) (2019) 14406–14421.
- [60] D. Voß, A. Kern, E. Traggiai, M. Eickmann, K. Stadler, A. Lanzavecchia, S. Becker, Characterization of severe acute respiratory syndrome coronavirus membrane protein, *FEBS Lett.* 580 (3) (2006) 968–973.
- [61] L. Kuo, K.R. Hurst-Hess, C.A. Koetzner, P.S. Masters, Analyses of coronavirus assembly interactions with interspecies membrane and nucleocapsid protein chimeras, *J. Virol.* 90 (9) (2016) 4357–4368.