

Sars-CoV-2 Envelope and Membrane proteins: structural differences linked to virus characteristics?

Martina Bianchi¹, Domenico Benvenuto², Marta Giovanetti³, Silvia Angeletti⁴, Massimo Ciccozzi² and Stefano Pascarella^{1*}

¹Department of Biochemical sciences “A Rossi Fanelli”, Sapienza University of Rome, 00185 Rome, Italy

²Unit of Medical Statistics and Molecular Epidemiology, University Campus Bio-Medico of Rome, Rome, Italy.

³Flavivirus Laboratory, Oswaldo Cruz Institute, Oswaldo Cruz Foundation, Rio de Janeiro, Brazil.

⁴Unit of Clinical Laboratory Science, University Campus Bio-Medico of Rome, Rome, Italy

Corresponding author

Stefano Pascarella
Department of Biochemical sciences “A Rossi Fanelli”
University of Rome Sapienza
00185 Rome, Italy
e-mail: Stefano.Pascarella@uniroma1.it

Abstract

The Coronavirus Disease 2019 (COVID-19) is a new viral infection caused by the severe acute respiratory coronavirus 2 (SARS-CoV-2). Genomic analyses have revealed that SARS-CoV-2 is related to Pangolin and Bat coronaviruses. In this report, a structural comparison between the Sars-CoV-2 Envelope and Membrane proteins from different human isolates with homologous proteins from closely related viruses is described. The analyses here reported show the high structural similarity of Envelope and Membrane proteins to the counterparts from Pangolin and Bat coronavirus isolates. However, the comparisons have also highlighted structural differences specific of Sars-CoV-2 proteins which may be correlated to the cross-species transmission and/or to the properties of the virus. Structural modelling has been applied to map the variant sites onto the predicted three-dimensional structure of the Envelope and Membrane proteins.

Introduction

COVID-19 has become a planetary emergency which is seriously threatening human health (Lai et al. 2020; Benvenuto, Giovanetti, Salemi, et al. 2020). Development of effective therapeutic and prevention strategies is significantly hampered also by the lack of detailed structural information on virus proteins, although several crystallographic structures of Sars-CoV-2 proteins are now available (Wrapp et al. 2020; Benvenuto, Giovanetti, Ciccozzi, Spoto, Angeletti, Ciccozzi, et al. 2020; Walls et al. 2020). In this report, a structural comparison between the Sars-CoV-2 surface proteins from different isolates with homologous proteins from closely related viruses such as those from Bat and Pangolin is described. This work has been focussed onto the Envelope (E) and Membrane (M) proteins that, along with the Spike, form the virus protein interface to the external environment. The Spike glycoprotein has been already extensively studied and a few crystallographic structures are available in the Protein Data Bank (Walls et al. 2020; Benvenuto, Giovanetti, Ciccozzi, Spoto, Angeletti, and Ciccozzi 2020; Wrapp et al. 2020; Yan et al. 2020); consequently, this protein has not been specifically addressed within this note. Identification of local structural differences, even minimal, to the closest virus proteins may indicate the mutations that enabled Sars-CoV-2 to cross species and/or to acquire its peculiar pathogenic properties (Angeletti et al. 2020; Ji et al. 2020). Indeed, a number of examples have been reported in the scientific literature suggesting how even single point mutations in virus proteins can significantly alter their biology and pathogenesis (André et al. 2019; Sakai et al. 2017). Therefore, comparative studies may shed light on the molecular mechanisms through which epidemic of epizootic origin can emerge and may also suggest molecular targets for therapeutics or reverse vaccinology experiments.

Material and Methods

Nucleotide and protein sequences have been taken from GenBank (Benson et al. 2018) data repository. Blast suite (Altschul et al. 1997) has been used for databank searches, Jalview (A. M. Waterhouse et al. 2009) and MAFFT (Kato and Standley 2013) have been used for multiple sequence display and alignment respectively. Transmembrane helix prediction has been obtained by TMHMM (Chen et al. 2003), MEMSAT (McGuffin, Bryson, and Jones 2000) and Protter (Omasits et al. 2014). Cd-hit program (Fu et al. 2012) has been used for sequence clustering. Homology modelling relied on Swiss-Model (A. Waterhouse et al. 2018), Modeller (Webb and Sali 2017) or HHpred (Zimmermann et al. 2018) and structure display and analysis on Open-Source PyMOL (Schrodinger LLC 2015). When necessary, I-Tasser (Yang et al. 2015) has been used as an alternative source of ab-initio homology models.

Results

Databank searches and structure modelling

From the GenBank repository, 797 complete genomes of Sars-CoV-2 have been collected (the full list is reported in **Supplementary Data**). The TblastN program has been used to extract the sequences of E and M proteins from each genome. To remove redundancy within each E and M protein set, cd-hit clustering has been applied at 100% sequence identity level: identical sequences have been assigned to the same group for which only one representative has been considered for further analysis. The Sars-CoV-2 E and M protein sets have been grouped into three and seven clusters, respectively. This finding suggests that within the 797 genomes three and seven variants of the E and M proteins can be observed, respectively. E and M homologous proteins from closely related virus have been retrieved from the GenBank using the TblastN tool.

Envelope protein

The E protein is conserved across β -coronaviruses. Only three variants have been found in the Sars-CoV-2 E set collected. Sequence comparisons show that the Sars-CoV-2 E protein from the reference genome (RefSeq code YP_009724392) is identical to the sequences from Pangolin CoV MP798 and Bat CoV CoVZXC21, CoVZC45 and RaTG13 isolates. The multiple sequence alignment reported in **Figure 1** demonstrates that a distinguishing feature of Sars-2-CoV E variants is the presence of Arg at position 69 that substitutes Glu, Gln, Asp in other homologous Sars-CoV E proteins. This site is followed by a deletion in position 70 corresponding to Gly or Cys in the other proteins. Sars-CoV-2 E sequences differ from the homologous proteins also at positions 55-56, where the dyad Ser-Phe replaces Thr-Val (except in Bat coronavirus isolate BtKY72, accession code KY352407). Variants of the Sars-CoV-2 E protein differ at positions 37 and 72 where His substitutes a Leu and Leu replaces a conserved Pro, respectively. Size of each Envelope variant cluster is reported in **Supplementary Data Table 1** along with accession codes and definitions of the isolates. A homology model of the E protein has been built with Modeller using as a template the pentameric ion channel structure of Sars-CoV protein identified by the PDB code 5X29. This sequence shares 91% identity to Sars-CoV-2 E protein and covers the segment encompassed by positions 8-65. **Figure 2** displays the structure of the homology model of the Sars-CoV-2 E protein assembled as a pentameric viroporin-like protein. **Figure 2** displays also the position of the variant sites onto the three-dimensional model. Prediction of the transmembrane helices is difficult in a short protein. Therefore, transmembrane topology cannot be assigned reliably. Likewise, experiments have not clarified definitively which portions of the E protein are exposed to the external or internal side of the virus membrane (Schoeman and Fielding 2019).

Membrane glycoprotein

The M glycoprotein is conserved across the β -coronaviruses. However, seven variants of Sars-CoV-2 M protein were identified in the collected set while only three variants were observed for the E protein (**Figure 3**). The multiple sequence alignment shows a remarkable similarity (98% identity)

among the Sars-CoV-2 M variants and the sequences from Bat and Pangolin isolates. However, a difference at the N-terminal position (**Figure 3**) can be observed: the insertion of a Ser residue at position 4 of human Sars-CoV-2 seems to be a unique feature of this protein. In the corresponding position, the RaTG13 Bat M protein displays a deletion while Bat CoVZXC21, CoVZC45, and Pangolin MP789 proteins have an Asp residue. The seven M protein variants differ at positions 2, 3, 57, 70, 85, 89 and 175. Size of each Membrane variant cluster is reported in **Supplementary Data Table 1** along with accession codes and definitions of the isolates. Noteworthy, the protein from the Sars-CoV-2 NIHE isolate (accession code MT127115) possesses an Arg instead of a conserved Gly at position 89 (**Figure 3**). The mutation occurs within a predicted transmembrane helix and, if confirmed, may have a significant impact on the protein properties (**Figure 3**).

The three-dimensional model of the M protein has been taken from I-Tasser server (code QHD43419) as other methods failed to find any suitable template. However, it should be mentioned that HHpred found a weak local affinity, albeit below the statistical significance level, to 4N31, a peptidase-like protein from *Streptococcus pyogenes* essential for pilus polymerisation. **Figure 4** displays the positions of the variant sites onto the model structure. This model has been predicted by ab-initio techniques. Therefore, it should be considered with great caution and should be treated as a low-resolution approximation of the real structure. According to the prediction of the transmembrane helix topology, the N- and C-terminal portions of the M protein are exposed outside and inside the virus particle, respectively (**Figure 4**).

Discussion

Previous studies pointed out that E and M proteins could be important for viral entry, replication and particle assembly within the human cells (Schoeman and Fielding 2019; J Alsaadi and Jones 2019). According to the accepted theories, the current COVID-19 pandemic has been caused by the cross-species transmission of a β -coronavirus normally hosted by Bats and, perhaps, Pangolin to humans (Benvenuto, Giovanetti, Ciccozzi, Spoto, Angeletti, and Ciccozzi 2020; Lu et al. 2020). In this paper, E and M proteins from 797 Sars-CoV-2 genomes have been compared to the counterparts taken from the most closely related virus also to evaluate the potential role of amino acid mutations in epizootic origin of COVID-19. E protein is a minor component of the virus membrane though it is deemed to be important for many stages of virus infection and replication (Schoeman and Fielding 2019; J Alsaadi and Jones 2019). Sequence comparison has shown that this protein is identical to the counterparts of specific Bat and Pangolin coronavirus isolates, even though Sars-CoV-2 sequence seems to possess specific modifications and characteristics with respect to other Sars CoV. In particular Arg69, a positively charged amino acid, replaces Glu or Gln residues, negatively charged and neutral respectively, in the homologous CoV proteins. Moreover, a deletion specific to Sars-CoV-2 proteins flanks this position. Unfortunately, it is not possible to predict reliably whether the sites of these modifications are exposed to the internal or external side of the membrane. In any case, the

substitution and the deletion appear a rather drastic change and may have significant impact on conformational properties and possibly on protein-protein interactions. Further structural studies are needed. However, it may be hypothesized that these changes can also affect the oligomerization process necessary to form a transmembrane ion channel.

It has been demonstrated that M protein is more prevalent within the virus membrane and it is deemed to be important for the budding process of coronaviruses. Indeed, during the process of virus particle assembly, this protein interacts with the Nucleocapsid, Envelope, Spike and Membrane glycoprotein itself (J Alsaadi and Jones 2019). Moreover, in Alphacoronaviruses it has been demonstrated that this protein cooperates with the Spike during the cell attachment and entry (Naskalska et al. 2019). Therefore, mutations occurring at the N-terminus region, which is exposed to the virus surface, could play a key role in the host cell interaction.

In conclusion, the analyses here reported show the structural similarity of E and M proteins to the counterparts from Pangolin and Bat coronavirus isolates. At the same time, comparisons have highlighted structural differences specific of Sars-CoV-2 proteins which may be correlated to the cross-species transmission and/or to the properties of the virus. Although further studies are needed, it is clear that these amino acid variations have been important for the virus evolutionary history and the results may hint at how similar mutations within the coronavirus family can lead in the next years to other epizootic epidemic events similar to the one that we are experiencing these days.

Conflicts of Interest

The authors declare that there is no conflict of interest regarding the publication of this paper.

Funding Statement

This work has been in part funded by a grant to SP from Sapienza University of Rome [RP11916B74B27C4D].

References

- Altschul, S., T L Madden, A A Schäffer, J Zhang, Z Zhang, W Miller, and D J Lipman. 1997. "Gapped BLAST and PSI-BLAST: A New Generation of Protein Database Search Programs." *Nucleic Acids Research* 25 (17): 3389–3402. <https://doi.org/10.1093/nar/25.17.3389>.
- André, Nicole M, Brieuc Cossic, Emma Davies, Andrew D Miller, and Gary R Whittaker. 2019. "Distinct Mutation in the Feline Coronavirus Spike Protein Cleavage Activation Site in a Cat with Feline Infectious Peritonitis-Associated Meningoencephalomyelitis." *JFMS Open*

Reports 5 (1): 2055116919856103. <https://doi.org/10.1177/2055116919856103>.

Angeletti, Silvia, Domenico Benvenuto, Martina Bianchi, Marta Giovanetti, Stefano Pascarella, and Massimo Ciccozzi. 2020. "COVID-2019: The Role of the Nsp2 and Nsp3 in Its Pathogenesis." *Journal of Medical Virology*, February. <https://doi.org/10.1002/jmv.25719>.

Benson, Dennis A, Mark Cavanaugh, Karen Clark, Ilene Karsch-Mizrachi, James Ostell, Kim D Pruitt, and Eric W Sayers. 2018. "GenBank." *Nucleic Acids Research* 46 (D1): D41–47. <https://doi.org/10.1093/nar/gkx1094>.

Benvenuto, Domenico, Marta Giovanetti, Alessandra Ciccozzi, Silvia Spoto, Silvia Angeletti, and Massimo Ciccozzi. 2020. "The 2019-New Coronavirus Epidemic: Evidence for Virus Evolution." *Journal of Medical Virology* 92 (4): 455–59. <https://doi.org/10.1002/jmv.25688>.

Benvenuto, Domenico, Marta Giovanetti, Alessandra Ciccozzi, Silvia Spoto, Silvia Angeletti, Massimo Ciccozzi, Roujian Lu, et al. 2020. "Membrane Protein of Human Coronavirus NL63 Is Responsible for Interaction with the Adhesion Receptor." *Journal of Virology* 93 (10224): 455–59. <https://doi.org/10.1128/JVI.00355-19>.

Benvenuto, Domenico, Marta Giovanetti, Marco Salemi, Mattia Prosperi, Cecilia De Flora, Luiz Carlos Junior Alcantara, Silvia Angeletti, and Massimo Ciccozzi. 2020. "The Global Spread of 2019-NCov: A Molecular Evolutionary Analysis." *Pathogens and Global Health*, February, 1–4. <https://doi.org/10.1080/20477724.2020.1725339>.

Chen, Yunjia, Peng Yu, Jingchu Luo, and Ying Jiang. 2003. "Secreted Protein Prediction System Combining CJ-SPHMM, TMHMM, and PSORT." *Mammalian Genome : Official Journal of the International Mammalian Genome Society* 14 (12): 859–65. <https://doi.org/10.1007/s00335-003-2296-6>.

Fu, Limin, Beifang Niu, Zhengwei Zhu, Sitao Wu, and Weizhong Li. 2012. "CD-HIT: Accelerated for Clustering the next-Generation Sequencing Data." *Bioinformatics (Oxford, England)* 28 (23): 3150–52. <https://doi.org/10.1093/bioinformatics/bts565>.

J Alsaadi, Entedar A, and Ian M Jones. 2019. "Membrane Binding Proteins of Coronaviruses." *Future Virology* 14 (4): 275–86. <https://doi.org/10.2217/fvl-2018-0144>.

Ji, Wei, Wei Wang, Xiaofang Zhao, Junjie Zai, and Xingguang Li. 2020. "Cross-Species Transmission of the Newly Identified Coronavirus 2019-NCov." *Journal of Medical Virology* 92 (4): 433–40. <https://doi.org/10.1002/jmv.25682>.

Katoh, Kazutaka, and Daron M Standley. 2013. "MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability." *Molecular Biology and Evolution* 30 (4): 772–80. <https://doi.org/10.1093/molbev/mst010>.

- Lai, Chih-Cheng, Tzu-Ping Shih, Wen-Chien Ko, Hung-Jen Tang, and Po-Ren Hsueh. 2020. "Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) and Coronavirus Disease-2019 (COVID-19): The Epidemic and the Challenges." *International Journal of Antimicrobial Agents* 55 (3): 105924. <https://doi.org/10.1016/j.ijantimicag.2020.105924>.
- Lu, Roujian, Xiang Zhao, Juan Li, Peihua Niu, Bo Yang, Honglong Wu, Wenling Wang, et al. 2020. "Genomic Characterisation and Epidemiology of 2019 Novel Coronavirus: Implications for Virus Origins and Receptor Binding." *Lancet (London, England)* 395 (10224): 565–74. [https://doi.org/10.1016/S0140-6736\(20\)30251-8](https://doi.org/10.1016/S0140-6736(20)30251-8).
- McGuffin, L J, K Bryson, and D T Jones. 2000. "The PSIPRED Protein Structure Prediction Server." *Bioinformatics (Oxford, England)* 16 (4): 404–5. <https://doi.org/10.1093/bioinformatics/16.4.404>.
- Naskalska, Antonina, Agnieszka Dabrowska, Artur Szczepanski, Aleksandra Milewska, Krzysztof Piotr Jasik, and Krzysztof Pyrc. 2019. "Membrane Protein of Human Coronavirus NL63 Is Responsible for Interaction with the Adhesion Receptor." *Journal of Virology* 93 (19). <https://doi.org/10.1128/JVI.00355-19>.
- Omasits, Ulrich, Christian H Ahrens, Sebastian Müller, and Bernd Wollscheid. 2014. "Protter: Interactive Protein Feature Visualization and Integration with Experimental Proteomic Data." *Bioinformatics (Oxford, England)* 30 (6): 884–86. <https://doi.org/10.1093/bioinformatics/btt607>.
- Sakai, Yusuke, Kengo Kawachi, Yutaka Terada, Hiroko Omori, Yoshiharu Matsuura, and Wataru Kamitani. 2017. "Two-Amino Acids Change in the Nsp4 of SARS Coronavirus Abolishes Viral Replication." *Virology* 510 (October): 165–74. <https://doi.org/10.1016/j.virol.2017.07.019>.
- Schoeman, Dewald, and Burtram C Fielding. 2019. "Coronavirus Envelope Protein: Current Knowledge." *Virology Journal* 16 (1): 69. <https://doi.org/10.1186/s12985-019-1182-0>.
- Schrodinger LLC. 2015. "The AxPyMOL Molecular Graphics Plugin for Microsoft PowerPoint, Version 1.8."
- Walls, Alexandra C., Young Jun Park, M. Alejandra Tortorici, Abigail Wall, Andrew T. McGuire, and David Veasler. 2020. "Structure, Function, and Antigenicity of the SARS-CoV-2 Spike Glycoprotein." *Cell* 181 (2): 281-292.e6. <https://doi.org/10.1016/j.cell.2020.02.058>.
- Waterhouse, Andrew, Martino Bertoni, Stefan Bienert, Gabriel Studer, Gerardo Tauriello, Rafal Gumieny, Florian T Heer, et al. 2018. "SWISS-MODEL: Homology Modelling of Protein Structures and Complexes." *Nucleic Acids Research* 46 (W1): W296–303. <https://doi.org/10.1093/nar/gky427>.
- Waterhouse, Andrew M., James B. Procter, David M.A. Martin, Michèle Clamp, and Geoffrey J.

- Barton. 2009. "Jalview Version 2-A Multiple Sequence Alignment Editor and Analysis Workbench." *Bioinformatics* 25 (9): 1189–91. <https://doi.org/10.1093/bioinformatics/btp033>.
- Webb, Benjamin, and Andrej Sali. 2017. "Protein Structure Modeling with MODELLER." *Methods in Molecular Biology* 1654: 39–54. https://doi.org/10.1007/978-1-4939-7231-9_4.
- Wrapp, Daniel, Nianshuang Wang, Kizzmekia S Corbett, Jory A Goldsmith, Ching-Lin Hsieh, Olubukola Abiona, Barney S Graham, and Jason S McLellan. 2020. "Cryo-EM Structure of the 2019-NCov Spike in the Prefusion Conformation." *Science (New York, N.Y.)* 367 (6483): 1260–63. <https://doi.org/10.1126/science.abb2507>.
- Yan, Renhong, Yuanyuan Zhang, Yaning Li, Lu Xia, Yingying Guo, and Qiang Zhou. 2020. "Structural Basis for the Recognition of SARS-CoV-2 by Full-Length Human ACE2." *Science* 367 (6485): 1444 LP – 1448. <https://doi.org/10.1126/science.abb2762>.
- Yang, Jianyi, Renxiang Yan, Ambrish Roy, Dong Xu, Jonathan Poisson, and Yang Zhang. 2015. "The I-TASSER Suite: Protein Structure and Function Prediction." *Nature Methods* 12 (1): 7–8. <https://doi.org/10.1038/nmeth.3213>.
- Zimmermann, Lukas, Andrew Stephens, Seung-Zin Nam, David Rau, Jonas Kübler, Marko Lozajic, Felix Gabler, Johannes Söding, Andrei N. Lupas, and Vikram Alva. 2018. "A Completely Reimplemented MPI Bioinformatics Toolkit with a New HHpred Server at Its Core." *Journal of Molecular Biology* 430 (15): 2237–43. <https://doi.org/10.1016/j.jmb.2017.12.007>.

Figure captions

Figure 1

Multiple sequence alignment among Sars-CoV-2 envelope protein variants and a set of the most similar homologous proteins. The sequence labelled Sars-CoV-2 corresponds to the reference sequence identified by the RefSeq code YP_009724392. Red lines below the alignment indicate the changed sites discussed in the text. Blu background denotes conserved alignment positions.

Figure 2

Three-dimensional model of the viroporin-like tetrameric assembly of the E protein from Sars-CoV-2 represented as a cartoon model. Residues corresponding to the mutated sites indicated in Figure 1 are displayed as transparent space-filling spheres and labelled with the amino acid one-letter code. The C-terminal segments of the model are reported for completeness. However, they convey no structural information due to lack of a corresponding segment in the structural template used in homology modelling. Structure in panel B is rotated by approximately 180° along the x axis with respect to the orientation shown in panel A.

Figure 3

Multiple sequence alignment among Sars-CoV-2 M protein variants and a set of most similar homologous proteins. The sequence label Sars-CoV-2 indicates the reference sequence identified by the RefSeq code YP_009724393. Red box indicates the variant sites at the N-terminal discussed in the text. Numbered red bars under the multiple alignment mark the prediction of transmembrane helices. The location of the connect loop with respect to the virion surface is indicated as “in” or “out”. Blu background denotes conserved alignment positions.

Figure 4

I-Tasser model of the Membrane protein represented as cartoon model. Variant positions are displayed as transparent space filling spheres and labelled with the amino acid one-letter code.

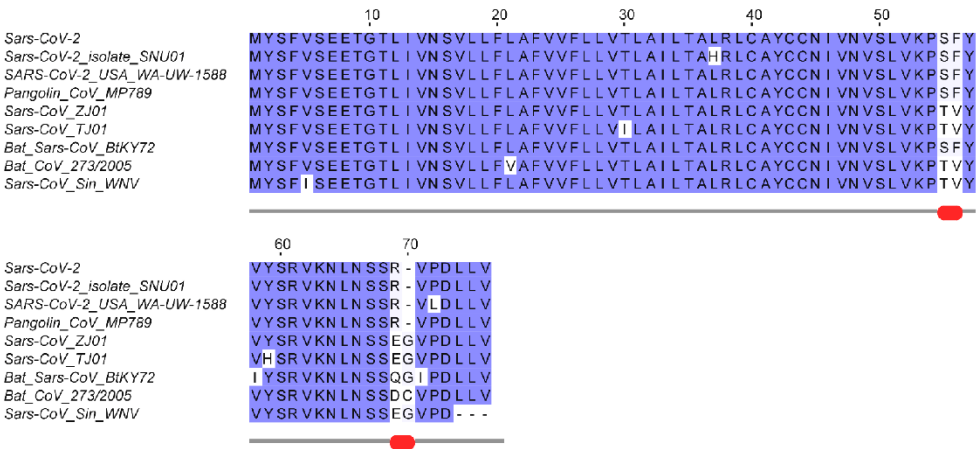


Figure 1

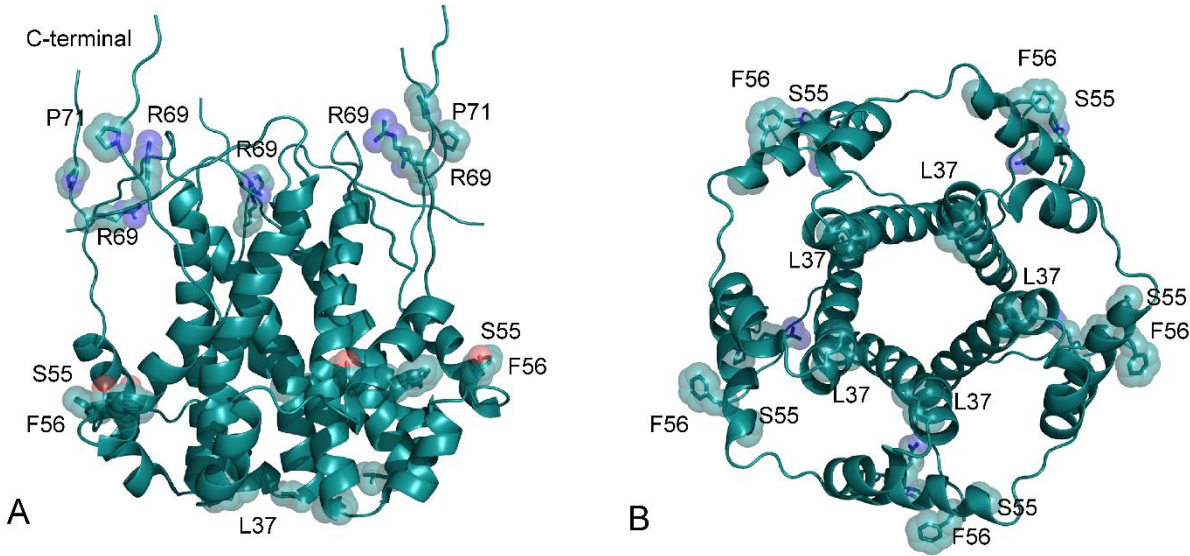


Figure 2

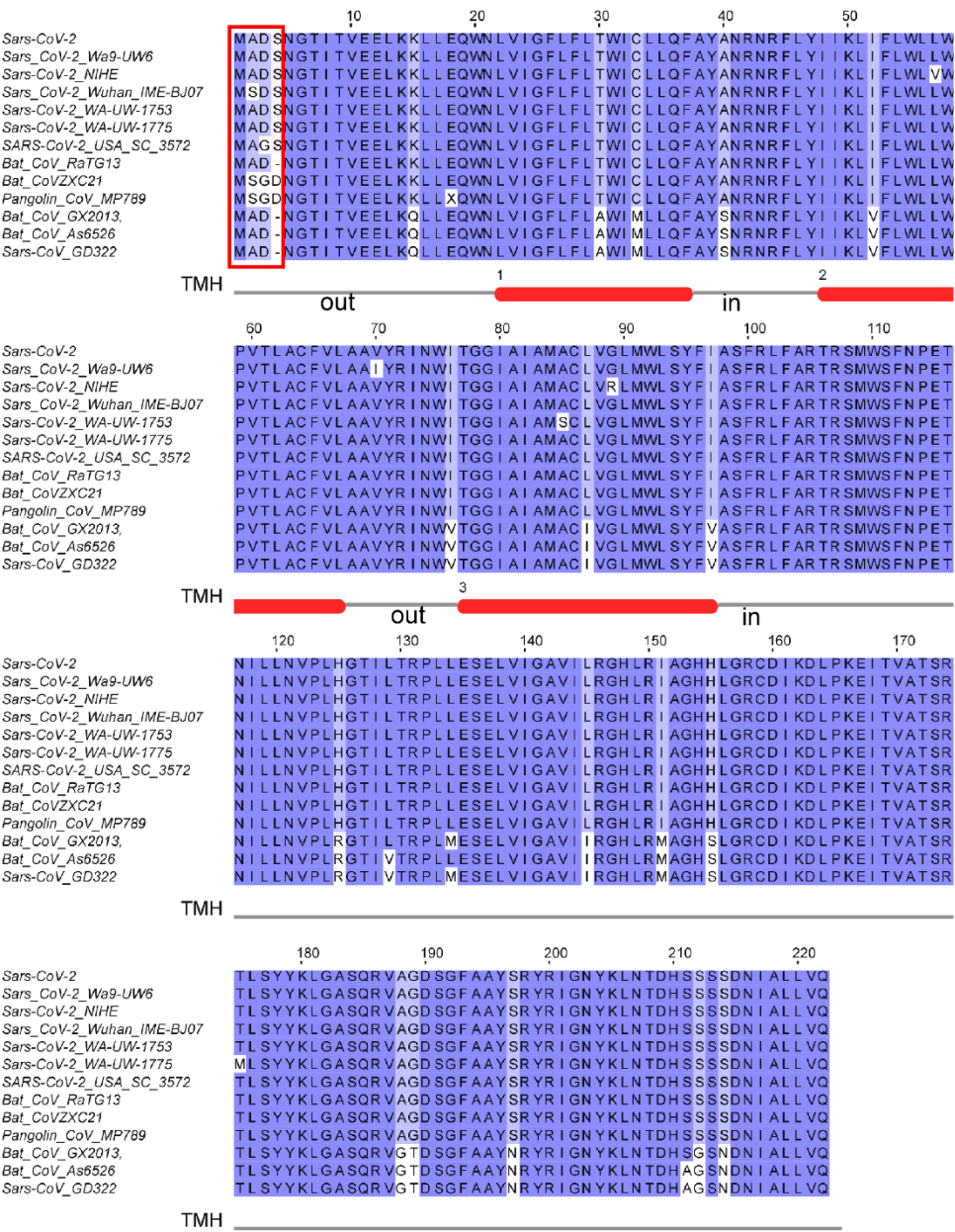


Figure 3

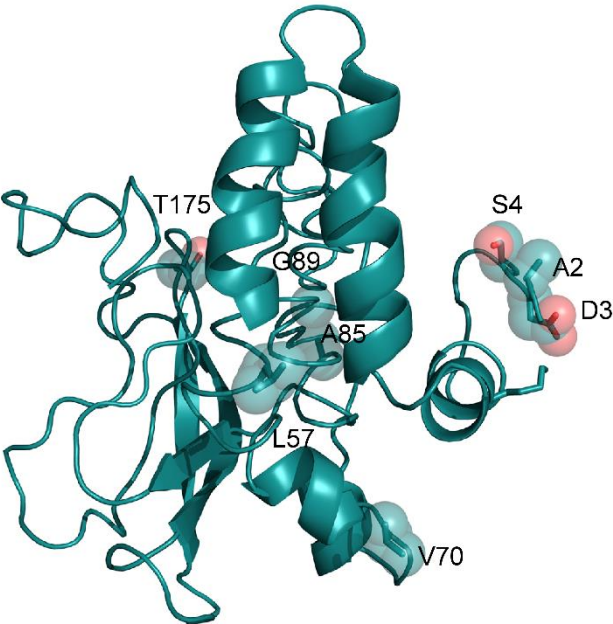


Figure 4