Sars-CoV-2 Envelope and Membrane proteins: differences from closely related proteins linked to cross-species transmission?

Martina Bianchi[1], Domenico Benvenuto[2], Marta Giovanetti[3], Silvia Angeletti[4], Massimo Ciccozzi[2] and Stefano Pascarella[1]*

[1]Department of Biochemical sciences "A Rossi Fanelli", Sapienza University of Rome, 00185 Rome, Italy
[2]Unit of Medical Statistics and Molecular Epidemiology, University Campus Bio-Medico of Rome, Rome, Italy.
[3]Flavivirus Laboratory, Oswaldo Cruz Institute, Oswaldo Cruz Foundation, Rio de Janeiro, Brazil.
[4]Unit 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 (COVID-19) is a new viral infection caused by severe acute respiratory coronavirus 2 (SARS-CoV-2) that was initially reported in city of Wuhan, China and afterwards spread globally. Genomic analyses revealed that SARS-CoV-2 is phylogenetically related to severe acute respiratory syndrome-like (SARS-like) Pangolin and Bat coronavirus specific isolates. In this study we focused on two proteins of Sars-CoV-2 surface: Envelope protein and Membrane protein. Sequences from Sars-CoV-2 isolates and other closely related virus were collected from the GenBank through TBlastN searches. The retrieved sequences were multiply aligned with MAFFT. The Envelope protein is identical to the counterparts from Pangolin CoV MP798 isolate and Bat CoV isolates CoVZXC21, CoVZC45 and RaTG13. However, a substitution at position 69 where an Arg replace for Glu, and a deletion in position 70 corresponding to Gly or Cys in other Envelope proteins were found. The Membrane glycoprotein appears more variable with respect to the SARS CoV proteins than the Envelope: a heterogeneity at the N-terminal position, exposed to the virus surface, was found between Pangolin CoV MP798 isolate and Bat CoV isolates CoVZXC21, CoVZC45 and RaTG13. Mutations observed on Envelope protein are drastic and may have significant implications for conformational properties and possibly for protein-protein interactions. Mutations on Membrane protein may also be relevant because this protein cooperates with the Spike during the cell attachment and entry. Therefore, these mutations may influence interaction with host cells. The mutations that have been detected in these comparative studies may reflect functional peculiarities of the Sars-CoV-2 virus and may help explaining the epizootic origin the COVID-19 epidemic.

Introduction

COVID-19 has become a planetary emergency which is seriously threatening human health (Benvenuto, Giovanetti, Salemi, et al., 2020; Lai, Shih, Ko, Tang, & Hsueh, 2020). Many aspects of the structure and biology of the Sars-CoV-2 virus are yet to be elucidated. Development of effective therapeutic and prevention strategies is significantly hampered by the lack of detailed structural information on virus proteins, although a few crystallographic structures of virus proteins are now available (Walls et al., 2020; Zhang et al., 2020). A contribution to the deciphering of virus properties may also come from careful comparative protein sequence and structure analysis to detect significant differences to similar viruses. In this report, we describe the results of a comparison of the Sars-CoV-2 surface proteins from different isolates of the virus to homologous proteins from the most closely related proteins such as those from Bat and Pangolin coronavirus. Our work has been focussed onto the Envelope (E) and Membrane (M) proteins that form along with the Spike, the virus protein interface to the external environment through which interacts initially with target human cells. The Spike glycoprotein has been already extensively studied and a crystallographic structure is available in the Protein Data Bank (Benvenuto, Giovanetti, Ciccozzi, et al., 2020; Walls et al., 2020); in consideration of this, the protein has not been specifically addressed within this note. Identification of local structural differences, even minimal, to the closest virus proteins may suggest the mutations that enabled Sars-CoV-2 to cross species and to acquire its peculiar pathogenic properties (Angeletti et al., 2020; Ji, Wang, Zhao, Zai, & Li, 2020). In fact, a number of examples have been published in the scientific literature showing how even single point mutations in virus proteins can significantly alter their biology and pathogenesis (André, Cossic, Davies, Miller, & Whittaker, 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, Procter, Martin, Clamp, & Barton, 2009) and MAFFT (Katoh & Standley, 2013) have been used for multiple sequence display and alignment. Transmembrane helix prediction has been obtained by TMHMM (Chen, Yu, Luo, & Jiang, 2003), MEMSAT (McGuffin, Bryson, & Jones, 2000) and Protter (Omasits, Ahrens, Müller, & Wollscheid, 2014). Homology modelling relied on Swiss-Model (A. Waterhouse et al., 2018), Modeller (Webb & Sali, 2017) or HHpred (Zimmermann et al., 2018) and structure display and analysis on 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 modelling

The Sars-CoV-2 E and M protein sequences (Table 1) have been used as TBlastN queries to search the GenBank nucleotide database restricted to the Viruses taxonomical division. Sequences have been collected separately for each Sars-CoV-2 proteins and aligned with MAFFT. At the time of access to the GenBank, 102 genomes from different Sars-CoV-2 isolates were retrieved.

Envelope protein

TblastN search confirms that, in general, E protein is well conserved across β-coronaviruses and particularly across SARS CoVs. In particular, the Sars-CoV-2 E protein is identical to that of Pangolin CoV MP798 isolate and Bat CoV isolates CoVZXC21, CoVZC45 and RaTG13 (**Table 1**). The multiple sequence alignment reported in **Figure 1** demonstrates that a distinguishing feature of Sars-2-CoV E protein is a substitution at position 69 where an Arg replace for Glu, Gln, Asp in other homologous SARS proteins, and a deletion in position 70 corresponding to Gly or Cys in other E proteins. Moreover, Sars-CoV-2 sequences differ from the others at positions 55-56, where Ser-Phe replace for Thr-Val (except in Bat coronavirus isolate BtKY72, accession code KY352407). Interestingly, Sars-CoV-2 isolate SNU01 (GenBank code MT039890) possesses an His at the hydrophobic position 57. A homology model of the E protein has been built with Modeller using as a template the pentameric ion channel structure of SARS coronavirus protein denoted by the PDB code 5X29. This sequence shares 91% identity to Sars-CoV-2 envelope protein and covers the segment encompassed by positions 8-65. **Figure 2** displays the structural mapping of the relevant sites onto the three-dimensional model of a putative pentameric envelope protein in the viroporin-like assembly. Prediction of the transmembrane helices and topology is difficult in such a short protein and therefore the internal and external portions cannot be assigned unambiguously. Experiments have not clarified definitively this point (Schoeman & Fielding, 2019).

Membrane glycoprotein

GenBank search confirms that, similarly to E protein, M glycoprotein is generally conserved across β-CoVs and specially across SARS CoVs. However, this protein appears more variable with respect to the SARS CoV proteins than the Envelope (**Figure 3**). Multiple sequence alignment points that there is a remarkable similarity among the Sars-CoV-2 sequences and those from the same Bat and Pangolin isolates referred to for the E protein. However, there is a heterogeneity at the N-terminal position **(Figure 3)** where an insertion of a Ser residue at position 4 of human Sars-CoV-2 seems to be a unique feature of this protein. RaTG13 Bat isolates displays a deletion while Bat CoVZXC21, CoVZC45, and Pangolin MP789 have an Asp residue. In these three cases, positions 2 and 3 are replaced by Ser and Gly. Interestingly, two Sars-CoV-2 isolates display a mutation at position 70 (WA9-UW6 isolate, GenBank code MT163721) and two mutations at positions 56 and 89 (NIHE

isolate, accession code MT127115). Substitutions at positions 56 and 70 occur in the predicted transmembrane and internal portions, respectively and conserve the hydrophobic characteristic of the site. At variance with them, substitution at position 89 replaces a transmembrane Gly with Arg. The mutation, if confirmed, should have a significant impact on the protein properties.

Three-dimensional model for the membrane protein has been taken from I-Tasser server since other methods failed to find any suitable template (code QHD43419). However, it should be mentioned that HHpred found a weak local affinity, well below the statistical significance level, to 4N31, a peptidase-like protein from *Streptococcus pyogenes* essential for pilus polymerisation. Mapping of the relevant sites onto the three-dimensional model is displayed in **Figure 4**. According to the transmembrane helix topology predictions, the N-terminal portion is located outside the virus particle while the C-terminal inside (**Figure 4**). As this model has been predicted by ab-initio techniques, it should be considered with great caution and only as a low-resolution approximation of the real structure.

Discussion

Previous studies highlighted that E and M proteins could be important for viral entry, replication and particle assembly within the human cells (J Alsaadi & Jones, 2019; Schoeman & Fielding, 2019). According to the most accepted theories, the current COVID-19 pandemic has been caused by the cross-species transmission of a Coronavirus normally hosted by Bats and, perhaps, Pangolin to humans (Benvenuto, Giovanetti, Ciccozzi, et al., 2020; Lu et al., 2020). In this paper, we have examined E and M proteins from a molecular point of view in order to evaluate also the potential role of aminoacidic mutations in the epizootic origin of SARS-CoV-2. 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 (J Alsaadi & Jones, 2019; Schoeman & Fielding, 2019). We have observed that this protein is very similar to the counteparts of other Bat and Pangolin coronavirus isolates, even though SARS-CoV-2 seems to possess unique modifications and characteristics. In particular Arg, a positively charged amino acid, replaces Glu or Gln, negatively charged and neutral respectively, present in the homologous CoV proteins at the C-terminal side of the sequence. Moreover, a unique deletion flanks this residue. Unfortunately, it is not possible to predict reliably whether this modification is exposed to the internal or external side of the membrane. In any case, this modification, substitution and deletion, appears rather drastic and may have significant implications for conformational properties and possibly for protein-protein interactions. Even though further structural studies are needed, it would not be surprising if this change affects also the envelope oligomerization to form the ion channel.

It has been demonstrated that M glycoprotein is more prevalent within the virus membrane and it is deemed to be important for the budding process of the Coronaviruses. Indeed, during the process of virus particle assembly, this protein interacts with the Nucleocapsid, Envelope, Spike and Membrane

glycoprotein itself (J Alsaadi & 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, mutation occurring at the N-terminus region, which is exposed to the virus surface, could probably play a key role in the host cell interaction.

In conclusion, with these analyses we have investigated the potential epizootic origin of the SARS-CoV-2 and the structural similarity of E and M proteins to the counterparts from Pangolin and Bat coronavirus isolates. Although further studies are needed, it is clear that these amino acid variations have been important for the virus evolutionary history and the results suggest 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.

References

Altschul, S., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W., & Lipman, D. J. (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, N. M., Cossic, B., Davies, E., Miller, A. D., & Whittaker, G. R. (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, S., Benvenuto, D., Bianchi, M., Giovanetti, M., Pascarella, S., & Ciccozzi, M. (2020). COVID-2019: The role of the nsp2 and nsp3 in its pathogenesis. *Journal of Medical Virology*. https://doi.org/10.1002/jmv.25719

Benson, D. A., Cavanaugh, M., Clark, K., Karsch-Mizrachi, I., Ostell, J., Pruitt, K. D., & Sayers, E. W. (2018). GenBank. *Nucleic Acids Research*, *46*(D1), D41–D47. https://doi.org/10.1093/nar/gkx1094

Benvenuto, D., Giovanetti, M., Ciccozzi, A., Spoto, S., Angeletti, S., & Ciccozzi, M. (2020). The 2019-new coronavirus epidemic: Evidence for virus evolution. *Journal of Medical Virology*, *92*(4), 455–459. https://doi.org/10.1002/jmv.25688

Benvenuto, D., Giovanetti, M., Salemi, M., Prosperi, M., De Flora, C., Junior Alcantara, L. C., … Ciccozzi, M. (2020). The global spread of 2019-nCoV: a molecular evolutionary analysis. *Pathogens and Global Health*, 1–4. https://doi.org/10.1080/20477724.2020.1725339

Chen, Y., Yu, P., Luo, J., & Jiang, Y. (2003). Secreted protein prediction system combining CJ-SPHMM, TMHMM, and PSORT. *Mammalian Genome : Official Journal of the International Mammalian Genome Society*, *14*(12), 859–865. https://doi.org/10.1007/s00335-003-2296-6

J Alsaadi, E. A., & Jones, I. M. (2019). Membrane binding proteins of coronaviruses. *Future Virology*, *14*(4), 275–286. https://doi.org/10.2217/fvl-2018-0144

Ji, W., Wang, W., Zhao, X., Zai, J., & Li, X. (2020). Cross-species transmission of the newly identified coronavirus 2019-nCoV. *Journal of Medical Virology*, *92*(4), 433–440. https://doi.org/10.1002/jmv.25682

Katoh, K., & Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Molecular Biology and Evolution*, *30*(4), 772–780. https://doi.org/10.1093/molbev/mst010

Lai, C.-C., Shih, T.-P., Ko, W.-C., Tang, H.-J., & Hsueh, P.-R. (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, R., Zhao, X., Li, J., Niu, P., Yang, B., Wu, H., … Tan, W. (2020). Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding. *Lancet (London, England)*, *395*(10224), 565–574. https://doi.org/10.1016/S0140-6736(20)30251-8

McGuffin, L. J., Bryson, K., & Jones, D. T. (2000). The PSIPRED protein structure prediction server. *Bioinformatics (Oxford, England)*, *16*(4), 404–405. https://doi.org/10.1093/bioinformatics/16.4.404

Naskalska, A., Dabrowska, A., Szczepanski, A., Milewska, A., Jasik, K. P., & Pyrc, K. (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, U., Ahrens, C. H., Müller, S., & Wollscheid, B. (2014). Protter: interactive protein feature visualization and integration with experimental proteomic data. *Bioinformatics (Oxford, England)*, *30*(6), 884–886. https://doi.org/10.1093/bioinformatics/btt607

Sakai, Y., Kawachi, K., Terada, Y., Omori, H., Matsuura, Y., & Kamitani, W. (2017). Two-amino acids change in the nsp4 of SARS coronavirus abolishes viral replication. *Virology*, *510*, 165–174. https://doi.org/10.1016/j.virol.2017.07.019

Schoeman, D., & Fielding, B. C. (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, A. C., Park, Y.-J., Tortorici, M. A., Wall, A., McGuire, A. T., & Veesler, D. (2020). Structure,

Function, and Antigenicity of the SARS-CoV-2 Spike Glycoprotein. *Cell*. https://doi.org/10.1016/j.cell.2020.02.058

Waterhouse, A., Bertoni, M., Bienert, S., Studer, G., Tauriello, G., Gumienny, R., … Schwede, T. (2018). SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Research*, *46*(W1), W296–W303. https://doi.org/10.1093/nar/gky427

Waterhouse, A. M., Procter, J. B., Martin, D. M. A., Clamp, M., & Barton, G. J. (2009). Jalview Version 2-A multiple sequence alignment editor and analysis workbench. *Bioinformatics*, *25*(9), 1189–1191. https://doi.org/10.1093/bioinformatics/btp033

Webb, B., & Sali, A. (2017). Protein structure modeling with MODELLER. *Methods in Molecular Biology*, *1654*, 39–54. https://doi.org/10.1007/978-1-4939-7231-9_4

Yang, J., Yan, R., Roy, A., Xu, D., Poisson, J., & Zhang, Y. (2015, January). The I-TASSER Suite: protein structure and function prediction. *Nature Methods*, Vol. 12, pp. 7–8. https://doi.org/10.1038/nmeth.3213

Zhang, L., Lin, D., Sun, X., Curth, U., Drosten, C., Sauerhering, L., … Hilgenfeld, R. (2020). Crystal structure of SARS-CoV-2 main protease provides a basis for design of improved α-ketoamide inhibitors. *Science (New York, N.Y.)*. https://doi.org/10.1126/science.abb3405

Zimmermann, L., Stephens, A., Nam, S.-Z., Rau, D., Kübler, J., Lozajic, M., … Alva, V. (2018). A Completely Reimplemented MPI Bioinformatics Toolkit with a New HHpred Server at its Core. *Journal of Molecular Biology*, *430*(15), 2237–2243. https://doi.org/10.1016/j.jmb.2017.12.007

Table 1

Percentage of identity to the most similar E and M sequences from CoV isolates

Virus isolates	Accession code	Percentage of identity	
		E[a] (YP_009724392)	M[a] (YP_009724393)
Bat coronavirus RaTG13	MN996532	100	99
Bat SARS-like coronavirus isolate bat-SL-CoVZXC21	MG772934	100	98
Bat SARS-like coronavirus isolate bat-SL-CoVZXC45	MG772933	100	98
Pangolin CoV MP789	MT084071	100	98

a) Sars-CoV-2 sequences. GenBank code in parentheses

Figure caption

Figure 1

Multiple sequence alignment among Sars-CoV-2 envelope proteins and a selection of the most similar homologous proteins. The single Sars-CoV-2 sequence is identical to all the isolates. The variant sequence is reported separately. Pangolin sequence is the representative of the identical sequences listed in Table 1. Red lines indicate the variant sites discussed in the text. Alignment blue hue is proportional to column percentage of identity.

Figure 2

Three-dimensional model of the viroporin-like tetrameric assembly of the envelope protein from Sars-CoV-2 represented as cartoon model. Residues corresponding to the mutated sites indicated in Figure 1 are displayed as transparent space filling spheres and labelled. The C-terminal segments are reported for completeness even though they have no conformational meaning for lack of a corresponding segment in the structural template.

Figure 3

Multiple sequence alignment among Sars-CoV-2 Membrane glycoproteins and a selection of the most similar homologous proteins. The single Sars-CoV-2 sequence is identical to all the isolates. The Sars-CoV-2 variant sequences are reported separately. Red box indicates the variant sites at the N-terminal discussed in the text. Red bars under the multiple alignment mark the consensus prediction of transmembrane helices. The location of the connect loop with respect to the virion surface is indicated as "in" or "out". Alignment blue hue is proportional to column percentage of identity.

Figure 4

I-Tasser model of the Membrane glycoprotein represented as cartoon model. Relevant residues are displayed as transparent space filling spheres and labelled.

Figure 1

Figure 2

Figure 3

Figure 4