AMLS Archives of Medical Laboratory Sciences

Review Article

Deep Characterization of SARS-CoV-2: An Overview

Ehsan Ul Haq[1*]

1. Department of Biotechnology, University of Sargodha (UOS), Sargodha, Punjab, Pakistan.
https://orcid.org/0000-0003-2938-8806

Received: January 6, 2021; Accepted: February 6, 2021

Abstract

The coronavirus epidemic has become one of the major health concerns all over the world recently. Like other strains of coronavirus, this strain also spreads through a droplet-based transmission that is the main cause of its worldwide spread. Several trials of antiviral medicines related to the control of the virus have already begun globally but still one of the main problems is the lack of a viable treatment option. An extensive amount of research is still taking place to organize the data associated with genomics and proteomics of its original strain SARS-CoV-2 alongside other mutant strains. This review summarizes the related up-to-date research that is going on the structural organization of the genome and proteome of the virus.

Keywords: SARS-CoV-2; Coronavirus; Structural Organization; Replication; Genomics, Proteomics.

[*]**Corresponding Author:** Ehsan Ul Haq; **Email:** ehsanulhaqsani5@gmail.com; BBTF17E116@uos.edu.pk

Please cite this article as: Ul Haq E. Deep Characterization of SARS-CoV-2: An Overview. Arch Med Lab Sci. 2021;7:1-6 (e2). https://doi.org/10.22037/amls.v7.33616

Introduction

The recent novel coronavirus belongs to the family of viruses called Coronaviridae that belongs to the order Nidovirales. It is an enveloped RNA virus. RNA present in this virus is single-stranded and positive sense. It has also been revealed on extensive phylogenetic analysis that it resembles 96.2% of the strain of coronaviruses found in bats (1). It spreads by respiratory droplets resulted in sneezing, cough and contact that may be direct or indirect. This droplet-based mode of transmission has been recently confirmed by modern researchers (30). It causes severe respiratory syndrome and attacks the same receptor site angiotensin-converting enzyme 2 (ACE2) (2).

The most important and obvious genomic characters of SARS CoV-2 include its binding with the human ACE2 receptor site and the other is its extensive arrangement of outside spike proteins (3). It is the seventh strain of coronavirus from its family that causes infectious diseases in human beings (4). It varies from spherical to pleomorphic forms with a length ranging from 80-160 nm. It comprises four structural proteins, which are Spike (S), Nucleocapsid (N), Envelope (E), and Membrane (M) that resembles both in structure and function with other members of its family. All of these proteins are responsible for a stable shape as well as exceptional virulent characters of the virus (5).

Genome and its organization

The genetic material of SARS CoV-2 is composed up of single-stranded positive-sense RNA having approximately 30,000 bases dispersed in a complex arrangement. The total number of nucleotides that the whole genome consists of is 29891 revealed after rigorous bioinformatics analysis. The percentage abundance of overall GC content is 38%. Its genome encodes 9860 total amino acids by utilizing the basic central dogma pathway. Functional open reading frames (ORFs) are 12 in number that is essential for proper replication of the virus. The sub-genomic mRNAs have a variety of conserved sequences out of which 9 are transcription-regulatory sequences and there are only two terminal UTRs (6). Its genome does not have a haemagglutinin-esterase gene found in other members of beta coronaviruses (7).

ORF1 has conserved about 66% of the total genome of the virus that is essential for the encoding of *pp1a* and *pp1ab* precursor proteins. It also translates 16 more NSP non-structural proteins. Other ORFs encode crucial elementary and other supplementary structural proteins. NSP includes viral cysteine proteases called NSP3 and NSP5, RNA associated RNA polymerase NSP13, and remaining NSPs, which play important functions in transcription and duplication and ultimately the whole replication process (8).

The genome organization of SARS-CoV-2 has been presented in figure 1 (9). The remaining viral genes encode S, M, E, and N proteins. The organization of the genome is 5'-leader-UTR-replicase-Spike-Envelope-Membrane-Nucleocapsid-3'UTR-polyA tail having basic genes that are structurally associated at the 3' end (29).

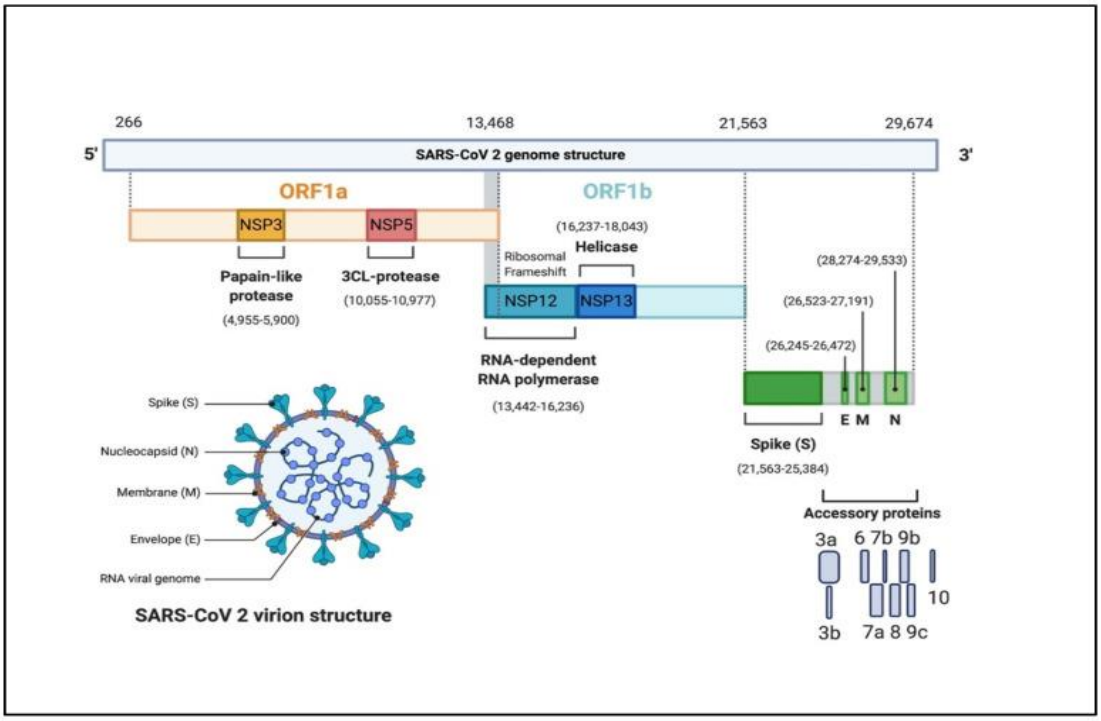

Figure 1. Organization of SARS-CoV-2 genome. Two large genes ORF1a (yellow) and ORF1b (blue). Structural proteins S, E, M, and N (Green). Accessory proteins (Gray Shading) (29).

Proteomics

There are four important structural proteins named S, E, M, and N proteins. They play different roles from shape maintenance to genome replication. All of these proteins are necessary to be present for the proper implication of each function. As we already know that SARS CoV-2 contains 12 functional ORFs in its genome which encodes different crucial proteins to enable the virus for the replication process (6, 10).

Terminal ORFs i.e., 1a and 1b generate a replicase gene, which is responsible for the transcription of *pp1a* and *pp1ab* having a size of 450 kD and 750 kD respectively. NSPs are proteolyzed by viral proteases and it is one of the most important steps in the replication process (11). Papain-like cysteine protease plays a vital role in the cleavage process of *pp1a* and *pp1ab*. Chymotrypsin like protease gives multi-subunits protein known as viral replicase-transcriptase (12). This is classified as 3C-like protease and, shortly written as 3CLpro because of its connection with 3C protease of β-coronaviruses (13). 3CLpro is a basic protease required for coronaviruses associated polyprotein cleavage and it controls the actions of replication complexes that are involved. 3CLpro is also known as the main protease, M-pro. It has become a focus of interest because of its inevitable part in the life cycle of viruses so that antivirals can be generated from the overall analysis of its kinetics (12).

Spike (S) protein is a type 1 membrane glycoprotein and it has two major domains to form its structure. The Ecto-domain is N-terminal linked and cysteine-rich domain is C-terminal linked. Translation of S protein occurs on specific regions of polysomes and then it is pushed into the rough endoplasmic reticulum where glycosylation and transfer to the Golgi complex take place. Viral particles are inserted during this overall movement. Other S particles that are free from the virus are pushed towards the membrane. S protein interacts with the receptor and starts viral entrance into the host cell (14). It comprises about 1273 amino acids lengthwise. It has 13 amino acids as signal peptides at N-terminal. C-terminus has 28 residues as a cytoplasmic tail (15). It is encoded in the form of a lengthy chain that is broken down into two linked subunits S1 and S2. S1 is termed as a marginal part and S2 is known to be a membrane-spanning subunit. Both of them cause cell fusion. They are believed to be involved in the initiation of their infectious process (16). N protein is a structural protein that is responsible for the identification of RNA stretches that ultimately generate signals and helps in the synthesis of ribonucleoprotein (RNP). N protein covers genomic RNA and forms a stable helical structure. The helical nucleocapsid is always transformed into a core having a characteristic icosahedral symmetry. It has 422 basic amino acids that are responsible for the overall basic charge. It has a low homological match-up with other coronaviruses (17). N protein has two major functions which are biochemical and thermodynamic (18). N protein can be easily denatured because it has no high permanency (19).

N-Terminal domain of M protein is always present on the surface of the virus and C-Terminus is open to the inner surface. It has a hydrophobic transmembrane that is comprised up of 12-37 characteristic residues. SMWSFNPE is a specifically conserved sequence present right after the end of hydrophobic transmembrane. It is the most abundant polypeptide present in the virus. Moreover, it is an elementary and basic component in the assembly and morphogenesis of viruses (20). It is also involved in forming the stable shape of the virus. The E proteins in different viruses of coronaviridae have a different structure but there is always some similarity in them (21). E proteins vary from 76-109 amino acids in size. It is the smallest protein in the genome of coronavirus (31). It has also a basic role in the formation of the envelope. E proteins are present in the intracellular membrane having C-terminus protruding to the cytoplasmic region in the target cell. The N-terminus is crusted in the membrane by 2/3 lengthwise (22). Studies show that E forms ion channels in the lipid bilayer of the membrane for monovalent cations selectively (23). It also adjusts cell membrane penetrability that is good for the replication of the virus in host cells. E type of proteins also demonstrates their role in viral replication (24, 32).

NSPs (Non-Structural Proteins)

The 5’-terminal two-third consists of the viral replicase gene, which in turn translates two large polyproteins *pp1a* and *pp1ab*. *Pp1a* shifts the ribosome into -1 upstream of the open reading frame (ORF) 1a where the termination codon is located. Proteolytic cleavage of polyproteins is courtesy of many non-structural viral proteins. NSPs also have a part in the replication and transcription of mRNAs ORF1 downstream process of replication. Proteases like 3CLpro and PL2pro are always represented by NSP5 and NSP3 respectively (10). PL2pro cuts at 3 characteristic sites and 3CLpro cuts polyproteins at 11 sites producing 16 NSPs at the end (24, 32).

Life Cycle Junctures

SARS CoV-2 has an exceptional replication process (Figure 2). It has different stages that are involved in the successful replication of the virus. It infects cells either by endosome or plasma membrane implication. Spike proteins S1 and S2 initiate attachment of the virus to the target cell by appealing the ACE2 site as their receptor (25). Virus after finishing entry tries to activate the spike proteins by signaling Cathepsin L that initiated the fusion process. Viral membrane fuses with the host cell membrane by cellular serine protease TMPRSS2 (25). Because of plasma membrane fusion, the host cell is now deprived of immunity against the virus (26). After the entry and attachment of the virus, it is time to synthesize viral polyproteins that takes the process even further.

Viral genomic RNA translates NSPs which have a critical role in RNA synthesis. *Pp1a* and *pp1ab* are proteolyzed into functional NSPs like helicase and RNA-replicase-transcriptase complex (27).

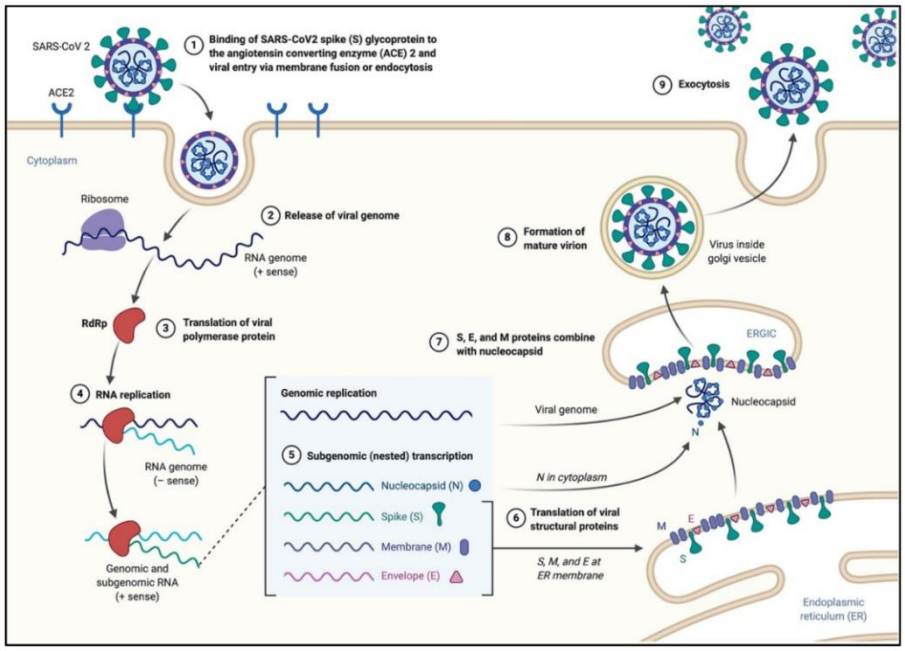

Figure 2. Host cells displaying the life cycle of SARS-CoV-2 (29).

Replication of spike, envelope, membrane, and nucleocapsid proteins is completed by the RdRp complex. They are translated by ribosomes and localized on the reticulum for assembly of the virus. The N protein remains in the cytoplasm for assembly with RNA. Virion precursors are exposed to N proteins and then they are moved through ER and Golgi complex by small vesicles. Now, the viruses inside the cell are in full mature form. All of these mature viruses are expelled out of the damaged cell via the process of exocytosis and again these newly produced viruses begin to search for a new target host cell where it can start the replication process again (28, 29).

Conclusion

All of this data is obtained from the initial researches that have been performed by several different research institutes internationally. However, this is a major issue of health concern around the globe nowadays, and more and more accurate and modern methods of research have been applied by the researchers by utilizing modern technology to construct a more accurate map of proteome and genome of SARS-CoV-2 is still in development. Hopefully, we may see better and more up-to-date research on the organization of the genome and proteome of SARS-CoV-2 in the future.

Conflict of Interest

The authors declared that they have no conflict of interest.

Acknowledgment

I thank Mr. Muhammad Shahzaib (University of Sargodha, Sargodha) for his useful discussion, editing, and proofreading services.

Funding/Support

The authors declared that there is no financial support for this work.

References

1. Paraskevis D, Kostaki EG, Magiorkinis G, Panayiotakopoulos G, Sourvinos G, Tsiodras S. Full-genome evolutionary analysis of the novel coronavirus (2019-nCoV) rejects the hypothesis of emergence as a result of a recent recombination event. Infect Genet Evol. 2020 Apr;79:104212. doi: 10.1016/j.meegid.2020.104212. Epub 2020 Jan 29. PMID: 32004758; PMCID: PMC7106301.

2. Wang Y, Zhou Y, Yang Z, Xia D, Hu Y, Geng S. Clinical Characteristics of Patients with Severe Pneumonia Caused by the SARS-CoV-2 in Wuhan, China. Respiration. 2020;99(8):649-657. doi: 10.1159/000507940. Epub 2020 Aug 25. PMID: 32841948; PMCID: PMC7490495.

3. Walls A, Park Y, Tortorici M, Wall A, McGuire A, Veesler D. Structure, function and antigenicity of the SARS-CoV-2 spike glycoprotein. 2020;.

4. Corman VM, Muth D, Niemeyer D, Drosten C. Hosts and Sources of Endemic Human Coronaviruses. Adv Virus Res. 2018;100:163-188. doi: 10.1016/bs.aivir.2018.01.001. Epub 2018 Feb 16. PMID: 29551135; PMCID: PMC7112090.

5. Phan T. Novel coronavirus: From discovery to clinical diagnostics. Infect Genet Evol. 2020 Apr;79:104211. doi: 10.1016/j.meegid.2020.104211. Epub 2020 Jan 30. PMID: 32007627; PMCID: PMC7129799.

6. Chan JF, Kok KH, Zhu Z, Chu H, To KK, Yuan S, et al. Genomic characterization of the 2019 novel human-pathogenic coronavirus isolated from a patient with atypical pneumonia after visiting Wuhan. Emerg Microbes Infect. 2020 Jan 28;9(1):221-236. doi: 10.1080/22221751.2020.1719902. Erratum in: Emerg Microbes Infect. 2020 Dec;9(1):540. PMID: 31987001; PMCID: PMC7067204.

7. Xia S, Liu M, Wang C, Xu W, Lan Q, Feng S, et al. Inhibition of SARS-CoV-2 (previously 2019-nCoV) infection by a highly potent pan-coronavirus fusion inhibitor targeting its spike protein that harbors a high capacity to mediate membrane fusion. Cell Res. 2020 Apr;30(4):343-355. doi: 10.1038/s41422-020-0305-x. Epub 2020 Mar 30. PMID: 32231345; PMCID: PMC7104723.

8. Fehr AR, Perlman S. Coronaviruses: an overview of their replication and pathogenesis. Methods Mol Biol. 2015;1282:1-23. doi: 10.1007/978-1-4939-2438-7_1. PMID: 25720466; PMCID: PMC4369385.

9. Marra MA, Jones SJ, Astell CR, Holt RA, Brooks-Wilson A, Butterfield YS, et al. The Genome sequence of the SARS-associated coronavirus. Science. 2003 May 30;300(5624):1399-404. doi: 10.1126/science.1085953. Epub 2003 May 1. PMID: 12730501.

10. Thiel V, Ivanov KA, Putics Á, Hertzig T, Schelle B, Bayer S, et al. Mechanisms and enzymes involved in SARS coronavirus genome expression. J Gen Virol. 2003 Sep;84(Pt 9):2305-2315. doi: 10.1099/vir.0.19424-0. PMID: 12917450.

11. Xiong B, Gui CS, Xu XY, Luo C, Chen J, Luo HB, et al. A 3D model of SARS_CoV 3CL proteinase and its inhibitors design by virtual screening. Acta Pharmacol Sin. 2003 Jun;24(6):497-504. PMID: 12791174.

12. Anand K, Ziebuhr J, Wadhwani P, Hilgenfeld R, Mesters JR. Coronavirus Main Proteinase ($3CL^{pro}$) Structure: Basis for Design of Anti-SARS Drugs. Science. 2003 Jun 13;300(5626)(5626):1763-1767. https://doi.org/10.1126/science.1085658

13. Ziebuhr J, Snijder EJ, Gorbalenya AE. Virus-encoded proteinases and proteolytic processing in the Nidovirales. J Gen Virol. 2000 Apr;81(Pt 4):853-79. doi: 10.1099/0022-1317-81-4-853. PMID: 10725411.

14. Han DP, Kim HG, Kim YB, Poon LL, Cho MW. Development of a safe neutralization assay for SARS-CoV and characterization of S-glycoprotein. Virology. 2004 Aug 15;326(1):140-9. doi: 10.1016/j.virol.2004.05.017. PMID: 15262502; PMCID: PMC7127165.

15. Huang Y, Yang C, Xu XF, Xu W, Liu SW. Structural and functional properties of SARS-CoV-2 spike protein: potential antivirus drug development for COVID-19. Acta Pharmacol Sin. 2020 Sep;41(9):1141-1149. doi: 10.1038/s41401-020-0485-4. Epub 2020 Aug 3. PMID: 32747721; PMCID: PMC7396720.

16. Li W, Moore MJ, Vasilieva N, Sui J, Wong SK, Berne MA, et al. Angiotensin-converting enzyme 2 is a functional receptor for the SARS coronavirus. Nature. 2003 Nov 27;426(6965):450-4. doi: 10.1038/nature02145. PMID: 14647384; PMCID: PMC7095016.

17. Rota PA, Oberste MS, Monroe SS, Nix WA, Campagnoli R, Icenogle JP, et al. Characterization of a novel coronavirus associated with severe acute respiratory syndrome. Science. 2003 May 30;300(5624):1394-9. doi: 10.1126/science.1085952. Epub 2003 May 1. PMID: 12730500.

18. Luo H, Ye F, Sun T, Yue L, Peng S, Chen J, et al. In vitro biochemical and thermodynamic characterization of nucleocapsid protein of SARS. Biophys Chem. 2004 Dec 1;112(1):15-25. doi: 10.1016/j.bpc.2004.06.008. PMID: 15501572; PMCID: PMC7116930.

19. Wang Y, Wu X, Wang Y, Li B, Zhou H, Yuan G, et al. Low stability of nucleocapsid protein in SARS virus. Biochemistry. 2004 Aug 31;43(34):11103-8. doi: 10.1021/bi049194b. PMID: 15323569.

20. Huang C, Wei P, Fan K, Liu Y, Lai L. 3C-like proteinase from SARS coronavirus catalyzes substrate hydrolysis by a general base mechanism. Biochemistry. 2004 Apr 20;43(15):4568-74. doi: 10.1021/bi036022q. PMID: 15078103.

21. Arbely E, Khattari Z, Brotons G, Akkawi M, Salditt T, Arkin IT. A highly unusual palindromic transmembrane helical hairpin formed by SARS coronavirus E protein. J Mol Biol. 2004 Aug 13;341(3):769-79. doi: 10.1016/j.jmb.2004.06.044. PMID: 15288785; PMCID: PMC7134595.

22. Surya W, Li Y, Verdià-Bàguena C, Aguilella VM, Torres J. MERS coronavirus envelope protein has a single transmembrane domain that forms pentameric ion channels. Virus Res. 2015 Apr 2;201:61-6. doi: 10.1016/j.virusres.2015.02.023. Epub 2015 Feb 27. PMID: 25733052; PMCID: PMC7114494.

23. Wilson L, McKinlay C, Gage P, Ewart G. SARS coronavirus E protein forms cation-selective ion channels. Virology. 2004 Dec 5;330(1):322-31. doi: 10.1016/j.virol.2004.09.033. PMID: 15527857; PMCID: PMC7111769.

24. Tan YJ, Lim SG, Hong W. Characterization of viral proteins encoded by the SARS-coronavirus genome. Antiviral Res. 2005 Feb;65(2):69-78. doi:

10.1016/j.antiviral.2004.10.001. PMID: 15708633; PMCID: PMC7114173.

25. Hoffmann M, Kleine-Weber H, Schroeder S, Krüger N, Herrler T, Erichsen S, et al. SARS-CoV-2 Cell Entry Depends on ACE2 and TMPRSS2 and Is Blocked by a Clinically Proven Protease Inhibitor. Cell. 2020 Apr 16;181(2):271-280.e8. doi: 10.1016/j.cell.2020.02.052. Epub 2020 Mar 5. PMID: 32142651; PMCID: PMC7102627.

26. Shirato K, Kawase M, Matsuyama S. Wild-type human coronaviruses prefer cell-surface TMPRSS2 to endosomal cathepsins for cell entry. Virology. 2018 Apr;517:9-15. doi: 10.1016/j.virol.2017.11.012. Epub 2017 Dec 6. PMID: 29217279; PMCID: PMC7112029.

27. He J, Hu L, Huang X, Wang C, Zhang Z, Wang Y, et al. Potential of coronavirus 3C-like protease inhibitors for the development of new anti-SARS-CoV-2 drugs: Insights from structures of protease and inhibitors. Int J Antimicrob Agents. 2020 Aug;56(2):106055. doi: 10.1016/j.ijantimicag.2020.106055. Epub 2020 Jun 11. PMID: 32534187; PMCID: PMC7286838.

28. McKimm-Breschkin JL. Influenza neuraminidase inhibitors: antiviral action and mechanisms of resistance. Influenza Other Respir Viruses. 2013 Jan;7 Suppl 1(Suppl 1):25-36. doi: 10.1111/irv.12047. PMID: 23279894; PMCID: PMC4942987.

29. Alanagreh L, Alzoughool F, Atoum M. The Human Coronavirus Disease COVID-19: Its Origin, Characteristics, and Insights into Potential Drugs and Its Mechanisms. Pathogens. 2020 Apr 29;9(5):331. doi: 10.3390/pathogens9050331. PMID: 32365466; PMCID: PMC7280997.

30. Rahman HS, Aziz MS, Hussein RH, Othman HH, Salih Omer SH, Khalid ES, Abdulrahman NA, Amin K, Abdullah R. The transmission modes and sources of COVID-19: A systematic review. International Journal of Surgery Open. 2020;26:125–36. doi: 10.1016/j.ijso.2020.08.017. Epub 2020 Sep 11. PMCID: PMC7484735.

31. Schoeman D, Fielding BC. Coronavirus envelope protein: current knowledge. Virol J. 2019 May 27;16(1):69. doi: 10.1186/s12985-019-1182-0. PMID: 31133031; PMCID: PMC6537279.

32. Katsnelson, A., 2021. What do we know about the novel coronavirus's 29 proteins? (online) Cen.acs.org. Available at: <https://cen.acs.org/biological-chemistry/infectious-disease/know-novel-coronaviruss-29-proteins/98/web/2020/04> (Accessed 3 January 2021).