Identity and Similarity Percentages of SARS-CoV-2 Proteins Can Be Used as Indicators of the Virus Origin

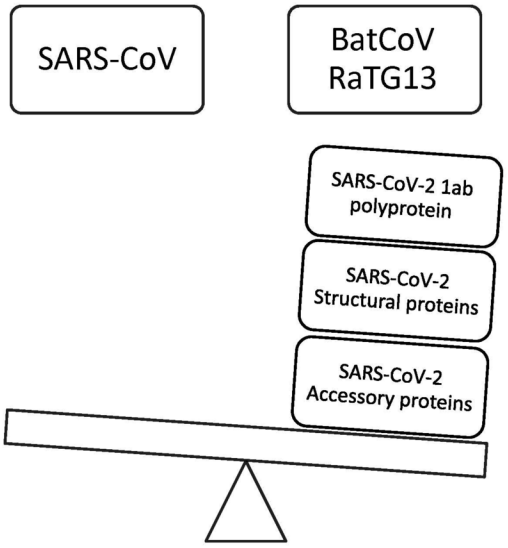
Mohammed Elimam Mohammed[1]

[1]King Khalid University

October 5, 2020

Abstract

There are three types of proteins in the coronaviruses; nonstructural, structural and accessory proteins. The coronaviruses proteins are essential for the viral replication and for the binding, invasion and regulation of the host cells metabolism and immunity. This article investigated the amino acid sequence similarity and identity percentages of 10 proteins present in SARS-CoV-2, SARS-CoV and the Rhinolophus affinis bat coronavirus (BatCoV RaTG13). The investigated proteins were 1ab polyprotein, spike protein, orf3a, envelope protein, membrane protein, orf6, orf7a, orf7b, orf8, and ncleocapsid protein. The online sequence alignment service of The European Molecular Biology Open Software Suite (EMBOSS) was used to determine the similarity and identity percentages of the three viruses proteins. The results showed that the similarity and identity percentages of the SARS-CoV-2 and BatCoV RaTG13 proteins are above 95% while the identity and similarity percentages of the SARS-CoV-2 and SARS-CoV are above 38%. The proteins of the SARS-CoV-2 and the BatCoV RaTG13 are of high identity and similarity percentages compared to those of the SARS-CoV-2 and the SARS-CoV.



The proteins of the SARS-CoV-2 are most identical and similar to those of BatCoV RaTG13 than to the proteins of SARS-CoV

1. Introduction

Coronavirus Disease- 2019 (COVID-19) originated from a sea food market at Wuhan city (the capital of Hubei province at the south east of China) and spread rapidly in more than 200 countries. On 2 Jul 2020, the total confirmed cases reached more than 10.5 million and 512000 deaths. The symptoms of COVID 19 include cough, fever, headache, fatigue, sore throat and malaise. The disease can be complicated to pneumonia and severe acute respiratory syndrome [1-3]. The COVID 19 is transmitted through direct or indirect contact with respiratory droplets and biological samples such as urine, saliva and stool [4]. However, some studies proved the presence of the virus in air samples and one study stated that the virus from air samples is viable for up to three hours [5-8].

Coronavirus 19 is named by the WHO and the International Committee on Taxonomy of Viruses (ICTV) as SARS-CoV-2 and is grouped in the same class of SARS-CoV [9]. The two viruses belong to the family: Coronaviridae; subfamily: Orthocoronavirinae; genus: betacoronavirus; subgenus: sarbecovirus; species: Severe Acute Respiratory Syndrome- related coronavirus. The bat coronavirus (BatCoV RaTG13) was isolated from the bat genus *Rhinolophus affinis*. As the SARS-CoV-2 and SARS-CoV, The bat coronavirus (BatCoV RaTG13) belongs to the beta coronaviruse and it has 96% genome sequence identity compared to the genome of SARS-CoV-2 [10].

The SARS-CoV-2, SARS-CoV, and the bat coronavirus (BatCoV RaTG13) have the same virion structure. They have an RNA with nucleocapsid protein and an envelope. The viral envelope contains a bi-lipid membrane and three proteins: the spike protein, envelope protein and membrane protein [11].

The three viruses contain two major genes: The orf 1ab and 1a (two third) and the structural and accessory proteins gene (one third). The Orf 1ab and 1a genes are translated and hydrolyzed to produced 16 non-structural proteins (nsp1- nsp16) while the translation of the second gene produces the structural proteins spike (S), envelope (E), membrane (M), and nucleocapsid (N); and the accessory proteins orf3a, orf3b, orf6, orf7a, orf7b, orf8a, orf8b, orf9b and orf10. The number and type of the accessory proteins differs according to the virus [10, 12-16].

Regarding the NS3, NS6, NS7a, NS7b and NS8 of the BatCoV RaTG13, some published articles named them as nonstructural proteins [16-17] and others named them as accessory proteins [18,19]. Because these proteins are located in the gene of the structural and accessory proteins, they are considered as accessory proteins and compared to the accessory proteins of SARS-CoV-2.

This article investigated the protein sequence identity and similarity percentages of SARS-CoV-2 compared to the proteins of SARS-CoV and the BatCoV RaTG13.

2. Material and methods

2.1. Study proteins

This article studied the 1ab polyprotein of SARS-CoV-2, SARS-CoV, and BatCoV RaTG13 . Also, the structural and accessory proteins found in SARS-CoV-2 and BatCoV RaTG13 were studied including the spike protein (S), orf3, envelope protein (E), membrane protein (M), orf6, orf7a, orf7b, orf8, and ncucleocapsid protein (N) (Table.2). The amino acid sequences were obtained from the National Center for Biotechnology Information (NCBI) site (https://www.ncbi.nlm.nih.gov/protein) (Table 1).

2.2. Sequence alignment

The online sequence alignment service of The European Molecular Biology Open Software Suite (EMBOSS) was used to determine the protein similarity and identity percentages of the SARS-CoV-2, SARS-CoV, and RaTG13. The matrix of the sequence alignment was EBLOSUM62 and the gap and extend penalties were 14 and 4, respectively. The sequence alignment service of the EMBOSS is available at: https://www.bioinformatics.nl/cgi-bin/emboss/matcher.

3. Results and discussion

3.1. The 1ab polyprotein

The 1ab polyproteins of SARS-CoV-2, SARS-CoV and BatCoV RaTG13 are composed of 7096, 7073, and 7095 amino acids, respectively (Table 1). The amino acid sequence identity and similarity of the 1ab polyprotein of the SARS-CoV-2 and the BatCoV RaTG13 were 98.5% and 99.1%, respectively. The identity and similarity percentages of the 1ab polyprotein of the SARS-CoV-2 and SARS-CoV were 86.2% and 92.9%., respectively. The results show that the SARS-CoV-2 is most probably originated from *Rhinolophus affinis* bat rather than laboratory modified SARS-CoV (Table.2).

After the production of the 1ab polyprotein, some endopeptidases in it produce the 1a polyprotein and 16 nonstructural proteins [20]. The cleavage products of the 1ab polyprotein carryout a wide range of activities associated with the replication of the virus. The activities include: binding and break down of ATP to produce ADP and phosphate, different endopeptidases leading to formation of nonstructural proteins (nonstructural proteins nsp3 and nsp5), production of ribose- 5-phosphate through exonuclease activity, synthesis of new nucleotides by the methyltransferase, RNA polymerase and helicase for the viral replication and removal of super twisting, and regulation of the transcription through zinc finger proteins [20].

3.2. The spike protein

The spike protein of SARS-CoV-2 contains 1273 amino acids while the spike protein of SARS-CoV contains 1255 amino acids and that of the BatCoV RaTG13 contains 1269 amino acids (Table 1). The spike protein of SARS-CoV-2 and SARS-CoV has identity percentage of 76% and similarity percentage of 86% (Table.2). The identity and similarity percentages of the spike proteins of SARS-CoV-2 and RaTG13 are 97.4% and 98.4%, respectively (Table.2). The identity and similarity percentages of the spike proteins of SARS-CoV-2 and RaTG13 are higher than those of SARS-CoV-2 and SARS-CoV.

The spike protein of the coronaviruses consists of three polypeptide chains with two domains; S1 and S2. The S1 and S2 domains are responsible for binding the host cell receptors (S1) and for fusing of the virus with the membrane of the host cell. There is a hinge region between the S1 and S2 which is a target for the host cells proteases [21,22]. The spike protein of SARS-CoV-2 has a furin cleavage site in the hinge region. The furin cleavage site is composed of four amino acids (681- 684). The presence of the furin cleavage site may be responsible for the high transmission rate of SARS-CoV-2 compared to the other coronaviruses [23].

3.3. Orf3a

The accessory protein orf 3a of SARS-CoV-2 contains 275 amino acids and its gene (25393..26220) is located between the spike and the E protein genes. The orf3a proteins of SARS-CoV contains 274 while the NS3 of BatCoV RaTG13 is composed of 275 amino acids (Table.1). The amino sequence alignment of the orf 3a of SARS-CoV-2 and SARS-CoV showed that the sequence identity is 72.4% and the sequence similarity is 85.1%. The similarity percentage of the orf3a of SARS-CoV-2 and SARS-CoV was 90.2 rather than 85.1% depending on the report of Yashimito (2020) [12] which may be due to the different software programs used by the two studies. The orf3a (SARS-CoV-2) and NS3 (BatCoV RaTG13) are characterized by 97.8% identity and 98.9% similarity (Table.2).

The orf 3a plays different roles for the virus including: 1) viral envelope assembling; 2) host cell binding and infusion through interacting with the structural proteins (M, S, and E) and the accessory protein (7a) of SARS-CoV [24]. In the host organisms, the highest immunogenicity of the N-terminal of orf3a is well known with strong protective effect on the humoral immunity [25]. The orf3a ha a cysteine rich domain which possess a potassium ion channel activity through interacting with the S and E protein [24,26]. The C- terminal of the orf3a arrests the host cells cell cycle through depleting the cyclin D3 and it facilitates the apoptosis of the host cells through interacting with the M protein [27-29].

3.4. Envelope protein (E protein)

The E proteins of SARS-CoV-2, SARS-CoV, and BatCoV RaTG13 are consisted of 75, 76, and 75 amino acids, respectively (Table 1). The identity and similarity percentages of the E protein of SARS-CoV-2 and SARS-CoV are 94.7 and 96.1 compared to 94.7 and 97.4 of Yashimito (2020) [12] (Table.2). The E proteins

of the SARS-CoV-2 and the BatCoV RaTG13 are 100% identical and similar. The results strongly favor the bat origin of SARS-CoV-2 over the SARS-CoV origin. The E protein contains three domains; C-terminal, N- terminal, and trans-membrane with different functions for the virus and in the host cells [30].

The E protein plays different functions for the viral replication and for the interaction of the virus with the host organisms and cells such as the assembly of the virion envelope; suppression of the host cells stress responses; facilitation of the viral replication and vitality; and acting as ion channel to induce the release of the virions from the host cells [31-35].

3.5. Membrane protein

The membrane proteins (M) of SARS-CoV-2, SARS-CoV, and BatCoV RaTG13 are composed of 222, 221, and 221 amino acids, respectively (Table.1). The identity and similarity percentages of the amino acid sequence of the M proteins of SARS-CoV-2 and SARS-CoV were 90.5 and 96.4 while those of the M protein of SARS-CoV-2 and BatCoV RaTG13 were 99.5% each (Table.2). The M protein has three domains; N-terminal, C-terminal, and Trans-membrane with different functions [36].

The M protein IS important for the assembly, transport, and release of the virus from the host cell organelles [37,38]. The M protein of SARS-CoV inhibits the transcription of interferon-1 which leads to the inhibition of the innate immunity of host organisms [39].

3.6. Orf6

The orf6 of the SARS-CoV-2 and the BatCoV RaTG13 contain 61 amino acids while it contains 63 amino acids in the SARS-CoV (Table 1). The orf6 proteins of the SARS-CoV-2 and SARS-CoV are characterized by identity percentage of 68.9% and similarity percentage of 88.5%. The identity and similarity percentages of the orf6 of the SARS-CoV-2 and BatCoV RaTG13 were 100% each (Table.2).

The functions of the orf6 include: 1) it participates the formation of replication\transcription to facilitate the viral replication, 2) it increases the number of the virions during the infection, 3) it contributes to the evasion of the virus to the host immune system and 4) it is involved in the formation of the Double Membrane Vesicle (DMV) in the host cells to ensure the virus assembling [40-42].

3.7. Orf7a

The number of amino acids of the orf7a of the SARS-CoV-2 and BatCoV RaTG13 are 121 amino acids while the orf7a of the SARS-CoV contains 122 amino acids (Table.1). The identity and similarity percentages of the orf7a of the SARS-CoV-2 and the SARS-CoV are 85.2 and 90.2 respectively. The orf7a of the SARS-CoV-2 and the BatCoV RaTG13 are with identity percentage of 97.5% and similarity percentage of 99.2% (Table.2).

The orf 7a of SARS-CoV is a trans-membrane protein divided into four regions from the N-terminal: 1) the first 15 amino acids is broken down by the infected host cells; 2) the amino acids 16-96 forms the intracellular domain; 3) the amino acids 97- 117 amino acids are transmembrane with hydrophobic nature; and 4) the C-terminal consists of the last five amino acids [43].

The orf7a plays a role in the binding and invasion of the virus to the host cells through interacting with the S, M, E, and the orf3a proteins [41,44]. The orf7a does not contribute for the replication of the virus [43-46]. The orf7a plays some functions in the host cells such as triggering of the apoptosis, down-regulation of protein synthesis, arrest of the cell cycle at the G0\G1 phase, and activation of the cytokine production [41,43,44,46].

3.8. Orf7b

The orf7b of both SARS-CoV-2 and BatCoV RaTG13 contains 43 amino acids while the orf7b of the SARS-CoV contains 44 amino acids (Table.1). The orf7b proteins of SARS-CoV-2 and SARS-CoV are characterized by identity percentage of 85.4 and similarity percentage of 90.2. On the other hand, the identity and similarity percentages of the orf7b proteins of SARS-CoV-2 and BatCoV RaTG13 are 97.7% each (Table.2).

The orf7b contains three domains; N-terminal domain (to the outside), C-terminal domain (to the cytoplasm), and a trans-membrane hydrophobic domain [43].

It is reported that the orf7b is not involved in the virus replication [43-46]. The anti orf7b antibody concentration increases in SARS-CoV patients which shows that it is highly immunogenic and it can be used in vaccination trials [46,47].

3.9. Orf8

The identity and similarity or orf8 of the SARS-CoV-2 and orf8a of the SARS-CoV are 38.9 and 77.8 respectively. The orf8 of SARS-CoV-2 is 44.4% identical and 66.7% similar to the orf8b of SARS-CoV [Fig.10]. The identity and similarity percentage of orf8 of the SARS-CoV-2 and the BatCoV RaTG13 are 95% and 95.9%, respectively (Table.2). However, the number of amino acids of the orf8 of SARS-CoV-2 is 121 amino acids compared to 39 amino acids for the orf8a of the SARS-CoV, 84 amino acids for the orf8b of the SARS-CoV, and 121 amino acids for the orf8 of the BatCoV RaTG13 (Table.1).

The orf8a and orf8b of SARS-CoV are not needed for viral replication. In the host cells, they are localized in vesicle like structures in the mitochondria, endoplasmic reticulum, cytosol and nucleus of host cells. The orf8a and orf8b of SARS-CoV stimulates cellular DNA synthesis and caspase-dependent apoptosis [48].

3.10. Nucleocapsid protein (N protein)

The N proteins of the SARS-CoV-2 and the BatCoV RatG13 contain 419 amino acids each while that of SARS-CoV contains 422 amino acids (Table.1). N proteins of the SARS-CoV-2 and the SARS-CoV are 90.5% identical and 94.3% similar while those of the SARS-CoV-2 and the BatCoV RaTG13 are 99% identical and similar (Table.2). The N- protein is an RNA binding protein with three domains: N- terminal domain that binds the RNA, C- terminal domain responsible for dimerization, and a disordered central region rich in serine and arginine (SR) [49].

The N protein is essential for the formation of the helical viral RNA, induction of the replication and transcription of the virus, and control of the host cells metabolism so as to ensure the viral replication process and to regulate the host cell apoptosis and cell cycle [49-51]. Moreover, the N protein is very immunogenic and it induces the host immune system to respond against the SARS-CoV [52].

Conclusion

The SARS-CoV-2 proteins compared to the BatCoV RaTG13 are of high identity and similarity percentages compared to those of the SARS-CoV. The findings of this study proved the usefulness of the protein identity and similarity in suggesting the origin of viruses.

Conflict of interest

The author declares no conflict of interest.

References

1- WHO. (2020). WHO Coronavirus Disease (COVID-19) Dashboard. Available at: https://covid19.who.int/. Accessed on 3 Jul 2020.

2- Ahmad, S., Hafeez, A., Siddqui, S.A., Ahmad, M., and Mishra, S., A Review of COVID-19 (Coronavirus Disease-2019)Diagnosis, Treatments and Prevention., EJMO., 2020, vol. 4 (2), pp. 116–125.

3- Velavan, T.P., and Meyer, C.G., The COVID-19 epidemic., Trop Med Int Health., 2020, vol. 25 (3), pp. 278–280.

4- Shereen, M.A., and Khan, S., COVID-19 infection: origin, transmission, and characteristics of human coronaviruses., J. Adv. Res., 2020, vol. 24, pp. 91–98.

5- Cheng, V.C.C., Wong, S.C., Chen, J.H.K., Yip, C.C.Y., Chuang, V.W.M., Tsang, O.T.Y., Sridhar, S., Chan, J.F.W., Ho, P., and Yuen, K., Escalating infection control response to the rapidly evolving epidemi-

ology of the coronavirus disease 2019 (COVID-19) due to SARS-CoV-2 in Hong Kong., Infect Control Hosp Epidemiol., 2020, vol. 41 (5), pp. 493–498.

6- Ong, S.W.X., Tan, Y.K., Chia, PY., Lee T.H., Ng, O.T., Wong, M.S.Y., and Marimuthu, K., Air, Surface Environmental, and Personal Protective Equipment Contamination by Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) From a Symptomatic Patient., JAMA., 2020, vol. 323 (16), pp. 1610–1612.

7- Liu, Y., Zhi, Z., Chen, Y., Guo, M., Liu, Y., Gali, N.K., Sun, L., Duan, Y., Cai, J., Westerdahl, D., Liu, X., Xu, K., Ho, K., Kan, H., Fu, Q., and Lan, K., Aerodynamic analysis of SARS-CoV-2 in two Wuhan hospitals., Nature., 2020, vol. 582, pp. 557–560.

8- van Doremalen, N., Bushmaker, T., Morris, D.H., Holbrook, M.G., Gamble, A., Williamson, B.N., Tamin, A., Harcourt, J.L., Thornburg, N.J., Gerber, S.I., Lloyd-Smith, J.O., de Wit, E., and Munster, V.J., Aerosol and Surface Stability of SARS-CoV-2 as Compared with SARS-CoV-1., N Engl J Med., 2020, vol. 382, pp. 1564–1567.

9- International Committee on Taxonomy of Viruses (ICTV). March 2020. ICTV 2019 Master Species List (MSL35). Available at: https://talk.ictvonline.org/files/master-species-lists/m/msl/9601.

10- Zhou, P., Yang, X.L., Wang, X.G., Hu, B., Zhang, L., Zhang, W., Si, H.R., Zhu, Y., Li, B., Huang, C.L., Chen, H.D., Chen, J., Luo, Y., Guo, H., Jiang, R.D., Liu, M.Q., Chen, Y., Shen, X.R., Wang, X., Zheng, X.S., Zhao, K., Chen, Q.J., Deng, F., Liu, L.L., Yan, B., Zhan, F.X., Wang, Y.Y., Xiao, G.F., and Shi, Z.L., A pneumonia outbreak associated with a new coronavirus of probable bat origin. Nature., 2020, vol. 579, pp. 270–273.

11- Perlman, S., and Netland, J., Coronaviruses post-SARS: update on replication and pathogenesis., Nat Rev Microbiol., 2009, vol. 7, pp. 439–450.

12- Yoshimoto, F.K., The Proteins of Severe Acute Respiratory Syndrome Coronavirus-2 (SARS CoV-2 or n-COV19), the Cause of COVID-19., Protein J., 2020, vol.39, pp. 198–216.

13- Wang, C., Liu, Z., Chen, Z., Huang, X., Xu, M., He, T., and Zhang, Z., The establishment of reference sequence for SARS-CoV-2 and variation analysis., J Med Virol., 2020, vol. 92, pp. 667–674.

14- Khailany, R.A., Safdar, M., and Ozaslan, M., Genomic characterization of a novel SARS-CoV-2., Gene Rep., 2020, vol. 19, 100682.

15- Wong, G., Bi, Y.H., Wang, Q.H., Chen, X.W., Zhang, Z.G., and Yao, Y.G., Zoonotic origins of human coronavirus 2019 (HCoV-19 / SARS-CoV-2): why is this work important?., Zool Res., 2020, vol. 41 (3), pp. 213-219.

16- GenBank: MN996532.1. Bat coronavirus RaTG13, complete genome. Available at: https://www.ncbi.nlm.nih.gov/nuccore/MN996532. Last updated on 24 March 2020. Accessed on 4 Jul 2020.

17- Fahmi, M., and Kubota, Y., Ito M., Nonstructural proteins NS7b and NS8 are likely to be phylogenetically associated with evolution of 2019-nCoV., Infect Genet Evol., 2020, vol. 81, 104272.

18- Tang, X., Wu, C., Li, X., Song, Y., Yao, X., Wu, X., Duan, Y., Zhang, H., Wang, Y., Qian, Z., Cui, J., and Lu, J., On the origin and continuing evolution of SARS-CoV-2., NATL SCI REV., 2020, vol. 7 (6), pp. 1012–1023.

19- Li, Y., Yang, X., Wang, N., Wang, H., Yin, B., Yang, X., and Jiang, W., The divergence between SARS-CoV-2 and RaTG13 might be overestimated due to the extensive RNA modification., Future Virol., 2020. https://doi.org/10.2217/fvl-2020-0066.

20- Snijder, E.J., Decroly, E., and Ziebuhr, J., The Nonstructural Proteins Directing Coronavirus RNA Synthesis and Processing., Adv Virus Res., 2016, vol. 96, pp. 59–126.

21- Li, F., Structure, Function, and Evolution of Coronavirus Spike Proteins., Annu Rev Virol., 2016, vol. 3 (1), pp. 237–261.

22- Bosch, B.J., van der Zee, R., de Haan, C.A., and Rottier, P.J., The coronavirus spike protein is a class I virus fusion protein: structural and functional characterization of the fusion core complex., J Virol., 2003, vol. 77 (16), pp. 8801–8811.

23- Walls, A.C., Park, Y., Tortorici, M.A., Wall, A., McGuire, A.T., and Veesler, D., Structure, Function, and Antigenicity of the SARS-CoV-2 Spike Glycoprotein., Cell., 2020, vol. 181 (2), pp. 281–292.

24- von Brunn, A., Teepe, C., Simpson, J.C., Pepperkok, R., Friedel, C.C., Zimmer, R., Roberts, R., Baric, R., and Haas, J., Analysis of intraviral protein-protein interactions of the SARS coronavirus ORFeome., PLoS One., 2007, vol. 2 (5), e459.

25- Zhong, X., Guo, Z., Yang, H., Peng, L., Xie, Y., Wong, T., Lai, S., and Guo, Z., Amino terminus of the SARS coronavirus protein 3a elicits strong, potentially protective humoral responses in infected patients., J. Gen. Virol., 2006, vol. 87, pp. 369–373.

26- Zeng, R., Yang, R.F., Shi, M.D., Jiang, M., Xie, Y., Ruan, H., Jiang, X., Shi, L., Zhou, H., Zhang, L., Wu, X., Lin, Y., Ji, Y., Xiong, L., Jin, Y., Dai, E., Wang, X., Si, B., Wang, J., Wang, H., Wang, C., Gan, Y., Li, Y., Cao, J., Zuo, J., Shan, S., Xie, E., Chen, S., Jiang, Z., Zhang, X., Wang, Y., Pei, G., Sun, B., and Wu, J., Characterization of the 3a protein of SARS-associated coronavirus in infected vero E6 cells and SARS patients., J Mol Biol., 2004, vol. 341 (1), pp. 271–279.

27- Yuan, X., Yao, Z., Wu, J., Zhou, Y., Shan, Y., Dong, B., Zhao, Z., Hua, P., Chen, J., and Cong, Y., G1 phase cell cycle arrest induced by SARS-CoV 3a protein via the cyclin D3/pRb pathway., Am J Respir Cell Mol Biol., 2007, vol. 37 (1), pp. 9–19.

28- Marra, M.A., Jones, S.J., Astell, C.R., Holt, R.A., Brooks-Wilson, A., Butterfield, Y.S.N., Khattra, J., Asano, J.K., Barber, S.A., Chan, S,Y., Cloutier, A., Coughlin, S.M., Freeman, D., Girn, N., Griffith, O.L., Leach, S.R., Mayo, M., McDonald, H., Montgomery, S.B., Pandoh, P.K., Petrescu, A.S., Robertson, A.G., Schein, J.E., Siddiqui, A., Smailus, D.E., Stott, J.M., Yang, G.S., Plummer, F., Andonov, A., Artsob, H., Bastien, N., Bernard, K., Booth, T.F., Bowness, D., Czub, M., Drebot, M., Fernando, L., Flick, R., Garbutt, M., Gray, M., Grolla, A., Jones, S., Feldmann, H., Meyers, A., Kabani, A., Li, Y., Normand, S., Stroher, U., Tipples, G.A., Tyler, S., Vogrig, R., Ward, D., Watson, B., Brunham, R.C., Krajden, M., Petric, M., Skowronski, D.M., Upton, C., and Roper, R.L., The Genome sequence of the SARS-associated coronavirus., Science., 2003, vol. 300 (5624), pp. 1399–1404.

29- Law, P.T.W., Wong, C.H., Au, T.C.C., Chuck, C., Kong, S., Chan, P.K.S., To, K., Lo, A.W.I., Chan, J.Y.W., Suen, Y., Chan, H.Y.E., Fung, K., Waye, M.M.Y., Sung, J.J.Y., Lo, Y.M.D., and Tsui, S.K.W., The 3a protein of severe acute respiratory syndrome-associated coronavirus induces apoptosis in Vero E6 cells., J Gen Virol., 2005, vol. 86 (Pt 7), pp. 1921–1930.

30- Schoeman, D., and Fielding, B.C., Coronavirus envelope protein: current Knowledge., Virol. J., 2019, vol. 16, 69.

31- Nieto-Torres, J.L., DeDiego, M.L., Álvarez, E., Jiménez-Guardeño, J.M., Regla-Nava, J.A., Llorente, M., Kremer, L., Shuo, S., and Enjuanes, L., Subcellular location and topology of severe acute respiratory syndrome coronavirus envelope protein., Virology., 2011, vol. 415 (2), pp. 69–82.

32- Álvarez, E., DeDiego, M.L., Nieto-Torres, J.L., Jiménez-Guardeño, J.M., Marcos-Villar, L., and Enjuanes, L., The envelope protein of severe acute respiratory syndrome coronavirus interacts with the non-structural protein 3 and is ubiquitinated., Virology., 2010, vol. 402 (2), pp. 281–291.

33- Corse, E., and Machamer, C.E., The cytoplasmic tails of infectious bronchitis virus E and M proteins mediate their interaction., Virology., 2003, vol. 312 (1), pp. 25–34.

34- Yuan, Q., Liao, Y., Torres, J., Tam, J.P., and Liu, D.X., Biochemical evidence for the presence of mixed membrane topologies of the severe acute respiratory syndrome coronavirus envelope protein expressed in mammalian cells., FEBS Lett., 2006, vol. 580 (13), pp. 3192–200.

35- Ruch, T.R., and Machamer, C.E., The coronavirus E protein: assembly and beyond., Viruses., 2012, vol. 4 (3), pp. 363–382.

36- Neuman, B.W., Kiss, G., Kunding, A.H., Bhella, D., Baksh, M.F., Connelly, S., Droese, B., Klaus, J.P., Shinji Makino, S., Sawicki, S.G., Siddell, S,G., Dimitrios G.Stamou, D.G., Wilson, I.A., Kuhn, P., Buchmeier, M.J., A structural analysis of M protein in coronavirus assembly and morphology., J Struct Biol., 2010, VOL. 174, PP. 11–22.

37- Ma, H., Fang, C., Hsieh, Y., Chen, S., Li, H., and Lo, S., Expression and membrane integration of SARS-CoV M protein., J Biomed Sci., 2008, vol. 15, pp. 301–310.

38- Siu, Y.L., Teoh, K.T., Lo, J., Chan, C.M., Kien, F., Escriou, N., Tsao, S.W., Nicholls, J.M., Altmeyer, R., Peiris, J.S.M., Bruzzone, R., and Nal, B., The M, E, and N Structural Proteins of the Severe Acute Respiratory Syndrome Coronavirus Are Required for Efficient Assembly, Trafficking, and Release of Virus-Like Particles., J Virol., 2008, vol. 82 (22), pp. 11318–11330.

39- Siu, K., Kok, K., Ng, M.J., Poon, V.K.M., Yuen, K., Zheng, B., and Jin, D., Severe Acute Respiratory Syndrome Coronavirus M Protein Inhibits Type I Interferon Production by Impeding the Formation of TRAF3*TANK*TBK1/IKK Complex., J. Biol. Chem., 2009, vol. 284, pp. 16202–16209.

40- Kumar, P., Gunalan, V., Liu, B., Chow, V.T.K., Druce, J., Birch, C., Catton, M., Fielding, B.C., Tan, Y.J., and Lal, S.K., The nonstructural protein 8 (nsp8) of the SARS coronavirus interacts with its ORF6 accessory protein., Virology., 2007, vol. 366 (2), pp. 293–303.

41- Narayanan, K., Huang, C., and Makino, S., SARS coronavirus accessory proteins., Virus Res., 2008, vol. 133, pp. 113–121.

42- Gunalan, V., Mirazimi, A., and Tan, Y., A putative diacidic motif in the SARS-CoV ORF6 protein influences its subcellular localization and suppression of expression of co-transfected expression constructs., BMC Res Notes., 2011, vol. 4, 446. https://doi.org/10.1186/1756-0500-4-446.

43- Liu, D.X., Fung, T.S., Chong, K.K., Shukla, A., and Hilgenfeld, R., Accessory proteins of SARS-CoV and other coronaviruses., Antiviral Res., 2014, vol. 109, pp. 97–109.

44- Tan, Y.J., Lim, S.G., and Hong, W., Understanding the accessory viral proteins unique to the severe acute respiratory syndrome (SARS) coronavirus., Antiviral Res., 2006, vol. 72 (2), pp. 78–88.

45- Yount, B., Roberts, R.S., Sims, A.C., Deming, D., Frieman, M.B., Sparks, J., Denison, M.R., Davis, N., and Baric, R.S., Severe acute respiratory syndrome coronavirus group-specific open reading frames encode nonessential functions for replication in cell cultures and mice., J Virol., 2005, vol. 79 (23), pp. 14909–14922.

46- Schaecher, S.R., Touchette, E., Schriewer, J., Buller, R.M., and Pekosz, A., Severe acute respiratory syndrome coronavirus gene 7 products contribute to virus-induced apoptosis., J Virol., 2007, vol. 81 (20), pp. 11054–11068.

47- Guo, J.P., Petric, M., Campbell, W., and McGeer, P.L., SARS corona virus peptides recognized by antibodies in the sera of convalescent cases., Virology., 2004, vol. 324 (2), pp. 251–256.

48- Keng, C.T., and Tan, Y.J., Molecular and Biochemical Characterization of the SARS-CoV Accessory Proteins ORF8a, ORF8b and ORF8ab., Molecular Biology of the SARS-Coronavirus., 2009, pp. 177–191.

49- Kang, S., Yang, M., Hong, Z., Zhang, L., Huang, Z., Chen, X., He, S., Zhou, Z., Zhou, Z., Chen, Q., Yan, Y., Zhang, C., Shan, H., and Chen, S., Crystal structure of SARS-CoV-2 nucleocapsid protein RNA binding domain reveals potential unique drug targeting sites., Acta Pharmaceutica Sinica B., 2020, https://doi.org/10.1016/j.apsb.2020.04.009.

50- Cong, Y.Y., Ulasli, M., Schepers, H., Mauthe, M., V'kovski, P., Kriegenburg, F., Thiel, V., de Haan, C.A.M., and Reggiori, F., Nucleocapsid protein recruitment to replication-transcription complexes plays a crucial role in corona viral life cycle., J Virol., 2020, vol. 94 (4), e01925-19.

51- Surjit, M., Liu, B., Chow, V.T., and Lal, SK., The nucleocapsid protein of severe acute respiratory syndrome-coronavirus inhibits the activity of cyclin-cyclin-dependent kinase complex and blocks S phase progression in mammalian cells., J Biol Chem., 2006, vol. 281, pp. 10669–10681.

52- Lin, Y., Shen, X., Yang, R.F., Li, Y.X., Ji, Y.Y., He, Y.Y., Shi, M.D., Lu, W., Shi, T.L., Wang, J., Wang, H.X., Jiang, H.L., Shen, J.H., Xie, Y.H., Wang, Y., Pei, G., Shen, B.F., Wu, J.R., and Sun, B., Identification of an epitope of SARS-coronavirus nucleocapsid protein., Cell Res., 2003, vol. 13, pp. 141–145.

Table 1: The studied proteins of the three viruses

RaTG13	SARS-CoV	SARS-CoV-2	Protein	Protein	
QHR63299.1	NP_828849.6	YP_-009724389.1	NCBI Code	1ab polyprotein	1
251..21537	265..21485	266..21555	Gene location		
7095	7073	7096	Amino acid number		
QHR63300.2	NP_828851.1	YP_-009724390.1	NCBI Code	S protein	2
21545..25354	21492..25259	21492..25259	Gene location		
1269	1255	1273	Amino acid number		
QHR63301.1	NP_828852.2	YP_-009724391.1	NCBI Code	Orf3	3
25363..26190	25268..26092	25393..26220	Gene location		
275	274	275	Amino acid number		
QHR63302.1	NP_828854.1	YP_-009724392.1	NCBI Code	E protein	4
26215..26442	26117..26347	26245..26472	Gene location		
75	76	75	Amino acid number		
QHR63303.1	NP_828855.1	YP_-009724393.1	NCBI Code	M protein	5
26493..27158	26398..27063	26523..27191	Gene location		
221	221	222	Amino acid number		
QHR63304.1	NP_828856.1	YP_-009724394.1	NCBI Code	Orf6	6
27169..27354	26913..27265	27202..27387	Gene location		
61	63	61	Amino acid number		
QHR63305.1	NP_828857.1	YP_-009724395.1	NCBI Code	Orf7a	7
27360..27725	27273..27641	27394..27759	Gene location		
121	122	121	Amino acid number		
QHR63306.1	NP_849175.1	YP_-009725318.1	NCBI Code	Orf7b	8
27722..27853	27638..27772	27756..27887	Gene location		

RaTG13	SARS-CoV	SARS-CoV-2	Protein	Protein	
43	44	43	Amino acid number		
QHR63307.1	NP_849176.1 NP_849177.1	YP_009724396.1	NCBI Code	Orf8	9
27860..28225	27779..27898, 27864..28118	27894..28259	Gene location		
121	39, 84	121	Amino acid number		
QHR63308.1	NP_828858.1	YP_-009724397.2	NCBI Code	N protein	10
28240..29499	28120..29388	28274..29533	Gene location		
419	422	419	Amino acid number		

The proteins and their amino acids numbers and sequences were obtained from the National Center for Biotechnology Information (NCBI) site available at: https://www.ncbi.nlm.nih.gov/protein.

Table 2. The identity and similarity percentages of the SARS-CoV-2 proteins compared to the SARS-CoV and the bat coronavirus (RaTG13)

Similarity%	Identity%	Protein	Protein	
92.9	86.2	SARS-CoV-2 and SARS-CoV	1 ab polyprotein	1
99.1	98.5	SARS-CoV-2 and RaTG13		
86	76	SARS-CoV-2 and SARS-CoV	Spike protein	2
98.4	97.4	SARS-CoV-2 and RaTG13		
85.1	72.4	SARS-CoV-2 and SARS-CoV	Orf3a (NS3)	3
98.9	97.8	SARS-CoV-2 and RaTG13		
96.1	94.7	SARS-CoV-2 and SARS-CoV	E protein	4
100	100	SARS-CoV-2 and RaTG13		
96.4	90.5	SARS-CoV-2 and SARS-CoV	M protein	5
99.5	99.5	SARS-CoV-2 and RaTG13		
88.5	68.9	SARS-CoV-2 and SARS-CoV	Orf6 (NS6)	6
100	100	SARS-CoV-2 and RaTG13		
90.2	85.2	SARS-CoV-2 and SARS-CoV	Orf7a (NS7a)	7
99.2	97.5	SARS-CoV-2 and RaTG13		

Similarity%	Identity%	Protein	Protein	
90.2	85.4	SARS-CoV-2 and SARS-CoV	Orf7b (NS7b)	8
97.7	97.7	SARS-CoV-2 and RaTG13		
77.8 66.7	38.9 44.4	SARS-CoV-2 and SARS-CoV	Orf8 (NS8)	9
95.9	95	SARS-CoV-2 and RaTG13		
94.3	90.5	SARS-CoV-2 and SARS-CoV	N protein	10