2020

Protein Structural Comparison between COVID-19 and other Coronaviruses

Vanessa Colmenares
Seton Hall University

Seton Hall Biochemistry II Annals of Student Reviews
Vol 1. No. 1
2020

Protein Structural Comparison between COVID-19 and other Coronaviruses

Author: Vanessa Colmenares
Reviewer/Editor: Dr. Gregory Wiedman

Introduction

The recent emergence of the novel coronavirus disease of 2019 (COVID-19) in Wuhan, China and its rapid spread across the world has posed a global health emergency. Scientists and researchers across the world are working together to improve the methods of detection and generate a vaccine that will cure the disease. However, in order to develop therapeutic strategies and techniques for the treatment of COVID-19, the structure of SARS-CoV-2 must be entirely scrutinized and comprehended. This review illustrates the similarities and differences between SARS-CoV-2 and other coronaviruses, while also indicating why the development of this lethal viral RNA can only be a result of natural evolution.

1. Spike Protein Structure

Coronaviruses (CoVs) are large enveloped viruses with a positive single-stranded RNA genome. They belong to the subfamily *Coronavirinae* within the family *Coronaviridae*, which is part of the Nidovirales order, and can be classified into four genera: *Alphacoronavirus*, *Betacoronavirus*, *Gammacoronavirus*, and *Deltacoronavirus*. Among all CoVs identified so far, seven have the ability to infect humans, including human coronavirus 229E (HCoV-229E) and human coronavirus NL63 (HCoV-NL63), which belong to alpha-CoVs[1], as well as human coronavirus OC43 (HCoV-OC43), human coronavirus HKU1 (HCoV-HKU1), severe acute respiratory syndrome coronavirus (SARS-CoV), Middle East respiratory syndrome coronavirus (MERS-CoV), and the newly emerged coronavirus (2019-nCoV), which are known to be beta-CoVs and are the focus of this study (Figure 1)[2].

Coronaviruses have the largest genome among all RNA viruses, typically ranging from 27 to 32 kb[3]. The genetic material of the virus is packed inside a helical capsid formed by the nucleocapsid protein (N) and further enclosed by an envelope that contains: the membrane protein (M) and the envelope protein (E), which are essential for cellular assembly, and the spike protein (S), which mediates virus entry into the host cells and therefore is the viral protein of interest of most scientific studies[3]. As explained in the literature, S is a viral fusion protein that promotes host attachment and synthesis of the membranes during entry[4]. Due to these qualities, S determines host range and cell tropism which differs significantly among coronaviruses and serves as the main point for comparison among the viruses. The S protein has been intensely scrutinized by scientists and researchers from all over the globe as it is the distinguishing feature that can be used to characterize coronaviruses; therefore, it will be a point of focus.

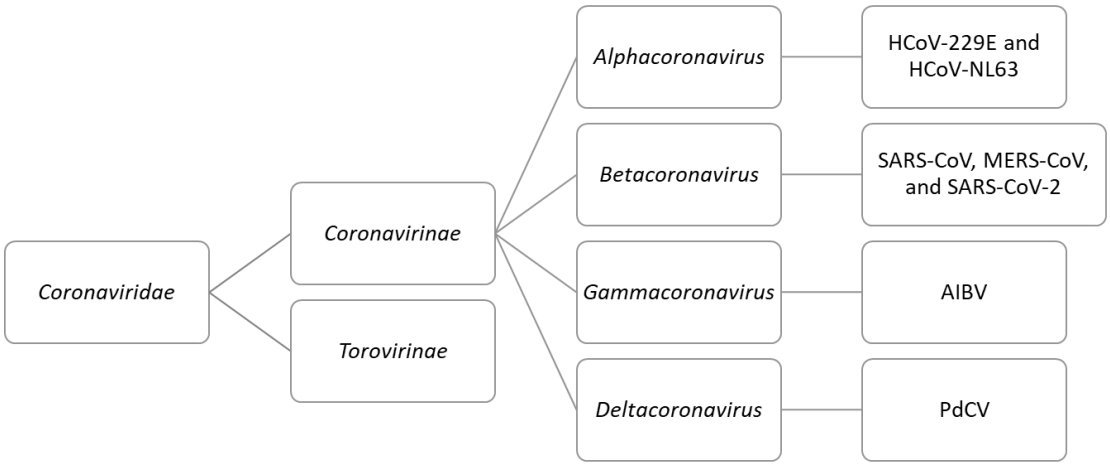

Figure 1. Classification of Coronaviruses. Representative coronaviruses in each genus are HCoV-229E and HCoV-NL63 for the genus Alphacoronavirus. SARS-CoV, MERS-CoV, and SARS-CoV-2 for the genus Betacoronavirus. AIBV for the genus Gammacoronavirus. PdCV for the genus Deltacoronavirus.

According to a study conducted in 2016, the coronavirus spike protein contains three segments: an ectodomain with a receptor-binding subunit S1 and a membrane-fusion subunit S2, a single-pass transmembrane anchor, and a short intracellular tail[5]. Electron microscopy experiments done by this same study revealed that the spike is a clove-shaped trimer with three S1 units and a trimeric S2 unit[5]. When entering the host cell, S1 of the virus binds to a receptor on the host cell surface for viral attachment, and the S2 unit assists with fusion between the host and viral membranes, allowing the viral genetic material to penetrate and enter host cells[3]. However, this binding interaction differs among the multiple coronaviruses. According to a recent study, the receptor-binding S protein of SARS-CoV-2 encoded by the *S* gene was highly divergent from other CoVs, with less than 75% nucleotide sequence identity to all SARSr-CoVs, except for a 93.1% nucleotide identity to RaTG13[6]. As portrayed in the results, the S genes of 2019-nCoV and RaTG13 are longer than the other SARSr-CoVs. The three short insertions in the N-terminal domain and differences in four out of the five key residues in the receptor binding motif (RBM) region is what distinguishes the S gene of SARS-CoV-2 from the sequence of SARS-CoV and makes the novel COVID-19 exceptional[6].

2. Receptor Binding Domains and Receptor Binding Motifs

Although the receptor binding domains (RBDs) among the *Betacoronaviruses* are very similar, their receptor binding motifs (RBMs) differ. This explains why coronaviruses identify different receptors and attack their targeted cells uniquely. Similar to HCoV-NL63[2], SARS-CoV and SARS-CoV-2 recognize the angiotensin-converting enzyme 2 (ACE2) as the receptor on the host cell through the RBD (CTP) region of its S1 unit[7], while MERS-CoV utilizes dipeptidyl peptidase 4 (DPP4) as a receptor through the RBD (CTD) region[8]. Not all coronaviruses attach to the host cells through the same receptors; it has been identified that the ACE2 is the preferred binding

structure of SARS-CoV and SARS-CoV-2, which differs from the other mechanisms of viral attachment and entry exhibited by other viruses.

The genome of SARS-CoV-2 encodes pp1a and pp1ab, and eight accessory proteins along with the four main proteins previously discussed[9] The MERS-CoV genome is about 30 kb in length and encodes pp1a, pp1ab and five accessory proteins along with the four main proteins[10.] Recent studies have found that SARS-CoV-2 virus is 79.5% identical to the sequence found in SARS-CoV[6] as it encodes pp1ab, and six accessory proteins along with the four main proteins. The similarity in structure and sequence supports the hypothesis that SARS-CoV-2 evolved from SARS-CoV RBDs in an attempt to improve binding affinity to the same ACE2 receptor[11]; however, the two viruses differ in their receptor binding motifs.

Structural studies have shown SARS-CoV-2 RBD has a unique Lys417 residue on its RBM region that forms salt-bridge interactions with the Asp30 in the ACE2 of its target cells; on the other hand, SARS-CoV lacks this amino acid and instead it has a valine residue that does not participate in ACE2 binding[11]. This lysine residue gives a positive-charged region on the SARS-CoV-2 RBD that is absent on the SARS-CoV RBD, an alteration that influences the binding affinity difference of the SARS-CoV-2 and SARS-CoV to the ACE2 receptor[11]. Although these studies were able to identify some residues that are potentially involved in the SARS-CoV-2-ACE2 interaction, the actual interaction remains elusive because computer modelling of the RBD, particularly for the RBM, appeared incomplete when a single spike protein was investigated[11]. Even though scientists have yet to crystallize the structure of the RBM in the SARS-CoV-2 RBD and the interaction still remains elusive, these findings illustrating the sequence similarities and discrepancies between SARS-CoV and SARS-CoV-2 will serve as a foundation for the development of the crystal structure of this region.

Not only does the evolution of the receptor binding motif distinguishes SARS-CoV-2 from any other coronavirus, but it also provides an account to counteract previous claims that the viral RNA was genetically modified in a laboratory. Early papers regarding the origin of the virus suggested that four insertions in the S protein of SARS-CoV-2 resembled amino acid residues in key structural proteins of HIV, the virus that causes AIDS[12]. These premature theories led to the speculation that 2019-nCoV had been genetically modified in the lab to use as a bioweapon, causing controversy within the scientific community. Most researchers believe that the experiments that suggest that the development of SARS-CoV-2 could not have been fortuitous in nature were rushed and provided miscalculated results. To be able to claim that 2019-nCoV derived from gene fragments of HIV-1, both viruses would need to co-infect the same cells. However, the host cells for bat CoV and HIV-1 are different, and therefore it is high unlikely that they exchange genetic material[13].

Conclusions

Despite the conspiracy theories circulating the origin of the virus, the scientific data illustrated in this review shows that the development of SARS-CoV-2 was a result of natural selection and not of purposeful manipulation. This review goes into detail regarding the structural comparison between 2019-nCoV and other coronaviruses, the genetic make-up discussed in here shows that SARS-CoV-2 has components that differ significantly from other known viruses. These discrepancies reduce the possibility of a man-made virus because these structures have not been identified in other viral RNAs, which demonstrates that SARS-CoV-2 is not derived from any previous used virus backbone and therefore is a product of nature itself and evolution. Nonetheless, further analysis of the viral structure needs to be accomplished to develop a successful and efficient therapeutic treatment for COVID-19. All of the scientific papers cited in this report offer valuable information regarding the similarities and differences between SARS-CoV-2 and other coronaviruses and establish the foundation of the assembly of the virus that may be used in future works to crystallize the final structure of the virus in its entirety.

References

(1) Chiu, S. S.; Chan, K. H.; Chu, K. W.; Kwan, S. W.; Guan, Y.; Poon, L. L., et al. Human coronavirus NL63 infection and other coronavirus infections in children hospitalized with acute respiratory disease in Hong Kong, China. *Clin. Infect. Dis.* 40, 1721–1729. doi: 10.1086/430301. 2005

(2) Wang, N.; Shang, J.; Jiang, S.; Du, L. (2020). Subunit Vaccines Against Emerging Pathogenic Human Coronaviruses. Front. Microbiol. doi: 10.3389/fmicb.2020.00298

(3) Li, F. Structure, Function, and Evolution of Coronavirus Spike Proteins. *Ann. Rev. of Virol.* Vol.3 237-261. doi: 10.1146/annurev-virology-110615-042301. 2016

(4) Bosch, B.; van der Zee, R.; de Hann, C.; Rottier, P. The Coronavirus Spike Proteins is a Class I Virus Fusion Protein: Structural and Functional Characterization of the Fusion Core Complex. J. Virol., p. 8801. doi: 10.1128/JVI.7716.8801-8811. 2003

(5) Kirchdoefer, R.; Cottrell, C.; Wang, N.; Pallesen, J.; Yassine, H.; Turner, H.; Corbett, K.; Graham, B.; McLellan, J.; Ward, A. Pre-fusion Structure of a Human Coronavirus Spike Protein. *Nature*. p. 118-121. doi: 10.1038/nature17200. 2016

(6) Zhou, P.; Yang, X.; Wang, X.; Hu, B.; Zhang, L.; Zhang, W.; Si, H.; Zhu, Y.; Li, B.; Huang, C.; Chen, H.; Chen, J.; Luo, Y.; Guo, H.; Jiang, R.; Liu, M.; Chen, Y.; Shen, X.; Wang, X.; Zheng, X.; Zhao, K.; Chen, Q.; Deng, F.; Liu, L.; Yan, B; Zhan, F.; Wang, Y.; Xiao, G; Shi, Z. A Pneumonia Outbreak Associated with a New Coronavirus of Probable Bat Origin. *Nature*. 2070-273. doi: 10.1038/s41586-020-2012-7. 2020

(7) Li, W., Moore, M. J., Vasilieva, N., Sui, J., Wong, S. K., Berne, M. A., et al. Angiotensin-converting enzyme 2 is a functional receptor for the SARS coronavirus. *Nature.* 426, 450–454. doi: 10.1038/nature02145. 2003

(8) Raj, V. S., Mou, H., Smits, S. L., Dekkers, D. H., Muller, M. A., Dijkman, R., et al. Dipeptidyl peptidase 4 is a functional receptor for the emerging human coronavirus-EMC. *Nature* 495, 251–254. doi: 10.1038/nature12005. 2013

(9) Marra, M. A., Jones, S. J., Astell, C. R., Holt, R. A., Brooks-Wilson, A., Butterfield, Y. S., et al. The genome sequence of the SARS-associated coronavirus. *Science* 300, 1399–1404. doi: 10.1126/science.1085953. 2003

(10) van Boheemen, S., de Graaf, M., Lauber, C., Bestebroer, T. M., Raj, V. S., Zaki, A. M., et al. Genomic characterization of a newly discovered coronavirus associated with acute respiratory distress syndrome in humans. mBio 3:e00473-12. doi: 10.1128/mBio.00473-12. 2012

(11) Lan, J.; Ge, J.; Yu, J.; Shan, S.; Zhou, H.; Fan, S.; Zhang, Q.; Shi, X.; Wang, Q.; Zhang, L.; Wang, X. Structure of the SARS-CoV-2 Spike Receptor-binding domain bound to the ACE2 receptor. *Nature*. doi: 10.1038/s41586-020-2180-5. 2020

(12) Pradhan, P.; Pandey, A.; Mishra, A.; Gupta, P.; Tripathi, P.; Menon, M.; Gomes, J.; Vivekanandan P.; Kundu, K. Uncanny Similarity of Unique Inserts in the 2019-nCoV Spike Protein to HIV-1 gp120 and Gag. *BioRxiv*. doi: 10.1101/2020.01.30.927871. 2020

(13) Xiao, C.; Li, X.; Liu, S.; Sang, Y.; Gao, S.; Gao, F. HIV-1 did not Contribute to the 2019-nCoV Genome. Emerg. Microbers. Infect. 9(1): 378-381. doi: 10.1080/22221751.2020.1727299. 2020