Protein Structure and Function Prediction of SARS-CoV 2: Prospective Antivirus Active Drug Binding Sites

Shivi Sharma 1,a), Shefali Maheshwari1,b)

Department of Biotechnology, Hindustan College of Science & Technology, Mathura (Affiliated to Dr. P.J. Abdul Kalam Technical University, Lucknow), Mathura-281122, Uttar Pradesh, India.(+91)

a)Corresponding author: shivisharma.hcst@sgei.org
b)Shefalisagar0315@gmail.com

Abstract: Today a newly emerged corona-virus known as SARS-CoV 2 has become a cause of global health concern and took away the lives of large number of people throughout the world. Coroanavirus are the enveloped virus with positive single stranded genome of 26.4 to 31.7 kb. Envelop of the corona-virus is made up of four structural proteins namely envelop protein (E), membrane protein (M), spike protein (S) and nucleocapsid protein (N). These four proteins are responsible for the overall shape and size of the virus (structure of virus). Envelop protein forms ion channels, membrane protein is responsible for the shape of the virus, spike protein is responsible for the entering inside the target host cell by binding to host receptor and nucleocapsid protein binds to the single stranded RNA genome of the virus forming multiple copies. We investigate the reliability and homogeneity among all the corona-virus species such as MERS-CoV, Bat-CoV HKU4, Transmissible gastro-enteritis coronavirus (TGEV), Porcine epidemic diarrhea virus (PEDV), HCoV-229E and Whale-CoV SW1, M-CoV, Hedgehog coronavirus 1, Bulbul-CoV HKU11 etc, using a binary graph which is helpful in the findings of sequence reliability, secondary and tertiary structure model prediction using advanced model builder, we build the model of various protein/gene products by selecting them from SARS-COV 2, which further helpful in the finding of target-ligand binding for future therapeutic applications. Consequently by modeling the structure of the proteins we bring into being that envelop protein have pentameric protein lipid pores that allow ion transportation and were able to depict the active drug binding sites.

INTRODUCTION

A coronavirus is a large group of RNA viruses that causes various respiratory tract infections in humans, birds, and animals. Coronaviruses are a single-stranded positive RNA virus that belongs to the coronaviridae family. Corona is a Latin word that means a crown; as the name suggests that these viruses have crown-like projections on their surface [10].

Since 8000 B.C.E., these viruses have been around. Warm-blooded flying vertebrates (bats and birds) serve as natural reservoirs for these viruses [37]. These Coronaviruses are further divided into four sub-species which have their origin from bats and birds namely alpha-coronavirus, beta-coronavirus, gamma-coronavirus, and delta-coronavirus. The origins of alpha and beta coronaviruses are from bat and of gamma and delta from birds.
Coronaviruses are spherical-shaped viruses with a crown-like projection on the surface and have an average diameter of 80 to 120 nm. Coronavirus has a molecular weight of 40,000 kDa on average. [10]. Coronavirus genome is of size 26.4 to 31.1 kb. 5' leader - UTR - replicase (ORF1ab) - spike (S) - envelope (E) - membrane (M) - nucleocapsid (N) - 3'UTR - poly (A) tail is the genome organization of these coronavirus. The open reading frames 1a and 1b constitute two third of the genome and encodes polyprotein (pp1ab). 16 non-structural proteins (nsp1 to nsp16) are formed by the cleavage of replicase polyprotein. Four structural proteins envelop protein (E), membrane protein (M), nucleocapsid protein (N) and spike protein (S) are encoded by open reading frames [27]. These 4 structural proteins are responsible for the structure of the virus along with entering the host and protecting the virus when it is outside the host. The lipid bilayer of the virus is composed of an envelope (E), spike (S), and membrane (M) protein in the ratio 1:20:300 [27].

The first coronavirus disease was experimental in 1930 in domesticated chickens. The coronavirus responsible for this disease was known as infectious bronchitis virus (IBV) [23]. Furthermore, in 1940, two more coronavirus species, transmissible gastroenteritis virus (TGEV) [43] and hepatitis mouse virus (MHV) [45], were identified for the infection in pigs and mice, respectively. Later, in 1960, the first human coronavirus, BB14, was discovered by isolating it from a boy suffering from the common cold. In 1961, BB14 Coronavirus was attempted to cultivate by using the same standard techniques that have previously been used successfully for the cultivation of other common cold viruses, such as rhinoviruses and adenoviruses, but the idea failed. Later, in 1965, the virus was successfully cultivated in human embryonic trachea organ culture by serially passing through it [25]. In 1962, the common cold virus was isolated from a medical student and cultivate in kidney tissue culture was named 229E [17].

All these viruses, including animals and humans, had relationships with each other, but it was unknown during that time. In 1967, the structure of IBV, BB14, and 229E using an electron microscope by Scottish virologists June Almeida and Tyrrell and found that these viruses were showing a relatable shape and distinctive club-like spikes. In the same year, the National Institution of Health isolated another more virus using the organ culture method and named it OC43 [48]. This virus was also causing cold and had a distinctive club-like spike on the surface has been observed in the electron microscope. Further, IBV showed morphological similarity with MHV, BB14, 229E, and OC43; thus, this group was termed coronavirus. Later, in 2003, 2004, and 2012, other human coronaviruses came into existence. SARS-CoV in 2003, HKU1 in 2004, MERS-CoV in 2012, and currently a new coronavirus strain, SARS-CoV 2 in 2019 have been discovered [47]. Among these viruses, SARS-CoV 2 spread among humans and becoming an outbreak.

At present, SARS-CoV 2, a newly emerged coronavirus, has been infecting the entire world and has become a reason for severe health concerns. This newly emerged coronavirus has taken away many lives, and no proper medication is available for the treatment of the disease. The spike protein binds to ACE2, cell surface receptor of host. ACE2 is also called as ACEH. It is a part of the dipeptidy l - carboxy dipeptidase angiotensin converting enzyme (ACE) family. Angiotensin 2 is converted to Ang 1−7 by ACE2 receptor. In many physiologic functions, such as the cell proliferation and hypertrophy, inflammatory response, blood pressure and the fluid balancing, ACE2 has high affinity to Type 1 and Type 2 Ang receptors. ACE2 is ubiquitously expressed in the lungs. And hence when Spike (S) protein binds to ACE 2 receptor to enter the host cell it directly targets the lung cells and causes respiratory dysfunction [50]. Cytokine release syndrome is a prominent characteristic of severe SARS-CoV-2 infections. A respiratory epithelial cell infection activates monocytes, macrophages, and dendritic cells resulting in a variety of proinflammatory cytokines, including interleukin-6, being secreted by SARS-CoV 2 (IL-6). IL-6 release causes a cascade of amplification, directly leading among other lymphocytic changes to T helper 17 (Th17). Circulation of IL-6 receptor complexes and soluble IL-6 receptor complexes indirectly activate several types of cells, including endothelial cells, which cause systemically produced cytokines to cause hypotension and acute respiratory distress syndrome and this result to the
SARS-CoV-2 to escape from the human immune system and cause the disease [51]. The risk factors for seriousness and mortality of COVID-19 are high for the persons who suffer from diabetes, hypertension, and cardiovascular diseases in particular. Vaccination is the only way to prevent the disease and stop it from spreading, for this work on the vaccine is in progress. But this virus is evolving and showing mutations and becoming more life-threatening.

Scientists have been discovering that this newly emerged coronavirus has become more virulent due to the occurrence of continuous mutation since November 2019. To know more about this deadly virus and its mutation, it is necessary to study its evolutionary relationship with other coronavirus species. And hence, it is required to study the structure and function prediction of the viral protein in detail.

**METHODLOGY**

**Data retrieving and analysis**

Twenty-four different coronavirus genome sequences were retrieved (as shown in TABLE 1) using NCBI for the analysis of the relationship among these different species of coronaviruses. Further, SARS-CoV-2 structural proteins known as envelope protein, membrane protein, nucleocapsid protein and spike protein amino acid sequences (as shown in TABLE 2) were used for secondary structure prediction and tertiary structure prediction of proteins.

| S.NO | CORONAVIRUS                                    | NCBI ID     | SUB-TYPE  |
|------|-----------------------------------------------|-------------|-----------|
| 01   | Avian Coronavirus                             | NC_001451  | Gamma-CoV |
| 02   | Beluga whale coronavirus SW1                  | NC_010646  | Gamma-CoV |
| 03   | Bovine coronavirus                            | NC_003045  | Beta-CoV  |
| 04   | Bulbul coronavirus HKU11                      | NC_011547  | Delta-CoV |
| 05   | Feline coronavirus                            | NC_002306  | Alpha-CoV |
| 06   | Hedgehog coronavirus 1                        | NC_039207  | Beta-CoV  |
| 07   | Human coronavirus 229E                        | NC_002645  | Alpha-CoV |
| 08   | Human coronavirus HKU1                        | NC_006577  | Beta-CoV  |
| 09   | Human coronavirus NL63                       | NC_005831  | Alpha-CoV |
| 10   | Human coronavirus OC43                       | NC_006213  | Beta-CoV  |
| 11   | Middle east respiratory syndrome (MERS)       | NC_019843  | Beta-CoV  |
| 12   | Miniopterus bat coronavirus 1                 | NC_010437  | Alpha-CoV |
| 13   | Miniopterus bat coronavirus HKU8              | NC_010438  | Alpha-CoV |
| 14   | Murine coronavirus                            | NC_048217  | Beta-CoV  |
| 15   | Pipisterellus bat coronavirus HKU5            | NC_009020  | Beta-CoV  |
| 16   | Porcine epidemic coronavirus                  | NC_003436  | Alpha-CoV |
| 17   | Porcine coronavirus HKU15                     | NC_039208  | Delta-CoV |
| 18   | Rhinolophus bat coronavirus HKU2              | NC_009988  | Alpha-CoV |
| 19   | Rousettus bat coronavirus HKU9                | NC_009021  | Beta-CoV  |
| 20   | Scotophilus bat coronavirus 512               | NC_009657  | Alpha-CoV |
| 21   | Sever acute respiratory syndrome (SARS)       | NC_004718  | Beta-CoV  |
| 22   | Sever acute respiratory syndrome              | NC_045512  | Beta-CoV  |
coronavirus 2 (SARS-CoV 2)

| S.NO | STRUCTURAL PROTEINS                  | ACCESSION NO. |
|------|--------------------------------------|---------------|
| 01   | Envelop protein (E)                  | YP_009724392  |
| 02   | Membrane protein (M)                 | YP_009724393  |
| 03   | Nucleocapsid protein (N)             | YP_009724397  |
| 04   | Spike protein (S)                    | YP_009724390  |

**TABLE 2:** List of structural proteins of SARS-CoV 2

**Phylogenetic analysis**

Analysis of all retrieved genome sequences and the relationship among these genome sequences had performed by aligning them using MultAlin and MEGA X [41]. The phylogeny has been derived using maximum likelihood method.

**Sequence analysis and Secondary structure prediction**

Sequence analysis and secondary structure evaluation of the SARS-CoV 2 structural proteins have been performed to know physiochemical parameters and packing of amino acid in alpha helix, beta-sheets, and beta turns using PortParam [13] and SOPMA, [06] respectively.

**Tertiary structure prediction and validation**

We had used SWISS-MODEL to perform template-based modeling. To evaluate the templates of SARS-CoV 2 four structural proteins, we had individually analyzed them using BLASTp [40] against all individual structural proteins to know the percentage identity and similarity. Based on BLASTp results suitable template was used to model the tertiary structure of the protein. For the predicted models, the overall quality was validated using ERRAT [05] and ProCheck servers [26] based on the Ramachandran plot. Then the predicted models were analyzed to know about the functional properties of the structural viral proteins.

**RESULTS AND DISCUSSIONS**

As per the following procedure discussed above, we predict the evolutionary relationship among the different coronavirus species. And modeled the three-dimensional structures of the four structural proteins of SARS-CoV 2 (envelop protein, membrane protein, nucleocapsid protein and spike protein), the virus responsible for the ongoing pandemic worldwide. We were also able to predict the physicochemical and biochemical functions of the proteins.
Phylogenetic analysis

We had found the close relationship by generating a guide tree on the basis of scoring matrix (FIGURE 1) using maximum likelihood method among the following species:– MERS coronavirus (NC_019843.3) and Tylonycteris bat coronavirus HKU4 (NC_009019.1); Transmissible gastroenteritis coronavirus (NC_038861.1) and Porcine epidemic diarrhea virus (NC_003436.1); Human coronavirus 229E (NC_002645.1) and Beluga whale coronavirus SW1 (NC_0106461); Murine coronavirus (NC_048217.1) and Hedgehog coronavirus 1 (NC_039207.1); Bulbul coronavirus HKU11 (NC_011547.1) and SARS (NC_004718.3); Avian infectious bronchitis virus (NC_001451.1) and Rousettus bat coronavirus HKU9 (NC_009021.1) sharing close relationships.

FIGURE 1. Scoring Matrix based alignment of the above genome sequences.
FIGURE 2. Phylogenetic tree for different coronavirus species

**Envelop Protein**

This protein is a minor structural protein of SARS-CoV 2, made up of 75 amino acid residues. The molecular weight of this protein is 8365.04 Daltons. The physiochemical prediction shows that envelop protein is rich in leucine (18.7%), valine (17.3%), and serine (10.7%) (FIGURE 3). The secondary structure analysis shows that the amino acid of enveloping protein [E] has 44% alpha helixes, 9% beta turns, 26.67% extended strands, and 20% random coil arrangement (FIGURE 4).

![Composition Graph](image)

FIGURE 3. SARS-CoV 2 envelop protein amino acid compositions.

![Secondary Structure Graph](image)

FIGURE 4. Secondary structure prediction for Envelop protein

During the modeling of the envelope protein, four templates had generated. After performing the BLASTp template used for modeling showed 91.38% percentage identity and 0.55 sequence similarities, which was an envelope small membrane proteins of conserved Golgi complex targeting
signal in coronavirus envelope protein. The verified structure has a QMEAN score of -3.40 and a global score of 0.39. ProCheck verification revealed that 92.6% of residues are in the favored region (A, B, L) and 94.64% favored by the Ramachandran plot (FIGURE 5). Structure verification by ERRAT reveals the average quality of the model with an overall quality factor of 68%. The tertiary structure of envelope protein results in the pentameric ion channel and Golgi complex of protein which targeting the signals inside the SARS-CoV 2 envelope protein.

![FIGURE 5. Envelop protein model and favored Ramachandran Plot.](image)

**Membrane Protein**

This protein is a major structural protein that provides the shape and maintains the structure of SARS-CoV 2, made up of 222 amino acid residues. The molecular weight of this protein is 25146.62 Daltons. The physiochemical prediction shows that the membrane protein is rich in leucine (15.8%) (FIGURE 6). The secondary structure analysis shows that the amino acids of membrane protein have 34.68% alpha helix, 6.76% beta turns, 21.17% extended strands, and 37.39% is random coil arrangement (FIGURE 7).

![FIGURE 6. SARS-CoV 2 membrane protein amino acid compositions.](image)
During the modeling of the membrane protein, 14 templates had generated. After performing the BLASTp template used for modeling showed 15.38% percentage identity and 0.28 sequence similarities. The template was an E.coli core signaling unit carrying receptor mutation. The verified structure has a QMEAN score of -2.48 and a global score of 0.3. ProCheck verification revealed that 83.3% of residues are in the favored region (A, B, L) and 93.88% favored by the Ramachandran plot (FIGURE 8). ERRAT reveals the average quality of the model with an overall quality factor of 57.14%. The tertiary structure of membrane protein results in the triple helix bundle that forms a 3-trans membrane domain in SARC-CoV 2.

Nucleocapsid Protein

This protein binds to the single-stranded RNA genome forming multiple copies in SARS-CoV 2, made up of 419 amino acid residues. The molecular weight of this protein is 45625.70Daltons. The physiochemical prediction
shows that the nucleocapsid protein is rich in glycine (10.3%) (FIGURE 9). The secondary structure analysis shows that the amino acids of nucleocapsid protein have 21.24% alpha-helix, 6.92% beta turns, 16.71% extended strands, and 55.13% random coil arrangement (FIGURE 10).

During the modeling of the nucleocapsid protein, 36 templates had generated. After performing the BLASTp template used for modeling showed 92.06% percentage identity and 0.61 sequence similarities, which was an RNA binding domain of SARS-CoV nucleocapsid. The verified structure has a QMEAN score of 1.06 and a global score of 0.16. ProCheck verification revealed that 89.4% of residues are in the favored region (A, B, L) and 99.19% favored by the Ramachandran plot (FIGURE 11). ERRAT reveals a good quality of the model with an overall quality factor of 89.56%. The tertiary structure shows RNA binding domains present on the nucleocapsid protein of SARS-CoV 2.
Spike Protein

This protein binds to the host receptor for entering the host cell, made up of 1273 amino acid residues. The molecular weight of this protein is 141178.47 Daltons. The physiochemical prediction shows that the spike protein is rich in leucine (8.5%) (FIGURE 12). The secondary structure analysis shows that the amino acids of spike protein have 28.59% alpha-helix, 3.38% beta turns, 23.25% extended strands, and 44.78% is random coil arrangement (FIGURE 13).
During modeling of spike protein, 34 templates had generated. After performing the BLASTp template used for modeling showed 76.79% percentage identity and 0.55 sequence similarities were selected, which was trypsin cleaved low pH treated SARA-CoV spike glycoprotein. The verified structure has a QMEAN score of -3.72 and a global score of 0.60. ProCheck verification revealed that 80% of residues are in the favored region (A, B, L) and 87.11% favored by the Ramachandran plot (FIGURE 14). ERRAT reveals a good quality of the model with an overall quality factor of 84.53%. The tertiary structure of membrane protein results that spike protein has two functional subunits S1 and S2. S1 is globule and has receptor binding domains (RBD) that bind to the target host receptor and enters the cell. While S2 is the stem that is for membrane fusion (Fig14)

CONCLUSION
The evolutionary relation generated by the phylogenetic tree on the basis of its scoring matrix using maximum likelihood method, concludes that the four genera, α, β, γ, and δ coronaviruses are likely to occur in relationship with each other. For many years these coronaviruses are emerging and evolving. Their roots are from avian coronaviruses. The new classification system shows of bat coronaviruses are dominant in α-coronavirus sub- species. While in
β, γ, and δ-coronaviruses, bird coronaviruses are dominant. The enormous variety of coronaviruses originated from bats and birds make an excellent gene pool of alpha, beta, gamma and delta-coronaviruses [37].

Analyzing the SARS-CoV 2 structural proteins shows that the envelope protein (E) is 8.4 to 12kDa in size and consists of 76 amino acid residues [30]. This protein is an integral protein that is embedded into the lipid layer. The transmembrane domain and extra membrane C-terminal domain are the two domains of the E protein. This protein is vital for creating a viral assembly. There are almost 20 E proteins in the virus and forms an alpha-helical structure with a single alpha-helical transmembrane domain and forms a five-molecular ion channel in lipid bilayer [12]. Membrane protein (M) is a 7.8nm thick structural protein, consists of 222 amino acid residues. This protein is responsible for providing the overall shape of the virus [24]. M protein has three domains. One is N-terminal ectodomain, the second one is triply spanning trans-membrane domain, and the third one is C-terminal endodomain [04]. Nucleocapsid protein (N) is 43-50kDa in size, consists of 419 amino acid residues. Its function is the packaging of approx 30kb, 5’-capped positive-strand viral genome of single stranded RNA molecule into aribonucleoprotein (RNP) complex known as capsid. This protein is highly phosphorylated, due to which structural changes take place in the virus and enhances the affinity of SARS-CoV 2 [07]. Spike protein (S) is a 1273 amino acid residue protein. That is responsible for the corona's halo-like surface. There are 20nm-sized 74 surface spike composed of trimmer of S protein on the virus surface approximately [36]. There are two sub-units of spike protein S1 and S2 subunits. S1 subunit contributes as head of the protein and act as a receptor binding domain. S2 subunit contributes to stem of the protein which harbors the spike on the virus envelope. In the functionally active SARS-CoV 2 virus, three S1 subunits get attached with two S2 subunits. When the virus binds to the receptor of the host, the subunit complexes split into individual subunits [14]. Hence this study helps to develop a potential candidate for drug designing to fight against the viral infection.

ACKNOWLEDGEMENT
We acknowledge Hindustan College of Science and Technology, Farah Matura to provide a workplace in their Bioinformatics lab.

REFERENCES
[1] A.Tabibzadeh, M.Eshghaei, S.Soltani, P.Yousefi, M.Taherizadeh, F.S.Tameskel, M.Golahdooz, M.Hadi "Evolutionary study of COVID-19, sever acute respiratory syndrome coronavirus 2 (SARS-CoV 2) as an emerging coronavirus: phylogenetic analysis and literature review." Vet Med Sci. 2020;00:1–13; DOI: 10.1002/vms3.394
[2] A.A.T Naqvia, K.Fatimab, T. Mohammada, U. Fatimac, I. K. Singhd, A. Singhe, S.M. Atiff, G.Hariprasadg, G.M. Hasanh, M.I Hassan "Insights into SARS-CoV-2 genome, structure, evolution, pathogenesis and therapies: Structural genomics approach" BBA - Molecular Basis of Disease 2020 1866:165878; https://doi.org/10.1016/j.bbadis.2020.165878
[3] A.Sapkota, “Structure and Genome of SARS-CoV-2 (COVID -19) with diagram” April 5, 2020. Microbialnotes. https://microbenotes.com/structure-and-genome-of-sars-cov-2/
[4] Benjamin W. Neuman,a Gabriella Kiss,a Andreas H. Kunding,d,f David Bhella,i M. FazilBaksh,b Stephen Connelly,c Ben Droese,c Joseph P. Klaus,l Shinji Makino,h Stanley G. Sawicki,j Stuart G. Siddell,k Dimitrios G. Stamou,d,e,f Ian A. Wilson,c Peter Kuhn,c and Michael J. Buchmeierg, “A structural
analysis of M protein in coronavirus assembly and morphology" [Struct Biol. 2011 174(1): 11–22. doi: 10.1016/j.jsb.2010.11.021
[5] C Colovos, T O Yeates "Verification of protein structures: patterns of nonbonded atomic interactions" Protein Sci 1993 Sep;2(9):1511-9. doi: 10.1002/pro.5560020916
[6] C. Geourjon, G. Deléage "SOPMA: significant improvements in protein secondary structure prediction by consensus prediction from multiple alignments" Bioinformatics 1995; 11: 6, 681–684, https://doi.org/10.1093/bioinformatics/11.6.681
[7] C.K.Changa, M.H.Houb, C.F.Changc, C.D.Hsiaod, T.H.Huangace, "The SARS coronavirus nucleocapsid protein – Forms and functions" Antiviral Research 2014 (103) 39-50; https://doi.org/10.1016/j.antiviral.2013.12.009
[8] C.Baruah, P.Devi, D.Sharma “Sequence analysis and structure prediction of SARS-CoV 2 accessory proteins 9b and ORF14: evolutionary analysis indicates close relatedness to bat coronavirus" BioMed Research International Volume (2020); https://doi.org/10.1155/2020/7234961
[9] C.Wu, Y.Liu, Y.Yang, P.Zhang, W.Zhong, Y.Wang, Q.Wang, Y.Xu, M.Li, X.Li, M.Zheng, L.Chen, H.Li "Analysis of therapeutic targets for SARS-CoV-2 and discovery of potential drugs by computational methods." Acta Pharmacutica Sinica B (2020) 10(5):766e788; https://doi.org/10.1016/j.apsb.2020.02.008
[10] C. S. Goldsmith, K. M. Tatti, T. G. Ksiazek, P. E. Rollin, J.A. Comer, W. W. Lee, P. A. Rota, B.Bankamp, William, J. Bellini, and S. R. Zaki, "Ultrastructural Characterization of SARS Coronavirus" Emerging Infectious Disease 2004. 10(2): 320–326. doi: 10.3201/eid1002.030913
[11] D. E. Gordon, G. M. Jang, N. J. Krogan "A SARS-CoV-2 protein interaction map reveals targets for drug repurposing" Nature 2020 (583)459–468; https://doi.org/10.1038/s41586-020-2286-9
[12] D. Schoeman and B. C. Fielding, "Coronavirus envelope protein: current knowledge" Schoeman and Fielding Virology Journal (2019) 16:69 https://doi.org/10.1186/s12985-019-1182-0
[13] E. Gasteiger, C. Hoogland, A. Gattiker et al., "Protein identification and analysis tools on the ExPASy server," in The Proteomics Protocols Handbook, pp. 571–607, Humana Press, 200
[14] E. A J Alsaaid, I.M Jones, "Membrane binding proteins of coronaviruses" Future Virol. 2019 14(4):275–286; doi: 10.2217/fvl-2018-0144
[15] F. Hufsky, K. Lamkiewicz, A. Almeida, A.Ouacheria, C. Arighi, A. Bateman, J. Baumbach, N. Beerweninkel, C. Brandt, M. Cacciabue, S. Chuguransky, O. Drehsel, R. D. Finn, A.Fritz, S. Fuchs, et al. "Computational strategies to combat COVID-19: useful tools to accelerate SARS-CoV-2 and coronavirus research" Briefings in Bioinformatics, 2020, 1–22; doi: 10.1093/bib/bbaa232
[16] F.K Yoshimoto "The Proteins of Severe Acute Respiratory Syndrome Coronavirus-2 (SARS CoV-2 or n-COV19), the Cause of COVID-19" The Protein Journal (2020) 39:198-216 https://doi.org/10.1007/s10930-020-09901-4
[17] Geller C, Varbanov M, Duval RE, "Human coronaviruses: insights into environmental resistance and its influence on the development of new antiseptic strategies"
[18] Hamre D, Procknow JJ, "A new virus isolated from the human respiratory tract" Society for Experimental Biology and Medicine, 121(1), 190-193; https://doi.org/10.3181/00379727-121-30734

[19] J.A. Jaimes, N.M. André, J.S.Chappie, J.K.Millet, G.R.Whittaker "Phylogenetic Analysis and Structural Modeling of SARS-CoV-2 Spike Protein Reveals an Evolutionary Distinct and Proteolytically Sensitive Activation Loop" Journal of Molecular Biology (2020) 432, 3309-3325; https://doi.org/10.1016/j.jmb.2020.04.009

[20] J.Xu, S.Zhao, T.Teng, A.E.Abdalla, W.Zhu, L.Xie, Y.Wang, X.Guo "Systematic Comparison of Two Animal-to-Human Transmitted Human Coronaviruses: SARS-CoV-2 and SARS-CoV." Viruses 2020, 12, doi:10.3390/v12020244

[21] J.F.W. Chan, K.H.Kok, Z.Z. Hin Chu, K.K. W. To, S. Yuan, K.K Yuen, "Genomic characterization of the 2019 novel human pathogenic coronavirus isolated from a patient with atypical pneumonia after visiting Wuhan" January 28, 2020 Emerging Microbes & Infections. 10.1080/22221751.2020.1719902

[22] J.O. Wertheim, D. K. W. Chu, J. S. M. Peiris, S. L. K. Pond, L. L. M. Poon, "A Case for the Ancient Origin of Coronaviruses" Journal of Virology 2013, 7039–7045; doi:10.1128/JVI.03273-12

[23] Kahn, J. S. McIntosh, Kenneth, "History and Recent Advances in Coronavirus Discovery" The Pediatric Infectious Disease Journal: 2005 24(11) S223-S227; doi: 10.1097/01.inf.0000188166.17324.60

[24] K. Lalchhandama, "The chronicles of coronaviruses: the electron microscope, the doughnut, and the spike" June 30, 2020 Journal of the Mizo Academy of Sciences https://doi.org/10.33493/scivis.20.02.03

[25] Knapp, Alex. "The Secret History Of The First Coronavirus". Forbes. https://www.forbes.com/sites/alexnknapp/2020/04/11/the-secret-history-of-the-first-coronavirus-229e/#300a62da71d6

[26] Laskowski R A, MacArthur M W, Moss D S, Thornton J M (1993). PROCHECK - a program to check the stereochemical quality of protein structures. J. App. Cryst., 26, 283-291

[27] L. F. García, “Immune Response, Inflammation, and the Clinical Spectrum of COVID-19” June 16, 2020frontiers in immunology. https://doi.org/10.3389/fimmu.2020.01441

[28] M. Sarkar, S. Saha "Structural insight into the role of novel SARS-CoV-2 E protein: A potential target for vaccine development and other therapeutic strategies" PLoS ONE 15(8): e0237300. https://doi.org/10.1371/journal.pone.0237300

[29] M. Bianchi, D.Benvenut , M. Giovanetti , S. Angeletti , M. Ciccozzi, and S. Pascarella “Sars-CoV-2 Envelope and Membrane Proteins: Structural Differences Linked to Virus Characteristics?” Hindawi BioMed Research International 2020, 6, 4389089; https://doi.org/10.1155/2020/4389089

[30] M.M.C.Lai,D.Cavanagh, "The Molecular Biology of Coronaviruses" Advance in virus research.https://doi.org/10.1016/S0065-3527(08)60286-9
[31] N.Kaur a, R.Singh, Z.Dar, R.K.Bijarnia, N.Dhingra, T.Kaur “Genetic comparison among various coronavirus strains for the identification of potential vaccine targets of SARS-CoV2.” Elsevier B.V. (2020)1567-1348; https://doi.org/10.1016/j.meegid.2020.104490

[32] N. Decaro “Betacoronavirus”, The Springer Index of Viruses, DOI 10.1007/978-0-387-95919-1

[33] N. Decaro “Alphacoronavirus” The Springer Index of Viruses, DOI 10.1007/978-0-387-95919-1

[34] N. Decaro “Gammacoronavirus” The Springer Index of Viruses, DOI 10.1007/978-0-387-95919-1

[35] N. K. Dutta, K. Mazumdar, J. T. Gordyc “The Nucleocapsid Protein of SARS–CoV-2: a Target for Vaccine Development” Journal of Virology 2020, 94:3 https://doi.org/10.1128/JVI.00647-20.

[36] P.S.Masters, “The Molecular Biology of Coronaviruses” Advance in virus research https://doi.org/10.1016/S0065-3527(06)66005-3

[37] P.C.Y.Woo, S.K.P.Lau, C.S.F.Lam, C.Y.Lau, A.K.L.Tsang, J.H.N.Lau, R.Bai, J.L.L.Teng, C.C. C. Tsang, M. Wang, B.J. Zheng, K.H. Chan, and K.Y. Yuen “Discovery of Seven Novel Mammalian and Avian Coronaviruses in the Genus Deltacoronavirus Supports Bat Coronaviruses as the Gene Source of Alphacoronavirus and Betacoronavirus and Avian Coronaviruses as the Gene Source of Gammacoronavirus and Deltacoronavirus” Journal of Virology 2011, 3995–4008; doi:10.1128/JVI.06540-11

[38] P. C. Y. Woo, Y. Huang, S. K. P. Lau and K.Y. Yuen “Coronavirus Genomics and Bioinformatics Analysis” Viruses 2010, 2, 1804-1820; doi:10.3390/v2081803

[39] Q. Li, W. Peng, Y. Ou “Prediction and analysis of key protein structures of 2019-nCoV” Future Virol. (2020)15(6), 349–357; doi: 10.2217/fvl-2020-0020

[40] S. F. Altschul, T. L. Madden, A. A. Schaffer et al., “Gapped BLAST and PSI-BLAST: a new generation of protein database search programs,” Nucleic acids research, vol. 25, no. 17, pp. 3389-3402, 1997

[41] S. Kumar, G. Stecher, M. Li, C. Knyaz, and K. Tamura, “MEGA X: molecular evolutionary genetics analysis across computing platforms,” Molecular Biology and Evolution, vol. 35, no. 6, pp. 1547–1549, 2018

[42] S.Dong, J.Sun, Z.Mao, L.Wang, Yi‐Lin Lu, Jiesen L “A guideline for homology modeling of the proteins from newly discovered betacoronavirus, 2019 novel coronavirus (2019-nCoV).” J Med Virol. 2020;1–7; DOI: 10.1002/jmv.25768

[43] S. Williams, “A Brief History of Human Coronaviruses” June 2, 2020 https://www.the-scientist.com/news-opinion/a-brief-history-of-human-coronaviruses-67600

[44] S.Kanga, M. Yang , Z.Hong , L. Zhang , Z. Huang , X. Chen , S. He , Z. Zhou , Z. Zhou , Q. Chen
Y.Yan, C.Zhang, H.Shan, S.Chen “Crystal structure of SARS-CoV-2 nucleocapsid protein RNA binding domain reveals potential unique drug targeting sites” Acta Pharmaceutica Sinica B 2020;10(7):1228e1238; https://doi.org/10.1016/j.apsb.2020.04.009

[45] S.Chaplin, “COVID-19: a brief history and treatments in development” prescriber 2020; https://doi.org/10.1002/psb.1843

[46] T.Schwede, J.Kopp, N.Guex and M.C.Peitsch “SWISS-MODEL: an automated protein homology-modeling server” Nucleic Acids Research, 2003, 31: 13 33381–3385 DOI: 10.1093/nar/gkg520

[47] WebMd, “Coronavirus 2020 Outbreak: Latest Updates” https://www.webmd.com/lung/news/20200124/coronavirus-2020-outbreak-latest-updates

[48] WHO https://www.who.int/news-room/detail/07-04-2020-who-lists-two-covid-19-tests-for-emergency-use

[49] Z.Zhu, X.Lian, X.Su, W.Wu, G.A. Marraro and Y. Zeng “From SARS and MERS to COVID-19: a brief summary and comparison of severe acute respiratory infections caused by three highly pathogenic human coronaviruses” Respiratory Research (2020) 21:224 https://doi.org/10.1186/s12931-020-01479-w

[50] Y.Zhang, X.Geng, Y.Tan, Q.Li, C.Xu, J.Xu, L.Hao, Z.Zeng, X.Luo, F.Liu, H.Wang “New understanding of the damage of SARS-CoV-2 infection outside the respiratory system” Biomedicine & Pharmacotherapy (2020); 127.https://doi.org/10.1016/j.biopha.2020.110195

[51] Kanta Subbarao, and Siddhartha Mahanty "Respiratory Virus Infections: Understanding COVID-19" Immunity, (2020); 52(6): 905–909. 10.1016/j.immuni.2020.05.004