The Devil in its Details: Unravelling the Epitopes in COVID-19 Surface Glycoprotein with the potential for Vaccination and Antibody Synthesis

Abdul Mannan Baig (✉ abdul.mannan@aku.edu)

  Department of Biological and Biomedical Sciences, Aga Khan University

Abstract

With over 75762 patients infected and 2130 deaths been reported, the mortality and morbidity caused by the recent outbreak COVID-19 infections are proving colossal when compared to similar epidemics caused by SARS and MERS variant of coronavirus in the past. It was aimed to identify a receptor-binding domain (RBD) in surface glycoprotein (sGP) of COVID-19 and predict epitopes that are capable of interacting with major histocompatibility complex (MHC) alleles to evoke antibody production *in vivo*. Bioinformatic computational tools were used to analyze the well-studied sGP and the RBD in SARS-CoV and compare it with their homologs in COVID-19. In silico methods that predict epitopes capable of binding MHC allele were used to identify sequences in RDB of sGP in COVID-19 with the intention of discovering sequences that can be used for vaccination and production of monoclonal antibody (mAbs) against COVID-19. The results indicate that COVID-19 has a conserved RBD in the sGP with differences in its sequences that can be exploited for vaccination and manufacturing of specific antibodies against this variant of coronavirus. Reported are 10 sequences of epitopes that are predicted to bind the MHC class I and class II alleles and that do not cross-react with human proteins. Testing *in vitro* and in animal models can accelerate the translational utility of vaccinating and efficacy of mAbs against the COVID-19 virus.

Introduction

What gives chill and scare to humans in theatres while watching movies on viral pandemics happened in real-time in mainland Wuhan China to reflect the reality that such outbreaks of mammoth proportions are not only possible but can cause loss of human life and a global economic recession to come (1). It appears that we haven't learned from the past experiences with the emergence of a cross-species viral infections and dangers of its transmission with diseases like Ebola, MERS and SARS (2, 3), as investment by even first world countries into the research to prevent and tackle the emergence of outbreak of lethal viral epidemics has remained insignificant. The scientific and healthcare community on the other hand, at times like past SARS outbreak and recent COVID−19 epidemic, has always come forward to help with their expertise to combat such occurrences (4, 5, 6, 7) without any material ravenousness. Management of the outbreaks and attempts to contain the current COVID−19 crisis has proven to be monumental with the risk of it outbreaks that could go global (8) along with no vaccine and specific drugs that target the COVID−19 virus, the feeling of being helpless to combat the COVID−19 epidemic is frustrating and ominous. The fast and dedicated work of the scientists (9, 10) and devoted organization (11, 12, 13) have managed to archive the genome of COVID−19 and its different mutants in a very short time that gives an opportunity to explore and elucidate the molecular targets encoded by the COVID−19. Being taxonomically related to coronaviridae group of viral pathogens (9, 10), the COVID−19 shares a significant similarity with other members that are known to have caused similar if not identical cross-species viral diseases in humans (9, 14, 15). Examples of the latter include SARS (2002 to 2004) and MERS that affected the human population in 2012 (14, 15). The SARS genome (NCBI Reference Sequence: NC_004718.3) is known to encode 13 proteins, of which the surface glycoprotein (sGP) (NCBI Reference Sequence: NP_828851.1) is known to be essential to dock the virus to angiotensin-converting

enzyme 2 (ACE2) receptor on host cells in the pulmonary parenchyma (9, 10, 14). It was inferred that like both, SARS-nCoV and COVID−19, belong to the same betacoronavirus (14, 15) they would have a similar if not identical composition of the sGP that help to dock on ACE2 receptors (16, 17). in human cells before gaining entry into the host cells. It is important to mention here that an antigenic similarity between both the sGP of COVID−19 and SARS-CoV can be exploited to synthesize vaccine and monoclonal antibody (mAb) against the surface glycoprotein of the COVID−19, as previous knowledge on SARS-CoV sGP (9, 14) would help in achieving the understanding and steps needed to mass production of a vaccine and mAb to target COVID−19. That being said, the efficient use of the bioinformatic computational tools in molecular biology like computer-assisted generation prediction of microbial epitope prediction, antigenic sequence generation from genomic data and homology modelling are few examples (11, 12, 13, 14) that had helped in targeting SARS-CoV and are in mass use to enable us to fight the current COVID−19 outbreak. The reports of the genome sequencing (11), studying the mutations evolving (4), prediction of proteins encoded (11, 14), drawing the phylogenetic tree of COVID−19 (4, 14) are few examples of how the scientific community has started its battle against the COVID−19. As the outbreak continues and has worsened in recent weeks (1, 8), for humanity it's a race against time with COVID−19 to generate a vaccine, mAbs and an anti-viral drug that proves to be efficacious in the patients infected with COVID−19. Here, first, we compare the protein sequences of sGP of both SARS-CoV and COVID−19 to spot the similarities and contrast between the glycoproteins. Secondly, based on a decade long research finding on SARS-CoV sGP (4, 9, 17), the segments within the sGP of COVID−19 were planned to be tested in silico for their epitope attributes which can help generate a vaccine for COVID−19. Using segments of sGP with maximal similarities and contrast with SARS-CoV, prediction of monoclonal antibodies (mAbs) against the unique segments was attempted. The findings of antigenic components in COVID−19 sGP is expected to lay down the basis of further research on the segments discovered and *in vitro* testing in COVID−19 in labs and animal models to validate their utility. If proven to be effective in a small-scale animal trial, the vaccine generation and mAbs can prove useful in our massive ongoing research to combat COVID−19.

Methods

The genome of COVID−19 and the proteins it encodes

The genome of COVID−19 was retrieved from at the NCBI database (11). With the NCBI Reference Sequence: NC_045512.2, the genome consists of 21290 bp ss-RNA. The genome assembly was mapped for encoding protein sequences in general and sGP of COVID−19 in particular. The attributes of the sGP of COVID−19 was retrieved and the FASTA sequence of the sGP was downloaded from the NCBI database (11, 12).

COVID−19 BLASTn search, Multiple Sequence Alignments and Protein encoded

The sequence of sGP was searched for homologs in NCBI automated server and the results were retrieved with sequence identities, e-values and scores. In BLASTn search for homologs, the search was optimized for highly similar sequences (megablast) and SARSn-CoV was selected as an organism in which homologs were to be searched. From an array of the homologs of SARS-CoV genomes retrieved, the matches with sequence identities between 89% to 82% were selected for Multiple Sequence Alignments (MSA) and building the distance evolutionary tree. The NCBI automated server was used to uncover the proteins encoded by the CONVID−2019 nucleotides.

The surface glycoprotein encoded by COVID−19 searched for homologs

The sequence of sGP of CONVID−2019 was retrieved from NCBI database. Composed of 1273 amino acids this glycoprotein was searched for homologs in EMBL EBI database (18). The Uniprot database (12) was used to align the sequences of COVID−19 and Human SARS coronavirus (SARS-CoV) spike glycoprotein. The site, motif, transmembrane regions and sites of mutagenesis were compared for similarities and contrast in UniProt automated server.

Tracing a SARS-CoV like Receptor binding domain in CONVID−2019 sGP.

The SARS-CoV has been extensively studied for its receptor-binding domain (RBD) which is a known region in the sequence of the sGP of this virus (4, 9, 10, 17). Because of the novelty of the sequence of COVID−19, an RBD region in the sequence of encoded for sGP is not established yet. The 318 to 510 amino acids in the sequence of the sGP which is the RBD for SARS-CoV was compared with a similar sequence in the sGP of COVID−19. As the amino acid of COVID−19 is made up of 1273 amino acids as compared to 1255 amino acids of SARS-CoV, sGP and the RBD sequences of SARS-CoV and COVID−19 coronavirus were downloaded and compared.

Homology modelling of COVID−19

The sequence of the sGP was submitted to the automated severs that develop template-based models of novel proteins as is the case with sGP of COVID−19 virus. We used the SWISS- MODEL server (13) for this purpose. The FASTS sequence of CONVID−2019 that was retrieved from the NCBI database with GenBank ID: QHO62877.1was used to develop a template-based model. The template-based model was studied for its quality by parameters like GQME and QMEAN. Models were analyzed in-depth by using advanced automated tasks in the SWISS-MODEL server. Ligand binding sites were also studied in the model aligned with the template built for the target sGP of COVID−19.

Search for a putative antigenic epitope in sGP sequence encoded by COVID−19

Though the COVID−19 is declared to be a novel variant of betacoronavirus, its sGP has homologs (evolutionarily related) in the taxon of betacoronaviruses. The Immune Epitope Database and Analysis Resource (IEDB) (19) and DTU bioinformatic servers (20, 21) were used for epitope mapping in the sGP sequence in COVID−19 virus. Filters were applied to search within betacoronavirus, SARS and MERS members of the taxa only. Other filters used were selected from epitope type, host, assays type, MHC restrictions and infectious disease (19). Also, the amino acid sequence of sGP of COVID−19 was used to generate putative antibody against epitopes of sequence lengths between 5mer to 9mer where a peptide-MHC class I binding could occur for provoking antibody production *in vivo.* For the prediction of MHC Class II epitopes, that has a consensus approach which combines NN-align, SMM-align and combinatorial library methods, automated server from Immune Epitope Database Analysis Resource (19) was selected and predictions were generated for RBD of COVID−19 and SARS-CoV and compared.

Results

The genome of COVID−19 and the proteins it encodes.

Wuhan seafood market pneumonia virus isolate Wuhan-Hu−1, complete genome deposited in the NCBI database constitutes of 29903 bp ss-RNA with NCBI Reference Sequence: NC_045512.2 (Fig.1A and supplementary File. Fig 1). The linage of the COVID−19 shows that it belongs to a yet unclassified betacoronavirus group (Fig.1 B). The proteins encoded by COVID−19 are shown (Fig.1B), which includes the details of the locus YP_009724390 encoding sGP sequence (Fig.1 B, C) in COVID−19 (Fig.1 - black-arrow) to be composed of 1273 amino acids.

BLASTn search, Multiple Sequence Alignments and Protein encoded

A BLASTn search (Fig.1A) for homologs of COVID−19 nucleotide sequence showed identical sequences in Bat CoV, SARS-CoV and a large series of coronaviruses. Of the vast series of coronaviruses, the top ten homologs had sequence identities in a range of 89% to 81% and were selected for sequence alignments and to develop an evolutionary tree in form of a cladogram (Fig. 2 B). The evolutionary tree showed COVID−19 to share a common ancestor with all the coronaviruses compared. The Bat coronavirus complete genome and SARS-CoV showed to share a common ancestor as shown in the cladogram.

The Surface glycoprotein in COVID−19 and SARS-CoV family are homologs.

The alignment of the entire sequence of 1237 amino acids against 1255 amino acids of SARA-CoV sGP showed a sequence identity of 76%, and an e-value 0.0. The alignment of the complete sequences

(Supplementary File−1) and in-depth study of the sequences for similarities and contrast in transmembrane regions, motif, site and points of amino acid mutagenesis showed (Fig.3) that in transmembrane region (Fig 3 A1-light blue rectangles) the amino acids valine (V) and leucine (L) in sGP sequence of SARS-CoV is substituted by isoleucine (I) and methionine (M) in sGP sequence of COVID−19.

Within other regions like motifs (Fig.3 A -brick red row) and sites (Fig.3 A3), the amino acids remain identical. Interestingly the aligned sequences for points of mutagenesis (Fig.3 alphabets highlighted in brown color) showed three points of mutagenesis where SARS-CoV and COVID−19 were found to differ (Fig 3. A2-red arrows in between rows). Amino acids glycine (G), serine (S) and glutamine (Q) at 476th, 494th and 690th positions in COVID−19 (Fig.3 A2- top row) were seen to replace aspartic Acid (D), aspartic acid (D) and lysine (K) respectively in SARS-CoV (Fig.3 A2- bottom row).

Receptor binding domain in sGP of CONVID−2019 and SRAS-CoV showed them to be homologs.

The SARS-CoV composed of 1255 amino acids has an established RBD which stretches between its 318th to 510th amino acids (Fig.4) (22). The BLASTp results of RBD of SARS- CoV (Fig.4 A) showed it to be a homolog of sGP and spike protein of COVID−19 with 73.7% sequence identities. The evolutionary tree developed by MSA showed sGP and spike protein of COVID−19 to have their origins from a common ancestor (Fig. 4 A branching cladogram). A reverse BLASTp search result (Fig.4 B) of COVID−19 putative RBD (318−510 aa) fetched the SARS-CoV spike protein RBD domain in results with around 97% sequence identities (Fig.4B). The evolutionary tree developed showed links to common origins of RBD of two variants of SARS-CoV and the putative RBD of COVID−19. Five other viruses were found to have a similar sequence with the putative RBD of COVID−19.

As the RBD of SARS-CoV proved to be homolog of spike protein and sGP of COVID−19, a pairwise alignments of amino acids of the COVID−19 of similar range (318th −510th aa) was made to unravel any contrasts that could explain the novelty of COVID−19 and identify epitopes that could be identified in COVID−19 for vaccine development and mAbs synthesis.

We show that the range of RBD amino acid sequence of SARS-CoV (Fig.5 A top- row blue rectangle) 318th amino acid to the 510th amino acid (Fig.5 A, the last row) have a sequence identity of 62.622% with the same number of amino acids of the COVID−19 sGP sequence (Fig.5 B)

Homology modelling of COVID−19 Surface Glycoprotein and RBD

The FASTA sequence of sGP [Wuhan seafood market pneumonia virus] GenBank: QHD43416.1 and putative RBD (318th to 5110th aa) (of COVID−19 were submitted to the automated server in the SWISS-MODEL database. The database developed a 3 template- based model of SARS-CoV (Fig 6. C) spike

glycoprotein and ACE2 complex for the sGP sequence of COVID−19 (Fig.6 A-B) and for the putative RBD of CONVID−2019 a model of spike glycoprotein (Fig. 6 B2) was built with 71.88 % sequence identities.

Epitopes of COVID−19 with MCH Class I allele binding predictions

The prime intention of the comparison of the sGP and RBD of both the viruses in this study was to uncover sequences in COVID−19 that can serve as a possible epitope for vaccination and mAbs synthesis. We show that in silico methods identified segments in COVID−19 that can act as an antigen when injected in nM concentrations. The predictions made by the in silico methods (mentioned in the methodology section) also recognizes the ability of the segment to bind to MHC class 1 and class II allele that is known to be mandatory for antigen presentation to human lymphocytes *in vivo (*23, 24). The Table−1 and Fig. 7 A shows epitopes in COVID−19 an order that ranks them with the highest MHC binding probability to the least.

Epitopes of COVID−19 with MCH Class II allele binding predictions

Unique epitope sequences were generated (Fig.7 A) from input sequence of putative RBD of COVID−19 that predicted their binding to MHC class -II DRB allele that possibly could facilitate the synthesis of antibody (Ig) and killer cell T-cells against the COVID−19. When compared with SARS-CoV RBD the sequences appeared to be distinct (Fig.7 B). No homology of the peptide epitope sequences (Table−1 and Fig 7) generated for COVID−19 were found in human of BLASTp search (supplementary file. Fig 2 and 3).

Discussion

The objective of this study was to reflect the forte of the bioinformatics computational tools, target genome of coronaviruses and in silico methods of epitope prediction in inferring molecular targets in COVID−119 that has threatened a global pandemic (1, 4, 8). It is important to mention here that finding of sGP homologs (evolutionary related) of COVID−19 in other related coronaviruses like SARS and MERS was equally important as it was to identify the contrasts, in this regard the benefits of the research done in the past on the latter two coronaviruses can be very important in uncovering molecular targets in COVID−19. It is shown in this study that the genome of COVID−19 virus (Fig.1 A) encodes a structural sGP spike protein of 1237 amino acid chain length (Fig.1B, C). The COVID−19 isolate Hu−1 complete genome has a vast series of homologs with 89% −81% homology with bat coronavirus and SARS coronavirus SARS-CoV. The MSA of the genomes with development of evolutionary tree (Fig.2 A) for the first ten best matches shown in the form of a cladogram highlighted the origins of COVID−19/bat coronavirus from a common ancestor. The 13 proteins encoded by the COVID−19 are shown which includes an sGP (Fig 2 C), that in case of SARS-CoV has shown to be cardinal for infection of the host cell via its binding to ACE2 receptor (14, 16, 17). In a comparative analysis of the sequences of sGP of COVID−19 and SARS-CoV (Fig.3 A, A1-A3) the areas of similarities and contrast are shown. Two amino acids substitutions were noticed in the transmembrane regions (Fig.3 A1- Blue rectangles) and three

differences in the site of mutagenesis (A2-red arrows). Though the significance of these substitutions remains to established but it is inferred that it could provide an explanation for the difference seen in the infectivity patterns of COVID−19 when compared to SARS-CoV. The COVID−19 on BLASTp results has been shown to have 76% sequence identities with SARS-CoV (Fig.3 B), the regions like motif and site were found to be identical (Fig.3 A, A3). The BLASTp results of the sequence 318th to 510th amino acids which is an RBD of SARS- CoV (22) showed a homolog in COVID−19 protein sequence identified as spike protein (Fig 4 A). The evolutionary tree developed showed them to have a common ancestor. Results of reverse BLASTp and the evolutionary tree of putative RBD (318th −510th aa) of COVID−19 fetched the RBD sequence of SARS-CoV and common origins of both sequences (Fig 4 B).

As the total amino acid length difference of both COVID−19 and SARS-CoV is 18 aa, with COVID−19 exceeding by this digit, we inferred that RBD has remained conversed both viruses as can be seen in evolutionary tree in form of a cladogram developed by NCBI automated server (Fig.4 A-B) for the RBD regions of both the coronaviruses. Also, sequence alignment of the entire length of sGP of both viruses shows the difference in the early 1 to 317 (COVID−19) as compared to the sequence after that (supplementary file-fig.1-rectangle area). As spotting the difference in the RBD regions of both viruses was important to infer the amino acid segments that could qualify for epitope predictions, this segment was studied in- depth. It was found that in-between the RBS sequences there were areas of differences which at this stage were inferred to serve as unique sequences (latter confirmed—detailed below) for being recognized by in silico methods of MHC allele binding and therefore provoking antibody production. Rows of the segment of RBD of SARS-CoV (Fig. 5 A) that is already known and stretches between 318th amino acid (Fig.5 A- marked in a blue rectangle in the top- rows) and 510th amino acid (Fig.5 A -blue rectangle bottom-rows) is shown. When aligned with COVID−19, areas in between segments are highlighted (Fig. 5 A-B horizontal red arrows) where mutations (substituted amino acids) has occurred in the putative RBD sequence of COVID−19 (Fig. 5, A-B) bottom row when compared to SARS-CoV (Fig.5 A-B top row). Pairwise aligned with similarities and identities shaded in grey (Fig.5 B), the zone of mutations can be seen clearly.

Homology modelling for COVID−19 sGP and putative RBD confirmed our earlier findings, as the SWISS-MODEL automated server for development of template-based models generated the model of SARS-CoV sGP and spike glycoprotein respectively for the above two proteins of COVID−19 (Fig.6). The reason the SWISS-MODEL automated server developed these models and not a model of COVID−19 is the fact that the crystal structure of COVID−19 is not yet deposited in the protein database (PDB) and the archives of structural protein models in SIWSS-MODEL template library (STML) database.

The last task of this study was to show the significance of the contrasts and similarities in the sequences of molecules like sGP and RBD of COVID−19 in the generation of epitopes that can be predicted for their bindings with MHC class-I and class-II allele. The latter interaction is central in provoking an antibody response *in vivo* in macrophages and when vaccinations and antibody production is induced in mammalian animal models (23, 24) which are known hosts for COVID−19 (4, 10, 14, 15, 16). Findings of this study show that robust in silico epitope predicting methods (19, 20, 21) can pick segments RBD

regions of COVID−19 virus (Table−1 and Fig.7) which can potentially bind the MHC-Class I and II allele and therefore play a significant role in vaccine development and mAbs synthesis against COVID−19.

Conclusion And Future Directions

With the recent outbreak of CVID−19 in China and its spread into 24 countries, there is an urgent need to develop vaccine and mAbs to combat the outbreak. Though this study doesn't report the results of lab testing of the proposed epitopes in provoking an immunological response in macrophages and lymphocytes due to unavailability of COVID−19 samples, it shows that unrevealing the similarities and contrasts of COVID−19 with SARS-CoV have aided our understanding in understanding this deadly pathogen in a shorter time. Though a cross-reacting antigen with SARS-CoV is not an issue in vaccinations of people in the recent COVID−19 outbreak, this study has identified epitopes that are so unique to COVID−19 that they do not show any cross-reactivity with the proteins and peptides expressed by human cells as was seen in BLASTp results (supplementary File- Fig.2 and Fig 3). I am optimistic that the finding of this study would be taken to its next step, that is the testing these epitopes for provoking antibody synthesis *in vitro* and neutralization of the infectivity of COVID−19 in subsequent assays performed in labs where the virus samples are investigated.

References

1. Coronavirus sends ripples through the global economy Financial Times https://www.ft.com/content/e9fbbb78-4901-11ea-aeb3-955839e06441 Date accessed (16th February 2020)

2. Suwantarat N, Apisarnthanarak A. Risks to healthcare workers with emerging diseases: lessons from MERS-CoV, Ebola, SARS, and avian flu. Curr Opin Infect Dis. 2015 Aug;28(4):349-61. doi: 10.1097/QCO.0000000000000183. Review. PubMed PMID: 26098498.

3. Madhav N, Oppenheim B, Gallivan M, et al. Pandemics: Risks, Impacts, and Mitigation. In: Jamison DT, Gelband H, Horton S, et al., editors. Disease Control Priorities: Improving Health and Reducing Poverty. 3rd edition. Washington (DC): The International Bank for Reconstruction and Development / The World Bank; 2017 Nov 27. Chapter 17. Available from: https://ncbi.nlm.nih.gov/books/NBK525302/ doi: 10.1596/978-1-4648-0527- 1/pt5.ch17

4. Huang C, Wang Y, Li X, Ren L, Zhao J, Hu Y, Zhang L, Fan G, Xu J, Gu X, Cheng Z, Yu T, Xia J, Wei Y, Wu W, Xie X, Yin W, Li H, Liu M, Xiao Y, Gao H, Guo L, Xie J, Wang G, Jiang R, Gao Z, Jin Q, Wang J, Cao B. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. Lancet. 2020 Jan 24. pii: S0140-6736(20)30183-5. doi: 10.1016/S0140-6736(20)30183-5. [Epub ahead of print] Erratum in: Lancet. 2020 Jan 30;:. PubMed PMID: 31986264.

5. Smith N, Fraser M. Straining the System: Novel Coronavirus (COVID-19) and Preparedness for Concomitant Disasters. Am J Public Health. 2020 Feb 13:e1-e2. doi: 10.2105/AJPH.2020.305618.

[Epub ahead of print] PubMed PMID:

6. Velavan TP, Meyer CG. The Covid-19 epidemic. Trop Med Int Health. 2020 Feb doi: 10.1111/tmi.13383. [Epub ahead of print] PubMed PMID: 32052514.

7. Lum LH, Tambyah PA. Outbreak of COVID-19 - an urgent need for good science to silence our fears? Singapore Med J. 2020 Feb 13. doi: 10.11622/smedj.2020018. [Epub ahead of print] PubMed PMID: 32052064

8. Michael Le Page. Coronavirus: How well prepared are countries for a COVID-19 pandemic? New Scientist: https://newscientist.com/article/mg24532693-500- coronavirus-how-well-prepared-are-countries-for-a-covid-19-pandemic/ Date accessed (16th February 2020)

9. Lu R, Zhao X, Li J, Niu P, Yang B, Wu H, Wang W, Song H, Huang B, Zhu N, Bi Y, Ma X, Zhan F, Wang L, Hu T, Zhou H, Hu Z, Zhou W, Zhao L, Chen J, Meng Y, Wang J,Lin Y, Yuan J, Xie Z, Ma J, Liu WJ, Wang D, Xu W, Holmes EC, Gao GF, Wu G, ChenW, Shi W, Tan W. Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding. Lancet. 2020 Jan 30. pii: S0140-6736(20)30251-8. doi: 10.1016/S0140-6736(20)30251-8. [Epub ahead of print] PubMed PMID: 32007145.

10. Wang M, Cao R, Zhang L, Yang X, Liu J, Xu M, Shi Z, Hu Z, Zhong W, Xiao Remdesivir and chloroquine effectively inhibit the recently emerged novel coronavirus (2019-nCoV) in vitro. Cell Res. 2020 Feb 4. doi: 10.1038/s41422-020-0282-0. [Epub ahead of print] PubMed PMID: 32020029.

11. National Center for Biotechnology Information, U.S. National Library of Medicine 8600 Rockville Pike, Bethesda MD, 20894 USA. https://guides.nnlm.gov/tutorial/ncbi-gene-sequence- Date accessed (16th February 2020)

12. UniProt Consortium. UniProt: a hub for protein information. Nucleic Acids Res. 2015;43(Database issue): D204–D212. doi:10.1093/nar/gku989, (2015)

13. Waterhouse, A., Bertoni, M., Bienert, S., Studer, G., Tauriello, G., Gumienny, R., Heer, F.T., de Beer, T.A.P., Rempfer, C., Bordoli, L., Lepore, R., Schwede, T. SWISS-MODEL: homology modelling of protein structures and complexes. Nucleic Acids Res. 46(W1), W296-W303 (2018).

14. Zumla A, Chan JF, Azhar EI, Hui DS, Yuen KY. Coronaviruses - drug discovery and therapeutic options. Nat Rev Drug Discov. 2016 May;15(5):327-47. doi: 10.1038/nrd.2015.37. Epub 2016 Feb 12. Review. PubMed PMID:

15. Woo, P. C., Lau, S. K., Huang, Y. & Yuen, K. Y. Coronavirus diversity, phylogeny and interspecies jumping. Exp. Biol. Med. (Maywood) 234, 1117–1127 (2009).

16. Jia HP, Look DC, Shi L, et al. ACE2 receptor expression and severe acute respiratory syndrome coronavirus infection depend on differentiation of human airway epithelia. J Virol. 2005;79(23):14614–14621. doi:10.1128/JVI.79.23.14614-14621.2005

17. Wang H, Yang P, Liu K, Guo F, Zhang Y, Zhang G, Jiang C. SARS coronavirus entry into host cells through a novel clathrin- and caveolae-independent endocytic pathway. Cell Res. 2008 Feb;18(2):290-301. doi: 10.1038/cr.2008.15. PubMed PMID:

18. Madeira F, Park YM, Lee J, et al. The EMBL-EBI search and sequence analysis tools APIs in 2019. Nucleic Acids Research. 2019 Jul;47(W1):W636-W641. DOI: 10.1093/nar/gkz268

19. Kim Y, Ponomarenko J, Zhu Z, et al. Immune epitope database analysis resource. Nucleic Acids Res. 2012;40(Web Server issue):W525–W530. doi:10.1093/nar/gks438

20. Andreatta M, Nielsen M .Gapped sequence alignment using artificial neural networks: application to the MHC class I system., Bioinformatics (2016) Feb 15;32(4):511-7

21. Nielsen M, Lundegaard C, Worning P, Lauemoller SL, Lamberth K, Buus S, Brunak S Lund O. Reliable prediction of T-cell epitopes using neural networks with novel sequence representations. .Protein Sci., (2003) 12:1007-17

22. He Y, Li J, Heck S, Lustigman S, Jiang S. Antigenic and immunogenic characterization of recombinant baculovirus-expressed severe acute respiratory syndrome coronavirus spike protein: implication for vaccine design. J Virol. 2006;80(12):5757–5767. doi:10.1128/JVI.00083-06

23. Laurence L. Brunton, Bruce A. Chabner, Björn C. Knollmann. 2011. Goodman &Gilman's 'The Pharmacological basis of Therapeutics' – 12th edition, ISBN 978-0- 07-162442-8. Copyright © 2011by The McGraw-Hill Companies, Inc:

24. BARRETT, K. E., & GANONG, W. F. (2012). Ganong's review of medical physiology. New York, McGraw-Hill

Table 1

Due to technical limitations, Table 1 can be accesses via the supplemental files section.

Table–1: COVID–19 peptides predicted for vaccine and mAbs synthesis. The peptide sequences shown were unique to COVID–19 that were predicted to interact with human MHC-Class-I allele and elicit an immune response in the host. Strong binding (SB) prediction and nM affinity are shown
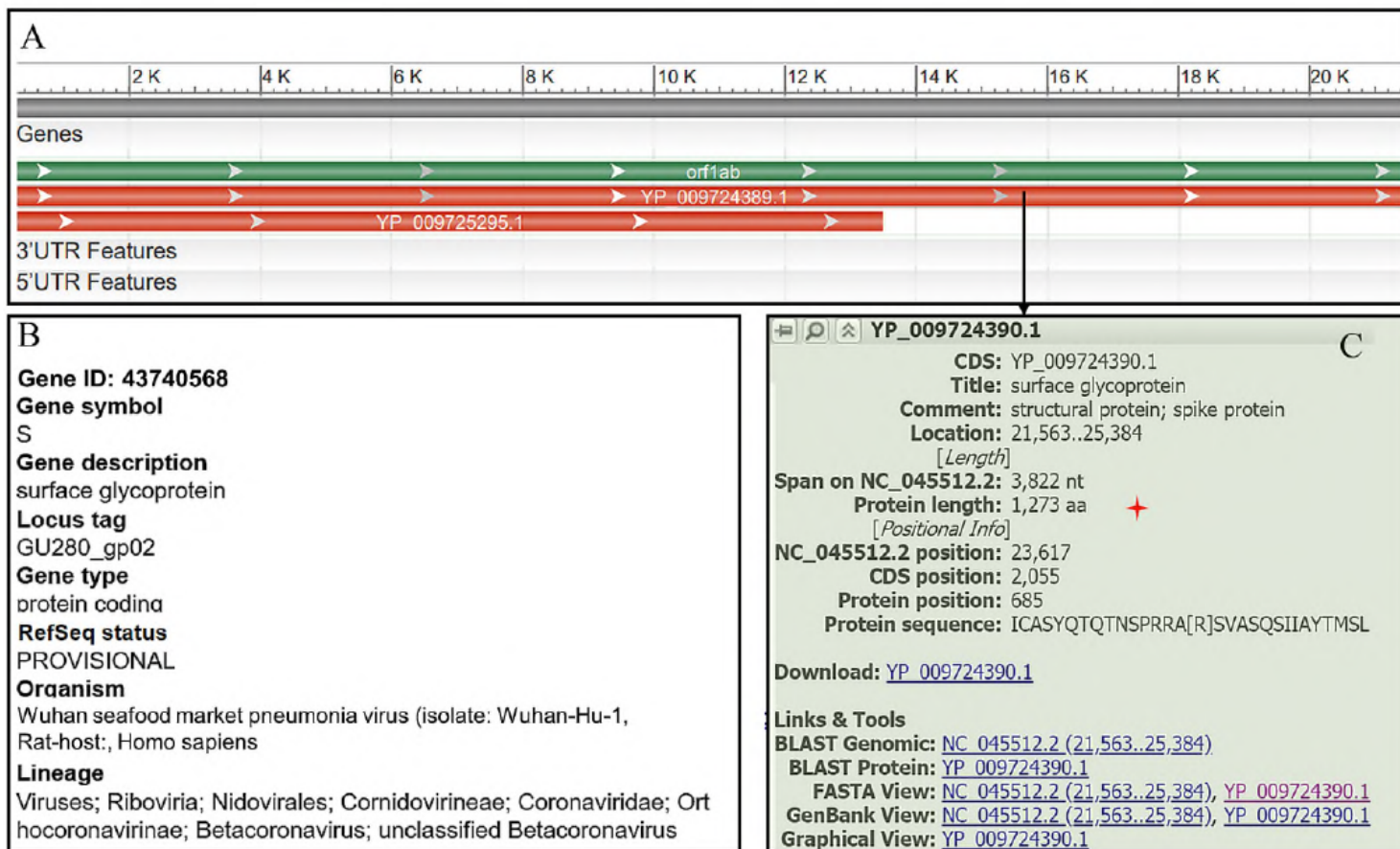
Figures

Figure 1

(A) Details of the genome of COVID-19 genome organization. The genomic regions are shown that encode for surface glycoprotein. (B) The gene ID and symbol with the details of the gene showed that COVID-19 is an unclassified betacoronavirus. (C). Shown are the details of COVID-19 encoding the surface glycoprotein which is a structural protein composed of 1237 amino acid chain length.
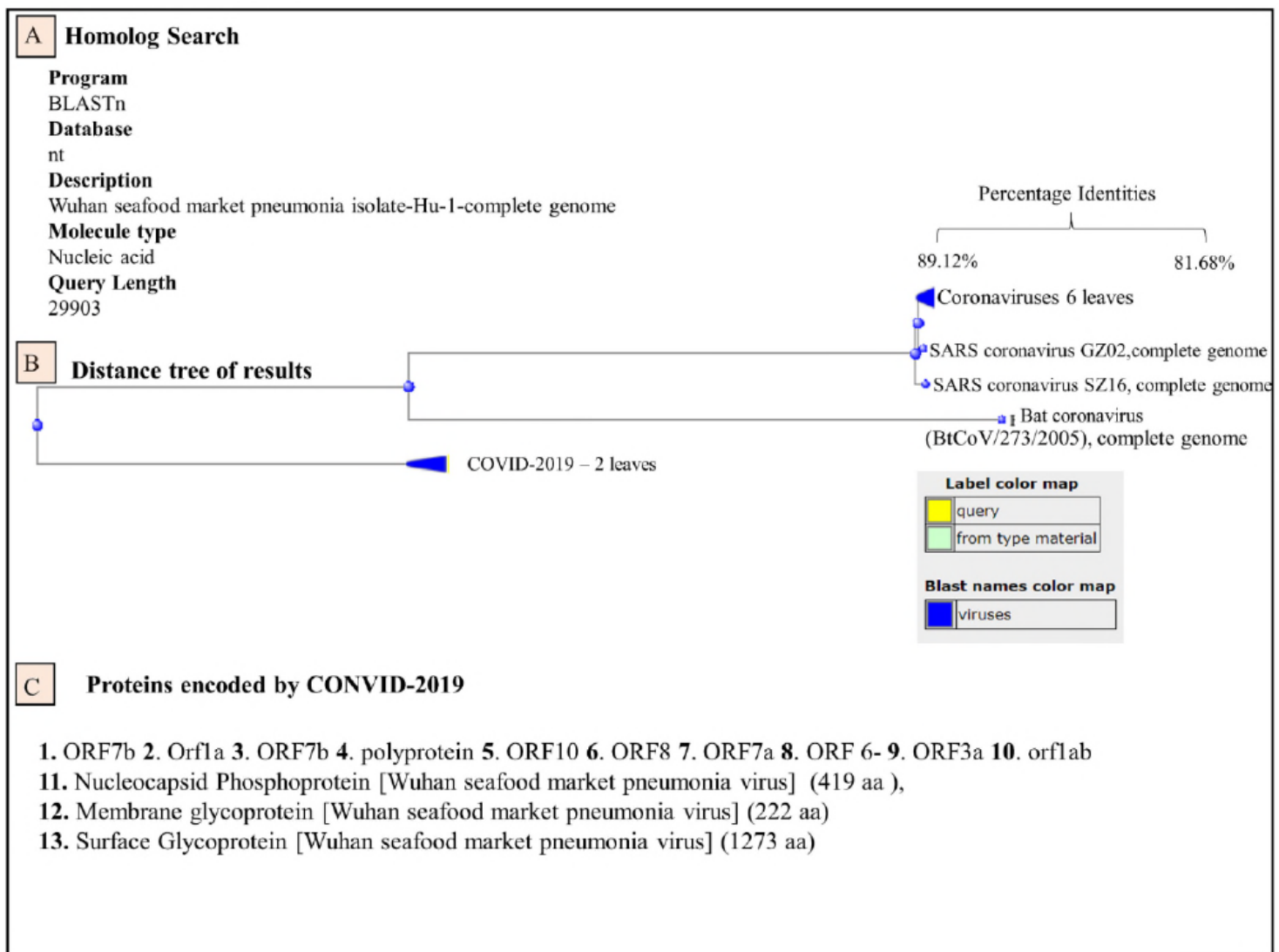
Figure 2

(A) Results of BLASTn search of COVID-19 isolate Hu-1 complete genome in NCBI database server fetched a vast series of homologs identified with 89% -81% homology with bat coronavirus and SARS coronavirus SARS-CoV (B) MSA with development of evolutionary tree for first ten best matches showed in form of a cladogram highlights the origins of COVID-19 and bat coronavirus from a common ancestor. (C) The 13 proteins encoded by the COVID-19 are shown.
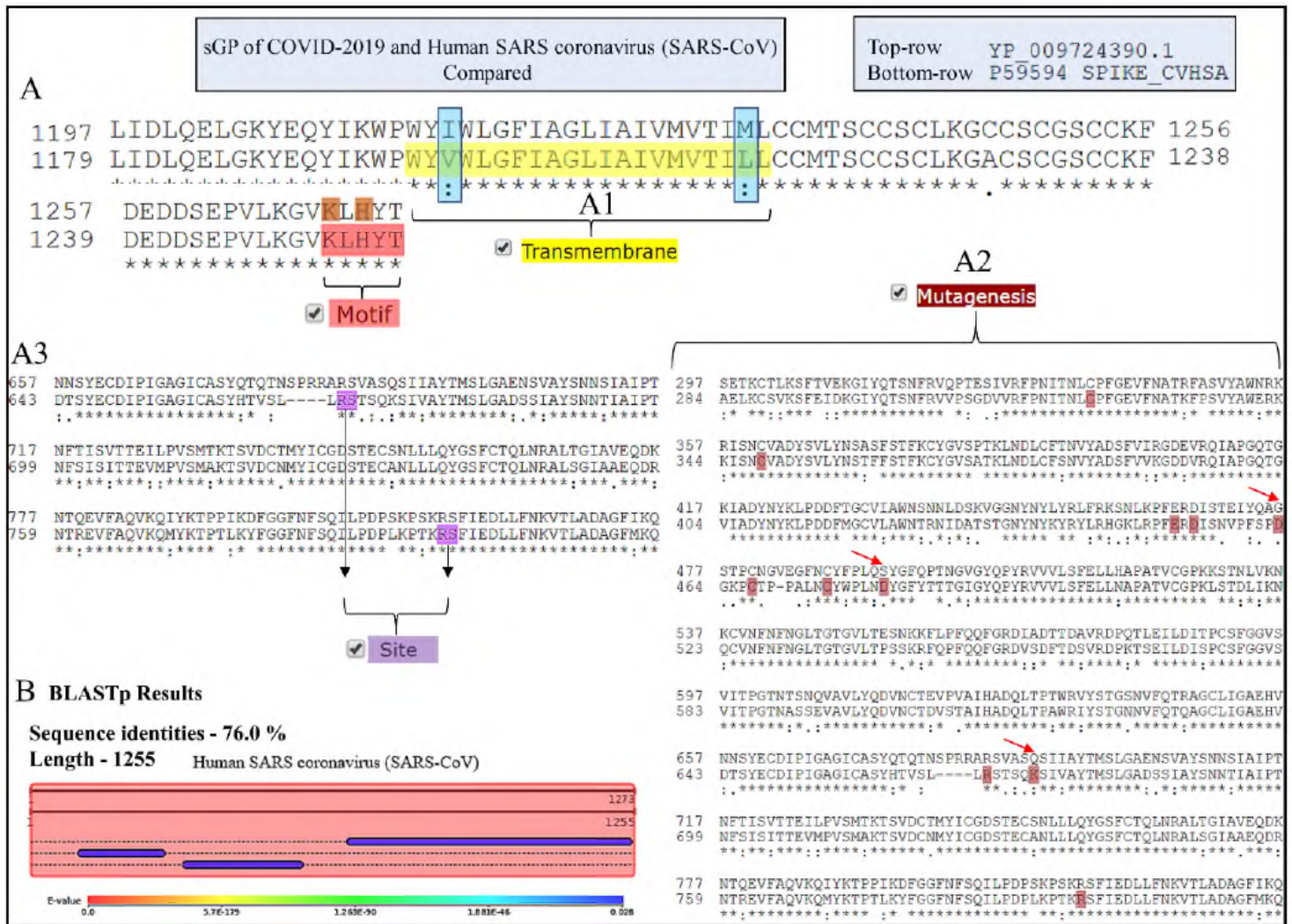
Figure 3

The sequences of sGP of COVID-19 and SARS-CoV (A, A1-A3) showed similarities and contrast. Two amino acids substitutions were noticed in the transmembrane regions (A1- Blue rectangles) and three differences in the site of mutagenesis (A2-red arrows). The COVID-19 on BLASTp search had been noticed to have 76% sequence identities with SARS-CoV (B), the regions like motif and site were found to be identical. (*) denoted identical sites, while (:) represent similarity of amino acids)
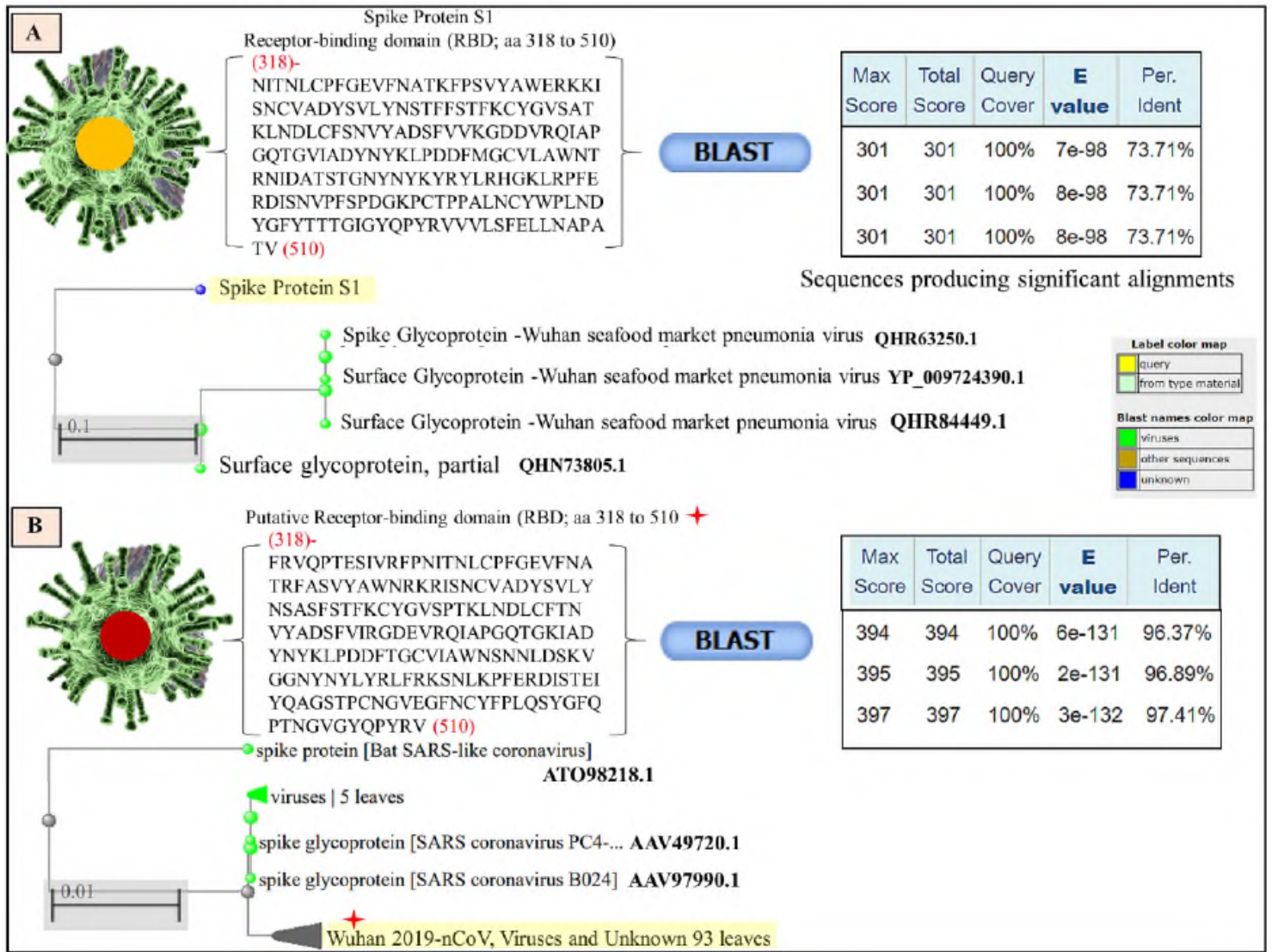
Figure 4

The BLASTp and reverse BLASTp results of RBD domain of SARS-CoV and COVID-19. (A) BLASTp results of the sequence 318-510th amino acids of RBD of SARS- CoV found a homolog in COVID-19 protein sequence identified as spike protein and sGP. The evolutionary tree developed showed them to have a common ancestor. (B) Results of reverse BLASTp and the evolutionary tree of putative RBD (318-510 aa) fetched the RBD of SARS-CoV and common origins of both sequences.
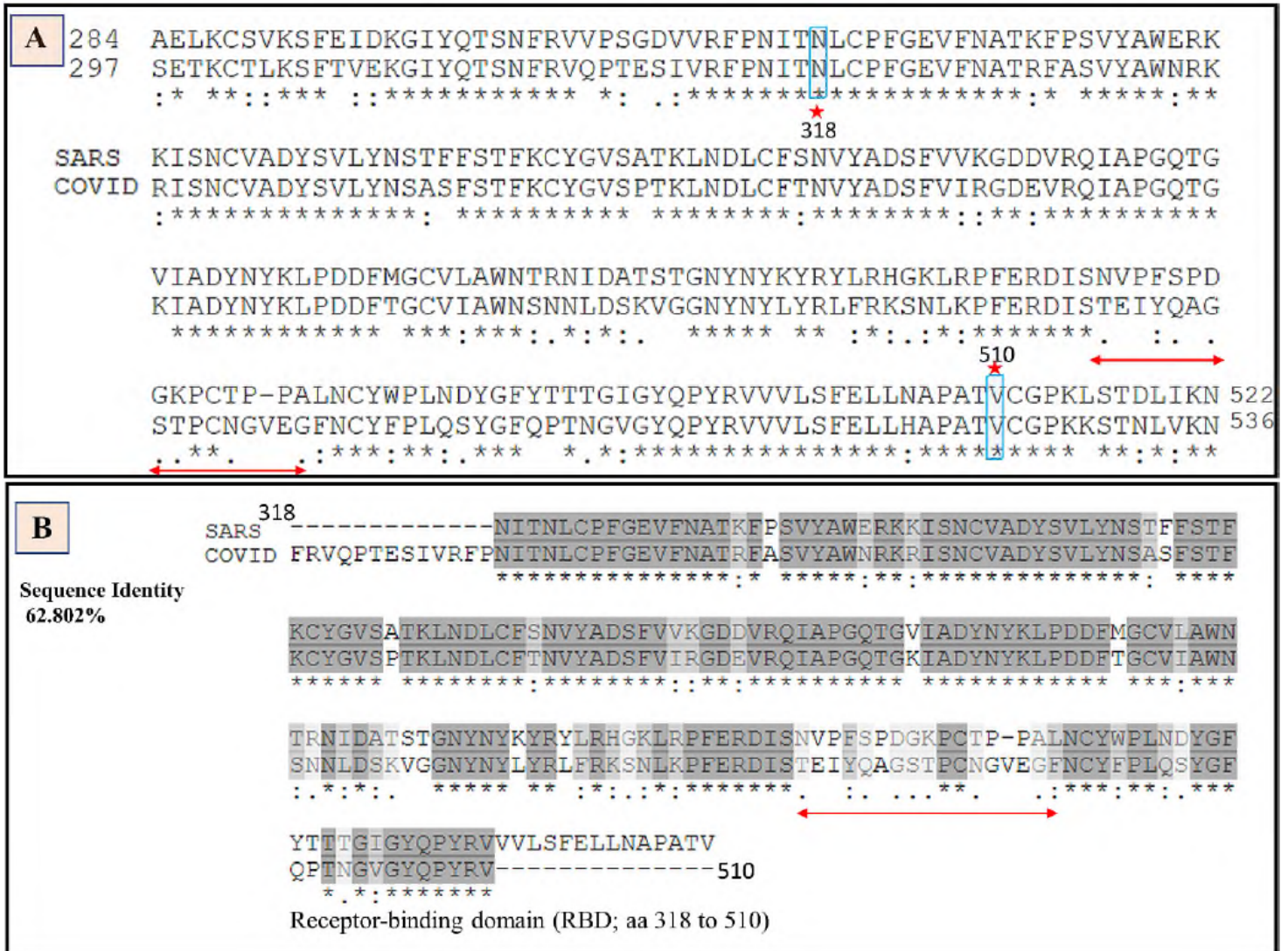
Figure 5

Sequence alignment of RBD of SARS-CoV (top-row) and COVID-19 (bottom-row). (A) Rows show the segment of RDB of SARS-CoV already known and stretches between 318th amino acid (marked in a blue rectangle in the top-rows) and 510th amino acid (blue rectangle bottom-rows). In between segments are highlighted areas (horizontal red arrows) where mutations have occurred in the putative RBD sequence of COVID-19.
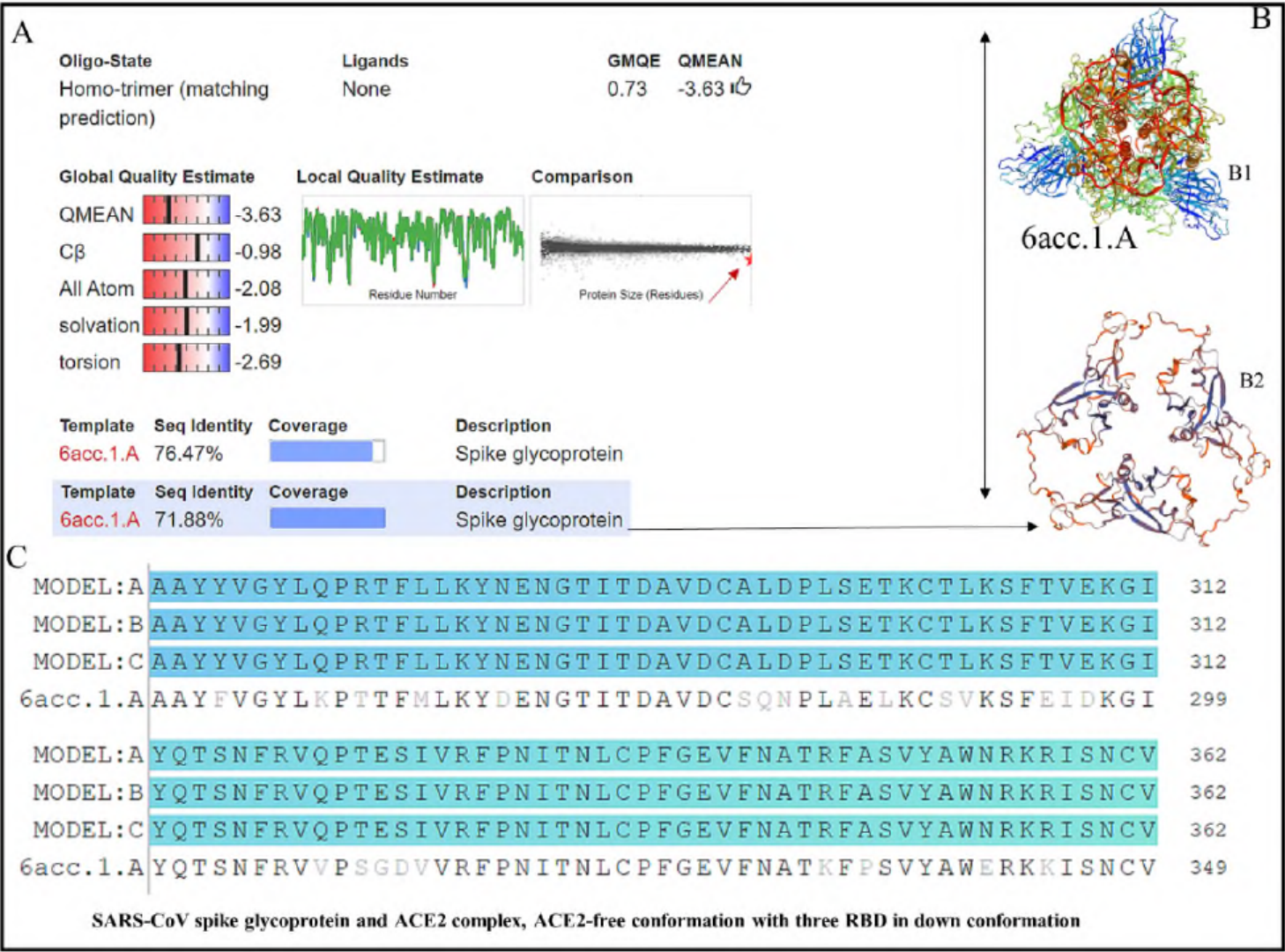
Figure 6

(A) The sGP of the COVID-19 in homology modelling developed spike glycoprotein and ACE2 complex with 76.47% identities. The quality estimation by GMQE and QMEAN values of 0.73 and -3.63 respectively showed it to be near-identical to SARS- CoV sGP. (B) The model built used the template 6acc.1.A (B1). The putative RBD of COVID-19 (318-510 aa) with 71.88% sequence identities developed a model (B2) of the spike protein of SARS-CoV.

Figure 7

The sequences of RBD of SARS-CoV and putative RBD of COVID-19 were submitted for MHC-Class-II allele binding needed in antigen presentation. Note the difference in sequences of the peptides predicted and their scores as predicted percentile ranks. A small numbered percentile rank indicates high affinity, which is the case for COVID-19 peptides when compared with SARS-CoV RBD.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Supplementalfigures.pdf
- Table1.png