

COVID-2019 Genome Sequence Analysis: Phylogenetic Molecular Evolution and Docking of Structural Modelling of Receptor Binding Domain of S Protein in Active Site of ACE2

Abaysew Ayele^{1*}, Baba Abdissa¹, Dereje Taye¹, Bereket Yemane², Rita Singh Majumdar³

¹Department of Startup and Innovations Promotion, Technology and Innovation Institute (TECHIN), Addis Ababa, Ethiopia

²Department of Diplomacy and International Relation, Addis Ababa University, Addis Ababa, Ethiopia

³Department of Biotechnology, School of Engineering and Technology, Sharda University, Greater Noida, India

Email: *betre.auroun@gmail.com

How to cite this paper: Ayele, A., Abdissa, B., Taye, D., Yemane, B. and Majumdar, R.S. (2020) COVID-2019 Genome Sequence Analysis: Phylogenetic Molecular Evolution and Docking of Structural Modelling of Receptor Binding Domain of S Protein in Active Site of ACE2. *Computational Molecular Bioscience*, 10, 95-110.

<https://doi.org/10.4236/cmb.2020.103007>

Received: June 28, 2020

Accepted: September 13, 2020

Published: September 16, 2020

Copyright © 2020 by author(s) and Scientific Research Publishing Inc.
This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

Meanwhile the outbreak of the Covid-19 since December, 2019 in China, it has killed more than a hundred thousand of people of all ages and sex across the globe in a short span of time. On the bases of this study the nearest family member of the virus and its receptor binding domain of S protein including its model structure and function of its active sites were naked through Multiple Sequence Alignment, modelling and molecular docking software accordingly its repository genome databases. The virus was genetically associated and molecular evolutionary related with (*RaTG13*) and it scores 96.12% homology with 99% query coverage followed by *bat-SL-CoVZC45* and *bat-SL-CoVZXC21* notch 89.12% and 88.65% respectively. However, SARS and MERS corona type virus those outbreak earlier respectively less likely family members of 2019-nCoV. Though the virus has a close genetic association with those previous SARS coronaviruses, and certainly the spike protein used as a binding receptor to fight against human receptor protein of ACE 2, but on the basis of FRODOC and HDock server analysis multi favorable active sites of S protein was discovered such GLN493 shown as a finest key in both model and possessed a unique traits on it resulting unexpected rate of transmission and number of people died while compared to the previous one. TYR500, ASN501, GLN498 and others residues preferably contemplate site also. In particular, the diversity of the virus in the world may be due to the genome structure of the virus and S gene changed over the time, across the world against to host of human genetic diversity, which may be more robust, and may be a new and unique feature. This is because it is characterized close to contact with

distance divergence between wild type novel coronavirus which was risen from China against to the genomes from Lebanon, India, Italy, and USA and so on. Thus, the World Health Organization and its researchers should focus on immunologic research and effective drug and vaccine development that will help to address the epidemiology of the virus, which can provide a long-term solution.

Keywords

MSA, Phylogenetic Construct, Genome Seq, RBD, Conserved Gene, ACE2, S Spike Protein, Genetic Mutation and Protein-Protein Docking

1. Introduction

Novel Coronavirus (2019-nCoV) according to the world report, the virus was discovered in a seafood market in the Chinese state of Wuhan, Hubei since late December, 2019 and now the virus classified a pandemic as the outbreak has spread to around the world and is currently infected in over 188 countries. Thus, up-to-date August-05-/2020, approximately 18.5 million total confirmed cases and 701,455 fatality cases have been verified (<https://coronavirus.jhu.edu/map.html>). Coronavirus causes pathogenicity for the various of animals and including humankind, and resulting severe lethal respiratory associated cases were recorded against to pigs, dogs, avian, bovine, masked palm civet, camels horse as well humans for the last decades [1] [2]. Primarily coronavirus incidence as SARS in 2002 and HKU3-1 to HKU3-3 were identified in the horseshoe bats (*rhinolophus-non-cave* species) in 2005 from Hong Kong and was thought as reservoir of the virus and will be responsible for future epidemic [3] [4]. Himalayan masked palm civets (*paguma larvata*) also and it was considered to be the natural reservoir of the virus and may have served as intermediate host between bat and the first human cases.

Consequently, remarkably two of coronavirus SARS (Severe Acute Respiratory Syndrome) and MERS (Middle East Respiratory Syndrome) are going to invade the human population across as epidemic in world. In November 2002, SARS coronavirus was outbreak in Guangdong, southern part of China and then spread to 29 countries in 2003. Due to this adverse plague, more than 774 fatalities were recorded out of 8000 confirmed cases and noted low transmission rate against to its high 9.5% mortality rate. Followed by, since 2012 the outbreak of MERS originally from Saudi Arabia and cause for the severe death of 858 out of total 2494 total confirmed cases and were registered as high 35% mortality rate in the world [5] [6] [7]. But (2019-nCoV) with a 3.8% fatality cases rate even high transmission proportion has been recorded globally [8].

We did not find it easy to know the meticulous origin of the virus, the receptor binding proteins (RBD) and its mutational change to talk about the distinc-

tiveness or diversity of the virus, as opposed to telling it even though the clinical symptom of a novel coronavirus-2019 (COVID-19) has much associated to the well-known respiratory tract infections including SARS and MERS those happened previously as of symptomized of high fever, frequent caught and as radiology report an invasive lesion of bronchial in both lungs [9] [10] [11].

Coronavirus belongs to the genus *Betacoronvirus* and positive sense RNA respected proteins can directly translated from mRNA, which contains in the range of 28 - 32 kbp genome size and capable to code around 15 notable genes. Due to exceedingly motivation of researches intercontinentally, many open genomes data from unlike patients (COVID-19 positive) from numerous countries has been sequenced and deposited at public repositories, but not yet fully accepting of the origin of the virus, receptor binding domains (RBD), polar contact site and the ways of mechanism of entry to receptor proteins in the host species. The clinical spectrum of (2019-nCoV) so high and in this study revealed the genetic family members of animals to avoid close contacts of them and mechanism of inhibitors against to its entry which gives us a very vital information for preventives measurement to minimize an existing fast transmission rate of human to human, mortality rate and keep strategic measurement in the future.

2. Result

2.1. Homology of Genomic Sequence and Genetically Mutation Divergent Analysis

From NCBI-gene bank 5000 expected target multiple sequences were retrieved and computed through blastn using standard databases (GenBank + EMBL + DDBJ + PDB + RefSeq sequences, but excludes EST, STS, GSS, WGS, TSA, patent sequences as well as phase 0, 1, and 2 HTGS sequences and sequences longer than 100 Mb). The database is *non-redundant*. Identical sequences have been merged into one entry, while preserving the accession, GI, title and taxonomy information for each entry. Based on alignment score, statically value (E-value) and query coverage, high quality of 25 genomes were selected for further genomic analysis, and we found that the genome sequence of 2019-nCoV were much identical to 96.12% with genome query coverage of 99% against to bat coronavirus (*RaTG13*); accession number: MN996532.1. Followed by Bat SARS-like Cov bat-SL-CoVZC45; accession number: MG772933.1 and Bat SARS-like Cov bat-SL-CoVZXC21; accession number: MG772934.1 took genomic homology score around 89.12% and 88.65% with query coverage of 95% and 94% nucleotide sequence respectively. Even if the divergence among 2019-nCoV and SARS like coronavirus are looking small percentage, more than 3% dissimilarity is too significance in genome study. It has just shown how 2019-nCoV family to SARS corona virus. Though the first submitted wild genome sequence which were isolated from USA and Wuhan-Chain novel coronavirus 2019 positive patients were practically around 99.98% nucleotide sequence identity with query coverage of 100% of each genome with accession number MT159722.1 and NC 045512.2

correspondingly, but right now it has keeping going on divergent too across the world as the geographical locations, over elapsed time and as host human genetics is changed.

If takes a trajectory of genomes sequence deposited from Lebanon, India, Taiwan, Uganda, Bangladesh, USA and so on countries had shown high S protein amino acid mutation took place and more than 15% significance divergent of genome sequence was analyzed through next strains hCoV-2019 virtual platform in dataset of epidemiology center (**Figure 1**).

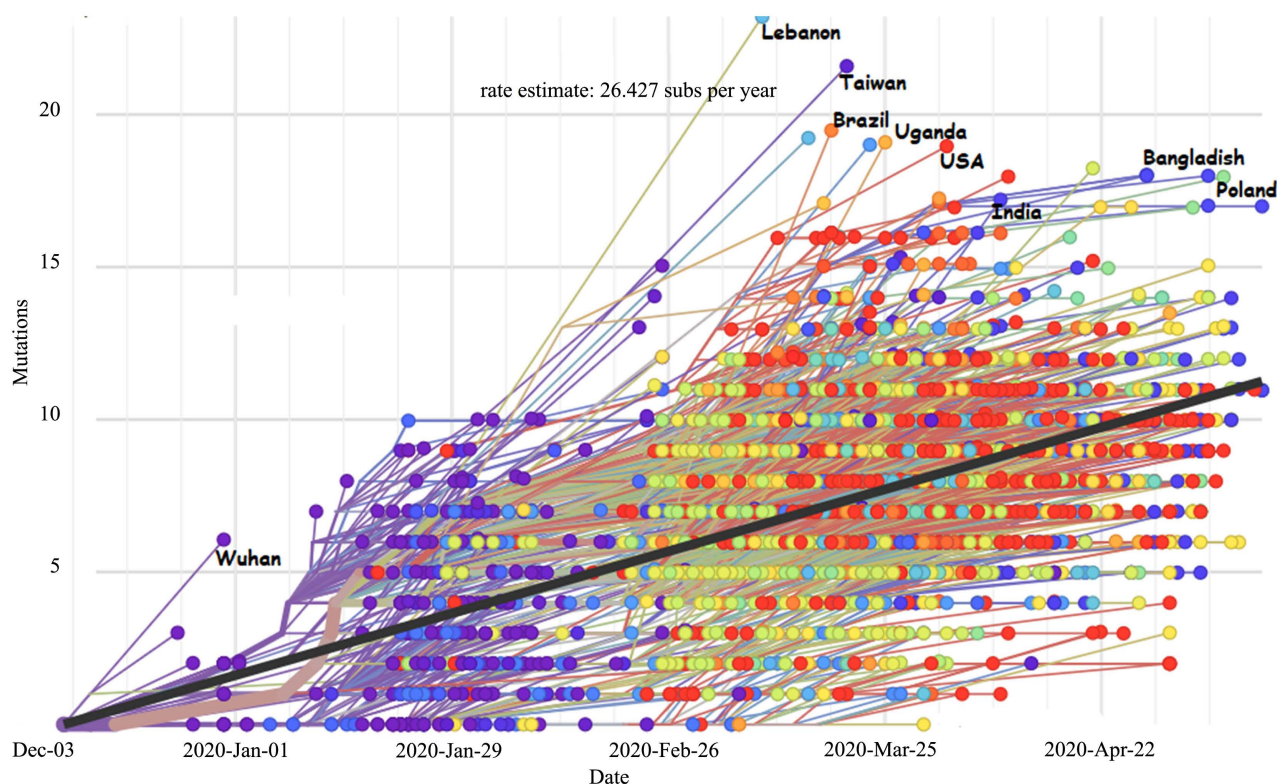


Figure 1. Up-to-date of 22-Apr-2020, Over 4000 novel coronavirus genome sample and its phylogeny and genetically mutational divergent across geographical locations, over elapsed time and host human genetic diversity globally. In particular, the countries whose names appear on the graph are taken as a sample where the virus genetically changed in fast.

(Source: <https://www.gisaid.org/epiflu-applications/next-hcov-19-app/>).

2.2. Prevalence of COVID-19 and Its Vaccine Development Program

Due to its rapid genetic variation, the distribution of Covid-19 across the world is so exceptional and unprecedented numbers of cases are going on reporting from different countries into world health organizations (WHO). Except a few countries, almost all countries has suffered from this pandemic as a map presented in **Figure 2**. Evidence suggests that it has been growing rapidly since 2020-June in contradiction of in the first few months. On the basis of WHO reports, Today, the number has risen to 18.5 million cases and at around 0.7 million fatality cases was verified at all age, sex and any genetic variations (exceptional study is needed).

Although diagnostic capabilities and diagnosis kits have been increasing over time, it has not been able to stop the spread of the virus because it has allowed it to spread rapidly with its genetic properties. Therefore, researches into the urgency of the development of effective vaccine ought to be considered and should be taken as the main solution [12]. There are currently no Food and Drug Authorization (FDA) licensed vaccines to prevent COVID-19, but repurposing drug research is going on through chemo informatics. Mainly, commercial vaccine manufacturers and other entities are developing COVID-19 vaccine candidates using different technologies including DNA vaccine, RNA and protein, or viral vectored vaccines so far, a number of pharmaceutical companies and academic institutions worldwide have launched their programs on vaccine development against to 2009-nCoV. Consequently, developing of effective vaccine/drugs should be targeted on receptor binding proteins or arrested transcriptional and replication process of the virus because the virus genetic makeup is going on altered. On the other hand, the prevalence of the virus contagion impact on the global economy is not insignificant in addition to its infections troubles. Thus, till the development of effective vaccines and medicines, the people and the government of the world community should work together to halt the spread of the virus by identifying the prevalence of the disease and related vulnerable areas.

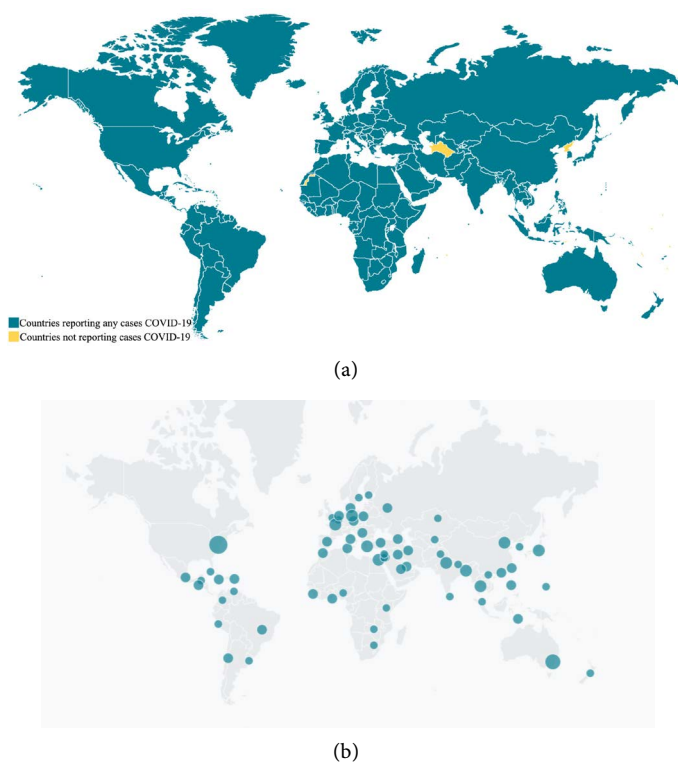


Figure 2. (a) The distribution of SARS_2 pandemic cases reported from world countries vs countries not reporting cases of COVID-19 (Source: CDC-Center for Disease Control and Prevention 20 July 2020); (b) The geographical distribution of genome sequence coverage those marked countries have submit to NCBI and those unmarked countries have not submit (Source: <https://www.ncbi.nlm.nih.gov/labs/virus/vssi/#/sars-cov-2>).

2.3. Phylogenetic and Molecular Evaluation Genetic Analysis

Phylogenetic and molecular evolutionary genetic analysis were conducted using MEGA software (version 10.1.7) using neighbor-joining statistical approaches and a single amino acid substitution [13]. Poisson model along phylogeny test of 500 bootstrap replication methods were shown on coding region of genomes only and found out (bat coronavirus *RaTG13*: accession number: MN996532.1) which where perversely isolated from Yunnan Province is consider to be the nearest family member of novel coronavirus (2019-nCoV) followed by bat-*SL-CoVZC45*; accession number: MG772933.1, and Bat SARS-like CoV bat-*SL-CoVZXC21*; accession number MG772934.1). However, Middle East Respiratory Syndrome (MERS: accession number: KT806053.1 and MK564474.1) coronavirus was disclosed as a genetic distance relationship among other than corona types based on its phylogenic inferred (**Figure 3**). This finding provisionally took a scientific evidence as per [4] [14] reported *Rhinolophus affinis* (intermediate horseshoe bat) was consider a natural reservoir of the coronavirus and supposed to a source of future epidemic. Moreover, the genomes of RNA-dependent RNA polymerase gene of *RaTG13* at 88% homology with 2019-nCoV, and *Rhinolophus affinis* or *Rhinolophus sinicus* was consider to be a source of novel corona virus.

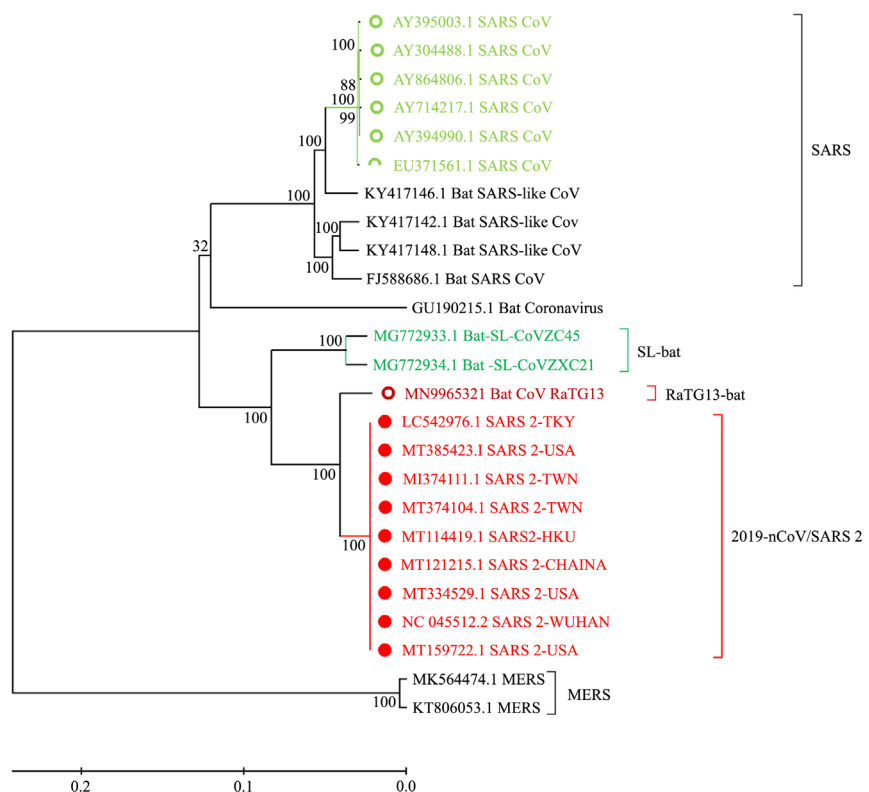


Figure 3. Phylogenetic molecular evolutionary relationship of novel coronavirus (2019-nCoV) sequenced from different countries (red color) against to corona type virus genome isolated for the last years such as bat SARS-like CoV: MERS and SARS coronavirus: (Note: **SARS-2**: Severe Acute Respiratory Syndrome-2; **MERS**: Middle East Respiratory Syndrome and **Bat CoV**: Bat coronavirus).

2.4. Multiple Sequence Alignment (MSA) of COVID-19

High quality of 25 genome sequence of COVID-19 and including corona type those were happened as plague around the world for the last years were selected for further analysis. Each selected genome sequence of 2019-nCoV against to any corona type targeted multiple sequence alignment (MSA) was performed through ClustalX software (version 2.0.10) [15]. Though the conserved regions segmented entire complete genome, there was nucleotide substitutions, deletion and insertion along whole genomes, specific a large deletions had recorded in the range of 3340 up to 3470 nucleotides position crossways all genome sequence of corona type such as SRAS, MERS and bat like SARS counter to novel coronavirus (COVID-19). Mainly, there is high conserved region was noted between SRAS 2 (2019-nCoV) and RaTG13 (MN996532.1) shadowed by SL-bat (MG772933.1) bat coronavirus the sample was collected from China. Unlikely MERS (MK564474.1 and KT806053.1) poor conserved region was recorded against to COVID-19 (Figure 4). We found the above homology identity and phylogenic to be a good confirmation as it has stretched out in the order of a single nucleotide genome map. It has shown clear gap, miss gap and deviation at all and 2019-nCoV contains substitute nucleotide unlike SRAS and MERS in the range of 3363 to 3422 was charted a big miss gap.

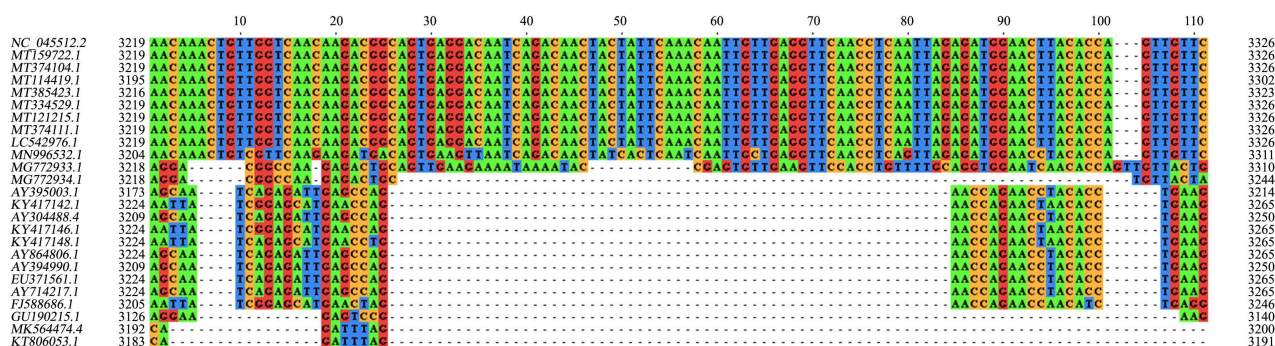


Figure 4. Multiple Sequence Alignment of each selected 2019-nCoV genome sequence against to corona type those were outbreak before many years back. **SARS** (AY395003.1 AY304488.1, AY864806.1, AY714217.1, AY394990.1 and EU371561.1); **Bat SARS-like** (KY417146.1, KY417142.1, KY417148.1 and FJ588686.1); **RaTG13** (GU190215.1), **SL-bat** (MG772933.1 and MG772934.1); **2019-nCoV** (MN996532.1, LC542976.1, MT385423.1, MT374111.1, MT374104.1, MT114419.1, MT121215.1, MT334529.1, NC 045512.2 and MT159722.1); **MERS** (MK564474.1 and KT806053.1).

2.5. Targeted S Spike-Surface Glycoprotein of COVID-19

Amongst all notable genes, S-gene translated and expressed into S or spike surface glycoproteins which support the binding affinity and fusion of the virus to pass in to the surface of host cells particular where a receptor protein Angiotensin Converting Enzyme II (ACE2) which expressed in alveolar cells of the lung, esophagus upper and stratified epithelial cells, absorptive enterocytes from ileum and colon [16]. 13 annotated and curated spike proteins sequence from Uniprot and NCBI were retrieved and computed Multiple Sequence Alignment (MSA) techniques through Mafft software (version 7.463) all parameters devour as per L-INS-I (Accurate oriented) [17], and each amino acid position was

viewed over Jalview (version 2.10.0.) [18]. The identity score was done and functional spike surface glycoproteins those sequenced from novel coronavirus those are submitted from different countries such that surface glycoprotein of SARS 2 (QJF75467.1) and surface glycoprotein SARS 2 QJS39567.1) had score 1.0 identity. Furthermore, synthetic construct (QJE37812.1) and spike glycoprotein of *RaTG13* (QHR63300.2: “from China” collection date was 24-Jul-2013) was accounted 0.97 and 0.98 distinctiveness respectively followed by bat-SL-CoVZC45 “from China” collection date was 2017, accession number: AVP78031.1 was scored 0.82 conserved sequence respectively. Despite of an average proteins sequence takes more than 0.8 identity, an accession number of YP_009047204.1 spike glycoprotein a MERS related coronavirus isolated on 13-Jun-2012 was scored 0.34 distinctiveness. In somehow, spike proteins of 2019-nCoV is consider to homology to SARS and descendent from the same gene family and function of spike protein could binding receptor for its attachment into host cells.

Phylogenetic molecular evolution analysis of 13 proteins were also conducted and noted that bat coronavirus *RaTG13* which isolated from China with accession number: QHR63300.2 was the nearest family members of the novel coronavirus followed by spike protein (accession number QIA48632.1) were isolated from PCoV_GX-P4L Malayan pangolin coronavirus from China since 2017. Bat SARS-like CoV of bat-*SL-CoVZC45* (accession number: MG772933.1) and Bat SARS-like CoV of bat-*SL-CoVZXC21* (accession number: MG772934.1) were also registered as close family member of novel coronavirus, but distance slightly related to Bat SARS-like coronavirus of spike protein (accession number: ATO98205.1 and ATO98157.1) (**Figure 5**). Thus, rendering to the genomic corresponds, the

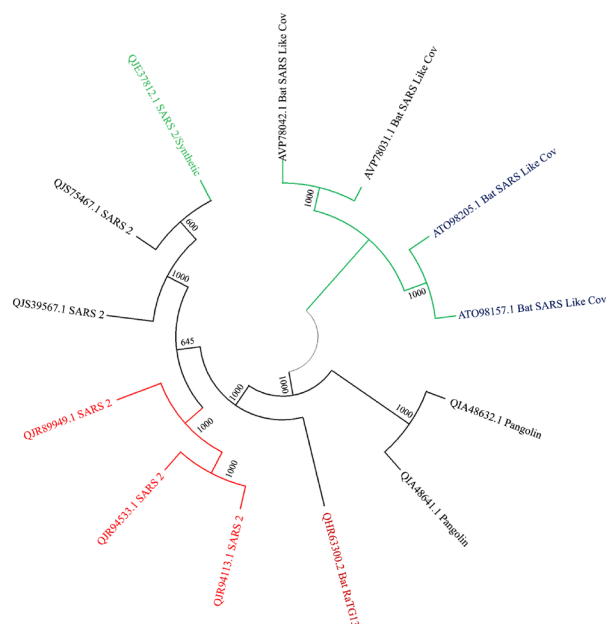


Figure 5. Phylogenetic molecular evolution analysis of 13 annotated spike surface glycoproteins isolated from **SRAS 2** (2019-nCoV of each), **MERS** (Middle East Respiratory Syndrome) and **Bat like SARS** CoV which were measured as epidemic diseases for the last many years.

molecular structure and function of proteins of 2019-nCoV descended much towards to bat SARS-like coronavirus which was outbreak before and declared as a cause of thousand fatality cases in the world. It also has strengthen our finding such that spike protein or S gene of novel coronavirus functioning as receptor binding for host receptor cells Angiotensin Converting Enzyme II (ACE 2) due to its ancestral phylogenic functional genomic analysis although it has its own unique character and will continue to alter its genetic makeup. Hence, we can confidently conclude that SARS particular, *RaGT13*, *SL*-bat corona virus was the origin of 2019-nCoV. Although the exact cause of the virus may not be ascertained, it might had been an unknown animal contact where a market sold sea food in Wuhan city or an unfamiliar contact with a domestic animals or on the other hand, it means that there is no clear indication of the exact origin and condition of the onset of viral infections.

2.6. Conserved Regions of Receptor Binding Domain (RBD) of S Proteins

S-spike protein of coronavirus is an envelope glycoprotein that plays the most important role in viral attachment, fusion, and entry into host cells, and serves as a major target for the development of neutralizing host antibodies, inhibitors of viral entry, and vaccines. It is synthesized as a precursor protein that is cleaved into two parts an amino or N-terminal S1 subunit and carboxyl or C-terminal S2 subunit that mediates attachment and membrane fusion, respectively. Among selected four of homology amino acid sequence of coronavirus, it was recorded highly conserved region at most, exclusively complete conserved in the range of 928 and 1023 was noted, but visible variation of amino acid, some substitution and deletion was recorded among selected spike protein of 2019-nCoV. In the range of 375 and 510 amino acid position among bat SARS-like CoV (AVP78042.1 and AVP78031.1), *RaTG13* (QHR63300.2) and 2019-nCoV (QJF75467.1) less likely conserved region and variation of codons was recorded. Even though spike proteins (Receptor Binding Domains and Receptor Binding Motifs) found out nearby members of bat like SARS coronavirus, definitely it has unique properties of binding affinity against to human Angiotensin converting enzyme-2 and it's contiguous due to those amino acid mutational changes at active site of its attachments. Depending on the coronavirus strains, C-terminal or N-terminal serve as receptor binding to their receptors. Utmost SARS and 2019-nCoV use C terminal to bind in host animal receptor protein. This binding affinity took place between in the range of 375 and 510 amino acid numbering of spike proteins of novel coronavirus according to their different homology modeling and the change of residues. In a nutshell, as it has shown in **Figure 6** there was shown frequently changed amino acid residues of targeted site of 2019-nCoV unlike to previous bat SARS-like coronavirus (AVP78042.1) such as VAL478 changed to ASN501; ALN435 changed to ASP439; SER470 changed to GLN; PHE486 and GLY485 were added; SER469

changed to DLN493; VAL477 changed to ASN500; ASN476 changed to TYR500; ASN476 changed to TYR500; ASN474 changed to GLN498; GLY439 changed to LYS44; HIS440 changed to ASN450; TYR481 changed to TYR505 and TYR449 and VAL445 were added into 2019-nCoV. It is difficult to say that this change was arbitrary and there was no definitive study yet to be confirmed the mechanism of this chaos amino acid change and this study should confirmed through in vitro laboratory using animal model and continue study the expression of mutational S gene for its phenotype trait.

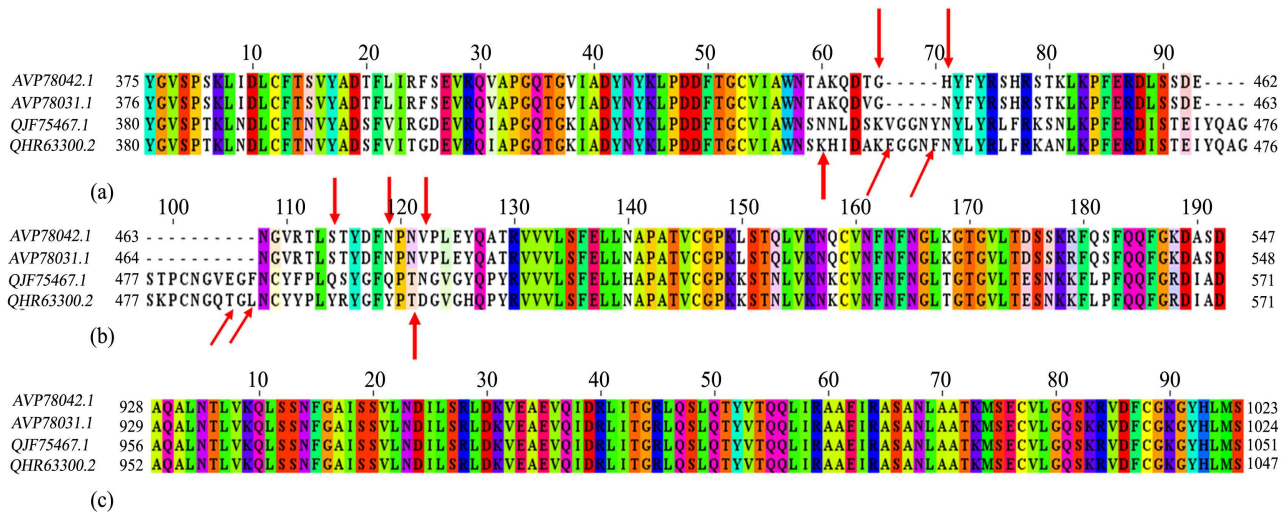


Figure 6. Conserved and non-conserved region between the four selected homology spike proteins of bat SARS-like CoV (AVP78042.1, AVP78031.1) and *RaTG13* (QHR63300.2) and 2019-nCoV (QJF75467.1) (a) Amino acid (aa) variations and deletions (b) Amino acid residues deletions and substitutions (c) Complete conserved region of spike proteins.

2.7. Homology Modeling of S Protein and Protein-Protein Interactions

Sequence of spike protein of 2019-nCoV (Acc: YP_009724390.1) which sample was collected from Wuhan sea food market positive patient as a reference and its homology modeling predictions and analysis were performed through Phyre2 server (Protein Homology/analogy Recognition Engine V 2.0) and high quality and $\geq 90\%$ confidence score of eight alignment was selected based on heuristics to maximise confidence, percentage identity and alignment coverage [19]. Such that out of them template c6xr8C_ of PDB: viral protein was scored 100% confidence and 99% alignment coverage and be candidate for further structural and functional analysis through protein-protein docking. The candidate of S spike protein of 2019-nCoV was characterized based on modeling prediction analysis and used for docking between S protein and ACE2. Out of hundred prediction, 10 paramount docking model predictions was generated from HDOCK online server [20]. Specifically two of them model (A) and Model (B) were selected based on minimum docking energy score (-237.70 k/J) and (-231.48 k/J) respectively and resulting high affinity interaction was record-

ed between them as it has shown in **Figure 7(a)**. The surface of S pike protein in accessible to solvent at (1.4 Angstrom probe) was done using FRODOCK server (Job ID 5613947) as the parameters published on [21] and the surface of area of the protein and its hydrophobic effects was shown in **Figure 7(b)**.

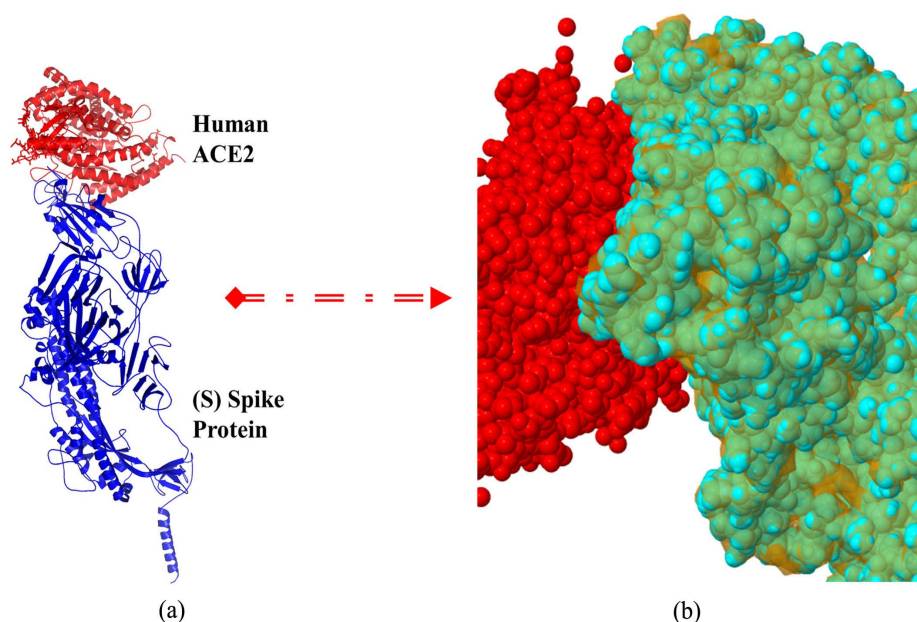


Figure 7. (a) Phyre2 structural homology modeling of surface glycoprotein of 2019-nCoV of template c6xr8C_ and human ACE2 of PDB: (1r42) (molecular docking (b)) Space filled molecular surface interaction of solvent accessible surface (1.4 Angstrom probe) (S spike protein of water shaded = Cyan color; ACE2 = Red color).

However, close contact between the molecules between them is very important to know the function of protein by identifying the particular active site or polar contacts. The surface of polar contact was visualized and analyzed using pyMol interface software (2.3.0) [22].

The polar contact of spike protein and ACE2 visualized in terms of model A and model B and active sites of amino acid was picked out. Although it has been published in previous editions [23], ASN501, ASN439, GLN493, GLY485 and PHE486 was supposed to active binding site of spike protein, it is now based on the results we have found exceptional LYS444, TYR505 and GLN493 from Model A or THR500, GLN498, GLN493 and VAL445 active residues from Model B was noted through pyMol analysis. Predominantly, GLN493 active site be could create more interactive bonds between the molecules and enhance the entry of the viral particles into human cells conceitedly in **Figure 8(a)** and **Figure 8(b)**. As a result, the mutational change of S gene impacted on translation of spike protein and makes it different variable bond interactions against to ACE2 human receptor protein such unique trait makes it easy to entry to respiratory cells.

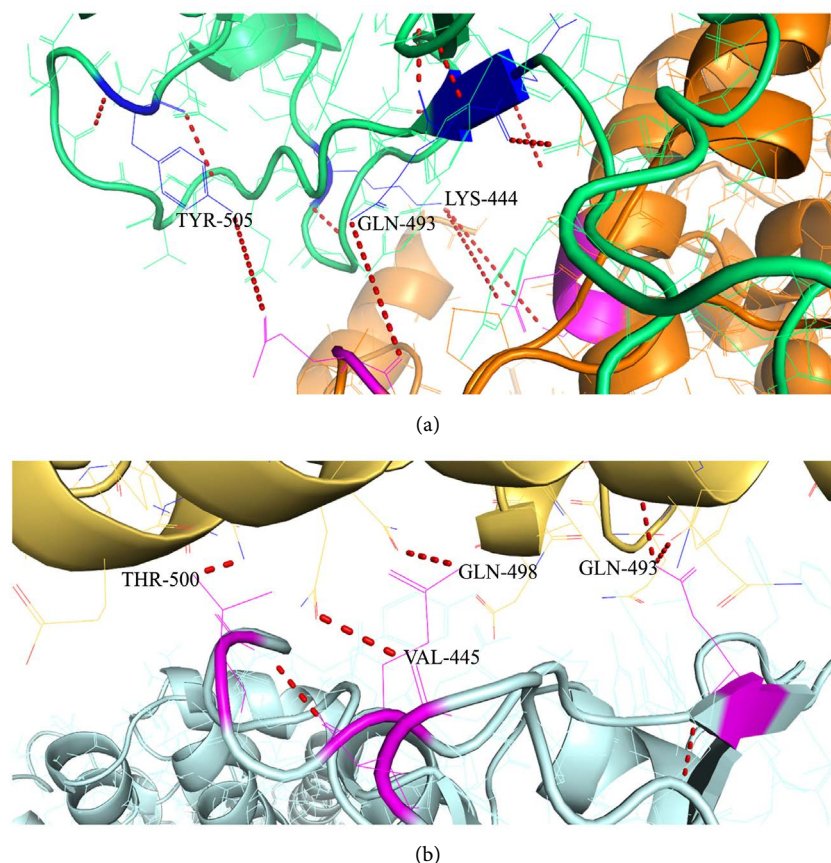


Figure 8. pyMol software interface visualization of molecular docking of protein-protein interactions of polar contact between homology structure of S Spike surface glycoprotein of 2019-nCoV template: c6xr8C_PDB: viral protein and ACE2 of PDB: _ (1r42); a) Model A: S protein = Green color; ACE2 = Orange Color; b) Model A: S protein = Grays color; ACE2 = Gold).

3. Methods

Open biological big genomic data of 2019-nCoV and any corona type virus virtually retrieved from NCBI gene bank (<https://www.ncbi.nlm.nih.gov/>) and related dataset as targeted of 5000 genome sequences of corona families out of hit blastn result 25 genome sequence of irredundant high percentage homology was selected based on their respected E-value, score identity and query coverage. Hit blastp at reference sequence of novel coronavirus and 13 annotated S spike surface glycoproteins sequence was nominated from NCBI and UniProt curated database (<https://www.uniprot.org/>) according to their genome percentage homology. Conserved domains were retrieved through CD-search (<https://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi>) and shown how the spike sub units conserved over the time. MEGA software used for set constructing of evolutionary phylogenetic tree for genomic sequence and protein amino acid residues. JLVIEW, CLUSTALX, and MAFFT software was accompanied for multiple sequence alignment of the proteins and genome sequence of coronavirus as well. Retrieved virtual genomic mutation analysis from virtual nextstrain

hCoV-2019 dataset and review the prevalence of the coronavirus from World Health Organization and CDC. Finally, protein homology modelling was performed and characterized through phyr2 online server and the S spike protein and ACE2 (PDB: 1r42) protein-protein interactions was plotted and analyzed over molecular HDock and FRODOCK server. Interface visualization of contact polar between each active sites of amino acid residues was done using pyMol software.

4. Conclusions and Discussions

Since the virus started in December 2019 in China, millions of people have been infected throughout the world, and hundreds thousands of deaths have resulted in deaths. While there are some deletions, substitutions and mutations that are generalized to the virus, the coding region genome of the virus is already in ancestral relationship with some of the previously created types of coronavirus, such as SARS, MERS and bat SRAS-like like CoV. In particular, it is strongly associated with the bat SARS-like corona virus (*Rhinolophus affinis*, *Rhinolophus sinicus* and *SL-bat strain*) based on evolutionary analysis through mega software. In addition to that the spike glycol surface protein used for adhesion and binding receptor into ACE2 has a high similarity to the previous *RaTG13* and *SL* strains of coronavirus as well synthetic spike proteins over finding of mafft Multiple Sequence Analysis. After all, though, it is believed that the bat SARS-like coronavirus will eventually become a pandemic in the world [14] and the present problem shows a lack of focus on early researches and a lack of preparedness. Bat SARS coronavirus eventually mutated and cause for amino acid change at site of S surface protein specifically GLN493 had strong bond while it was flipping the side in different position. Even though bat is not purchased or sold at market now, the people who used to play with it or consumed animal and its products should care on exclusively, but we strongly have suggested to keep on physical distance until full potential and official vaccine development because the virus already has adapted a human serology too much and able transmit easily rather than in animated. Altering of active amino acid of the binding protein makes the virus more contagious since it will have rotatable attachment or adherence capacity for human receptor protein. As a result the virus getting unconstrained chance to entry to the human cells and being contagious from human to human within fast transmission rate.

Hence, the study's facts indicate that the current distribution of the virus should be examined in terms of genomic diversity to realize its distance divergence across the world. Particularly, the mutation that takes place on S-gene would be complicated the viral for vaccine development program since its hot spot to be multivalent against to ACE2. The wild genome, which originally originated from Wuhan sea food market, China on the basis of geographical location, relapsed time, and host human genetic variation, is currently undergoing significant divergence, specifically, up to 25% of the variance in India, Italy and the United States has been documented through epidemiology dataset online analysis platforms. If it

continues with its virulence severity, the next most likely robust viral strain, SARS-3 coronavirus, is likely to be broader. It is therefore a major way to curb the spread by concentrating on the types of severe strains that are highly concentrated and studied in a comprehensive manner worldwide. Next, an unprecedented amount of financial and technical support should be made for researches undertaken globally to find the best drugs and vaccines internationally, with great emphasis on the World Health Organization. Somehow, we suggested to investigate the crystallography of spike proteins and would be proofed through in vitro lab and understand the whole genetic materials machinery process particularly those participated on RNA transcription, protein translations, surface binding and cell division as per its progressive mutations and revealed the exact common active binding sites or epitopes for drug design and vaccine development.

Acknowledgements

The study was support by Technology and Innovation Institute (TCHIN2) which providing sufficient internet access for virtual big data analysis without interruption at all. We would like to acknowledge the institute and Thanks to everyone who allowed us to access their database and software and kept the data secure on their server for some time.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

- [1] Fehr, A.R. and Perlman, S. (2015) Coronaviruses: An Overview of Their Replication and Pathogenesis. *Methods in Molecular Biology*, **1282**, 1-23.
https://doi.org/10.1007/978-1-4939-2438-7_1
- [2] Kahn, J.S. and McIntosh, K. (2005) History and Recent Advances in Coronavirus Discovery. *The Pediatric Infectious Disease Journal*, **24**, 223-227.
<https://doi.org/10.1097/01.inf.0000188166.17324.60>
- [3] Lau, S.K., Woo, P.C., Li, K.S., Huang, Y., Tsoi, H.-W., Wong, B.H., Wong, S.S., Leung, S.-Y., Chan, K.-H. and Yuen, K.-Y. (2005) Severe Acute Respiratory Syndrome Coronavirus-Like Virus in Chinese Horseshoe Bats. *Proceedings of the National Academy of Sciences of the United States of America*, **102**, 14040-14045.
<https://doi.org/10.1073/pnas.0506735102>
- [4] Cui, J., Li, F. and Shi, Z.-L. (2019) Origin and Evolution of Pathogenic Coronaviruses. *Nature Reviews Microbiology*, **17**, 181-192.
<https://doi.org/10.1038/s41579-018-0118-9>
- [5] Peiris, J.S.M., Lai, S.T., Poon, L.L.M., Guan, Y., Yam, L.Y.C., Lim, W., Nicholls, J., Yee, W.K.S., Yan, W.W., Cheung, M.T., *et al.* (2003) Coronavirus as a Possible Cause of Severe Acute Respiratory Syndrome. *The Lancet*, **361**, 1319-1325.
[https://doi.org/10.1016/S0140-6736\(03\)13077-2](https://doi.org/10.1016/S0140-6736(03)13077-2)
- [6] Chan-Yeung, M. and Xu, R.H. (2003) Sever Acute Respiratory Syndrome: Epidemiology. *Respirology*, **8**, 9-14. <https://doi.org/10.1046/j.1440-1843.2003.00518.x>

- [7] WHO (2015) Middle East Respiratory Syndrome Coronavirus (MERS-CoV): Summary of Current Situation, Literature Update and Risk Assessment. <https://apps.who.int/iris/handle/10665/179184>
- [8] WHO (2020) SARS 2: Coronavirus Disease (COVID-19) Dashboard. https://covid19.who.int/?gclid=EAIaIQobChMIgpr3uLo6QIVxbvVCh3mPw1vEAYASAAEgI13_D_BwE
- [9] Wang, Q., Wong, G., Lu, G., Yan, J. and Gao, G.F. (2016) MERS-CoV Spike Protein: Targets for Vaccines and Therapeutics. *Antiviral Research*, **133**, 165-177. <https://doi.org/10.1016/j.antiviral.2016.07.015>
- [10] Huang, C., Wang, Y., Li, X., *et al.* (2020) Clinical Features of Patients Infected with 2019 Novel Coronavirus in Wuhan, China. *The Lancet*, **395**, 497-506. [https://doi.org/10.1016/S0140-6736\(20\)30183-5](https://doi.org/10.1016/S0140-6736(20)30183-5)
- [11] Niu, P., Shen, J., Zhu, N., Lu, R. and Tan, W. (2016) Two-Tube Multiplex Real-Time Reverse Transcription PCR to Detect Six Human Coronaviruses. *Virologica Sinica*, **31**, 85-88. <https://doi.org/10.1007/s12250-015-3653-9>
- [12] Calina, D., Docea, A., Petrakis, D., Egorov, A., Ishmukhametov, A., Gabibov, A., Shtilman, M., Kostoff, R., Carvalho, F., Vinceti, M., Spandidos, D. and Tsatsakis, A. (2020) Towards Effective COVID19 Vaccines: Updates, Perspectives and Challenges (Review). *International Journal of Molecular Medicine*. http://file:///C:/Users/User/Downloads/ijmm.2020.4596_AOP_PDF.pdf <https://doi.org/10.3892/ijmm.2020.4596>
- [13] Kumar, S., Stecher, G., Li, M., Knyaz, C. and Tamura, K. (2018) MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. *Molecular Biology and Evolution*, **35**, 1547-1549. <https://doi.org/10.1093/molbev/msy096>
- [14] Li, W., Shi, Z., Yu, M., Ren, W., Smith, C., Epstein, J.H., Wang, H., Crameri, G., Hu, Z. and Zhang, H. (2005) Bats Are Natural Reservoirs of SARS-Like Coronaviruses. *Science*, **10**, 676-679. <https://doi.org/10.1126/science.1118391>
- [15] Larkin, M.A., Blackshields, G., Brown, N.P., Chenna, R., McGettigan, P.A., McWilliam, H., Valentin, F., Wallace, I.M., Wilm, A., Lopez, R., Thompson, J.D., Gibson, T.J. and Higgins, D.G. (2007) Clustal W and Clustal X Version 2.0. *Bioinformatics*, **23**, 2947-2948. <https://doi.org/10.1093/bioinformatics/btm404>
- [16] Su, S., Wong, G., Shi, W., *et al.* (2016) Epidemiology, Genetic Recombination, and Pathogenesis of Coronaviruses. *Trends in Microbiology*, **24**, 490-502. <https://doi.org/10.1016/j.tim.2016.03.003>
- [17] Nakamura, T., Yamada, K.D., Tomii, K. and Katoh, K. (2018) Parallelization of MAFFT for Large-Scale Multiple Sequence Alignments. *Bioinformatics*, **34**, 2490-2492. <https://doi.org/10.1093/bioinformatics/bty121>
- [18] Waterhouse, A.M., Procter, J.B., Martin, D.M.A., Clamp, M. and Barton, G.J. (2009) Jalview Version 2: A Multiple Sequence Alignment and Analysis Workbench. *Bioinformatics*, **25**, 1189-1191. <https://doi.org/10.1093/bioinformatics/btp033>
- [19] Kelley, A., Stefans, M., Christopher, M.Y., Wass, M.N. and Sternberg, M.J.E. (2015) Phyre2 Is a Web-Based Tool for Predicting and Analyzing Protein Structure and Function. *Nature Protocols*, **10**, 845-858. <https://doi.org/10.1038/nprot.2015.053>
- [20] Yan, Y., Tao, H., He, J. and Huang, S.Y. (2020) The HDock Server for Integrated Protein-Protein Docking. *Nature Protocols*, **15**, 1829-1852. <https://doi.org/10.1038/s41596-020-0312-x>
- [21] Garzón, J.I., López-Blanco, J.R., Pons, C., Kovacs, J., Abagyan, R., Fernández, J., Recio, P. and Chacón (2009) FRODOCK: A New Approach for Fast Rotational Protein-Protein Docking. *Bioinformatics*, **25**, 2544-2551. <https://doi.org/10.1093/bioinformatics/btp447>

- [22] Schrodinger, LLC (2015) The PyMOL Molecular Graphics System, Version 2.3.0.
- [23] Lu, R.J., Zhao, X., Li, J., Niu, P.H., Yang, B., Wu, H.L., *et al.* (2020) Genomic Characterization and Epidemiology of 2019 Novel Coronavirus: Implications for Virus Origins and Receptor Binding. *Lancet*, **395**, 565-574.
[https://doi.org/10.1016/S0140-6736\(20\)30251-8](https://doi.org/10.1016/S0140-6736(20)30251-8)