

# Analysis of COVID-19 Spike protein mutations and their effects on its affinity towards human cell receptors

Abir Elbeji (✉ [abirelbeji@gmail.com](mailto:abirelbeji@gmail.com))

National Institute of Applied Sciences and Technology <https://orcid.org/0000-0002-5293-4337>

---

## Research Article

**Keywords:** Covid-19, Spike protein, Protein-protein affinity, bioinformatics

**Posted Date:** March 1st, 2021

**DOI:** <https://doi.org/10.21203/rs.3.rs-271106/v1>

**License:**  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

---

# Abstract

The novel Coronavirus SARS-CoV-2 (2019-nCoV) is a member of the family Coronaviridae and contains a single-stranded RNA genome with positive-polarity. In order to reveal the evolution mechanism of the SARS-CoV2 genome, in particular its spike protein; the main driving force for host recognition, we conducted a comparative analysis with Coronaviruses of different strains, including MERS-CoV, SARS-CoV1 and Pangolin Coronavirus. In addition, a comparative analysis between the newly sequenced SARS-CoV2 from different regions of the world has been carried out in order to understand the evolution of this novel virus throughout its transmission. Among all sequenced strains, the latest France HCoV was the least identical to the reference. Further investigations have therefore been performed and it has been concluded that this strain has undergone mutations which have increased its binding affinity to the Angiotensin-Converting Enzyme 2 (ACE2) receptor, thus hypothetically increasing its infectivity.

## Introduction

In December 2019, an outbreak of apparently viral pneumonia of unknown etiology emerged in the city of Wuhan, in the Chinese province of Hubei. On January 9, 2020, the Chinese health authorities and the World Health Organization (WHO) officially announced the discovery of a novel coronavirus (first named 2019-nCoV, then officially SARS-CoV-2, which differs from the viruses SARS-CoV<sup>1</sup>, responsible for the SARS outbreak in 2003, and MERS-CoV, responsible for an ongoing outbreak that began in 2012 in the Middle East).

This new virus is the pathogen responsible for this infectious respiratory disease called Covid-19 (Coronavirus Disease). Globally, by July 1, 2020 (10.00 pm CET), according to the WHO, 10,357,662 cases of 508,055 Covid-19 have been confirmed and patients have died. In addition to confirmed cases, there are also suspected cases of Covid-19, the definition of which is evolving as times passes and the epidemic propagates.

Spike protein is a crucial recognition factor for virus attachment and entry to the host cell. Also, it's amino-acid sequence is shorter than the whole genome, which is time saving and storage saving in term of coding on Python. Therefore, our study is only based on Spike proteins<sup>3</sup>.

In this study, we focused on determining the strain responsible of the new pandemic COVID-19 and its origin, through comparative analysis with different strains of coronaviruses issued from animals and human coronaviruses. Also, we highlighted a new hypothesis concerning the 'Best-Fit' activity of this virus while comparing binding affinities of France H-CoV and the reference H-CoV issued from China (Wuhan) complexed with ACE2.

### Objectives:

To reveal the strain responsible of the new pandemic COVID-19 and its origin, a comparative analysis with different strains of coronaviruses issued from animals and human coronaviruses was performed.

Also, we highlighted a new hypothesis concerning the ‘Best-Fit’ activity of this virus while comparing binding affinities of France H-CoV and the reference H-CoV issued from China (Wuhan) complexed with Angiotensin-Converting Enzyme 2 (ACE2).

## Materials And Methods

All spike proteins were downloaded from NCBI. MSA of all sequences was performed using ClustalW's Python application and visualized on ESPript 3. The sequence identity between COVID-19 spike protein and each of the other coronavirus spike proteins was calculated with Pairwise sequence identity. France strain' spike protein was modelled using SWISS MODEL<sup>6,7,8,9,10</sup>. The spike protein (6Z43.1, Chain B), was the closest spike protein to France strain among all the human coronaviruses. Thus, it was used as a template for building the spike protein. 3D structure visualization of SARS-CoV-2 spike receptor-binding domain bound with ACE2<sup>3</sup> (PDB: 6M0J) was performed with Py3Dmol and NGLview and the binding residues of this complex were retrieved with the PDB Parser in Python. Haddock and Prodigy<sup>12</sup> were used for the docking of spike proteins with ACE2 and the prediction of binding affinities of the formed complexes.

Name	Accession number	S Protein accession number
WIV04 (ref)	MN996528	QHR63260.2
229E	NC 002645	NP 073551.1
PEDV (pig)	NC 003436	NP 598310.1
NL63	-	AFV53148.1
FarmA	MF094681	AVM80439.1
OC43	-	AOL02453.1
HKU1	NC 006577	YP 173238.1
BCoV-ENT (bovine)	NC 003054	NP150077.1
ToR2	AY274119	AAP41037.1
Civet007	AY572034	AAU04646.1
EMC-2012	JX869059	AFS88936.1
CAMEL-363	KJ713298	AHY22555.1
RaTG13 (bat)	MN996532	QHR63300.2
BTCov /133/2005 (bat)	-	AB47052.1
Pangolin CoV	-	QIG55945.1

Table 1 S protein accession numbers for coronavirus sequences

# Results

## 1. Comparative Analysis of Covid-19 S protein with other coronaviruses:

The sequences presented in Table1 correspond to spike proteins of animal coronaviruses<sup>2</sup> infecting humans (camel, bat and pangolin). Using ClustalW's Application ID on Python, we performed a Multiple Sequence Alignment was elaborated and for the visualization ESPrpt3 was used (Figure 1).

Identical residues are highlighted in red. While similar conserved residues are highlighted in yellow, among all the spike proteins. Alpha helices are shown by helix on the top of the MSA. The last sequence has the lowest sequence identity with the reference sequence.

Spike proteins	Pairwise alignment with SARS-CoV-2
MERS coronavirus	45.09%
SARS coronavirus	79.03%
Pangolin coronavirus	91.41%
BAT coronavirus	97.56%
HCoV in France (July 2020)	99.76%
Bat2005	47.05%
Saudi HCoV (July 2020)	99.92%

Table 2 Pairwise alignment results

MSA, the phylogenetic tree (Figure 2) and the percentages of sequence identities (Table 2) showed that COVID-19 belongs to the same clade as BAT-CoV, Pangolin and SARS-CoV representing respectively 97,41%, 91,12% and 78,48% in terms of sequence identity. It's far from being a new strain issued from Mers-CoV which is only 46,5% identical to COVID -19 against 78,48% with SARS-CoV, which confirms that COVID-19 is caused by a BAT coronavirus and issued from SARS-CoV and also highlights the possibility of the Pangolin being an intermediate host<sup>4,5</sup>.

## 2. Comparative analysis of COVID-19 strains emerging from different regions in the world:

MSA (Table 3) and phylogenetic tree (Figure 3) issued from the comparative analysis of COVID-19 strains emerging from different regions in the world have shown that France has the higher root length and is only 99.76% identical to the reference S Protein, thus France strain underwent some mutations that made it unique (Table 2). Therefore, we studied its affinity to ACE2 and compare it to the reference's affinity through analysis methods such as "Docking".

## 3. Affinity Analysis:

### a) Spike protein homology modelling:

In order to determine the affinity, we had to go through different steps: First starting out by modelling the spike glycoprotein using SWISS-MODEL<sup>6,7,8,9,10</sup>, different models were suggested. After comparing the QMEAN, the GMQE, (total coverage, Z score) and the sequence identity. The model presenting the highest score, was selected.

### b) Structure analysis of SARS-CoV-2 spike receptor-binding domain bound with ACE2:

Binding residues and their atoms
ASN 49 A.CB-ASN49 A.CG
ASN 330 A.CA-ASN 330 A.C
ASN 332 A.CG-ASN 322 A.OO1
ASN 61 A.CB-ASN 61 A.CG
ASN 487 E.CG- ASN 487 E.NO2
ASN D00 E.CG-ASN 400 E.NO2
ASN 322 A.CB-ASN 322 A.CG
ASN 330 A.N-ASN 330 A.CA
ASN 330 A.CG- ASN 330 A.NO2
ASN 501 E.CB-ASN 501 E.CG
ASN 501 E.CA-ASN 501 E.CB

Table 4 Binding residues for docking

A structure analysis between spike protein's E chain and ACE2's A chain was performed on a pre-modelled S protein and ACE2 complex<sup>11</sup>, using python.

The goal behind this analysis is to determine the binding residues between these two entities. The complex and the residues were visualized with NGLviewer on python, as shown on the (Figure 5); Red illustrating ACE2's A chain, Blue illustrating spike protein' E chain and Yellow representing the binding residues (Table 4).

### c) Docking and determination of binding affinities:

Having already determined the residues within our binding site, the next step was to perform a docking. In docking, only the conformation of the ligand is fully explored, in our case, the ACE2's A chain. This docking, gave out multiple clusters for each complex (Reference's S protein docked with ACE and France strain' S protein docked with ACE2<sup>12</sup>).

In fact, the smaller the dissociation constant and the binding affinity are, the more tightly bound the ligand is, or the higher the affinity between ligand and protein. The dissociation constant is influenced by non-covalent intermolecular interactions between the two molecules such as hydrogen bonding, electrostatic interactions, hydrophobic and van der Waals forces.

Spike protein (Chain ID B) ACE2 (Chain ID A)	Reference strain Spike protein docked with ACE2	France strain Spike protein docked with ACE2
Haddock score	85,8 +/- 4,2	76,7 +/- 11,9
Binding Affinity(Kcal.mol <sup>-1</sup> )	-8.6	-9.3
Dissociation constant Kd (M) at 37°C	8.1e-7	2.6e-7
RMSD from the overall lowest-energy structure	13.8+/- 0.3	19.3+/-0.2
Van der Waals energy	-55.8+/-9.7	-53.8+/-6.6
Electrostatic energy	-329.9+/-15.7	-363.0+/-15.3

Table 5 Docking results

The docking of the Spike protein of France and China with the ACE2 receptor and the determination of binding affinities of each complex (Table 5) have shown that the France strain S protein has a higher affinity to the ACE2 receptor ( $\Delta G = -9.3 \text{ Kcal.mol}^{-1}$ ,  $K_d = 2,6.10^{-7} \text{ M}$ ) compared to the affinity of the reference strain S protein with the same receptor ( $\Delta G = -8,6 \text{ Kcal.mol}^{-1}$ ,  $K_d = 8.10^{-7} \text{ M}$ ).

These results highlight the hypothesis concerning the 'Best-Fit' activity of the COVID-19. In fact, France strain' S protein presented several mutations, making it less identical to the common COVID-19's S protein. These mutations seem to have an effect on its affinity towards the ACE2 receptor according to the theoretical results presented above.

## Conclusion

Many host cell receptors are the target for this virus, including ACE2. Study of binding affinities with the ACE2 receptor, is essential to better understand its viral mechanism. The French Strain S protein presented several mutations that influenced its affinity to ACE2 receptors. Future work involving the modeling of Spike proteins and experimental validation is needed to study the 'Best-Fit' activity of COVID-19.

## Declaration

### Competing interests

The author declares no competing interests.

## Bibliography

- [1] Muhammad Adnan Shereen , Suliman Khan , Abeer Kazmi , Nadia Bashir , Rabeea Siddique. COVID-19 infection: Origin, transmission, and characteristics of human coronaviruses, *Journal of Advanced Research*. <https://doi.org/10.1016/j.jare.2020.03.005>.
- [2] Longxian Lv, Gaolei Li, Jinhui Chen<sup>3</sup>, Xinle Liang, Yudong Li. Comparative genomic analysis revealed specic mutation pattern between human coronavirus SARS-CoV-2 and Bat-SARSr-CoV RaTG13. <https://doi.org/10.1101/2020.02.27.969006>.
- [3] Ibrahim M. Ibrahima, Doaa H. Abdelmaleka, Mohammed E. Elshahata, Abdo A. Elkya,b COVID-19 spike-host cell receptor GRP78 binding site prediction, *Journal of Infection*. <https://doi.org/10.1016/j.jinf.2020.02.026>.
- [4] Tao Zhang, Qunfu Wu, Zhigang Zhang. Probable Pangolin Origin of SARS-CoV-2 Associated with the COVID-19 Outbreak, *Current Biology* <https://doi.org/10.1016/j.cub.2020.03.022>.
- [5] Kristian G. Andersen, Andrew Rambaut, W. Ian Lipkin, Edward C. Holmes& Robert F. Garry The proximal origin of SARS-CoV-2, *Nature Medicine*. <https://doi.org/10.1038/S41591-020-0820-9>.
- [6] 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).
- [7] Bienert, S., Waterhouse, A., de Beer, T.A.P., Tauriello, G., Studer, G., Bordoli, L., Schwede, T. The SWISS-MODEL Repository - new features and functionality. *Nucleic Acids Res.* 45, D313-D319 (2017).
- [8] Guex, N., Peitsch, M.C., Schwede, T. Automated comparative protein structure modeling with SWISS-MODEL and Swiss-PdbViewer: A historical perspective. *Electrophoresis* 30, S162-S173 (2009).
- [9] Studer, G., Rempfer, C., Waterhouse, A.M., Gumienny, G., Haas, J., Schwede, T. QMEANDisCo - distance constraints applied on model quality estimation. *Bioinformatics* 36, 1765-1771 (2020).
- [10] Bertoni, M., Kiefer, F., Biasini, M., Bordoli, L., Schwede, T. Modeling protein quaternary structure of homo- and hetero-oligomers beyond binary interactions by homology. *Scientific Reports* 7 (2017).
- [11] Yushun Wan, Jian Shang, Rachel Graham, Ralph S. Baric, Fang Lia. Receptor Recognition by the Novel Coronavirus from Wuhan: an Analysis Based on Decade-Long Structural Studies of SARS Coronavirus, *Journal of Virology* <https://doi.org/10.1128/JVI.00127-20>.
- [12] G.C.P van Zundert, J.P.G.L.M. Rodrigues, M. Trellet, C. Schmitz, P.L. Kastiris, E. Karaca, A.S.J. Melquiond, M. van Dijk, S.J. de Vries and A.M.J.J. Bonvin (2016). "The HADDOCK2.2 webserver: User-

## Additional Table

Table 3 is available in the Supplementary Files.

## Figures

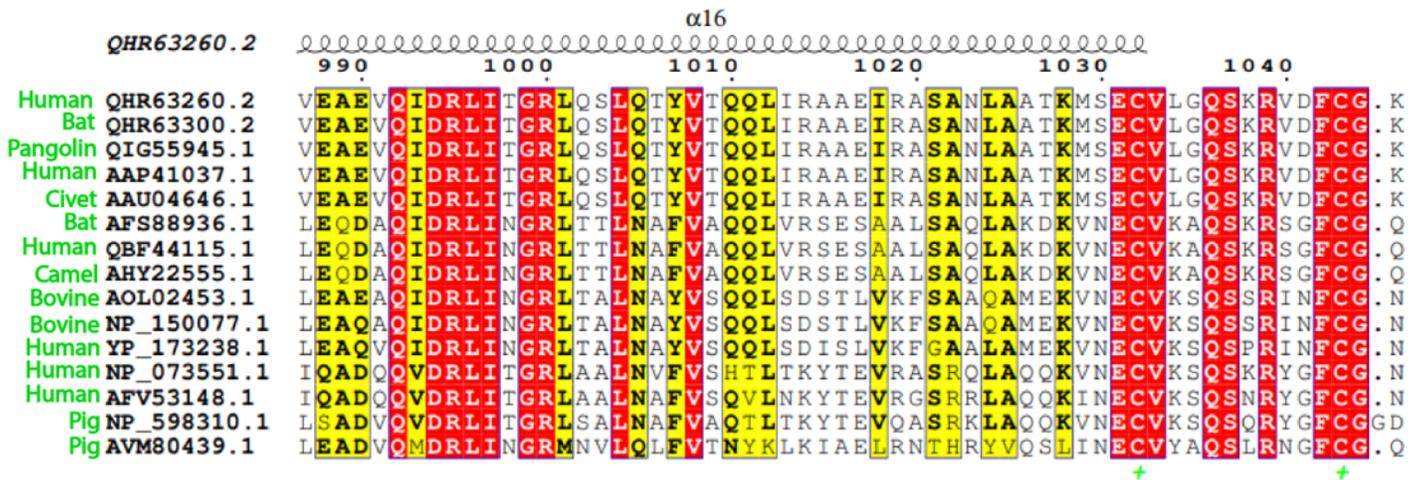


Figure 1

MSA Visualization with ESprit3.1

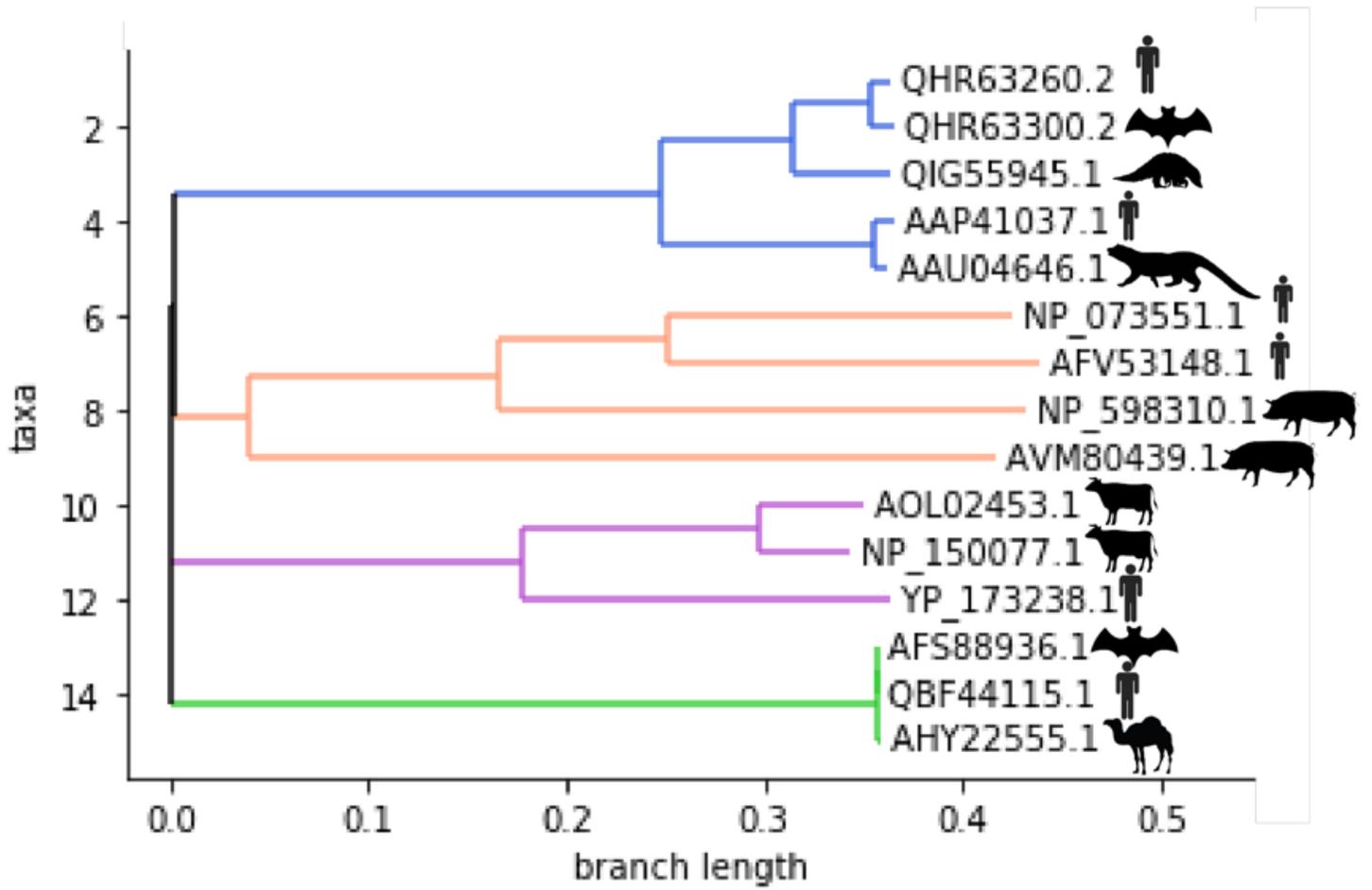


Figure 2

Phylogenetic tree

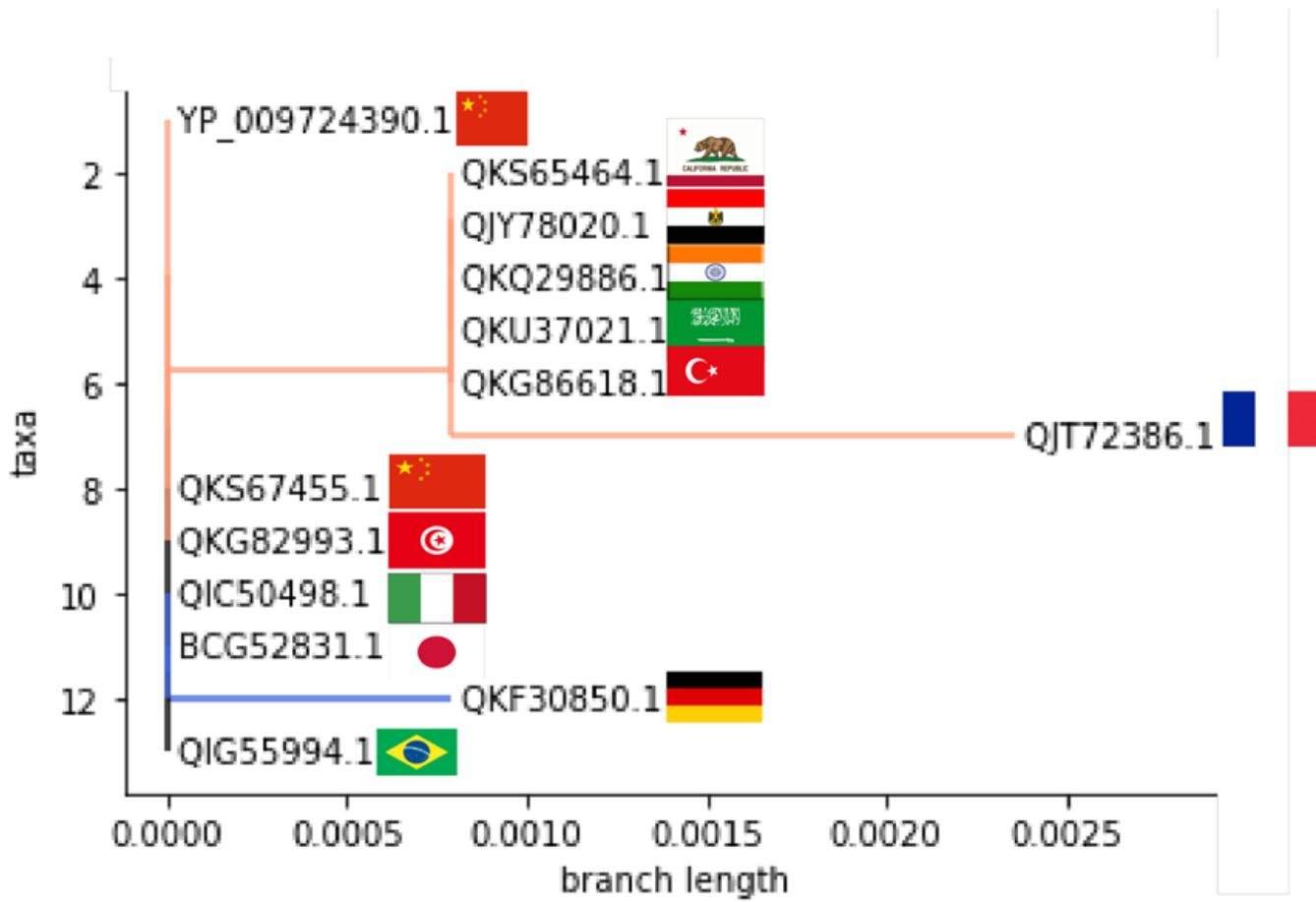


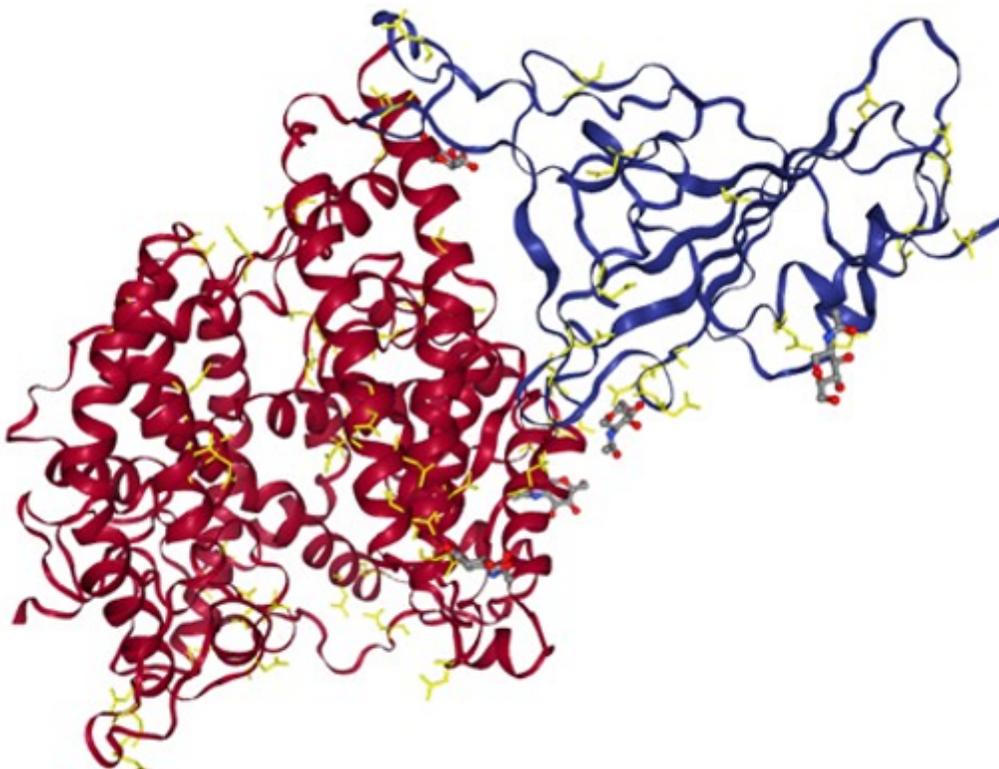
Figure 3

Phylogenetic Tree (2)



**Figure 4**

3D structure of the modeled spike protein



## Figure 5

ACE2-Spike protein docking complex

## Supplementary Files

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

- [Table3.png](#)