

Genetic diversity and structural characterization of spike glycoprotein of newly emerged SARS-CoV-2

Abhishek Das

ICMR-Department of Molecular microbiology, ICMR-National Institute of Cholera and Enteric Diseases, Kolkata-700010, West Bengal

M. Bhattacharya

Department of Zoology, Fakir Mohan University, Vyasa Vihar, Balasore 756020, Odisha

R. Roy

ICMR-Department of Molecular microbiology, ICMR-National Institute of Cholera and Enteric Diseases, Kolkata-700010, West Bengal

P.Ghosh

Department of Zoology, Vidyasagar University, Midnapore-721102, West Bengal

N.Mondal

Department of Biology, Indiana State University, TerreHaute, Indiana, USA

T. Das

ICAR-Central Inland fisheries Research institute, Barrackpore, Kolkata-700120, West Bengal

B.K.Behera

ICAR-Central Inland fisheries Research institute, Barrackpore, Kolkata-700120, West Bengal

B.C. Patra

Department of Zoology, Vidyasagar University, Midnapore-721102, West Bengal

S. Acharya (✉ sagar.gene@gmail.com)

Department of Zoology, Vidyasagar University, Midnapore-721102, West Bengal

Research Article

Keywords: SARS-CoV-2, COVID19, homology modeling, molecular docking, Phylogenetic analysis.

Posted Date: May 26th, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-31235/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

Abstract

A new beta Coronavirus (SARS-CoV-2) infection was first identified in the Wuhan City, China in December 2019 and after that it had spread rapidly throughout the globe and subsequently WHO have announced it as a pandemic. So, SARS-CoV-2 has now become a global threat to human civilization. Recent studies showed that the proteomic data of SARS-CoV-2 is closely related with other beta Coronavirus. The phylogenetic tree revealed the closeness of recently reported SAR-CoV2 with SARS-CoV by using MEGA 7 along with the suitable protocol of Neighbor joining algorithm. The spike glycoprotein plays the most important role during the onset of infection. Several mutations have been reported across the globe in the S Proteins. In this research, molecular docking between the SARS-CoV-2 spike glycoprotein and ACE2 protein was carried out in PatchDock web servers. WEBnm@ calculated the molecular simulation using Normal Mode Analysis (NMA) along with lowest deformation energy value which signifies the domain motions. Also during multiple sequence analysis, variations were observed within the Spike protein reported globally. 3- Dimensional structure of protein molecules were designed using homology modeling and the structure were validated through Q mean score and Ramachandran plot. All of the designed sequences were having around 91% of the amino acid in the favored region of Ramachandran plot. In order to check the binding affinity difference between the mutated and non-mutated strains, the generated models were docked with human ACE2 molecules. The non mutated strains have given the similar ACE value. However, there were variations in ACE value of the mutated strains. This observation provides evidence of Phylogenetic diversity and evolution.

Introduction

During last few decades, there were around 6 reported strains of Coronaviruses viz. HCovOC43, HCoV-HKU1, HCoV-229E, HCoV-NL63, SARS-CoV and MERS-CoV responsible for causing diseases among humans (Zhang et al. 2020). However, among them the two beta Coronaviruses which were commonly known to infect humans includes the Middle East respiratory syndrome Coronavirus or MERS-CoV and SARS-CoV. The Middle East respiratory syndrome (MERS) was first identified in Saudi Arabia in 2012 and it causes viral respiratory disease (Fehr et al. 2017). In the year 2002, the outbreak of SARS was first observed in China and it was found to be associated with the death of around 774 infected patients (de Wit et al. 2016). Lau and co-workers have reported that Chinese horseshoe bats acts as host reservoirs for the SARS-COV (Lau et al. 2005). Various intermediate hosts like dogs, cats which are commonly sold as food in those Chinese meat markets were found to be responsible for transmitting the virus to human (Guan et al. 2003).

As of December 2019, the world is being ravaged by a new strain of Corona virus named Severe acute respiratory syndrome– Coronavirus 2 (SARS-CoV-2). The origin of this virus was reported from Wuhan in the Hubei province of China (Tian et al. 2020), and this raised intense attention not only within China but across the globe. Corona viruses belong to those virus categories which have been reported for causing Severe Acute respiratory. In Wuhan, on 7th January, 2020, Chinese researchers have isolated a novel Corona virus (CoV) from patients. Among other Coronaviruses, SARS-CoV-2 was found to be the seventh viral strain that can infect humans. On January 30th, 2020, the World Health Organization declared the SARS-CoV-2 outbreak as a public health emergency of international concern. SARS-CoV-2 is a positive strand RNA virus which shares about 80% identity with that of SARS-CoV and is about 96% identical to the bat coronavirus BatCoV RaTG13 isolate (Yan et al. 2020).

The surface glycoprotein also known as spike (S) protein of Corona viruses facilitates viral entry into human cells. For viral entry and attachment to target cell, the S1 subunit of S protein binds to cellular receptor. The process of entry requires S protein priming by cellular proteases, wherein the S protein is cleaved at the S1/S2 sites and S2 site allows fusion of viral and cellular membranes. Spike protein uses ACE2 receptor molecule as an entry receptor and for its priming it used cellular serine protease TMPRSS2 (Hoffmann et al. 2020). After studying the SARS-S/ACE2 interface at atomic level, researchers have found that ACE2 is a key molecule for the transmission of the virus. SARS-S und SARS-2-S share around 76% amino acid identity.

A specific region on the S glycoprotein; termed as Receptor Binding Domain (RBD) on the virion surface mediates receptor recognition and membrane fusion (Yan et al. 2020). The receptor binding domain (RBD) is the most variable part of the coronavirus genome (Zhang et al. 2020). There are around Six RBD amino acids which are found to be responsible for binding to

ACE2 receptor and also for identifying the host. ACE2 is a type I membrane protein expressed in lungs, heart, kidneys, and intestine its primary physiological role is in the maturation of angiotensin (Ang), a peptide hormone that controls vasoconstriction and blood pressure. Structural studies have shown that SARS-CoV-2 is having a RBD which can bind to ACE2 molecule of various organisms viz. humans, ferrets, cats and other species. SARS-CoV-2 may bind human ACE2 with high affinity, however Anderson and co-workers through computational analyses have predicted that due to difference of RBD sequence of SARS-CoV-2 and SARS-CoV, the interaction between RBD and ACE2 is not that much ideal (Zhang et al. 2020). Another noticeable feature of SARS-CoV-2 is the presence of a polybasic site at the (RRAR) at the S1-S2 junction (Zhang et al. 2020). This help in cleavage by furin and other protease agents. Also it plays an important role in characterizing host range and viral infection.

Phylogenetic tree analysis have always contributed for studying the occurrence, transmission and progression of different RNA viruses like have largely contributed to a better understanding of the emergence, spread and evolution of many RNA viruses avian influenza virus (Lam et al. 2008, Cattoli et al. 2009) and Ebola virus (Gire et al. 2014). After the outbreak of SARS-CoV2, there have been different reports for the evolution of the virus from SARS-CoV reported from Bat and there have studies going on to check the variation of spike protein of SARS-CoV2. So the aim of the present study is to corroborate the genetic diversity with its functional efficacy of SARS-CoV2 to bind with target cell with special emphasis to spike protein by using proteomics tools of molecular docking. Also, the stability of docked complex was verified through NMA dynamic simulation focusing on Ca atoms mobility

Materials And Methods

GIS mapping

In the present study, Qgis V 2.18.26 software was used for the mapping and visualization of the reported cases of SARS-COV-2 infection reported globally till 16.4.2020 retrieved from NCBI database. We did not use the exact co-ordinate for framing as we do not have the exact GPS coordinates of the cases, who have been participated for the whole genome sequencing. All the cases are tagged with a different legend to show the vast spreading of the Covid-19 outbreak and the NCBI accession numbers are assigned to each legend (Fig. 1).

Phylogenetic analysis of spike protein of SARS-CoV- 2

To study the evolutionary pathway of the SARS-CoV-2, three phylogenetic trees were constructed. In the first Phylogenetic analysis, the spike protein (Surface glycoprotein) amino acid sequences of various SARS virus were retrieved from NCBI-Genbank database (Table 1). The sequences were aligned using Clustal W alignment program and the tree was constructed using Neighbor joining alignment. Bootstrapping was performed at 1000 replications of MEGA 7 program (Kumar et al. 2016). Tilapia lake virus hypothetical protein having genebank accession number [MN094791](#) was kept as an out group

The second Phylogenetic tree was constructed to determine the evolutionary relationship of mostly reported human Coronavirus viz. MERS, SARS-CoV and SARS-CoV-2. The S protein amino acid sequences were retrieved from NCBI database and were aligned. The evolutionary tree was generated using Neighbor joining program of MEGA 7 (Kumar et al. 2016) and bootstrapping was performed at 1000 replications. Tilapia lake virus hypothetical protein was kept as an out group.

Since SARS-CoV2 infection have been reported to be pandemic by WHO (<https://www.who.int/emergencies/diseases/novel-coronavirus-2019>) so the third Phylogenetic tree was constructed to check whether there is any variation between the S protein that have been reported globally. For the construction of the evolutionary tree, almost every representative sequence of Spike protein reported till 16.4.2020 from various geographical locations including Asia, Africa, North America, South America and Europe were retrieved from the new NCBI virus database (Table. 2) providing inclusive information regarding SARS-CoV2 (https://www.ncbi.nlm.nih.gov/labs/virus/vssi/#/virus?VirusLineage_ss=Severe%20acute%20respiratory%20syndrome%20coronavirus%202,%20taxid:2697049&SeqType_s=Nucleotide). The sequences were aligned using Neighbor joining program and bootstrapping was performed at 1000 replications of MEGA 7 program (Kumar et al. 2016). Tilapia lake virus hypothetical protein was kept as an out group.

Molecular Docking between SARS-Cov-2 spike glycoprotein and ACE2 receptor

The protein structure of SARS-Cov-2 spike glycoprotein having PDB ID-6VXX was retrieved from PDB database. Molecular docking between the SARS-CoV-2 spike glycoprotein (PDB ID: 6VXX) protein and ACE2 (PDB ID: 108A) protein are furthermost vital process to characterize the molecular interaction and accurate bonding pattern of the proteins. This is an essential process for supporting the binding affinity towards the outer spike glycoprotein and human ACE2 receptor. Molecular docking determines the cellular functions of the proteins (Lavi et al. 2013). In the present study, PatchDock docking server was used for carrying out molecular docking analysis (Duhovny et al. 2002). The server is having a geometric complimentary based algorithm to achieve the exact molecular docking (Yadav et al. 2017). From this server we also evaluated interface area, geometric score and Atomic Contact Energy (ACE) of the docking complexes along with the generated PDB file for analysis the proper molecular interactions. The RasMol software v.2.7.4.2 was helped to visualize the generated PDB file of protein-peptide docking complex (Chen et al. 2009).

Molecular Dynamic Simulation

Normal mode analysis (NMA) is a powerful method for predicting the possible large-scale movements of a specified bio-macromolecule (Prabhakar et al. 2016). This specific analysis has used in broad field of structural biology, like the study of conformational changes of protein binding to ligand, changing of conformation opening and closure structural stability of membrane channel protein, potential actions of the ribosome, and also calculate viral capsid maturation. The iMODS server was employed for representation of the dynamic motion of the docking complex accordingly with the help of NMA (López-Blanco et al., 2014). This server measures the different parameters regarding molecular structural dynamics like; deformability plot, B-factor, eigen value, Covariance matrix. Result of deformability plot reveals the specific region of protein (coiled regions) while the B-factor shows atomic deformation. Eigen value calculates rigidity of molecular motion. Co-variance matrix signifies the correlated atomic pairs depending on specific color code.

Homology modeling and molecular docking of Spike protein

In order to generate the 3D structure of spike protein, homology modeling was carried out at Swiss Model server (Waterhouse et al. 2018). Amino acid sequence having accession number viz. [MT049951](#); [MT12098](#), [MT007544](#), [NC045512](#), [MT019532](#) were retrieved from NCBI Genbank server. Spike protein structure having PDB ID- 6VXX was used as template for the generation of the model. Validation of the generated model was carried through Q-mean score (Benkert et al. 2011) and Ramachandran plot analysis.

To characterize the molecular interaction and accurate bonding pattern of the designed proteins, the generated 3D protein structure were docked with human ACE2 (PDB ID: 108A) receptor protein. PatchDock docking server was used for carrying out molecular docking analysis (Duhovny et al. 2002). The server can also evaluate interface area, geometric score and Atomic Contact Energy (ACE) of the docking complexes.

Results

Phylogenetic analysis of spike protein of SARS-COV 2

The first Phylogenetic tree represents the relationship of SARS-CoV2 with other members of Corona virus family. From the evolutionary relationship tree, it was found that the spike protein of SARS-CoV2 is evolutionary close to SARS-CoV sharing the same node with highest bootstrap value of 100 (Fig. 2). However, MERS was found to be evolutionary closer to these two strains of corona virus.

The second phylogenetic tree exhibit the evolutionary correlation between the of mostly reported human Corona virus viz. MERS, SARS-CoV and SARS-CoV2. From the tree it can be observed that the spike protein of MERS reported from various organism were form a single cluster and sharing some similarity with SARS-CoV reported from Bat. However, recently reported SAR-CoV2 from various infected humans was found to be sharing the same node with the SARS-CoV reported from Bat with highest bootstrap value (Fig. 3).

The third Phylogenetic tree was constructed to check whether there is any variation in Spike protein of SARS-CoV2 isolated from different patient globally. From the tree it can be easily observed that there is not much variation observed in the spike protein sequence and they are sharing the same node (Fig. 4). However, in multiple sequence alignment file (Supplementary, S1), a substitution of a single amino acid Aspergine (N) in place of Tyrosine (Y) at position number 28 have been observed in the sequence reported from Yunan, China having NCBI Accession No.- [MT049951](#). Also deletion of amino acid Tyrosine (Y) at 145 position and substitution of Isoleucine in place of Lysine have been observed at 408 position in the sequence reported from Kerala, India (NCBI Accession No: [MT12098](#)). Substitution of Glycine in place of Aspartic acid has been observed in sequence having accession number: [MT292570](#), [MT263074](#), [MT328032](#), [MT323033](#) [MT324062](#), [MT320538](#) have been observed at 614 position. Also substitution of serine by arginine was observed in sequence Accession number -[MT007544](#) at 247 position and substitution of phenylalanine by cystine was also observed in sequence ID- [MT093571](#) at 797 position.

Molecular Docking

Molecular interaction between SARS-CoV-2 spike glycoprotein protein and ACE2 protein was analyzed by Patch Dock web servers. Top 20 docked complexes were gained in the PDB format with absolute clustering root-mean-square deviation (RMSD) factor of value 4.0. In this work, the output models were visualized with the help of pre discussed software and the most acceptable model was selected by their (Atomic Contact Energy) ACE value, geometric shape complementarity score and complex interface area. The ACE value of the selected complex was -422.22, the docking complex with the lowest ACE value (negative) was selected for the spontaneous reactivity, whereas the geometric shape complementarities score was 15390 and complex interface area for the best model was 2804.00. The selected docking complex was shown in figure 5.

Molecular Dynamic Simulation

Normal mode analysis (NMA) is now a promising method to investigate the slowest motions in macromolecule. NMA is found to be useful for studying large molecular docking stability of SARS-CoV-2 virus spike protein and Human ACE2 protein. Deformation energies and eigen values indicates the energy related with normal mode and inversely related to the amplitude of the dynamic motion. WEBnm@ calculated 14 normal mode indexes along with their deformation energies and here, we have selected the lowest deformation energy value indicating a mode with large rigid regions, which has a good chance of demonstrating domain motions (Fig. 6A) The selected deformation energy value of the present complex is found to be 1236.54. In this plot, anti-correlated, uncorrelated and correlated movement C α in the protein represent via blue, white and orange colour gradient respectively showing in figure-6B. The square of the fluctuation of each C α is calculated a lowest non-trivial modes as eigen value plot (Fig. 6C). The fluctuations of the atomic displacements in selected mode were inverse of their corresponding eigen values. The normalized squared atomic displacement plot is shown in figure-6D.

Structure prediction by homology modeling

The generated 3D structure of spike protein of SARS-COV-2 through homology modeling was predicted by using SARS-CoV-2 spike protein having PDB ID-6VXX as template. The identified template was an electron microscopic structure of spike protein having a resolution of 2.8 Å. The generated models were shown in Fig. 7 A-E.

Structural quality assessment

The generated protein structure through homology modeling was validated through different protein structure validation tools. The quality of the generated model was evaluated based on their Q mean score and Ramachandran plot (Fig. 8A-E). The result of both the analysis tool was shown in Table. 3.

Molecular Docking

The molecular interaction between the designed 3D structure of SARS-CoV-2 spike glycoprotein and ACE2 (PBD ID:108A) protein was analyzed by using Patch Dock web servers. Top 20 docked complexes were gained in the PDB format with absolute clustering root-mean-square deviation (RMSD) factor of value 4.0. The output models were visualized with the help of pre

discussed software and the most acceptable model was selected by their (Atomic Contact Energy) ACE value, geometric shape complementarity score and complex interface area. The results of molecular docking have been shown in Table. 4 and Fig. 9A-E.

Discussion

Globally, a new Coronavirus infection has been declared pandemic by WHO (Phan. 2020). In December 2019, its outbreak was first reported from Wuhan, China from a sea food market (Zhu et al. 2020). The infection was having pneumoniae like symptoms but of unknown etiology. Later, it was found to be a new member of beta Coronavirus and is found to be similar with the SARS like Coronavirus which have been reported earlier from Bat. The International Committee on Taxonomy of Viruses (ICTV) has named the viral agent as SARS-CoV-2 and the associated disease as COVID-19 (Cui et al. 2019). After the first outbreak in China, the infection of SARS-CoV-2 have been reported continuously from different geographical areas like Europe, North America, South America, Asia and some parts of Africa and the infection cases are increasing drastically day by day (Velavan and Meyer 2020). There have been many reports of Coronaviruses infection caused by SARS-CoV and MERS where wild animals like bat were being reported as reservoir. However, Menachery and coworker have reported that Coronaviruses have the property of cross the species barrier and can also infect humans (Menachery et al. 2015). Since the first outbreak was reported from sea food market so, animals were thought to be the transmission agent. However, there were reports of infection in person who have not visited the sea food market. So, it had provided the insight of human to human transmission in more than 100 countries in the world (Nishiura et al. 2020; Shereen et al. 2019). The virus has been detected from the lungs, saliva, nasopharyngeal swabs and sputum of the infected patient (Zhu et al. 2019; Bastola et al. 2020; Lin et al. 2020). The transmission of COVID 19 particles involves the close contact with the infected persons having symptoms of sneezing, coughing and high fever. The inhaled aerosol particles act as a medium for transmitting the virus (Shereen et al. 2020). The mutation rate of RNA viruses is generally found to be higher than their host which enable them to evolve and adapt to changing conditions (Pachetti et al. 2020). In SARS-CoV-2 genome, Wang and coworkers have reported around 13 variation sites in the ORF3a, ORF8, ORF1ab, N and S regions. They have found around 30.53% mutation in ORF8 region and 29.47% in ORF1a region (Wang et al. 2020). Earlier results showed that SARS-CoV-2 is moving fast from one country to another and new hotspots of mutations are emerging in SARS-CoV2 genome (Pachetti et al. 2020). Since the viral entry and host tropism is majorly defined by the spike protein it is thus important to study the structure and biochemical nature of this protein. So the present study was conducted to find the origin of SARS-CoV-2 and also the study the genetic diversity and structural characterization of spike protein of SARS-CoV-2.

The molecular docking of SARS-CoV-2 spike glycoprotein and ACE2 were significant with a negative ACE value of -422.22 indicating strong bonding affinity. The receptor-ligand docking regulates the cellular activities and triggered cellular serine protease TMPRSS2 for S protein priming. So complex binding will be also able to elicit a specific immune response. Molecular Dynamic simulation through NMA method revealed the lowest deformation energy value (1236.54) with normal mode explaining the domain motions. The actual correlated movement of the C α atom of the docked complex was calculated via correlation matrix index with their specific range between -1 to 1. The lowest non-trivial (non-zero) value of eigen value indicates the atomic fluctuation is inversely related with deformability and lastly the normal squared atomic displacement plot demonstrated the actual atomic fluctuation along with eigen value (Azim et al. 2019). The normalized square atomic displacements signify square displacement of each C- α atom in protein. The X and Y axes symbolize the residue index of amino acid sequences and normal mode of square atomic displacement, correspondingly.

Presently three phylogenetic trees were constructed to check the genetic diversity of the SARS-CoV-2 spike protein. The first and second phylogenetic tree revealed that SARS-CoV-2 spike protein is highly similar with SARS-CoV spike protein reported from Bat. Recent studies have also shown 79% similarity between spike protein of SARS-CoV-2 and spike protein of SARS-CoV reported from Bat (*Rhinolophus sinicus*) (Phan. 2020; Zhou et al. 2020). Also Ceraola and coworkers have reported 96% homology between *R. affinis* SARS-CoV spike protein and SARS-CoV-2 spike protein (Ceraolo and Giorgi 2020). The third Phylogenetic tree revealed that globally reported spike proteins are not having much genetic variation and they are sharing the same branch. However, from multiple sequence analysis, deletion of single nucleotide has been observed in SAR-CoV-2 spike protein reported from Kerala, India. Also a substitution of amino acids have been observed from the sequence submitted Australia, Peru, Yunan (China), Greece, Spain and South Africa. Phan have also reported three deletions in nucleotide sequence reported from Australia, USA and Japan. He had also reported around 93 mutations in SARS-CoV2 genome (Phan. 2020).

According to Robson. 2020, a single variation in single amino acid sequence may lead to change in characteristic of coronavirus strain which will lead to generation of a new strain. Presently, five protein structures were generated using homology modeling. Two structures were generated using sequence from Wuhan and USA having no variation whereas three protein structures were generated by using sequence reported from Yunan, Australia and India. These three sequences were having variation from rest of the sequences. After docking with human ACE2 receptor it was very clear that the variation in single amino acid sequence will lead to change in binding affinity of the protein. The docking score of the model generated with the sequence from Wuhan and USA were having the same Atomic Contact Energy, score and area whereas the ACE value, area and score were found to be different for the docked protein reported from India, Yunan and Australia. Phan have also predicted that mutation in spike glycoprotein can induce major conformational changes and referred this protein as major protein of interest. However, due to unavailability of the amino acid sequences he was not able to find out the changes (Phan. 2020). The infection rate is increasing daily and mutations are becoming a barrier towards development of therapeutics. So much more sequencing data is needed to find out the different type of mutations and strains. Also the mapping carried out in our study will be helpful for estimating the number of cases using geo-location and geospatial techniques. However, due to unavailability of the exact GPS coordinates, we were not able to show the route of transmission from one place to another. So in future, the GPS coordinates will be crucial for studying the transmission route of this disease globally. Our study provides a clear insight regarding the variation in SAR-CoV-2 spike protein and also the finding will helpful for the future researcher for combating against this pathogen.

Declarations

Conflict of Interest

The authors have declared no conflict of interest

Acknowledgement

The authors would like to express their sincere gratitude to Mr. Hirak Jyoti Chakrobarty, ICAR-CIFRI, Barrackpore for his immense guidance towards analysis of data. The authors are also thankful to Mr. Asim Kumar Jana, ICAR-CIFRI, Barrackpore for his assistance.

References

1. Azim KF, Hasan M, Hossain MN, Somana SR, Hoque SF, Bappy MN, Chowdhury AT, Lasker T (2019) Immunoinformatics approaches for designing a novel multi epitope peptide vaccine against human norovirus (Norwalk virus). *Infection, Genetics and Evolution* 74:103936
2. Bastola A, Sah R, Rodriguez-Morales AJ, Lal BK, Jha R, Ojha HC, Shrestha B, Chu DK, Poon LL, Costello A, Morita K (2020) The first 2019 novel coronavirus case in Nepal. *The Lancet Infectious Diseases* 20(3):279-80
3. Benkert P, Biasini M, Schwede T (2011) Toward the estimation of the absolute quality of individual protein structure models. *Bioinformatics* 1;27(3):343-50
4. Cattoli G, Monne I, Fusaro A, Joannis TM, Lombin LH, Aly MM, Arafa AS, Sturm-Ramirez KM, Couacy-Hymann E, Awuni JA, Batawui KB (2009) Highly pathogenic avian influenza virus subtype H5N1 in Africa: a comprehensive phylogenetic analysis and molecular characterization of isolates *PLoS One*;4(3).
5. Ceraolo C, Giorgi FM (2020) Genomic variance of the 2019-nCoV coronavirus. *Journal of medical virology*
6. Chen Q, Ma J, Yan M, Mothobi ME, Liu Y, Zheng F (2009) A novel mutation in CRYAB associated with autosomal dominant congenital nuclear cataract in a Chinese family. *Molecular vision* 15:1359
7. Cui J, Li F, Shi ZL (2019) Origin and evolution of pathogenic coronaviruses. *Nature reviews Microbiology* 17(3):181-92
8. de Wit E, van Doremalen N, Falzarano D, Munster VJ (2016) SARS and MERS: recent insights into emerging coronaviruses. *Nature Reviews Microbiology* 14(8):523
9. Duhovny D, Nussinov R, Wolfson HJ (2002) Efficient unbound docking of rigid molecules. In *International workshop on algorithms in bioinformatics* pp. 185-200 Springer, Berlin, Heidelberg

10. Fehr AR, Channappanavar R, Perlman S (2017) Middle East respiratory syndrome: emergence of a pathogenic human coronavirus. *Annual review of medicine* 14;68:387-99.
11. Gire SK, Goba A, Andersen KG, Sealfon RS, Park DJ, Kanneh L, Jalloh S, Momoh M, Fullah M, Dudas G, Wohl S (2014) Genomic surveillance elucidates Ebola virus origin and transmission during the 2014 outbreak. *Science* 12;345(6202):1369-72.
12. Guan Y, Zheng BJ, He YQ, Liu XL, Zhuang ZX, Cheung CL, Luo SW, Li PH, Zhang LJ, Guan YJ, Butt KM (2003) Isolation and characterization of viruses related to the SARS coronavirus from animals in southern China. *Science* 302(5643):276-8
13. Hoffmann M, Kleine-Weber H, Schroeder S, Krüger N, Herrler T, Erichsen S, Schiergens TS, Herrler G, Wu NH, Nitsche A, Müller MA (2020) SARS-CoV-2 cell entry depends on ACE2 and TMPRSS2 and is blocked by a clinically proven protease inhibitor. *Cell*
14. Kumar S, Stecher G, Tamura K (2016) MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Molecular biology and evolution* 22;33(7):1870-4
15. Lam TT, Hon CC, Pybus OG, Pond SL, Wong RT, Yip CW, Zeng F, Leung FC (2008) Evolutionary and transmission dynamics of reassortant H5N1 influenza virus in Indonesia. *PLoS pathogens* 4(8).
16. Lau SK, Woo PC, Li KS, Huang Y, Tsoi HW, Wong BH, Wong SS, Leung SY, Chan KH, Yuen KY (2005) Severe acute respiratory syndrome coronavirus-like virus in Chinese horseshoe bats. *Proceedings of the National Academy of Sciences*. 27;102(39):14040-5.
17. Lavi A, Ngan CH, Movshovitz-Attias D, Bohnuud T, Yueh C, Beglov D, Schueler-Furman O, Kozakov D (2013) Detection of peptide-binding sites on protein surfaces: The first step toward the modeling and targeting of peptide-mediated interactions. *Proteins: Structure, Function, and Bioinformatics*. 81(12):2096-105
18. Lin X, Gong Z, Xiao Z, Xiong J, Fan B, Liu J (2020) Novel coronavirus pneumonia outbreak in 2019: computed tomographic findings in two cases. *Korean Journal of Radiology* 1;21(3):365-8
19. López-Blanco JR, Aliaga JI, Quintana-Ortí ES, Chacón P (2014) iMODS: internal coordinates normal mode analysis server. *Nucleic acids research* 42(W1):W271-6
20. Menachery VD, Yount Jr BL, Debbink K, Agnihothram S, Gralinski LE, Plante JA, Graham RL, Scobey T, Ge XY, Donaldson EF, Randell SH (2015) A SARS-like cluster of circulating bat coronaviruses shows potential for human emergence. *Nature medicine* 21(12):1508
21. Nishiura H, Jung SM, Linton NM, Kinoshita R, Yang Y, Hayashi K, Kobayashi T, Yuan B, Akhmetzhanov AR (2020) The extent of transmission of novel coronavirus in Wuhan, China,
22. Pachetti M, Marini B, Benedetti F, Giudici F, Mauro E, Storici P, Masciovecchio C, Angeletti S, Ciccozzi M, Gallo RC, Zella D (2020) Emerging SARS-CoV-2 mutation hot spots include a novel RNA-dependent-RNA polymerase variant. *Journal of Translational Medicine* 18(1):1-9
23. Phan T (2020) Genetic diversity and evolution of SARS-CoV-2 [published online ahead of print, 2020 Feb 21]. *Infect Genet Evol* 81:104260.
24. Prabhakar PK, Srivastava A, Rao KK, Balaji PV (2016) Monomerization alters the dynamics of the lid region in *Campylobacter jejuni* CstII: an MD simulation study. *Journal of Biomolecular Structure and Dynamics* 34(4):778-91
25. Robson B (2020) COVID-19 Coronavirus spike protein analysis for synthetic vaccines, a peptidomimetic antagonist, and therapeutic drugs, and analysis of a proposed achilles' heel conserved region to minimize probability of escape mutations and drug resistance. *ComputBiol Med* 103749
26. Shereen MA, Khan S, Kazmi A, Bashir N, Siddique R (2020) COVID-19 infection: origin, transmission, and characteristics of human coronaviruses. *Journal of Advanced Research*.
27. Tian X, Li C, Huang A, Xia S, Lu S, Shi Z, Lu L, Jiang S, Yang Z, Wu Y, Ying T (2020) Potent binding of 2019 novel coronavirus spike protein by a SARS coronavirus-specific human monoclonal antibody. *Emerging microbes & infections* 1;9(1):382-5
28. Velavan TP, and Meyer CG (2020) The COVID-19 epidemic. *Tropical medicine & international health* 25(3), p.278
29. Wang D, Hu B, Hu C, Zhu F, Liu X, Zhang J, Wang B, Xiang H, Cheng Z, Xiong Y, Zhao Y (2020) Clinical characteristics of 138 hospitalized patients with 2019 novel coronavirus-infected pneumonia in Wuhan, China. *JAMA* 323 (11): 1061-9.

30. Waterhouse A, Bertoni M, Bienert S, Studer G, Tauriello G, Gumienny R, Heer FT, de Beer TA, Rempfer C, Bordoli L, Lepore R (2018) SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic acids research* 2;46(W1):W296-303
31. Yadav S, Pandey SK, Singh VK, Goel Y, Kumar A, Singh SM (2017) Molecular docking studies of 3-bromopyruvate and its derivatives to metabolic regulatory enzymes: Implication in designing of novel anticancer therapeutic strategies. *PloS one* 12(5)
32. Yan R, Zhang Y, Li Y, Xia L, Guo Y, Zhou Q (2020) Structural basis for the recognition of SARS-CoV-2 by full-length human ACE2. *Science* 27;367(6485):1444-8.
33. Zhang J, Ma K, Li H, Liao M, Qi W (2020) The continuous evolution and dissemination of 2019 novel human coronavirus. *Journal of Infection*.
34. Zhu N, Zhang D, Wang W, Li X, Yang B, Song J, Zhao X, Huang B, Shi W, Lu R, Niu P (2020) A novel coronavirus from patients with pneumonia in China, 2019. *New England Journal of Medicine*

Tables

Table 1. Spike proteins of human CoVs with their identification, locus, accession, version, locus and sequence length analyzed in this study

Sl. No.	Protein name	Other Information	Length
1	SARS-CoV-2	LOCUS: QIH45053 ACCESSION: QIH45053 VERSION: QIH45053.1	1273 aa
2	SARS-CoV	LOCUS: AAT74874 ACCESSION:AAT74874 VERSION :AAT74874.1	1255 aa
3	MERS-CoV	LOCUS :ATG84833 ACCESSION :ATG84833 VERSION:ATG84833.1	1353 aa
4	HCoV-229E	LOCUS: BAL45641 ACCESSION: BAL45641 VERSION: BAL45641.1	1170 aa
5	HCoV-NL63	LOCUS: AKT07952 ACCESSION:AKT07952 VERSION:AKT07952.1	1356 aa
6	HCoV-OC43	LOCUS: AMK59677 ACCESSION: AMK59677 VERSION: AMK59677.1	1359 aa
7.	HCoV-HKU1	LOCUS: BBA20983 ACCESSION:BBA20983 VERSION: BBA20983.1	1356 aa

Table 2. Spike glycoprotein sequence of SARS-CoV-2 reported globally till 16.4.2020

Sl. No	Genbank Accession number	Protein ID	Country of origin
1. 1.	NC045512	YP_009724390.1	China: Wuhan
1. 2.	MT188341	"QIK02964.1	USA
1. 3.	MT325561	QIZ15549.1	USA
1. 4.	MT325571	QIZ15669.1	USA
1. 5.	MT072688	QIB84673.1	Nepal
1. 6.	MT066175	QIA98596.1	Taiwan
1. 7.	MT066176	QIA98606.1"	Taiwan
1. 8.	MT324062	QIZ15537.1	South Africa
1. 9.	MT327745	QIZ16509.1	Turkey
1. 10.	MT304476	QIV15008.1	South Korea
1. 11.	MT304475	QIV14996.1	South Korea
1. 12.	MT050493	QIA98583.1	India: Kerala
1. 13.	MT012098	QHS34546.1	India: Kerala
1. 14.	MT240479	QIQ22760.1	Pakistan: Gligit
1. 15.	MT152824	"QIH55221	USA: Snohomish County
1. 16.	MT118835	QID98794	USA: CA
1. 17.	MT192772	QIK50438.1	Vietnam : Ho Chi Minh city
1. 18.	MT192773	QIK50448.1	Vietnam : Ho Chi Minh city
1. 19.	MT192759	"QIK50417.1	Taiwan
1. 20.	MT121215	QII57161.1	China :Shanghai
1. 21.	MT135041	QIH45023.1	China: Beijing
1. 22.	MT123292	QIE07471.1	China: Guangzhou
1. 23.	MT049951	QIA20044.1	China: Yunan
1. 24.	MT039890	QHZ00379.1	South Korea
1. 25.	MT039873	QHZ00358	China: Hangzhou

Table 3. Q mean Score and Ramachandran plot value of the generated spike protein three dimensional structure through Homology modelling

1. 26.	MT019533	QHU36864.1	China: Wuhan
1. 27.	MT019532	QHU36854.1	China: Wuhan
1. 28.	MT020880	"QHU79194.1	USA:WA
1. 29.	MT328032	QIZ16535	Greece
1. 30.	MT328033	QIZ16547	Greece
1. 31.	MT077125	QIC50498	Italy
1. 32.	MT320538	QIX12148	France
1. 33.	MT292570	QIU78719	Spain
1. 34.	MT292571	QIU78731	Spain
1. 35.	MT066156	QIA98554	Italy
1. 36.	MT093571	QIC53204.1	Sweden
1. 37.	MT007544	QHR84449	Australia: Victoria
1. 38.	MT263074	QIS60288	Peru
1. 39.	MT256924	QIS30054	Colombia: Antioquia
1. 40.	MT126808	QIG55994	Brazil

Table 4. Molecular Docking score of generated spike protein and ACE2 receptor

Figures

Sl. No	Country and NCBI Accession number	Q mean score	Ramachandran plot	
			Favoured	Outlier
1	China_ Wuhan (NC045512)	-2.07	91.15%	1.82%
2.	USA (MT325561)	-2.07	91.15%	1.82%
3.	India_Kerala (MT12098)	-2.07	91.33%	1.85%
4.	Australia_Victoria (MT007544)	-1.91	91.83%	1.46%
5.	China_Yunan (MT049951)	-2.01	91.45%	1.88%

Sl.No.	Country	ACE	Score	Area
1.	China_ Wuhan (NC045512)	-236.75	14310	2932.20
1.	USA (MT325561)	-236.75	14310	2932.20
1.	India_Kerala (MT12098)	-102.22	15144	2544.70
1.	Australia_Victoria (MT007544)	-275	16814	2616.50
1.	China_Yunan (MT049951)	-164.20	14592	3393.90

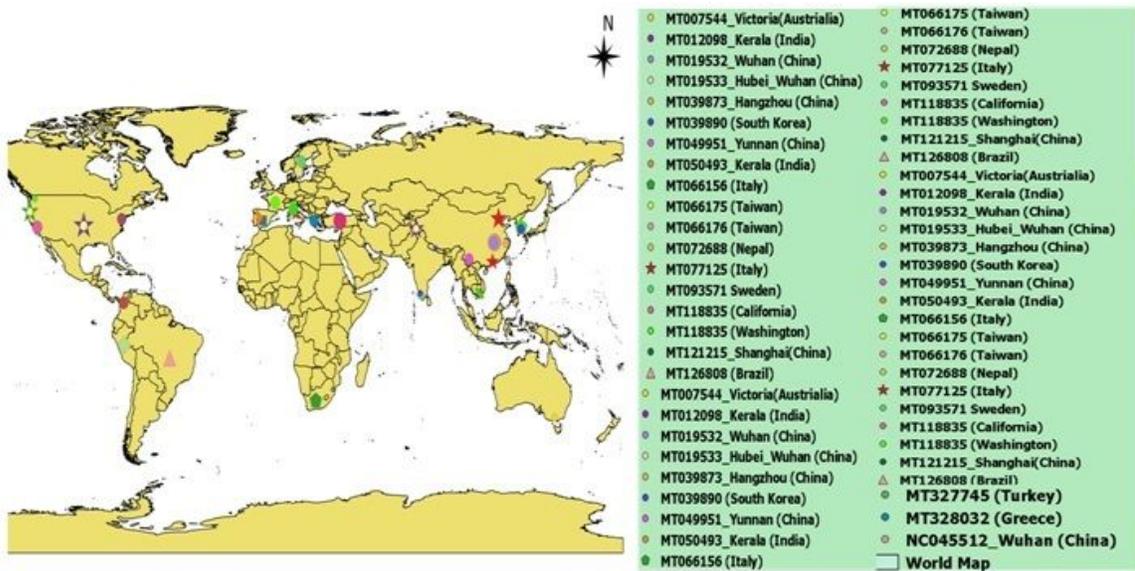


Figure 1

Mapping and visualization of the reported cases of SARS-CoV-2 infection reported globally till 16.4.2020. Note: The designations employed and the presentation of the material on this map do not imply the expression of any opinion whatsoever on the part of Research Square concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries. This map has been provided by the authors.

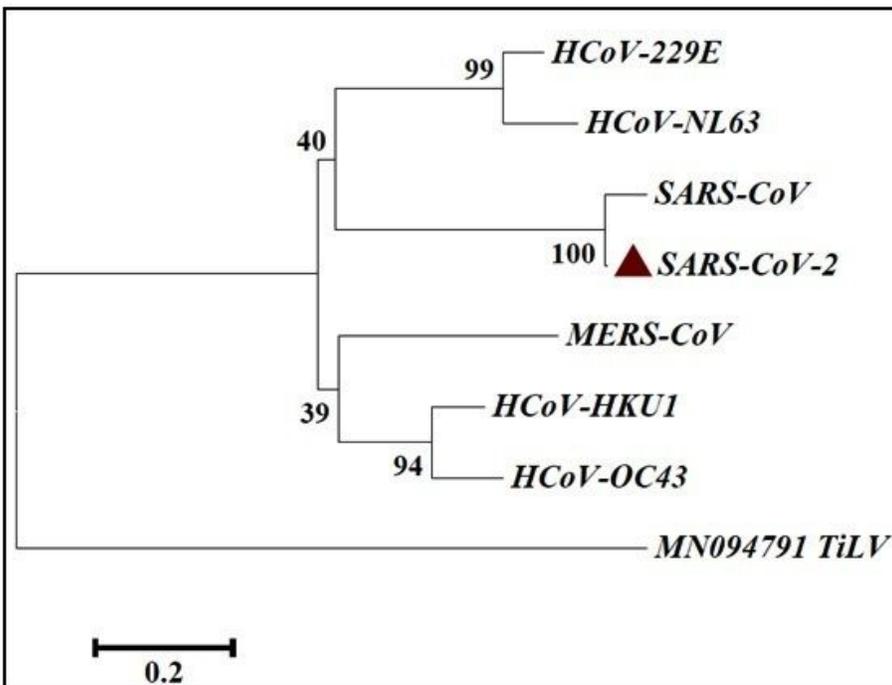


Figure 2

Evolutionary relationship of different Corona virus trains used in this study using Neighbor- joining method by the MEGA 7 software. The numbers next to the branches indicate percentage values for 1000 bootstrap replicates. Bootstrap values are shown at the nodes.

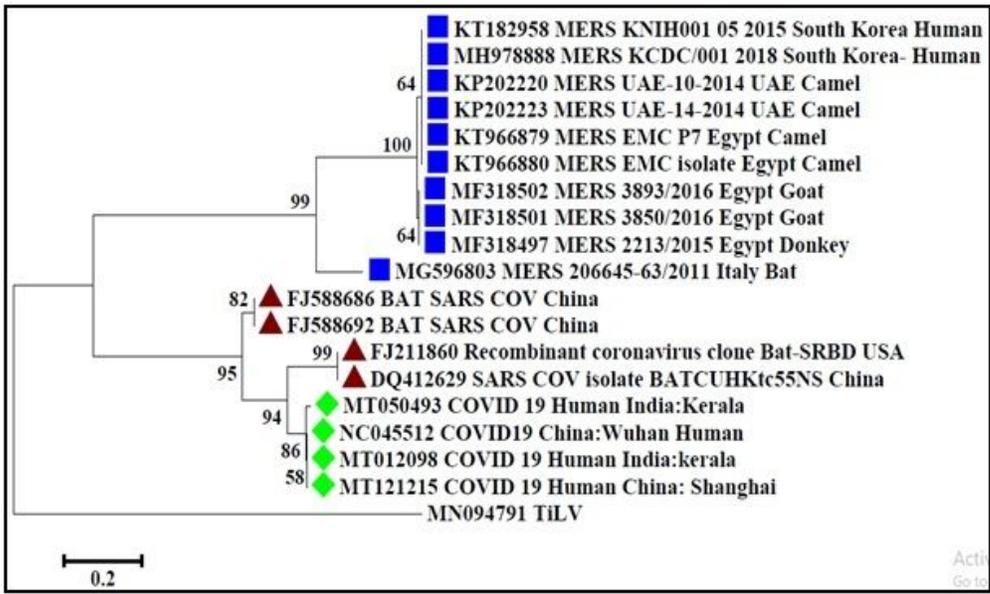


Figure 3

Phylogenetic relationship of mostly reported Corona virus responsible for human infection were used in this study using Neighbor-joining method by the MEGA 7 software. The numbers next to the branches indicate percentage values for 1000 bootstrap replicates. Bootstrap values are shown at the nodes.



Figure 4

Phylogenetic relationship of Spike protein of globally reported SARS-COV2 responsible for human infection were used in this study using Neighbor- joining method by the MEGA 7 software with bootstrap value of 1000. The percentage indicates the percentage identity. However, TiLV hypothetical protein protein have been there in the outgroup.

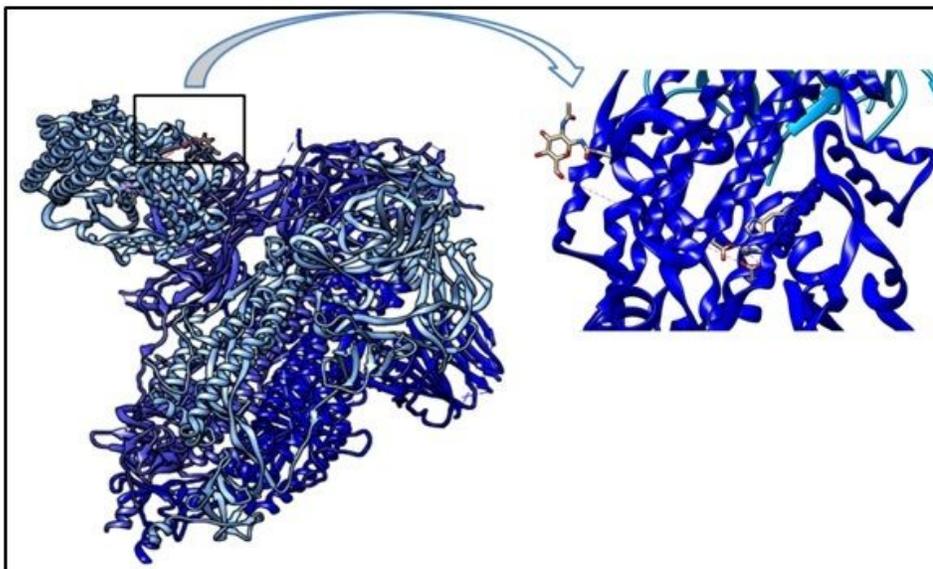


Figure 5

Molecular docking between SARS-CoV-2 spike glycoprotein (PDB ID: 6VXX) and ACE2 receptor (PDB ID:108A)

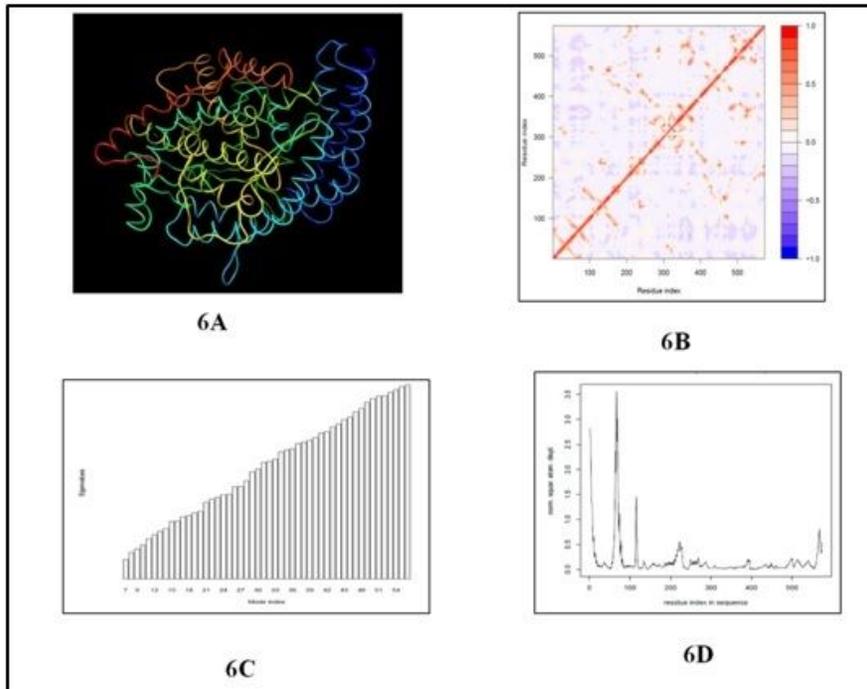


Figure 6

A-D. Molecular dynamic simulation of SARS-COV-2 virus spike protein and Human ACE2 protein. The mode of NMA mobility of the domains of docking complex upon binding (Fig. 6A); Correlation matrix plot represents anti-correlated, uncorrelated and correlated movement C α in the protein (Fig. 6B); Eigen value plot of the square of the fluctuation of each C α (Fig. 6C); The normalized squared atomic displacement plot represents atomic displacement (Fig. 6D)

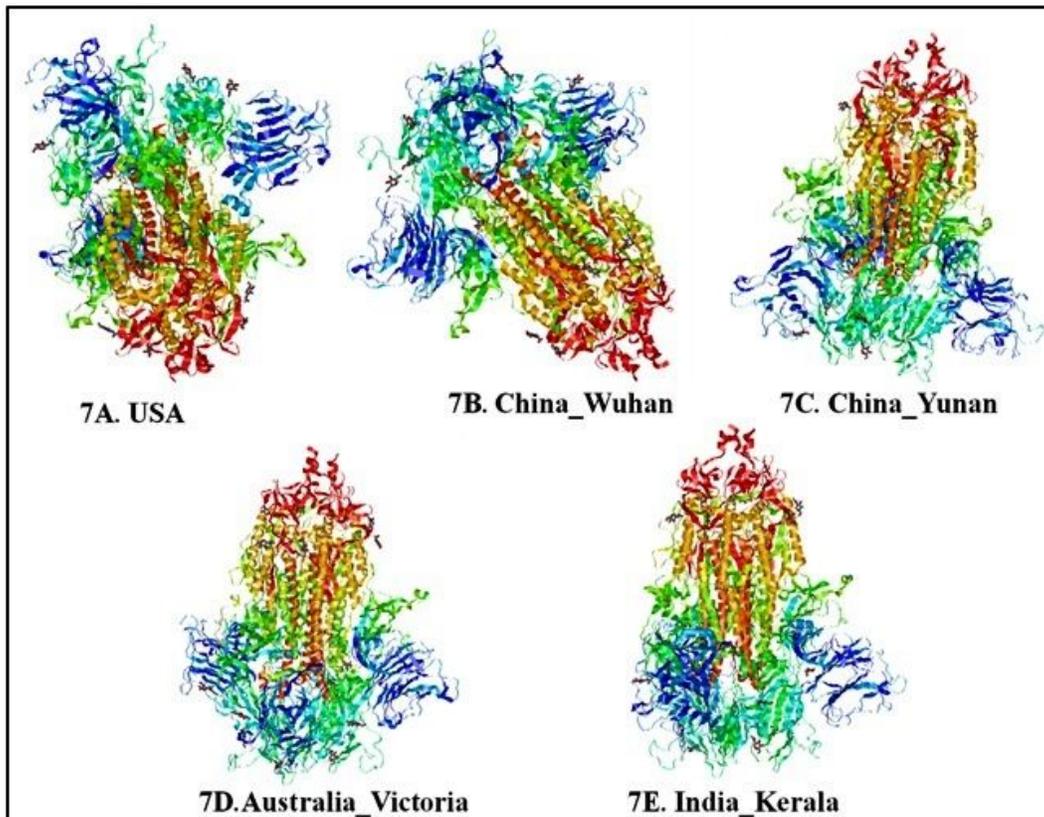


Figure 7

A-E. Three dimensional structure of spike protein reported from different countries 7A. USA; 7B. China_Wuhan; 7C. China_Yunan, 7D. Australia_Victoria ; 7E. India_Kerala. The 3D structure was generated in Swiss model server by using SARS-COV-2 spike protein (PDB ID-6VXX) as a template.

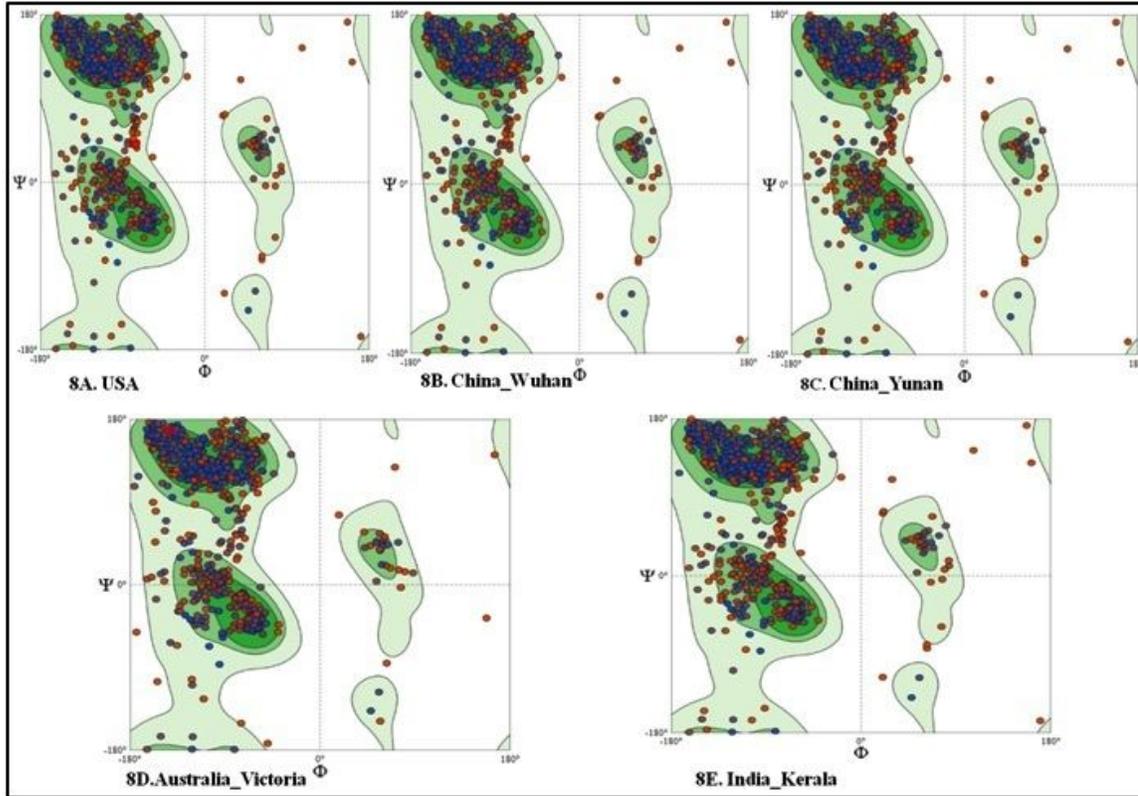


Figure 8

A-E. Ramachandran plot analysis of generated 3-D model of spike protein reported from different countries.

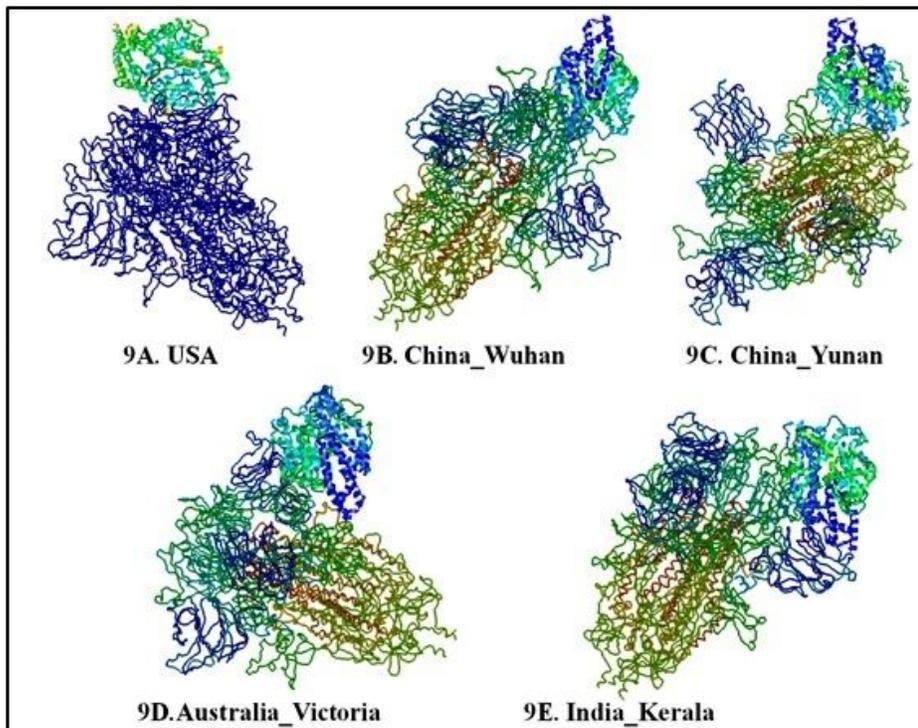


Figure 9

A-E. Molecular docking structure spike protein reported from various countries. The generated 3D structure was docked with Human ACE2 receptor molecule.

Supplementary Files

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

- [S1.xls](#)