



Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.



SARS-CoV 2 spike protein S1 subunit as an ideal target for stable vaccines: A bioinformatic study

Nagesha S.N.^{a,*}, Ramesh B.N.^a, Pradeep C.^a, Shashidhara K.S.^b, Thippeswamy Ramakrishnappa^c, Krishnaprasad B.T.^a, Jnanashree S.M.^a, Manohar M.^a, Arunkumar N.^a, Yallappa^a, Dhanush Patel D.^a, Rakesh T.V.^a, Girish E.^a, Mahantesh Bagoji^a, Shreeram S. Chandaragi^a

^a Department of Biotechnology, College of Agriculture, Hassan, University of Agricultural Sciences, Bangalore 560065, Karnataka, India

^b Department of Genetics and Plant Breeding, College of Agriculture, Hassan, University of Agricultural Sciences, Bangalore 560065, Karnataka, India

^c Department of Chemistry, BMS Institute of Technology and Management, Avalahalli, Yelahanka, Bengaluru 560064, Karnataka, India

ARTICLE INFO

Article history:

Available online 15 July 2021

Keywords:

SARS-CoV-2

Spike glycoprotein

Phylogenetic and 3D modelling

ABSTRACT

The Covid-19 a pandemic infectious disease and affected life across the world resulting in over 188.65 million confirmed cases across 223 countries, territories and areas with 4.06 million deaths. It is caused by a severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) and spike (S) protein of SARS-CoV-2, which plays a key role in the receptor recognition and cell membrane fusion process, is composed of two subunits, S1 and S2. The S1 subunit contains a receptor-binding domain (RBD) that recognizes and binds to the host receptor angiotensin-converting enzyme 2 (ACE2), while the S2 subunit mediates viral cell membrane fusion. Hence, it is a key target for developing neutralizing antibodies. Here, we have performed phylogenetic analysis and structural modeling of the SARS-CoV-2 spike glycoprotein, which is found highly conserved. The overall percent protein sequence identity from the SARS-CoV-2 spike protein sequences from the NCBI database was 99.68%. The functional domains of the S protein reveal that the S1 subunit was highly conserved (99.70%) than the S2 subunit (99.66%). Further, the 319–541 residues (RBD) of amino acids within the S1 domain were 100% similar among the spike protein. The 3D modeling of SARS-CoV-2 spike glycoprotein indicated that S protein has four domains with five protein units and the S1 subunit from 1 to 289 amino acid of domain 1 is highly conserved without any change in the ligand interaction site. This analysis clearly suggests that the S1 subunit (RBD 319–541) can be used as a target region for stable and safe vaccine development.

© 2021 Elsevier Ltd. All rights reserved.

Selection and peer-review under responsibility of the scientific committee of the Web International Conference on Accelerating Innovations in Material Science – 2020.

1. Introduction

The coronavirus disease 2019 (COVID-19), a global pandemic, represents an unprecedented public health, social and economic challenge [1,2]. In December 2019, an incident occurred in Wuhan, southern China where a series of pneumonia cases was reported. It was not long before the cases were classified as viral pneumonia and the virus was speculated to belong to β coronavirus. Primarily it was named as 2019-novel coronavirus (2019-nCoV) by World Health Organization (WHO), which later called it as the coronavirus disease 2019 (COVID- 19) [3]. The Coronavirus belongs to

the Coronaviridae family and Orthocoronaviridae subfamily, which is divided into four genera: Alpha coronavirus, beta coronavirus, gamma coronavirus and delta coronavirus. SARS-CoV, MERS-CoV and SARS-CoV 2 are all beta coronaviruses, a genus that includes many viruses that infect humans, bats and other domesticated and wild animals [4]. Coronaviruses (CoVs) are enveloped positive-sense RNA viruses [5]. Enveloped CoVs enter host cells and initiate infection through the fusion of viral and cellular membranes [6,7]. Membrane fusion is mediated by the large type I transmembrane S glycoprotein on the viral envelope and the cognate receptor on the surface of host cells [8–10]. The surface-exposed location of the S glycoprotein not only allows it to carry out membrane fusion but also render it a direct target for host immune responses, making it as the major target of neutralizing antibodies [11]. Because of its central roles in viral infection and

* Corresponding author.

E-mail address: nageshasn@uasbangalore.edu.in (N. S.N.).

¹ ORCID ID: <https://orcid.org/0000-0003-3279-6034>.

eliciting protective humoral and cell-mediated immune responses in hosts during infection [11], the S protein is the primary target for vaccine design as well as antiviral therapeutics [12]. The coronavirus spike protein (S) is the primary determinant of viral tropism and is responsible for receptor binding and membrane fusion. It is a large (approx. 180 kDa) glycoprotein that is present on the viral surface as a prominent trimer, and is composed of two domains, S1 and S2 [10]. The recent surfacing of the novel coronavirus SARS-CoV-2 (first identified on December 12, 2019) was initially detected in Wuhan, Hubei Province, China [13]. On March 11, 2020, the World Health Organization (WHO) officially declared it as a global pandemic. The rapidly evolving situation has prompted most affected countries to impose tight restrictions on border movements and clamp unprecedented statewide lockdown measures. The WHO official data online (<https://covid19.who.int/>) reveals that there are over 188.65 million confirmed cases across 223 countries, territories and areas with 4.06 million deaths globally till date (17/07/2021).

1.1. The spike (S) glycoprotein-A key target for vaccine

With a size of 180–200 kDa, the S protein consists of an extracellular N-terminus, a transmembrane (TM) domain anchored in the viral membrane, and a short intracellular C-terminal segment [14]. S normally exists in a metastable, prefusion conformation; once the virus interacts with the host cell, extensive structural rearrangement of the S protein occurs, allowing the virus to fuse with the host cell membrane. The spikes are coated with polysaccharide molecules to camouflage them, evading surveillance of the host immune system during entry [15]. The total length of SARS-CoV-2 S is 1273 amino acids and consists of a signal peptide (amino acids 1–13) located at the N-terminus, the S1 subunit (14–685 residues), and the S2 subunit (686–1273 residues); the last two regions are responsible for receptor binding and membrane fusion, respectively. In the S1 subunit, there is an N-terminal domain (14–305 residues) and a receptor-binding domain (RBD, 319–541

residues); the fusion peptide (FP) (788–806 residues), heptapeptide repeat sequence 1 (HR1) (912–984 residues), HR2 (1163–1213 residues), TM domain (1213–1237 residues), and cytoplasm domain (1237–1273 residues) comprise the S2 subunit [16]. S protein trimers visually form a characteristic bulbous, crown-like halo surrounding the viral particle. In the native state, the CoV S protein exists as an inactive precursor. During viral infection, target cell proteases activate the S protein by cleaving it into S1 and S2 subunits [17], which is necessary for activating the membrane fusion domain after viral entry into the target cells [18]. Similar to other coronaviruses, the S protein of SARS-CoV-2 is cleaved into S1 and S2 subunits by cellular proteases, and the serine protease TMPRSS2 is used as a protein primer. Although the cleavage site of SARS-CoV is known, that of SARS-CoV-2 S has not yet been reported [18,19]. The key role played by the S1 subunit of S protein of SARS-CoV-2 in receptor binding with host membrane makes it an excellent target for developing neutralizing antibodies for the pandemic virus. After knowing the seriousness of this virus and the need for developing stable and safe vaccines to contain it, an *In-silico* study was performed to check the suitability of the S1 subunit of SARS-CoV-2 as a target for stable and safe vaccine development. The SARS CoV 2 spike protein accessions (114) with alpha human SARS CoV (HCoV 229E), representing many countries of the world and seven affected countries, were used for analysis. The confirmed cases and deaths recorded (WHO) due to COVID19 were as follows: USA, India, Brazil, Mexico, UK, China, and Italy. Further, 3D models and ligand interaction sites were modeled for the selected accessions of seven countries

to check the suitability of the S1 subunit as the potential target for developing neutralizing antibodies.

2. Materials and methods

2.1. Collection of SARS CoV 2 spike protein amino-acid sequences

The amino acid sequence S proteins used in the study for phylogenetic analysis were obtained from NCBI viruses SARS-CoV-2 Data Hub (<http://www.ncbi.nlm.nih.gov>) and their accession numbers obtained were as follows:

QJH92179.1, QLA09688.1, QLA09760.1, QLA09784.1, QLA09810.1, QLA09822.1, QLA10152.1, QLF97939.1, QLF97987.1, QLF98071.1, QLF98083.1, QLF98095.1, QLF98119.1, QLF98143.1, QLF98236.1, QLF98277.1, QLH64815.1, QLH64863.1, QKY60165.1, QKY64614.1, QKY64792.1, QKY65277.1, QLA10164.1, QKM77204.1, QKM77276.1, QKM77288.1, QNN88070.1, QNN88094.1, QNN88214.1, QNN90137.1, QNN26432.1, QNN30872.1, QNN31208.1, QNN83662.1, QNL35926.1, QNL35962.1, QNN88130.1, QKJ68485.1, QOF19325.1, QQQ29796.1, QQP60983.1, QOI11511.1, QOI09591.1, QOF07709.1, QOF07733.1, QOF09113.1, QOF17405.1, QOF19133.1, QOF19145.1, QOH27654.1, QNM81021.1, QNL10802.1, QJY40181.1, QQQ47845.1, QQQ47905.1, QQQ47917.1, QQQ47941.1, QQQ47953.1, QQQ47965.1, QQQ47977.1, QQQ48001.1, QQQ48025.1, QPM98571.1, QND76238.1, QHU36834.1, YP009724390.1, QHX41667.1, QIG55994.1, QKE43703.1, QQS80838.1, QPK41427.1, BCM16174.1, QLG43114.1, QQS80850.1, QQQ16503.1, QQO75523.1, QQO75692.1, QQL13874.1, QPK91107.1, QPJ58622.1, QLJ57687.1, QQG33273.1, QQD90634.1, QPZ56530.1, QPZ33351.1, QPZ33363.1, QPP19204.1, QPN97030.1, QPN02379.1, QPN53404.1, QPK67600.1, QPJ77272.1, QNR60316.1, QPF48704.1, QPC41132.1, QPC41144.1, QOT69847.1, QPA20102.1, QOW18389.1, QKR86593.1, QKI36913.1, QJZ27921.1, QJX74523.1, QJT43392.1, QJS53338.1, QJT73046.1, QJQ04481.1, QJD23270.1, QHU79173.2, QNL36070.1, QNJ45106.1, QND78273.1, QMX86989.1, QIK50448.1, AOG74783.1

2.2. Alignment and phylogenetic analysis of spike glycoprotein

The coding sequences for the spike glycoprotein were retrieved from the NCBI virus SARS-CoV-2 Data Hub. Multiple sequence alignment of the CDS region was performed using ClustalW of BioEdit software (BioEdit version 7.2). The sequences were selected from different countries affected by the COVID19 pandemic. The phylogenetic analysis of SARS-CoV-2 spike proteins of different countries was performed using MEGAX software (MEGA-X Version 10.2) [20]. It was accomplished through multiple comparisons using the neighbor-joining algorithm in MEGA-X. The numbers at nodes indicate bootstrap support (100 bootstraps). Multiple comparisons were made by ClustalW multiple sequence alignment and the neighbor-joining phylogenies were estimated by maximum likelihood for 114 sequences collected from NCBI virus SARS-CoV-2 Data Hub with an out group HCoV 229E. More specifically, phylogenies were estimated for the seven spike protein sequences from affected countries of the world: USA (QOF19325.1), India (QJH92179.1), Brazil (QIG55994.1), Mexico (QPK41427.1), UK (QPC41144.1), China (QHU36834.1), Italy (QKE43703.1) employing HCoV-229E (AOG74783.1) spike protein amino-acid sequence considered in the study as an out group for comparison. The evolutionary history was inferred using the neighbor-joining method [21]. The optimal tree is shown. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (100 replicates) are shown next to

the branches [22]. The evolutionary distances were computed using the Poisson correction method [23] and are in the units of the number of amino-acid substitutions per site. This analysis involved 115 amino-acid sequences. All ambiguous positions were removed for each sequence pair (pairwise deletion option). There were a total of 2597 positions in the final dataset. Evolutionary analyses were conducted in MEGA X [20]. The number at the nodes indicates the bootstrap support (100 replicates), and the scale bar indicates the estimated number of substitutions per site.

2.3. 3D structural modelling of spike glycoprotein

SARS-CoV-2 S protein structures of amino acids were obtained from the NCBI virus SARS-CoV-2 Data Hub of seven countries (USA, India, Brazil, Mexico, UK, China, and Italy), which are affected by the virus with an out group HCoV 229E. The pairwise amino-acid sequence alignments among each of the SARS-CoV-2 were performed using ClustalW BioEdit software and exported as FASTA file extension for further application. The S protein models were built using PHYRE2 (Protein Homology/analogy Recognition Engine V 2.0) for structural modeling [24]. The domains of the S proteins of most affected countries selected with the out group HCoV 229E were determined by using Protein Peeling 3D (new tools for analyzing protein structures) with the parameters chosen: R-value-95, Minimal size of secondary structure-8, Minimal size of protein unit-16, CI cut off-0.2, Cut-off distance between atoms-8.0 and Delta value in logistic. Function-1.5 [25] and the ligand interaction sites on the S protein of the selected countries of SARS-CoV-2 were determined by using RCBS,

PDB (Research Collaboratory for Structural Bioinformatics, Protein Data Bank) [26] along with the out group HCoV 229E protein sequence following standard parameters.

3. Results and discussion

3.1. Comparison of amino acid identity of the spike (S) protein of SARS-CoV-2 with Alpha human SARS-CoV (HCoV-229E)

The COVID-19 pandemic caused by SARS-CoV-2 is evidence of the potential of coronaviruses to continuously evolve in wild reservoirs and jump to new species [27]. The ongoing pandemic of coronavirus disease 2019 (COVID-19), caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), poses a grave threat to global public health and imposes a severe burden on the entire human society. Like other coronaviruses, the SARS-CoV-2 genome encodes spike (S) glycoproteins, which protrude from the surface of mature virions. The S glycoprotein plays essential roles in virus attachment, fusion and entry into the host cell. Surface location of the S glycoprotein renders it a direct target for host immune responses, making it the main target of neutralizing antibodies. This *In-silico* study aims at contributing to view S1 subunit of spike protein of SARS CoV 2 as a potential target site for stable and safe vaccine development for the globe [28]. To obtain an initial assessment of shared and specific features of the SARS-CoV-2 spike (S) glycoprotein, a protein-sequence alignment was performed to compare the sequence of the SARS-CoV-2 spike protein collected from NCBI virus SARS-CoV-2 Data Hub from the seven affected countries (USA, India, Brazil, Mexico, UK, China, and Italy) of the world, where fatality was recorded (WHO) with HCoV-229E used as an out group. (Supplementary Information 1). The overall percent S protein sequence identity found by the alignment in the affected countries of the world was 99.68% (Fig. 1). The functional domains of the S protein reveal that the S1 subunit is highly conserved (99.70%) than the S2 subunit (99.66% identity). Further, within the S1 domain, (RBD 319–541)

of amino acids were 100% similar among the SARS-CoV-2 spike protein sequences collected from the NCBI database of the infected countries. Out of the 1273 amino-acid residues of SARS-CoV-2 spike protein, only four changes were recorded, and when aligned, the spike protein sequences collected from the NCBI database of seven countries were affected. At position 14H, (His-Histidine) is substituted in USA SARS-CoV-2 spike protein sequences, but Q (Gln-Glutamine) in the other countries' S protein sequences; at 614 position, D (Asp-Aspartic Acid) is substituted in China and Brazil, but G (Gly-Glycine) was noticed in the SARS-CoV-2 spike protein sequences of other countries. At 677 amino-acid position H, (His-Histidine) is noticed in India spike protein sequence which is substituted by Q (Gln-Glutamine) in other countries' spike protein sequence. Finally, at 1228 amino-acid position of spike protein sequence L, (Leu-Leucine) is recorded in Mexican spike protein sequence which is replaced by V (Val-Valine) in all other countries' spike protein sequences considered for the sequence similarity search by multiple sequence alignment.

3.2. Phylogenetic analysis of SARS-CoV-2 spike protein

Phylogenetic analysis centered on 114 S protein sequences of the world collected from NCBI virus SARS-CoV-2 Data Hub (Fig. 2) along with the out group HCoV-229E. The analysis confirmed that all SARS-CoV-2 S protein sequences clustered very closely with very little variation at amino acid level of S protein, and it was evident in phylogenetic tree constructed using SARS CoV 2 S protein sequences of affected countries with HCoV 229E considered as an out group.

3.3. SARS-CoV-2 S protein homology structure modeling

To understand the common and possibly distinguishing structural features found in SARS-CoV-2 S protein, homology modeling was undertaken. The analysis of modeled proteins provides a powerful tool to identify predicted structural characteristics, which can translate into structure function changes in the studied protein. To perform the modeling, it is first necessary to identify a suitable protein structure to be used as template, which will determine the accuracy of the predicted model. The S protein structure of the SARS-CoV-2 spike protein sequences collected for modeling was retrieved from the NCBI virus SARS-CoV-2 Data Hub of seven countries (USA, India, Brazil, Mexico, UK, China, and Italy) along with human SARS-CoV (HCoV-229E) for comparison. The 3D models of the S protein were determined using the online PHYRE2 tool and domains were determined for all the S protein sequences used for 3D modeling using Protein Peeling 3D splitting a protein structure into protein units and domains [25,29–31]. The number of splitting events (USA, India, Brazil, Mexico, UK, China, and Italy) (corresponding to five Protein Units) detected in spike protein are four (Domain 1 : [1–289], Domain 2 : [290–319]; [591–703], Domain 3 : [320–590] and Domain 4 : [704–1273], and for alpha human SARS-CoV(HCoV-229E) S protein, the domains detected were four with varied amino-acid residue range compared to S protein of selected countries (Domain 1 : [1–280], Domain 2 : [281–495], Domain 3 : [496–563] and Domain 4 : [564–1171]) (Fig. 3). These 3D modeling results clearly indicate that the spike protein is highly conserved. The ligand interaction residues were identified in all the spike protein sequences selected from the affected countries with HCoV-229E virus spike glycoprotein considered for comparison employing Pymol (<https://pymol.org/2/>) tool using pdb format files collected from RCSB PDB database. The ligand interaction sites remained the same {Threonine-Thr (T)} at the 108 position of amino-acid residue in all the selected countries spike protein but changed to {Tyrosine-Tyr(Y)} at the

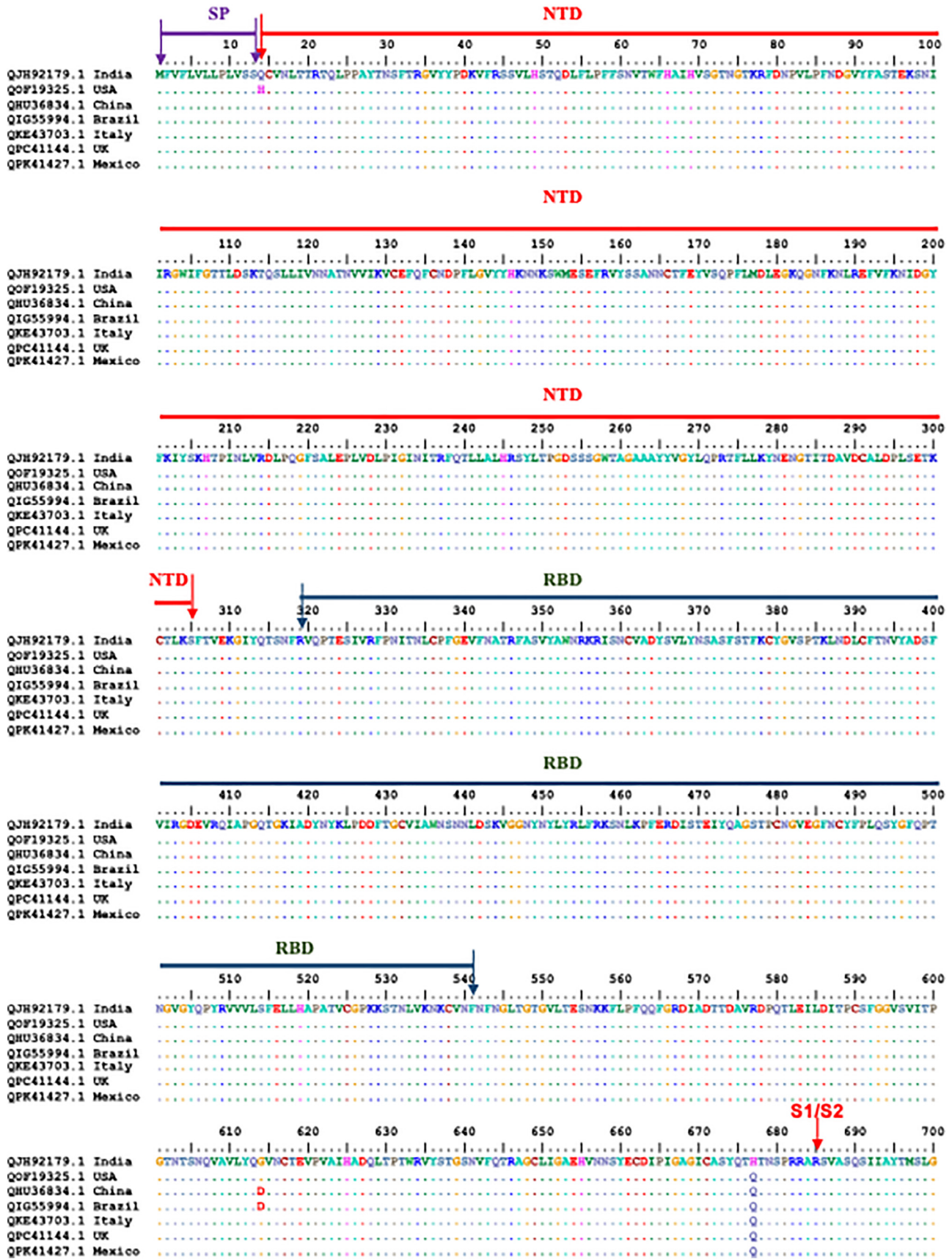


Fig. 1 (continued)

202 position of amino acid in S1 subunit of HCoV-229E virus spike glycoprotein considered for comparison. Further ligand interaction

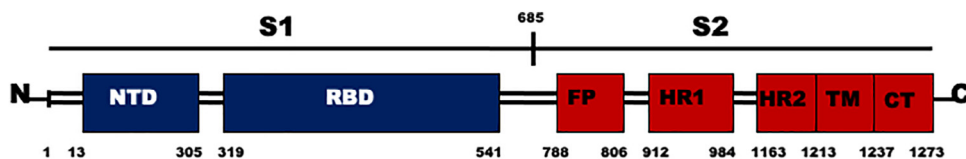


Fig. 1. Structure of the SARS-CoV-2 S protein. Annotation of the SARS-CoV-2 spike glycoprotein Multiple sequence alignment of SARS CoV 2 Spike protein S1 unit (685 Amino acid residues). S1 subunit, there is a signal peptide (SP, 1–13 residues shown in violet line), N-terminal domain (NTD, 14–305 residues shown in red line) and a receptor-binding domain (RBD, 319–541 residues shown in green line). The region corresponding to the S1/S2 cleavage site at 685 aminoacid position (red arrow) is shown. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

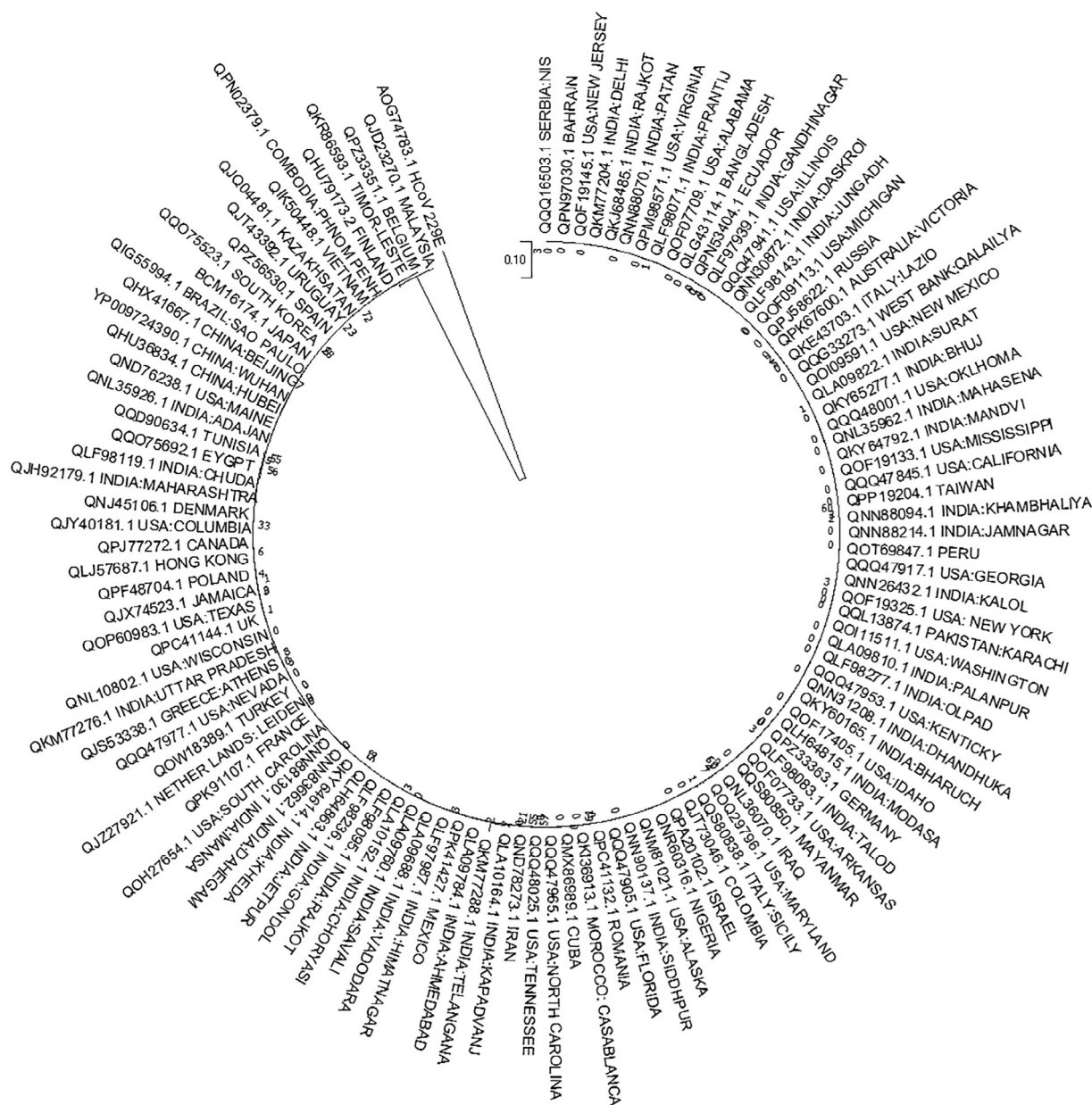


Fig. 2. Phylogenetic tree using SARS CoV2 Spike protein sequences retrieved from NCBI viruses SARS-CoV-2 Data Hub. This analysis involved 115 aminoacid sequences. All ambiguous positions were removed for each sequence pair (pairwise deletion option). There were a total of 2597 positions in the final dataset. Evolutionary analyses were conducted in MEGA X. Number at nodes indicates bootstrap support (100 replicates), and the scale bar indicates the estimated number of substitutions per site. Phylogenetic tree construction using SARS CoV 2 spike protein sequences of affected countries This analysis involved 8 amino-acid sequences. All ambiguous positions were removed for each sequence pair (pairwise deletion option). There were a total of 1310 positions in the final dataset. Evolutionary analyses were conducted in MEGA X. The number at nodes indicates bootstrap support (100 replicates) and the scale bar indicates the estimated number of substitutions per site.

sites were modeled for all the selected spike sequence country wise along with Alpha human SARS-CoV (HCoV-229E) to confirm

the similarity at ligand interaction site using RCSB PDB database by homology modeling. Interestingly, at the 108 amino-acid resi-

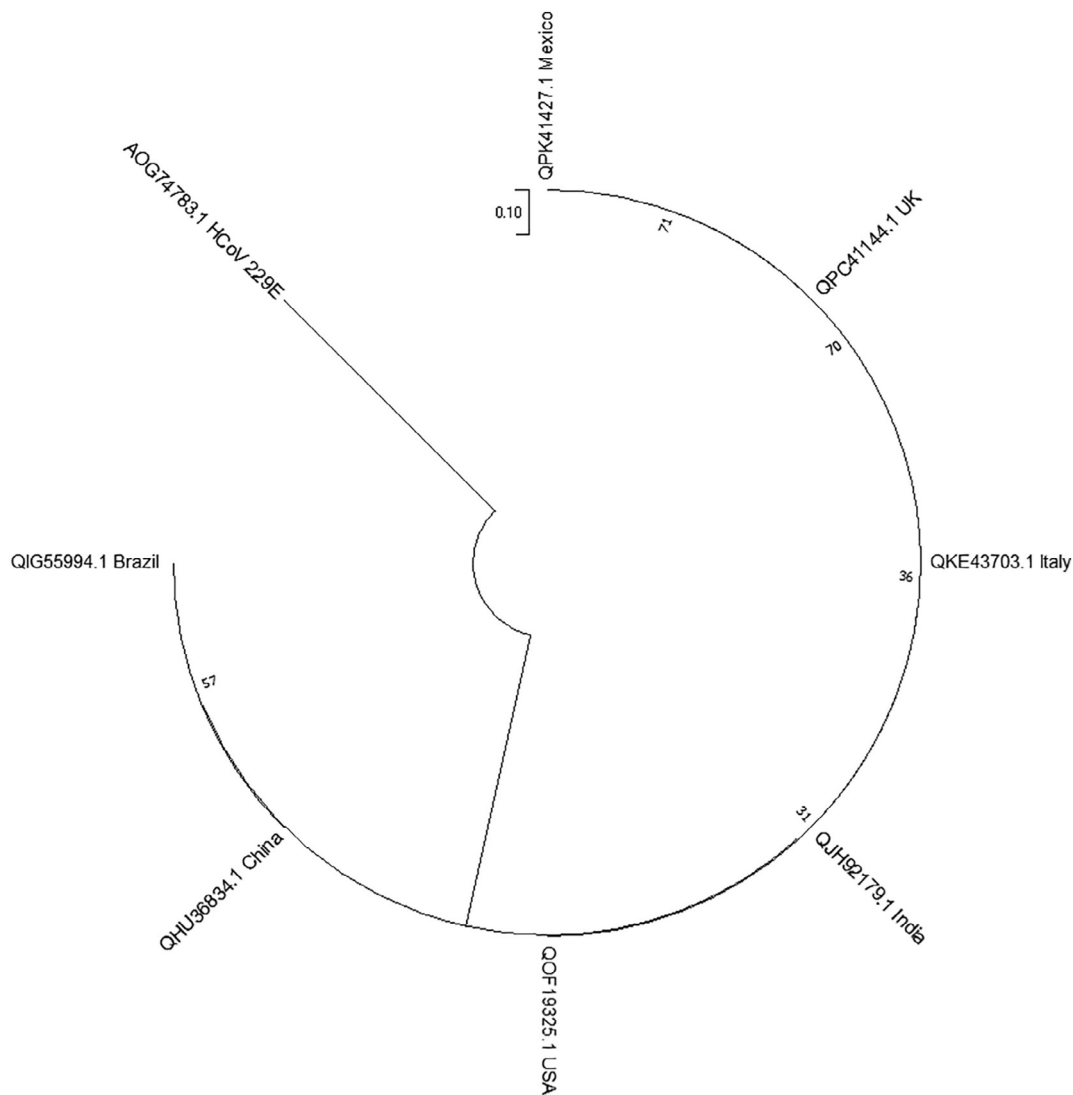


Fig. 2 (continued)

dues, {Threonine-Thr (T)} is involved in ligand interaction site, which is similar in all the sequences selected from the most affected countries but was changed {Tyrosine-Tyr (Y)} at the 202 position of amino acid in S1 subunit of HCoV-229E virus spike glycoprotein (Fig. 3) indicating that the S1 region of spike protein is an ideal target for developing neutralizing antibodies.

In this study, we show that the S1 subunit (319–541) can be considered as potential target site for stable and safe vaccine development across the globe. Since SARS-CoV-2 was identified as the causative agent of COVID-19, and its first genome sequence was released immediately for free by a Chinese research group [32], SARS-CoV-2 vaccine candidates have been developed, based on various vaccine platforms, such as inactivated or live attenuated vaccines, DNA and mRNA vaccines, viral vector-based vaccines, and recombinant protein-based vaccines [12,33]. Most of these vaccine strategies are based on the full-length S glycoprotein, the major viral surface antigen [12]. When a vaccine strategy requires that the SARS-CoV-2 S protein be recombinantly expressed in the human body, the ER retrieval

signal (ERRS) should be omitted to enhance the cell-surface expression level of the resulting protein. Theoretically, the native HIV-1 Envelope trimer present on the surface of intact virions is thought to be the most ideal immunogen [34], as most of the neu-

tralizing antibodies thus far described, could recognize and bind to the prefusion form of trimeric HIV-1 Envelope, although such neutralizing antibodies against this glycan-covered, sequence-variable native form were induced with great difficulty [35]. For SARS-CoV-2, different lines of research have shown that convalescent sera from SARS-CoV and SARS-CoV-2 patients showed no or limited cross-neutralization activity against these two viruses by pseudotyped and authentic viral infection assays, despite significant cross-reactivity in binding to the S glycoproteins of both the viruses [9,36–38]. Similar results were also observed in infected or immunized animals ([18,36,38]. Together with the finding that although the SARS-CoV-2 S protein shares a high degree of amino-acid sequence identity with that of SARS-CoV (~76% overall), the receptor-binding motif (RBM) is less conserved (~47% identity) than any other functional region or domain [27], It can thus be surmised that the RBM has the most immunodominant neutralizing epitope(s) of the whole S protein and is capable of readily eliciting strong neutralizing antibody responses. It should be noted that although the vaccine candidates based on the full-length S protein of the closely related SARS-CoV could elicit neutralizing antibody responses against infection of SARS-CoV, they may also induce harmful immune responses, including damage to the liver of the vaccinated animals, infection of human immune cells by

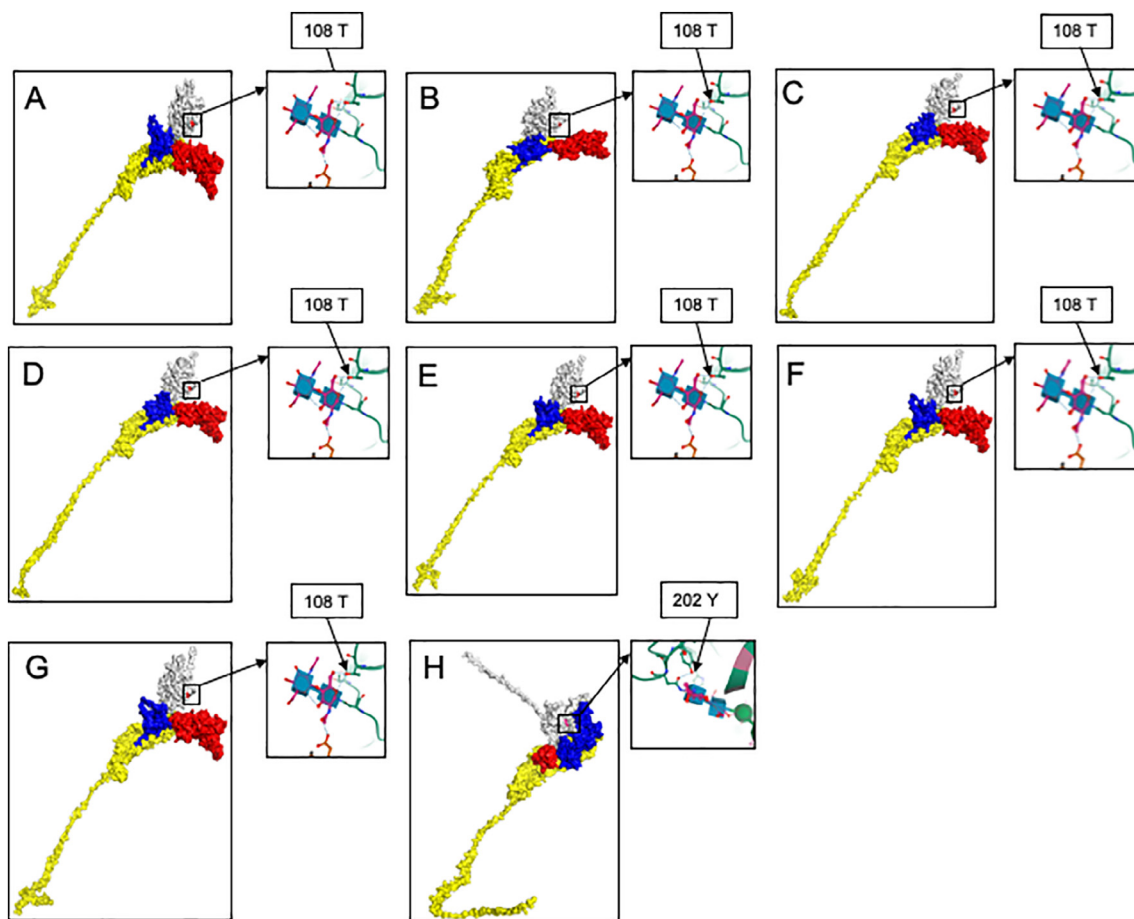


Fig. 3. 3D Models and Ligand interaction sites of SARS CoV 2 spike protein 3D Models of SARS CoV 2 spike protein (USA-A, India-B, Brazil-C, Mexico-D, UK-E, China-F, and Italy-G) (corresponding to 5 Protein Units) were detected in spike protein 4 (Domain 1[Grey]: [1–289], Domain 2[Blue]: [290–319];[591–703], Domain 3[Red] : [320–590] and Domain 4 [Yellow] : [704–1273] and for alpha human SARS-CoV(HCoV-229E)-H spike protein. The domains detected were 4 with varied amino-acid residue range compared to S protein of selected countries, which were (Domain 1[Grey]: [1–280], Domain 2[Blue]: [281–495], Domain 3[Red]: [496–563] and Domain 4[Yellow]: [564–1171]). The ligand interaction site is shown as red-colored residue along with the interaction site 3D image. Ligand interaction sites of spike protein. The ligand interaction site is at 108{Threonine (T)} in all at amino-acid residues of the affected countries and for HCoV 229E the ligand interaction site is at 202 amino-acid residue {Tyrosine (Y)} using Multiple sequence alignment of the SARS CoV2 spike protein. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

SARS-CoV, and antibody-dependent enhancement of SARS-CoV infection [39–43]. Therefore, although the S proteins of both SARS-CoV and SARS-CoV-2 are thought to be promising vaccine immunogens for generating protective immunity, optimizing antigen design is critical to ensure an optimal immune response through exposing more neutralizing epitopes and displaying fewer potentially weakly or non-neutralizing epitopes [44]. Vaccines containing or expressing the full-length S protein or its soluble ectodomain form should thus be engineered to sample an RBD(s) “up” conformation while the rest is still kept in prefusion state [45,46]. The RBD of SARS-CoV is highly immunogenic [47,48] and is targeted by most of the neutralizing monoclonal antibodies that have been characterized [49]. Based on the observation that a 193-amino-acid fragment (residues 318–510) was previously identified to be the minimal RBD region of SARS-CoV [50], a corresponding 194-amino-acid fragment (residues 331–524) can be readily selected as the minimal RBD region of SARS-CoV-2 and has already been characterized [51]. This the minimal form of RBDs of both viruses could serve as a vaccine candidate [51]. Even the phytochemicals present in the extracts of *Andrographis paniculata* and *Phyllanthus amarus* might have synergistic effect with action on multiple target sites of SARS-CoV-2 [52]. Considering the importance of designing stable vaccines for SARS CoV-2 (COVID19) and

the research work carried on the virus spike protein, we have evaluated the S1 subunit (319–541) amino-acid sequences retrieved from NCBI viruses SARS-CoV-2 Data Hub. The study revealed that 114 number of spike protein amino-acid sequences from seven affected countries retrieved from NCBI virus SARS-CoV-2 Data Hub of different countries in NCBI are highly conserved when compared to Alpha human CoV (HCoV 229E) considered as out group. Further, multiple sequence alignment of S1 subunit (319–541) of the affected countries considered in this study showed 100% similarity, which clearly indicated that this S1 RBD is a potential target for developing neutralizing antibodies and right choice for development of stable and safe vaccines against

coronavirus disease 2019. To confirm further the 3D models were developed by homology modeling using bioinformatic tools mentioned in material and methods, the 3D models of SARS CoV-2 spike protein were subjected to domain identification, which showed four domains in all the spike protein amino-acid sequences of the affected countries considered and their range remained conserved but varied when compared to HCoV- 229E. The ligand interaction sites modeled in the SARS CoV 2 spike protein amino-acid sequences from affected countries retrieved from NCBI using RCSB PDB database revealed that Threonine (T) at 108 position of amino-acid position in the spike protein is involved

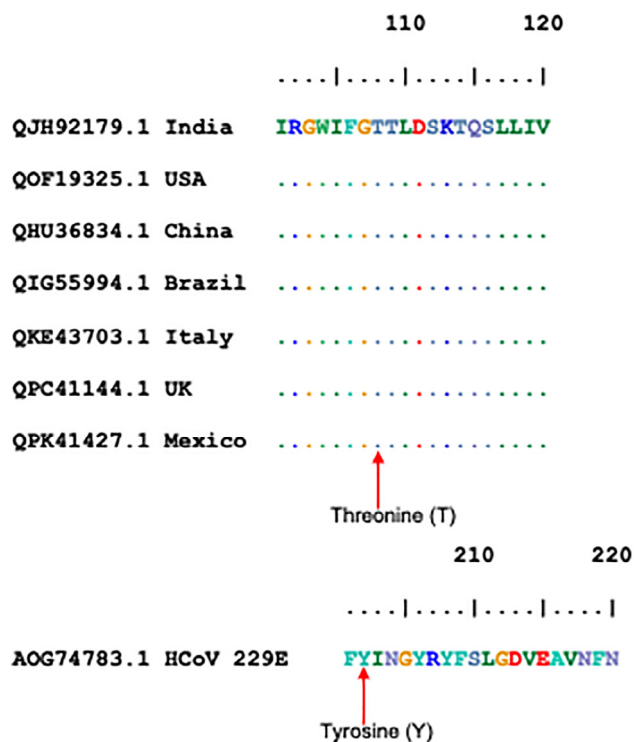


Fig. 3 (continued)

in ligand interaction and remained the same in all the spike protein amino-acid sequences of the seven affected countries considered but changed to Tyrosine (Y) at 202 position of amino-acid sequence in the spike protein of Alpha human CoV(HCoV-229E) considered for comparison. To conclude, the findings of our *In-silico* analysis of SARS CoV-2 spike protein amino-acid sequences retrieved from NCBI virus SARS-CoV-2 Data Hub indicates that the RBD S1 subunit (319–541) is a potential target for designing stable and safe vaccines against coronavirus disease 2019.

4. Conclusion

The SARS-CoV-2 is a contagious virus that continues to spread rapidly around the globe and caused one of the worst pandemics in history. A safe and efficacious vaccine represents one of the best ways to eliminate the COVID-19 pandemic. Although several potential SARS-CoV-2 vaccines have been licensed and used for vaccination across the globe, still there is a need for a stable and efficient vaccine. Vaccine candidates designed to elicit such neutralizing antibodies are feasible and it is widely accepted that the S protein of SARS-CoV-2 is the most promising immunogen for producing protective immunity. The spike protein S1 subunit (RBD), which is conserved and analyzed *in-silico* in this study based on the SARS-CoV-2 S glycoprotein research related to its biosynthesis, structure, function, antigenicity as well as immunogenicity will likely contribute to the success of safe and efficacious vaccine development against COVID-19.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

We thank all the members of the Department of Biotechnology, University of Agricultural Sciences, Bangalore, College of Agriculture, Hassan and the Indian Institute of Horticultural Research, Bangalore, for their comments and valuable inputs.

References

- [1] R. Keni, A. Alexander, P.G. Nayak, J. Mudgal, K. Nandakumar, COVID-19: Emergence, Spread, Possible Treatments, and Global Burden, *Front. Public Health* 8 (2020) 216, <https://doi.org/10.3389/fpubh.2020.00216>.
- [2] Y. Liang, M.-L. Wang, C.-S. Chien, A.A. Yarmishyn, Y.-P. Yang, W.-Y. Lai, Y.-H. Luo, Y.-T. Lin, Y.-J. Chen, P.-C. Chang, S.-H. Chiou, Highlight of Immune Pathogenic Response and Hematopathologic Effect in SARS-CoV, MERS-CoV, and SARS-Cov-2 Infection, *Front. Immunol.* 11 (2020), <https://doi.org/10.3389/fimmu.2020.01022>.
- [3] P. Mangar, S. Pradhan, S. Rai, K. Lepcha, V.K. Ranjan, A. Rai, Comparative Analysis Based on the Spike Glycoproteins of SARS-CoV2 Isolated from COVID 19 Patients of Different Countries, *Preprints* (2020), 2020040154 (doi: 10.20944/preprints202004.0154.v1).
- [4] E.J. Lefkowitz, D.M. Dempsey, R.C. Hendrickson, R.J. Orton, S.G. Siddell, D.B. Smith, Virus taxonomy: the database of the International Committee on Taxonomy of Viruses (ICTV), *Nucleic Acids Res.* 46 (D1) (2018) D708–D717, <https://doi.org/10.1093/nar/gkx932>.
- [5] J. Cui, F. Li, Z.-L. Shi, Origin and evolution of pathogenic coronaviruses, *Nat. Rev. Microbiol.* 17 (3) (2019) 181–192, <https://doi.org/10.1038/s41579-018-0118-9>.
- [6] S.C. Harrison, Viral membrane fusion, *Virology* 479–480 (2015) 498–507, <https://doi.org/10.1016/j.virol.2015.03.043>.
- [7] J. Shang, Y. Wan, C. Luo, G. Ye, Q. Geng, A. Auerbach, F. Li, Cell entry mechanisms of SARS-CoV-2, *Proc. Natl. Acad. Sci. U.S.A.* 117 (21) (2020) 11727–11734, <https://doi.org/10.1073/pnas.2003138117>.
- [8] T. Tang, M. Bidon, J.A. Jaimes, G.R. Whittaker, S. Daniel, Coronavirus membrane fusion mechanism offers a potential target for antiviral development, *Antiviral Res.* 178 (2020), <https://doi.org/10.1016/j.antiviral.2020.104792> 104792.
- [9] X. Ou, Y. Liu, X. Lei, P. Li, D. Mi, L. Ren, L. Guo, R. Guo, T. Chen, J. Hu, Z. Xiang, Z. Mu, X. Chen, J. Chen, K. Hu, Q. Jin, J. Wang, Z. Qian, Characterization of spike glycoprotein of SARS-CoV-2 on virus entry and its immune cross-reactivity with SARS-CoV, *Nat. Commun.* 11 (1) (2020), <https://doi.org/10.1038/s41467-020-15562-9>.
- [10] S. Belouzard, J.K. Millet, B.N. Licitra, G.R. Whittaker, Mechanisms of coronavirus cell entry mediated by the viral spike protein, *Viruses* 4 (2012) 1011–1033, <https://doi.org/10.3390/v4061011>.
- [11] A.C. Walls, Y.-J. Park, M.A. Tortorici, A. Wall, A.T. McGuire, D. Velesler, Structure, Function, and Antigenicity of the SARS-CoV-2 Spike Glycoprotein, *Cell* 181 (2) (2020) 281–292.e6, <https://doi.org/10.1016/j.cell.2020.02.058>.
- [12] F. Amanat, F. Krammer, SARS-CoV-2 Vaccines Status Report, *Immunity* 52 (4) (2020) 583–589, <https://doi.org/10.1016/j.immuni.2020.03.007>.
- [13] R. Lu, X. Zhao, J. Li, P. Niu, B.o. Yang, H. Wu, W. Wang, H. Song, B. Huang, N.a. Zhu, Y. Bi, X. Ma, F. Zhan, L. Wang, T. Hu, H. Zhou, Z. Hu, W. Zhou, Li. Zhao, J. Chen, Y. Meng, J.i. Wang, Y. Lin, J. Yuan, Z. Xie, J. Ma, W.J. Liu, D. Wang, W. Xu, E. C. Holmes, G.F. Gao, G. Wu, W. Chen, W. Shi, W. Tan, novel coronavirus: implications for virus origins and receptor binding, *Lancet* 395 (10224) (2020) 565–574.
- [14] B.J. Bosch, R. van der Zee, C.A.M. de Haan, P.J.M. Rottier, The coronavirus spike protein is a class I virus fusion protein: structural and functional characterization of the fusion core complex, *J. Virol.* 77 (16) (2003) 8801–8811.
- [15] Y. Watanabe, J.D. Allen, D. Wrapp, J.S. McLellan, M. Crispin, Site-specific glycan analysis of the SARS-CoV-2 spike, *Science* 369 (6501) (2020) 330–333.
- [16] S. Xia, Y. Zhu, M. Liu, Q. Lan, W. Xu, Y. Wu, T. Ying, S. Liu, Z. Shi, S. Jiang, L.u. Lu, Fusion mechanism of 2019-nCoV and fusion inhibitors targeting HR1 domain in spike protein, *Cell. Mol. Immunol.* 17 (7) (2020) 765–767.
- [17] S. Bertram, R. Dijkman, M. Habjan, A. Heurich, S. Gierer, I. Glowacka, K. Welsch, M. Winkler, H. Schneider, H. Hofmann-Winkler, V. Thiel, S. Pöhlmann, TMPRSS2 activates the human coronavirus 229E for cathepsin-independent host cell entry and is expressed in viral target cells in the respiratory epithelium, *J. Virol.* 87 (11) (2013) 6150–6160.
- [18] M. Hoffmann, H. Kleine-Weber, S. Schroeder, N. Krüger, T. Herrler, S. Erichsen, T.S. Schiergens, G. Herrler, N.-H. Wu, A. Nitsche, M.A. Müller, C. Drosten, S. Pöhlmann, SARS-CoV-2 Cell Entry Depends on ACE2 and TMPRSS2 and Is Blocked by a Clinically Proven Protease Inhibitor, *Cell* 181 (2) (2020) 271–280. e8, <https://doi.org/10.1016/j.cell.2020.02.052>.
- [19] L. Du, R.Y. Kao, Y. Zhou, Y. He, G. Zhao, C. Wong, S. Jiang, K.-Y. Yuen, D.-Y. Jin, B.-J. Zheng, Cleavage of spike protein of SARS coronavirus by protease factor Xa is associated with viral infectivity, *Bio.- Chem. Biophys. Res. Commun.* 359 (1) (2007) 174–179.
- [20] S. Kumar, G. Stecher, M. Li, C. Knyaz, K. Tamura, F.U. Battistuzzi, MEGA X: Molecular Evolutionary Genetics Analysis across computing platforms. *Molecular Biology and Evolution* 2018; 35:1547–1549.
- [21] N. Saitou, M. Nei, The neighbor-joining method: A new method for reconstructing phylogenetic trees, *Mol. Biol. Evol.* 4 (1987) 406–425.

- [22] J. Felsenstein, Confidence limits on phylogenies: An approach using the bootstrap, *Evolution* 39 (4) (1985) 783–791.
- [23] E. Zuckerkandl, L. Pauling, Evolutionary divergence and convergence in proteins. Edited in *Evolving Genes and Proteins* by V. Bryson and H.J. Vogel, 1965; 97–166. Academic Press, New York.
- [24] L.A. Kelley, S. Mezulis, C.M. Yates, M.N. Wass, M.J.E. Sternberg, The Phyre2 web portal for protein modeling, prediction and analysis, *Nat. Protoc.* 10 (6) (2015) 845–858, <https://doi.org/10.1038/nprot.2015.053>.
- [25] J.-C. Gelly, A.G. de Brevern, Protein Peeling 3D: new tools for analyzing protein structures, *Bioinformatics* 27 (1) (2011) 132–133, <https://doi.org/10.1093/bioinformatics/btq610>.
- [26] D. Sehnal, A.S. Rose, J. Kořánek, S.K. Burley, Velankar, S: Mol*: towards a common library and tools for web molecular graphics MolVA/EuroVis Proceedings. Proceedings of the Workshop on Molecular Graphics and Visual Analysis of Molecular Data, 2018; 29–33 doi:10.2312/molva.20181103.
- [27] J.A. Jaimes, N.M. André, J.S. Chappie, J.K. Millet, G.R. Whittaker, Phylogenetic Analysis and Structural Modeling of SARS-CoV-2 Spike Protein Reveals an Evolutionary Distinct and Proteolytically Sensitive Activation Loop, *J. Mol. Biol.* 432 (10) (2020) 3309–3325, <https://doi.org/10.1016/j.jmb.2020.04.009>.
- [28] L. Duan, Q. Zheng, H. Zhang, Y. Niu, Y. Lou, H. Wang, The SARS-CoV-2 Spike Glycoprotein Biosynthesis, Structure, Function, and Antigenicity: Implications for the Design of Spike-Based Vaccine Immunogens, *Front. Immunol.* 11 (2020), <https://doi.org/10.3389/fimmu.2020.576622>.
- [29] G. Faure, A. Bornot, A.G. de Brevern, Analysis of protein contacts into Protein Units, *Biochimie* 91 (7) (2009) 876–887, <https://doi.org/10.1016/j.biochi.04.008>. ISSN 0300-9084.
- [30] J.-C. Gelly, A.G. de Brevern, S. Hazout, 'Protein Peeling': an approach for splitting a 3D protein structure into compact fragments, *Bioinformatics* 22 (2) (2006) 129–133, <https://doi.org/10.1093/bioinformatics/bti773>.
- [31] Jean-Christophe Gelly, C. Etchebest, S. Hazout, A.G. de Brevern, Protein Peeling 2: a web server to convert protein structures into series of protein units, *Nucleic Acids Research* 34 (2) (2006) W75–W78, <https://doi.org/10.1093/nar/gkl292>.
- [32] F. Wu, S. Zhao, B. Yu, Y.-M. Chen, W. Wang, Z.-G. Song, Y. Hu, Z.-W. Tao, J.-H. Tian, Y.-Y. Pei, M.-L. Yuan, Y.-L. Zhang, F.-H. Dai, Y. Liu, Q.-M. Wang, J.-J. Zheng, L. Xu, E.C. Holmes, Y.-Z. Zhang, A new coronavirus associated with human respiratory disease in China, *Nature* 579 (7798) (2020) 265–269, <https://doi.org/10.1038/s41586-020-2008-3>.
- [33] T. Thanh Le, Z. Andreadakis, A. Kumar, R. Gómez Román, S. Tollefsen, M. Saville, S. Mayhew, The COVID-19 vaccine development landscape, *Nat Rev Drug Discov* 19 (5) (2020) 305–306, <https://doi.org/10.1038/d41573-020-00073-5>.
- [34] L. Duan, J. Du, X. Liu, Insights into vaccine development for acquired immune deficiency syndrome from crystal structures of human immunodeficiency virus-1 gp41 and equine infectious anemia virus gp45, *Protein Sci.* 24 (10) (2015) 1549–1559, <https://doi.org/10.1002/pro.2750>.
- [35] M. Pancera, T. Zhou, A. Druz, I.S. Georgiev, C. Soto, J. Gorman, et al., Structure and immune recognition of trimeric pre-fusion HIV-1 Env, *Nature* 514 (2014) 455–461, <https://doi.org/10.1038/nature13808>.
- [36] R. Yang, J. Lan, B. Huang, R. A. M. Lu, W. Wang, W. Li, Y. Deng, G. Wong, W. Tan, Lack of antibody-mediated cross-protection between SARS-CoV-2 and SARS-CoV infections, *EBioMedicine* 58 (2020) 102890, <https://doi.org/10.1016/j.ebiom.2020.102890>.
- [37] D.E. Anderson, C.W. Tan, W.N. Chia, B.E. Young, M. Linster, J.H. Low, et al., Lack of cross-neutralization by SARS patient sera towards SARS-CoV-2, *Emerg Microbes Infect* 9 (2020) 900–902, <https://doi.org/10.1080/22221751.2020.1761267>.
- [38] H. Lv, N.C. Wu, O.-Y. Tsang, M. Yuan, R.A.P.M. Perera, W.S. Leung, R.T.Y. So, J.M. C. Chan, G.K. Yip, T.S.H. Chik, Y. Wang, C.Y.C. Choi, Y. Lin, W.W. Ng, J. Zhao, L.L. M. Poon, J.S.M. Peiris, I.A. Wilson, C.K.P. Mok, Cross-reactive Antibody Response between SARS-CoV-2 and SARS-CoV Infections, *Cell Rep.* 31 (9) (2020) 107725, <https://doi.org/10.1016/j.celrep.2020.107725>.
- [39] L. Du, Y. He, Y. Zhou, S. Liu, B.-J. Zheng, S. Jiang, The spike protein of SARS-CoV-2: a target for vaccine and therapeutic development, *Nat. Rev. Microbiol.* 7 (3) (2009) 226–236, <https://doi.org/10.1038/nrmicro2090>.
- [40] S. Jiang, Y. He, S. Liu, SARS vaccine development, *Emerg. Infect. Dis.* 11 (2005) 1016–1020, <https://doi.org/10.3201/1107.050219>.
- [41] C. Ma, S. Su, J. Wang, L. Wei, L. Du, S. Jiang, From SARS-CoV to SARS-CoV-2: safety and broad-spectrum are important for coronavirus vaccine development, *Microbes Infect.* 2 (2020) 245–253, <https://doi.org/10.1016/j.micinf.2020.05.004>.
- [42] K. Karthik, T.M.A. Senthilkumar, S. Udhayavel, G.D. Raj, Role of antibody-dependent enhancement (ADE) in the virulence of SARS-CoV-2 and its mitigation strategies for the development of vaccines and immunotherapies to counter COVID-19, *Hum Vaccin Immunother* (2020) 1–6, <https://doi.org/10.1080/21645515.2020.1796425>.
- [43] A.M. Arvin, K. Fink, M.A. Schmid, A. Cathcart, R. Spreafico, C. Havenar-Daughton, A. Lanzavecchia, D. Corti, H.W. Virgin, A perspective on potential antibody-dependent enhancement of SARS-CoV-2, *Nature* 584 (7821) (2020) 353–363, <https://doi.org/10.1038/s41586-020-2538-8>.
- [44] N. Lurie, M. Saville, R. Hatchett, J. Halton, Developing Covid-19 Vaccines at Pandemic Speed, *N. Engl. J. Med.* 382 (21) (2020) 1969–1973, <https://doi.org/10.1056/NEJMp2005630>.
- [45] M. Mandolesi, D.J. Sheward, L. Hanke, J. Ma, P. Pushparaj, L.P. Vidakovic, et al., SARS-CoV-2 protein subunit vaccination elicits potent neutralizing antibody responses, *BioRxiv.* (2020), <https://doi.org/10.1101/2020.07.31.228486>.
- [46] K.S. Corbett, D.K. Edwards, S.R. Leist, O.M. Abiona, S. Boyoglu-Barnum, R.A. Gillespie, et al., SARS-CoV-2 mRNA vaccine design enabled by prototype pathogen preparedness, *Nature* (2020), <https://doi.org/10.1038/s41586-020-2622-0>.
- [47] L. Premkumar, B. Segovia-Chumbez, R. Jadi, D.R. Martinez, R. Raut, A. Markmann, C. Cornaby, L. Bartelt, S. Weiss, Y. Park, C.E. Edwards, E. Weimer, E.M. Scherer, N. Roupael, S. Edupuganti, D. Weiskopf, L.V. Tse, Y.J. Hou, D. Margolis, A. Sette, M.H. Collins, J. Schmitz, R.S. Baric, A.M. de Silva, The receptor binding domain of the viral spike protein is an immunodominant and highly specific target of antibodies in SARS-CoV-2 patients, *Sci. Immunol.* 5 (48) (2020) eabc8413, <https://doi.org/10.1126/sciimmunol.abc8413>.
- [48] L. Dai, T. Zheng, K. Xu, Y. Han, L. Xu, E. Huang, Y. An, Y. Cheng, S. Li, M. Liu, M. Yang, Y. Li, H. Cheng, Y. Yuan, W. Zhang, C. Ke, G. Wong, J. Qi, C. Qin, J. Yan, G.F. Gao, A Universal Design of Betacoronavirus Vaccines against COVID-19, MERS, and SARS, *Cell* 182 (3) (2020) 722–733.e11, <https://doi.org/10.1016/j.cell.2020.06.035>.
- [49] S. Jiang, C. Hillyer, L. Du, Neutralizing Antibodies against SARS-CoV-2 and Other Human Coronaviruses, *Trends Immunol.* 41 (5) (2020) 355–359, <https://doi.org/10.1016/j.it.2020.03.007>.
- [50] S.K. Wong, W. Li, M.J. Moore, H. Choe, M. Farzan, A 193-amino acid fragment of the SARS coronavirus S protein efficiently binds angiotensin-converting enzyme 2, *J. Biol. Chem.* 279 (5) (2004) 3197–3201, <https://doi.org/10.1074/jbc.C300520200>.
- [51] W. Tai, L. He, X. Zhang, J. Pu, D. Voronin, S. Jiang, Y. Zhou, L. Du, Characterization of the receptor-binding domain (RBD) of 2019 novel coronavirus: implication for development of RBD protein as a viral attachment inhibitor and vaccine, *Cell. Mol. Immunol.* 17 (6) (2020) 613–620, <https://doi.org/10.1038/s41423-020-0400-4>.
- [52] S. Hiremath, H.D.V. Kumar, M. Nandan, et al., In silico docking analysis revealed the potential of phytochemicals present in *Phyllanthus amarus* and *Andrographis paniculata*, used in Ayurveda medicine in inhibiting SARS-CoV-2, *Biotech* 11 (2021) 44, <https://doi.org/10.1007/s13205-020-02578-7>.