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.
In silico identification of RBD subdomain of spike protein from Pro\(^{322}\)–Thr\(^{581}\) for applications in vaccine development against SARS-CoV2

Nataraj S Pagadala\(^a,\)*, Abdolamir Landi\(^b\), Paramahamsa Maturu\(^c\), Jack Tuszyński\(^b\)

\(^a\) Carnegie Mellon University, Department of Chemistry, Mellon Institute Bldg, 4400 Fifth Avenue, Pittsburgh, PA 15213-2683, United States
\(^b\) Li Ku Shing Applied Virology Institute, Department of Medical Microbiology and Immunology, University of Alberta, Edmonton, AB, Canada
\(^c\) Department of Pediatrics, Baylor College of Medicine, Houston, TX, United States

**A R T I C L E   I N F O**

Article history:
Received 11 January 2021
Revised 17 April 2021
Accepted 19 April 2021
Available online 30 April 2021

**Keywords:**
SARS-CoV 2
Spike protein
ACE2

**A B S T R A C T**

The three-dimensional hybrid structures of coronavirus spike proteins including the C-terminal sequence and receptor binding motif (RBM) was remodeled and energy minimized. Further, protein–protein docking show that Receptor Binding Domain (RBD) of SARS-CoV 2 Lys\(^{577}\)–Pro\(^{590}\) bind on the surface of ACE2 receptor near N-terminal helices to form host-pathogen attachment. In this binding interface, SARS-CoV 2 shows a tight network of hydrogen bonds than other spike proteins from BtRsRaTG13–CoV, SARS-CoV, BtRsBeta–CoV, BtRsCoV–related, Pangolin–CoV (PCoV), human–CoV (hCoV), MERS–CoV (MCoV), Avian–CoV (AcoV) and PEDV1–CoV. Further studies show that subdomains from SARS-CoV 2 RBD Pro\(^{577}\)–Thr\(^{581}\), SARS-CoV RBD Pro\(^{308}\)–Pro\(^{312}\), BtRsRaTG13 RBD Thr\(^{581}\)–Thr\(^{585}\), BtRsBeta–CoV RBD Ser\(^{311}\)–Thr\(^{315}\), BtRsCoV–related Arg\(^{308}\)–Pro\(^{312}\) and PCoV RBD Gin\(^{308}\)–Ser\(^{312}\) show binding conformations with ACE2 like their full-length structures of spike proteins. In addition, the subdomains MCoV RBD Gly\(^{312}\)–Val\(^{316}\), ACoV RBD Gly\(^{312}\)–Val\(^{316}\) and PEDV1–CoV RBD Ala\(^{151}\)–Tyr\(^{175}\) also binds on the surface of ACE2 similar to their full-length spike proteins. The B-Cell epitope mapping also identified main antigenic determinants predicting that these nine subdomains are highly useful in recombinant vaccine development in inducing cross-neutralizing antibodies against SARS-CoV 2 spike protein and inhibits its attachment with ACE2.

© 2021 The Authors. Published by Elsevier B.V.
This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/)

1. Introduction

Coronaviruses (CoVs) are the largest group of RNA viruses, which belong to the genus coronavirus, the family coronaviridae, and the order nidovirales. The genome of CoV is a single stranded positive-sense RNA (+ssRNA) with a size of 27–32 kb [1,2]. There are 4 structural proteins including spike (S), envelope (E), membrane (M), and nucleocapsid (N) in coronaviruses which are embedded into envelope, contributing to the crown-like feature of viral particles [3]. This family of viruses include human coronaviruses (hCoV)—229E, hCoV-OC43, hCoV HKU-1, and hCoV NL63, causing mild upper respiratory infection, known as common cold and are constantly circulating among 70% of the human population [4]. In contrast, two fatal coronaviruses, severe acute respiratory syndrome SARS-CoV and Middle East respiratory syndrome MERS-CoV that are causing severe upper and lower respiratory diseases leading to fatal pneumonia are transmitted from animals to humans [5]. SARS-CoV, which resides in Chinese horseshoe bats as a natural reservoir, was associated with 8096 cases and 774 deaths globally in 2002–3, started in Guangdong province, China, [6]. The virus had been transmitted humans through civet cats and raccoon dogs that were consumed as food and sold in Chinese wet markets [7]. Due to lack of specific antivirals or approved vaccines for the SARS-CoV in 2002–3, conventional measures had been taken to stop the spread of the disease, including travel restrictions and patient isolation. MERS-CoV infection that was first reported in Saudi Arabia in 2012, was mainly spread in the Middle East and later self-controlled with ~2000 infected cases with a fatality rate of ~35%. Both SARS and MERS had limited spread and are not a health concern anymore. In 2019 December, a novel form

[https://doi.org/10.1016/j.molstruc.2021.130534](https://doi.org/10.1016/j.molstruc.2021.130534)

0022-2860/© 2021 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/)
of coronavirus named as SARS-CoV 2 emerged in China at Wuhan city, causing severe pandemic affecting the global public health resulting in progressive respiratory failure due to alveolar damage and death [8–10]. According to daily report by Centre for Systems Science and Engineering at Johns Hopkins University, as of February 25, 2021, there have been over 112 million confirmed SARS-CoV 2 infections with more than 2,5 million global fatal cases, exceeding any former epidemics by coronaviruses.

The infections by coronavirus are mainly due to the process of receptor binding by the spike membrane glycoprotein (S protein) in mediating membrane fusion [11], resulting in the high virulence of SARS-CoV 2. The mechanism of spike-mediated membrane fusion, which is similar to that of class I virus fusion proteins have been previously studied in murine coronavirus (mouse hepatitis virus; MHV) [12,13]. This mechanism of membrane fusion is due to attachment of S1 subunit of spike protein to the cellular receptor, facilitating viral attachment to the surface of target cells. Similarly, studies have shown angiotensin converting enzyme 2 (ACE2), which regulates blood pressure acts as a cellular receptor for viral entry by SARS-CoV and hCoV NL63 [14–18], where cellular serine protease TMPRSS2 is used for cleavage and conformational changes of S protein, called priming [15,19–21]. Current studies on SARS-CoV 2 have also demonstrated that ACE2 receptor is utilized as the entry point in Chinese horseshoe bats, civet, swine, but not in mouse [10]. These observations clearly reveal that ACE2 plays a key role in SARS-CoV spread. As shown for SARS-CoV, the virus binds to the peptidase domain of ACE2 and both spike and ACE2 (primarily expressed on pulmonary epithelium) are cleaved by cellular proteases such as TMPRSS2. This results in conformational change in spike and allows it to insert its S1 subunit into the membrane, facilitating viral entry. During the entry process, spike cleavage is critical for virus entry and blocking the cleavage would reduce viral entry [22].

Comparative studies on the viral sequences have demonstrated a similarity of ~80% between SARS-CoV 2 and SARS-CoV with major difference to be in three regions. These differences exist in open reading frame (ORF) 1a/b, ORF8, expressing a protein involved in immune evasion, and more importantly spike region. This similarity is even higher with BtRsRaTG13-CoV, which is 96% identical to SARS-CoV 2 in its amino acid (AA) sequence. However, since the mismatch is localized at Receptor Binding Domain (RBD) of S protein, BtRsRaTG13 CoV does not infect humans due to lack of binding to ACE2. Conversely, the RBD domain of S is highly identical to that of another Bat coronavirus detected in Pangolin; however, pangolin CoV does not infect humans either, because of significant differences in other parts of spike protein [23]. Accordingly, it has been hypothesized by other authors, that during a cross-species recombination, the RBD in BtRsRaTG13-CoV might have been substituted by that of PCoV to produce SARS-CoV 2 that can infect humans. The other unique feature of SARS-CoV 2 is the cleavage domain between S1 and S2. This domain seems to be acquired by adding a number of amino acids, making the region more susceptible to a wide range of proteases, facilitating the conformational change in S protein and insertion of its S1 subunit into membrane [24,25]. Although it is not yet known whether SARS-CoV 2 and SARS-CoV sequence similarities correlates with similar biological properties, including pandemic potential [26], the interface details for Spike/ACE2 elucidated that SARS-CoV 2 transmissibility is due to efficient use of ACE2 as a key determinant at the atomic level [27,28].

Regardless of strict health measures such as social distancing, lockdown in many parts of the world, the high transmissibility of the virus still results in a significant number of infected cases around the world, which makes a fatality rate of 2% a very significant loss. To tackle this crisis, scientists have started lots of efforts in two major paths to first develop a vaccine to control transmission and spread of the infection and second to manufacture antivirals to treat the infected cases. As of now, more than 10 vaccines are approved for SARS-CoV 2 while over 250 teams are still working to develop vaccines against the virus using different methods (https://www.who.int/publications/m/item/draft-landscape-of-covid-19-candidate-vaccines). These includes development of inactivated/weakens virus particles, nucleic acid (DNA or RNA) vaccines, non/replicating viral vectors, and protein based vaccines including recombinant subunit proteins or virus-like particles [29]. Although the non-protein developed vaccines may help with the urgent need to protect at risk population, vaccine previous experiences suggest that recombinant protein-based vaccine would be likely the most efficient and safest vaccine for long-term use as a prophylactic vaccine for public. Current evidence almost unanimously recommends spike protein as the best candidate to develop an optimal vaccine with respect to humoral and cellular immune responses. Since antibody dependent enhancement is also a potential concern for SARS-CoV 2 vaccine, it is reasonable to pick as small as possible part of spike protein that is critical target to be used as vaccine. In this study, we have studied the spike protein from 10 different coronaviruses of animals and humans, including SARS-CoV and SARS-CoV 2 to pinpoint the most critical region of S protein to be used as an antigen for vaccine development.

2. Methodology

Protein sequences of spike proteins from, SARS-CoV 2 (QHD43416.1), BtRsRaTG13-CoV (QHR63300.2), SARS-CoV (AAP134411.1), BtRsBeta-CoV (QDF43825.1), BtRsCoV-related (AT098157.1), PChoV (QIQ54048.1), hCoV-HKU1 (BBA209561.1), MCoV (YP_000947204.1), AcCoV (~ACV87265.1) and PEDV-CoV (ALB835885.1) were obtained from the National Center for Biotechnology Information (NCBI) database. The initial homology models of full-length spike protein from SARS-CoV 2 was remodeled including the missing C-terminal sequence and receptor binding motif using the crystal structure of 2019-ncov chimeric receptor-binding domain (PDB ID: 6VV1) with MODELLER 9v7 on windows operating system [30]. The co-ordinates for the structurally conserved regions (SCRs) of RBD SARS-CoV 2 sequence were assigned from the template using pair wise sequence alignment, based on the Needleman-Wunsch algorithm [31,32]. In addition, BtRsRaTG13-CoV, SARS-CoV, BtRsBeta-CoV and BtRsCoV-related, PCoV, hCoV, MCoV, AcCoV, and PEDV-CoV homology models were developed with the same methodology as described above using the build homology model of SARS-CoV 2 as the template. Further, protein-protein docking studies and their interactions of the full-length SARS-CoV 2, BtRsRaTG13-CoV, SARS-CoV, BtRsBeta-CoV, BtRsCoV-related, PCoV, hCoV, MCoV, AcCoV and PEDV-CoV spike proteins with its receptor ACE2 (PDBID: 6VV1) was performed with the online server ZDOCK in which proteins were treated as rigid objects and 6-dimensional rotational and translational degrees of freedom were explored. Similarly, protein–protein docking was also performed using the spike RBD subdomains, SARS-CoV 2 RBD Pro322-Thr581, BtRsRaTG13-CoV RBD Thr381-Thr322, SARS-CoV Pro309-Pro575, BtRsBeta-CoV RBD Ser311-Thr508, BtRsCoV-related Arg306-Pro575, PCoV RBD Gln319-Ser589, hCoVrBRD subdomain Ala315-Tyr675, MCoV RBD Gly372-Val616, AcCoV RBD -Asp250-Gln489 and PEDV-CoV-RBM Ala315-Tyr675 using the above method and their top ten conformations were extracted. For protein–protein docking, the residues from Lys417-508 of the exposed loop regions of the SARS-CoV 2 RBD and Ser19-Met60 of ACE2 receptor were specified in a filter, feature blocking all other residues to involve in the binding interface with the receptor cavity of the ACE2. Finally, ZRANK, a scoring algorithm that relies on the usage of
a combination of three atom-based terms, i.e., Van der Waals, electrostatics, and desolvation was used to rank the structures [33–38]. Out of top 10 conformations that were generated, the top conformation of the SARS-CoV 2 spike protein RBM binds on the surface of the receptor ACE2 for viral host interaction was analysed. In addition, the binding conformations of full-length BrRsTaTG13-CoV, SARS-CoV, BrtRsBeta-CoV, BrtRsCoV-related, PCoV, hCoV, MCoV, ACoV and PEDV1-CoV full-length spike proteins and their subdomains SARS-CoV 2 RBD Pro309-Pro431, BrRsTaTG13-CoV RBD Thr381-Thr332, BrtRsBeta-CoV RBD Ser311-Thr358, PCoV RBD Gln319-Ser359, hCoV RBD Ala315–Tyr475, BrtRsCoV-related Arg306–Pro375, MCoV RBD Gly372–Val416, ACoV RBD Asp250–Gln488 and PEDV1-CoV RBD Ala315–Tyr473 with its receptor ACE2 (PDBID: 6VV1) were also analyzed. The docking conformations of receptor binding spike subdomains that replicates the docking conformation of the full-length spike proteins, still able to bind with higher affinity and with similar hydrogen bonding network with the receptor ACE2 were extracted. Further, the B-Cell antigenic determinants or epitopic sequences of these isolated sub domains were predicted using “ElliPro”, a web-based tool for the prediction of antibody epitopes in protein antigens of a given sequence or structure [39,40].

3. Results and discussion

3.1. Phylogeny

Phylogenetic analysis of the CoV spike proteins falls under five subfamilies. Sequences from PCoV, SARS-CoV 2 and BrRsTaTG13-CoV fall under cluster I with SARS-CoV 2. - On the other hand, MCoV, hCoV, ACoV and PEDV1-CoV are closely related falling under cluster II where MCoV and hCoV falls under subfamily-I while ACoV and PEDV1-CoV falls under subfamily-II. Finally, SARS-CoV, BrtRsCoV-related and BrtRsBeta-CoV falls under cluster III where BrtRsBeta-CoV is too divergent showing separate branch in the phylogenetic tree. The percentage of identity between the sequences reveals that SARS-CoV 2 has 97%, 92%, 76%, 76%, 75%, 26%, 24%, 21% and 19% identity with, BrRsTaTG13-CoV, PCoV, BrtRsCoV-related, BrtRsBeta-CoV, SARS-CoV, MCoV, hCoV, ACoV and PEDV1-CoV, respectively. This shows that SARS-CoV 2, BrtRsTaTG13-CoV and PCoV are very closely related to each other compared to others in the evolution (Fig. 1). Futhers structural studies shows that the RMSD of the full-length SARS-CoV 2 with other species - showed a wide range of deviation from 2.6 to 17.2 Å while the super pose structures of CoV spike subdomain-ACE2 complexes show a least back bone RMSD difference with its full-length spike protein-ACE2 complexes within a range of 0.1-4.1 Å. However, the superimposition of SARS-CoV 2 RBM (Receptor Binding Motif) with others show a least RMSD’s ranging from 0.16 to 0.85 Å where both BrRsTaTG13-CoV and PCoV RBM’s are close related to SARS-CoV 2 RBM with 0.16 and 0.18 Å indicating a clear evolutionary ship between these three species (Table 1 & Fig 2A-2J).

3.2. Spike-protein subdomains-ACE2 interactions

Protein-Protein docking studies of spike protein subdomains shows that Arg403, Lys417 and Tyr453 from SARS-CoV 2 RBD Pro322-Thr358 in the middle of the bridge shows strong network of hydrogen bonds and ionic interactions with His43, Asp30, Asp31, Lys31 and Glu35 of α1. In addition, Tyr449, Glu464, Gly485, Pro491, Glu493, Ser494 and Tyr495 forms another set of strong hydrogen bonding network with six hydrogen bonds and α-stacking interaction in the middle of the bridge with the residues of ACE2 α2, Asn64, Ala71, Lys74, Glu75 and Lys94. In this network of hydrogen bonds, the positively charged Arg403 forms two hydrogen bonds with His34 and Asp38 (3.4 and 3.7 Å), while Lys417 shows contacts with Asp30, Glu35 and Lys31 with two ionic and a hydrogen bond (3.1, 3.0 and 3.8 Å). In addition, the hydroxylphenyl ring of Tyr453 shows contacts with Glu35 in the middle of ACE2 α1 (2.6 Å). Apart from these interactions at α1 of ACE2, Tyr469 towards the N-terminal end of spike protein α2 shows π-interaction with Asn94. In the middle of the α2, the backbone oxygens of Tyr495 and Ser494 and terminal nitrogen of Glu493 shows contacts with Lys48 with three hydrogen bonds (2.9, 3.4 and 2.8 Å). Towards the C-terminal end, Pro491, Glu484 and Gly485 of α2 shows another three hydrogen bonds with the positively charged Lys74 (2.8, 3.4 and 3.5 Å).

Similarly, SARS-CoV 2 RBD Pro322-Thr358 also indicates similar type of interactions with seven hydrogen bonds and two ionic interactions including one π-stacking on the surface of ACE2. At the N-terminal of α1, Leu455 and Tyr473 shows both hydrogen and π-stacking interactions with Lys31 (3.8 and 3.2 Å). The residues, Glu406 and Lys417 in the middle of the bridge forms one hydrogen and two ionic interactions with His34, Asp30 and Glu35 (2.9, 2.6 and 3.5 Å). At the C-terminal end of α1, Asp501 shows a hydrogen bond with Glu442 (2.5 Å). In addition, the residues Glu445, Tyr495, Glu494, Asp497 and Asn498 shows four hydrogen bonds with Glu57 at the N-terminal of ACE2 α2, Lys58 at the middle of the bridge and Thr78 at the C-terminal of α2 (3.0, 3.7, 2.9 and 3.2 Å).

In SARS-CoV-Pro309-Pro431 the residues, Asp30, Arg405, Asp406 and Asp480 forms ionic interactions with Lys74, Glu80, Lys353 and Lys31 (3.5, 3.4, 2.5 and 3.6 Å). In addition, Tyr449 and Tyr492 shows two hydrogen bonds with Glu35 and Glu57 (3.6 and 2.7 Å). The subdomain, BrtRsBeta-CoV Ser311-Thr358 forms a network of hydrogen bonds along with π-interaction on the receptor surface. The residues, Gly483, Asn488 and Tyr492 in the middle of the bridge interact with Asp30 and Lys353 (3.3, 3.0 and 3.9 Å). However, BrtRsCoV-related Arg306-Pro375 binds far from N-ter Met62 of ACE2 with similar orientation like BrtRsBeta-CoV Ser311-Thr358 on the receptor surface predicting different type of interactions between the protein and the receptor. This allows the subdomain to form five hydrogen bonds, including one π-stacking on the receptor surface. The residues, Cys475 at the C-terminal end of spike protein α2 and Asp481 in the middle of the bridge shows two hydrogen bonds with Thr78 and Lys58 (2.7 and 3.6 Å). In extension to these hydrogen bonds, Gly483, Asn488, and Asp492 forms another set of strong hy-
drgen bonds and π-stacking in the middle of α1 with Asp38 and Lys353 (3.3, 3.0, and 3.9 Å).

Oppositely, PcoV RBD Gln310-Ser589 shows similar binding site and orientation compared to SARS-CoV 2 with six hydrogen bonds on the surface of the receptor ACE2. The residues, Tyr419 and Tyr471 at the C-terminal and in the middle of the α1 contacts with two hydrogen bonds (3.6 and 3.4 Å). However, Ala473 shows contacts with Gln75 and Thr78 at the middle of the α2 (3.2 and 2.9 Å) while both Gly444 and Glu491 also show two hydrogen bonds with both negatively and positively charged Glu57 and Lys608 (3.4 Å). In MCoV RBD Gly372-Val516 out of a total of six hydrogen bonds and an ionic interaction with the receptor ACE2, two of them show contacts with N-acetyl-D-glucosamine through Thr492 and Lys493 (1.5 and 3.3 Å). The residue, Lys493 also shows contacts with Glu57 at the N-terminal of α2 (3.3 Å), while Ser524 at the C-terminal end of the α2 contacts with Gln75 (2.9 Å). However, Tyr541 and Lys543 in the middle of the α1 shows contacts with Asp543 through both hydrogen and ionic interactions (2.7 and 3.7 Å). Surprisingly, hCoV RBD Ala315-Tyr475 shows contacts with N-acetyl-D-glucosamine with eleven hydrogen bonds with the receptor ACE2. The residue, Asn452 shows contacts with Asn49 and Thr52 at the C-terminal end of α1 through two hydrogen bonds (3.8 and 3.0 Å). At the N-terminal end of α2, Pro396 and Pro491 shows two hydrogen bonds with Gln36 (2.8 and 3.7 Å). In addition, the residue Ser494 show hydrogen bond with Thr125 at the C-terminal of α5 (2.9 Å). Similar to SARS-CoV 2 subdomain, ACoV RBD Asp250-Gln489 - also shows similar binding orientation with six hydrogen bonds and an ionic interaction with ACE2. The residue, Ser382 at the C-terminal end of α2 shows contacts with Glu75 (3.1 Å) while Tyr390 and Val392 in the middle of the helix α2 also shows contacts with Glu75 and Lys608 (3.0 Å). However, Arg477 at the N-terminal end of the α2 contacts with Glu56 with an ionic interaction (3.9 Å). In extension to these contacts, Gin385 and Cys388 shows contacts with the same Lys51 towards the N-terminal end of the α1 (3.0 Å). Finally, PEDV1-CoV RBD Ala135-Tyr475 shows similar binding site and orientation compared to SARS-CoV 2 subdomain on the ACE2 receptor surface with three hydrogen bonds. The residues, Asn508 and Thr558 at the C-terminal and in the middle of the α1 contacts with two hydrogen bonds (3.5 and 2.8 Å) while Ile451 shows contacts with Thr78 at the middle of the α2 with the distance (3.5 Å).

These protein-protein docking studies of spike protein subdomains reveal that the residues Arg403, Lys417, Tyr449, Tyr453, Glu484, Gly485, Pro489, Gin491, Lys493, Ser504 and Tyr495 of SARS-CoV 2 RBD Pro252-Thr581 (Fig. 3A); Gin460, Lys467, Lys484, Lys495, Tyr527, Glu485, Asn487, Tyr495 and Asp501 of BtRsRatG13-CoV RBD Thr581, Thr323 (Fig. 3B); Asp392, Arg395, Tyr442, Asp463, Thr476, and Asp480 of SARS-CoV 2 RBD Pro398-Pro475 (Fig. 3C); Cys475, Asp481, Gin483, Asn488, and Tyr492 of BetaCoV RBD Ser311-Thr368 (Fig. 3D); Cys475, Asp481, Gin483, Asn488 and Tyr492 of BtRsCoV-related RBD Arg365-Pro475 (Fig. 3E); Tyr491, Gin444, Tyr471, Ala473 and Glu491 of PcoV RBD Gln189-Ser589 (Fig. 3F); Thr492, Lys493, Ser524, Tyr541, and Lys543 of MCoV RBD Gly372-Val516 (Fig. 3G); Ser382, Gin385, Cys388, Tyr390, Val392, and Arg477 of ACoV RBD Asp250-Gln489 (Fig. 3H) and the residues Asn408, Ile551 and Thr558 of PEDV1-CoV RBD Ala135, Tyr475 (Fig. 3I) acts as common pharmacophores in viral host interactions with stronger hydrogen bonds with its receptor ACE2 (Table 2). However, the results reveal that N-acetyl-D-glucosamine from ACE2 plays an important role in viral host interactions with stronger hydrogen bonds in hCoV (Fig. 3I).

3.3. Full-length spike protein -ACE2 interactions

In comparison to the spike subdomains, the residues of the full-length SARS-CoV 2 RBD, Gin474, Gin488, Thr500, and Asp501 at the N and C terminus of α1 form a network of hydrogen bonds with Gin24, Tyr41, Gin42, Met82, Lys353 and Arg357 of ACE2 receptor (Fig. 4A). The residues of the BtRsRatG13-CoV RBD, Lys417, Tyr453, Arg494, Tyr498, Asp501 and His505 shows contacts through eight hydrogen bonds, while BtRsBeta-CoV RBD shows only four hydrogen bonds with ACE2 receptor surface (Fig. 4B and 4D). However, the residues of SARS-CoV 2 RBD, Thr433 Tyr475, Pro477 and Tyr481 at the N and C-terminus of α1 makes contacts with Asp38, Lys58, Glu57 and Glu73. In the middle of the bridge, Ser461 and Leu472 interacts with Met64 and Lys74, respectively, (Fig. 4C). Both Trp442 and Arg475 of BtRsCoV-related forms an hydrogen and ionic interaction with the same His34 while Arg299 also forms ionic interaction with Asp38. Apart from these interactions, Gin471, Asn473 and Tyr475 forms hydrogen bonds and π-interaction with Thr78, Gin24 and Lys31 with ACE2, respectively (Fig. 4E). The substitution of interface residues of PcoV RBD allow Tyr488 and Glu491 to form
Fig. 2. Superimposition of SARS-CoV 2 RBM (Fig 2A) with BtRSRatG13-CoV RBM (Fig 2B), SARS-CoV RBM (Fig 2C), BtRSBeta-CoV RBM (Fig 2D), BtRSCoV-related RBM (Fig 2E), PCoV RBM (Fig 2F), hCoV RBM (Fig 2 G), MCoV RBM (Fig 2H), ACoV RBM (Fig 2I) and with PEDV1-CoV RBM (Fig 2 J) with its receptor ACE2 predicted using MOE software suite [Molecular Operating Environment (MOE), 2014.01; Chemical Computing Group LLC, 1010 Sherbrooke St. West, Suite #910, Montreal, QC, Canada] H3A 2R7, 2021). SARS-CoV 2 RBM is represented in red color while the RBM of the other species is shown in green color.

Table 2

Amino acids of CoV spike protein subdomains involved in binding to its receptor ACE2 predicted using MOE software suite ([Molecular Operating Environment (MOE), 2014.01; Chemical Computing Group LLC, 1010 Sherbrooke St. West, Suite #910, Montreal, QC, Canada] H3A 2R7, 2021).

| Species | Spike Subdomain | Residues of Spike Subdomain involved in binding with the receptor ACE2 | Residues of ACE2 receptor involved in binding with the spike subdomains |
|---------|-----------------|---------------------------------------------------------------------|---------------------------------------------------------------------|
| SARS    | SARS-CoV 2 RBM  | Pro\(^{312}\)-Thr\(^{318}\) | Arg\(^{403}\), Lys\(^{417}\), Tyr\(^{469}\), Tyr\(^{473}\), Glu\(^{484}\), Gly\(^{486}\), Pro\(^{493}\), Gly\(^{498}\), Ser\(^{504}\), and Tyr\(^{506}\) | Asp\(^{300}\), Lys\(^{313}\), His\(^{314}\), Glu\(^{315}\), Asp\(^{316}\), Asn\(^{319}\), Lys\(^{323}\), and Glu\(^{326}\) |
| Bat     | BtRS RatG13-CoV | RBM Thr\(^{581}\)-Thr\(^{591}\) | Glu\(^{506}\), Lys\(^{517}\), Gly\(^{545}\), Leu\(^{571}\), Gly\(^{573}\), Asn\(^{575}\), Asp\(^{579}\), Tyr\(^{580}\), and Asp\(^{591}\) | Asp\(^{300}\), Arg\(^{309}\), Tyr\(^{462}\), Asp\(^{463}\), Trp\(^{476}\), and Asp\(^{480}\) |
| SARS    | SARS-CoV RBM    | Pro\(^{500}\), Pro\(^{504}\) | Cys\(^{473}\), Asp\(^{479}\), Gly\(^{507}\), Asn\(^{508}\), and Tyr\(^{512}\) | Asp\(^{300}\), Lys\(^{313}\), His\(^{314}\), Glu\(^{315}\), Asp\(^{316}\), Asn\(^{319}\), Lys\(^{323}\), and Glu\(^{326}\) |
| Bat     | BtRSBeta-CoV RBM| Ser\(^{511}\)-Thr\(^{581}\) | Tyr\(^{501}\), Gly\(^{544}\), Tyr\(^{571}\), Ala\(^{573}\) and Glu\(^{576}\) | Asp\(^{300}\), Lys\(^{313}\), His\(^{314}\), Glu\(^{315}\), Asp\(^{316}\), Asn\(^{319}\), Lys\(^{323}\), and Glu\(^{326}\) |
| Pangolin| PCoV RBM        | Gln\(^{513}\), Ser\(^{570}\) | Thr\(^{502}\), Lys\(^{503}\), Ser\(^{524}\), Tyr\(^{541}\) and Lys\(^{544}\) | Asp\(^{300}\), Arg\(^{309}\), Tyr\(^{462}\), Asp\(^{463}\), Trp\(^{476}\), and Asp\(^{480}\) |
| MERS    | MCoV RBM        | Gly\(^{372}\), Val\(^{516}\) | Ala\(^{315}\), Tyr\(^{467}\), Asp\(^{463}\), Pro\(^{490}\), Pro\(^{492}\), and Ser\(^{494}\) | Asp\(^{300}\), Lys\(^{313}\), His\(^{314}\), Glu\(^{315}\), Asp\(^{316}\), Asn\(^{319}\), Lys\(^{323}\), and Glu\(^{326}\) |
| Human   | hCoV- RBM       | Ala\(^{315}\), Tyr\(^{475}\) | Ser\(^{382}\), Gln\(^{385}\), Cys\(^{386}\), Tyr\(^{490}\), Val\(^{492}\), and Arg\(^{497}\) | Asp\(^{300}\), Glu\(^{315}\), Thr\(^{317}\), and Lys\(^{324}\) |
| Avian   | ACoV RBM        | Asp\(^{250}\), Gln\(^{489}\) | | Asp\(^{300}\), Glu\(^{315}\), Thr\(^{317}\), and Lys\(^{324}\) |
| Pig     | PEDV1-CoV RBM   | RBM Ala\(^{315}\), Tyr\(^{475}\) | | Asp\(^{300}\), Glu\(^{315}\), Thr\(^{317}\), and Lys\(^{324}\) |
Fig. 3. Protein-Protein interactions of SARS-CoV 2 RBD Pro\textsuperscript{322}-Thr\textsuperscript{581} (A), BtRnRATG13-CoV RBD Thr\textsuperscript{322}-Thr\textsuperscript{581} (B), SARS-CoV RBD Pro\textsuperscript{309}-Pro\textsuperscript{575} (C), BtRsBeta-CoV RBD Ser\textsuperscript{311}-Thr\textsuperscript{568} (D), BtRsCoV-related RBD Arg\textsuperscript{306}-Pro\textsuperscript{575} (E), PCoV RBD Gln\textsuperscript{319}-Ser\textsuperscript{589} (F), hCoV RBD subdomain Ala\textsuperscript{315}-Tyr\textsuperscript{675} (G), MCoV RBD Gly\textsuperscript{372}-Val\textsuperscript{616} (H), ACoV RBD Asp\textsuperscript{250}-Gln\textsuperscript{489} (I), PEDV1-CoV RBD Ala\textsuperscript{315}-Tyr\textsuperscript{675} (J) with its receptor ACE2 predicted using MOE software suite (Molecular Operating Environment (MOE), 2014.01; Chemical Computing Group ULC, 1010 Sherbrooke St. West, Suite #910, Montreal, QC, Canada) H3A 2R7, 2021). Spike protein is represented in maroon ribbons with residues in cyan color while the ACE2 receptor is represented in yellow ribbons with residues in blue colors.
Fig. 4. Protein-Protein interactions of SARS-CoV-2 (A), BtRsRaTG13-CoV (B), SARS-CoV (C), BtRsBeta-CoV (D), BtRsCoV-related (E), PCoV (F), hCoV (G), MCoV (H), ACoV (I), PEDV-CoV (J), with its receptor ACE2 predicted using MOE software suite. Spike protein is represented in maroon ribbons with residues in cyan color while the ACE2 receptor is represented in yellow ribbons with residues in blue colors.

Fig. 4. Continued
hydrogen and ionic interactions with the positively charged Lys58. However, another bulky residue Tyr471 shows hydrogen bond with the negatively charged Glu75. Apart from these interactions, both Arg455 and Arg492 forms hydrogen bonds with Lys31 and Asn61. In addition, both Phe554 and Ala473 also forms hydrogen bonds with Gln30 and Thr78 with ACE2 receptor surface (Fig. 4F).

Conversely, the residues of the hCoV forms hydrogen bonds with N-acetyl-D-glucosamine around the surface of the receptor. The positively charged Arg446 and Arg447 and the backbone of Phe550 shows hydrogen bonds with NAG711 oxygens. In addition, both the Tyr448 and Gly449 also shows hydrogen bonds with NAG710 oxygens (Fig. 4G). Similar to hCoV, MCoV also shows a different mode of binding and allows the residues Pro471, Gly483, Thr487 to form hydrogen and π-interactions at the attachment site with the residues of the receptor ACE2. The phenyl ring of Phe418 and Asn421 at the middle of the bridge shows both π-stack and a hydrogen bond with Gln24. However, His486 at the C terminus of α1 forms a hydrogen bond with Asn103 (Fig. 4H). Moreover, ACoV-ACE2 also shows a different mode of binding and no alignment was seen from Phe490-Pro495 of SARS-CoV-2-RBD with hCoV-RBD. The positively charged Arg370, shows ionic interaction with Asp191 while Ser258 and Thr412 of RBD forms hydrogen bonds with Glu110 and Met82 of ACE2 respectively (Fig. 4I). Likewise, PEDV-1CoV with different mode of binding allows Asn358, Ile551, and Thr558 to form three hydrogen bonds with Ser19, Thr78 and Asp38 of ACE2 respectively (Fig. 4J). These network of hydrogen bonds with different amino acids in different species with the receptor ACE2 is due to amino acid variations of RBM in comparison to SARS CoV 2 RBM (Table 3).

Over all the critical residues at receptor binding motif of spike proteins shows that both positively charged Arg403, Lys17, Lys444 and negatively charged Glu406 play an important role in formation of ionic interactions/salt bridges in attaching to the host receptor ACE2 and is only discussed further. The residue Arg403 of SARS-CoV 2 and Lys308 of SARS-CoV shows an ionic interaction with the nearby residue Asp38. However, Lys391 of BtRsBeta-CoV deviates and forms interaction with His34 instead of Asp38 at the interface of the receptor ACE2. In comparison, ε-amine of Lys390 in BatBtRs-CoV related contacts on the surface of the receptor ACE2. On the other hand, SARS-CoV 2 RBD Pro322-Thr581 also shows contacts with Asp30, Glu35 and Lys31 through interactions and a hydrogen bond with −8.2, −17.9 and −1.2 kcal/mol higher than the energies obtained with full length protein. On the other side, BtRsRaTG13-CoV RBD Thr581-, Thr323 also contacts with Asp30 and Glu15 similar to full length protein with −9.2 and −10.3 kcal/mol. However, no contacts were seen with Lys304 of ACoV and ACoV RBD Asp250-Gln489 due to its positioning far away from the receptor surface similar to full length protein.

In addition, the mutants Arg403Thr in both full-length and subdomains of BtRsRaTG13-CoV and BtRsRaTG13-CoV RBD RBD Thr581-, Thr323, ACoV and ACoV RBD Asp250-Gln489, Arg403Ser, Arg403Pro and Arg405Tyr in MCoV and MCoV RBD Gly372-Val616, hCoV and hCoV-RBD Ala131-Tyr675, PEDV1-CoV and PEDV1-CoV RBD Ala131-Tyr675 resulted in a loss of this ionic interaction due to their smaller side chain and lack of positive charge on the receptor surface ACE2. This confirms that higher stability of Arg403 in SARS-CoV 2 is due to its internal network of hydrogen bonds with Ile402, Gly404 and Asp405 that allows to orient and makes stronger interaction with the receptor ACE2 are not seen with other viral species.

Furthermore, another ionic interaction/salt bridge is formed at the nearby residues between Glu406-His134 of SARS-CoV 2 with the binding energy of −14.4 kcal/mol. The residue also contacts internally with Asp405, Val407, Arg408 and Gln409 through hydrogen bonds predicting to be highly stable in this orientation. Although the residue Glu406 is highly conserved in BtRsRaTG13-CoV and shows internal contacts with Arg403, Arg408 and Gln409, no hydrogen bonds are seen with receptor surface. This predicts that Glu406 to show lesser contribution in receptor binding compared to SARS-CoV 2. Similar type of internal contacts was also seen with mutant Glu406Asp, where two hydrogen bonds are seen in BtRsBeta-CoV and BtRsCoV-related with Arg493 and Gln409 while two hydrogen bonds are seen only with Arg404 through terminal oxygens in SARS-CoV. In extension to these mutant, Glu406Met, Glu406Arg, Glu406Gly and Glu406Gln mutants also shows similar effect in MCoV, hCoV, PEDV1-CoV and ACoV without any internal and external hydrogen bonds predicting to show lesser contribution in binding affinity with the receptor surface.

On the other hand, the subdomains of BtRsBeta-CoV RBD Ser311-Thr568 and SARS-CoV RBD Pro309-Pro575 also maintains the ionic interaction/salt bridge with Lys353 with −5.1 and −4.2 kcal/mol showing no sign of formation in BtRsBeta-CoV-related. Similarly, the Glu406Met, Glu406Arg, Glu406Gly and Glu406Gln in other viral species like MCoV RBD Gly372-Val616, hCoV-RBD Ala131-, Tyr575, PEDV1-CoV RBD Ala131-Tyr675 and ACoV RBD Gly372-Val616 has not shown any sign of ionic interaction/salt bridge with surface receptor showing an order SARS-CoV RBD Pro309-Pro575 > BtRsBeta-CoV RBD Ser311-Thr568 > BtRsRaTG13-CoV RBD Thr581, Thr583 > SARS-CoV 2 RBD Pro322-Thr581, respectively.

In extension to these charge interactions, the residue Lys17 of SARS-CoV 2 which is conserved in BtRsRaTG13-CoV and ACoV result in tighter association because of the two ionic interactions/salt bridges formed between the terminal amino group of Lys17, Asp30 and Glu31 of ACE2 with a total energy of −14.3 kcal/mol. In addition, Asp30 also accepts electrons with the terminal carbons of Lys17 to form additional bonds with −1.6 and −0.5 kcal/mol. In this orientation, the backbone oxygen of Lys417 also forms internal hydrogen bond by accepting electrons with the near by residue Tyr421 for higher stability. In comparison, Lys417 of BtRsRaTG13-CoV donates electrons to both Asp30 and Glu35 to form two hydrogen bonds with lesser binding energies of −4.2 and −9.5 kcal/mol less than the energies formed in SARS-CoV 2. However, the terminal carbons donate electrons to leu455 while backbone oxygen accepts electrons with the same Tyr421 to form two internal hydrogen bonds. On the other hand, SARS-CoV 2 RBD Pro322-Thr581 also contacts with Asp30, Glu35 and Lys31 with −8.2, −17.9 and −1.2 kcal/mol through ionic interactions and a hydrogen bond higher than the energies obtained with full length protein. In addition, BtRsRaTG13-CoV RBD Thr581-, Thr323 also contacts with Asp30 and Glu35 similar to full length protein with −9.2 and −10.3 kcal/mol. However, no contacts were seen with Lys304 of ACoV and ACoV RBD Asp250-Gln489 due to its positioning far away from the receptor surface similar to full length protein. Replacing with hydrophobic Valine (Lys17Val) shows no contacts between the protein and the receptor both in BtRsBeta-CoV and BtRsBeta-CoV RBD Ser311-Thr568, BtRsCoV-related and BtRsCoV-related Arg406-Pro575, SARS and SARS-CoV RBD Pro309-Pro575 due to its hydrophobic environment and inducing structural changes to stay away from the charged residues on the receptor surface. In addition, elimination of the positively charged group and introducing the bulky side chain of phenylalanine at position Lys17 restrict the conformational changes which in turn greatly decreases the affinity of RBD showing no direct contact with the receptor ACE2 both in hCoV and hCoV RBD subdomain Ala131-, Tyr675. However, mutations of Lys417Pro result in the appearance of the fastest phase in folding at RBM of MCoV and disrupt the local structure near the binding interface. This causes the shift at the RBM to bind away from the surface showing internal hydrogen bond between the α-amino group and the nearby residue Gly462, both in MCoV and MCoV RBD Gly372-Val616. Similarly, replacing the uncharged
amino acid Thr423 with Lys417 (Lys417Thr) decreases the net charge near the negatively charged amino acids on the receptor surface showing no sign of contacts with the viral protein both in PEDV1-CoV and PEDV1-CoV RBD Ala315-Tyr675. This indicates that introduction of a positive charge of the e-NH2 of Lys417 may be generating charge attraction with spatially closely to negatively charged residues to form hydrogen bond with Asp35 and Glu35 of ACE2 predicting to be the hot spot for protein-protein interactions. Previous structural studies also show that Lys317 of the RBD in the middle of the “bridge may result in tight hydrogen bonds with Asp35 and Glu35 of ACE2 [41]. In addition, Lys441 shows another salt bridge with Glu468 with −0.5 and −2.0 Kcal/mol indicating lesser contribution in SARS-CoV-2 than in BtRsRaTG13-CoV. However, no salt bridge was seen in aCoV although the positively charged Lys423 (Lys423 in SARS CoV-2) is highly conserved due to its positioning far away from the surface receptor ACE2. In turn, the salt bridge between the carboxylic group of Glu27 and the Lys44 amino is disrupted in the mutants Lys444Thr in BtRsBeta-CoV, BtRsCoV, hCoV, and MCoV, where the amino group of threonine probably cannot make such an ionic interaction with the receptor due to its longer distance with Glu27.

Protein-protein docking studies of SARS-CoV-2 RBD Pro322-Thr581 BtRsRaTG13-CoV RBD Thr581-Thr323, SARS-CoV RBD Pro309-Pro575, BtRsBeta-CoV RBD Ser311-Thr568, BtRsCoV-related Arg306-Pro575 and PCoV RBD Gin319-Ser589 subdomains binds similar to full length spike proteins of SARS-CoV-2, BtRsRaTG13-CoV, SARS-CoV, BtRsBeta-CoV, BtRsCoV-related and PCoV to the N-terminal helices of ACE2 receptor predicting to be the important subdomains that may induce antibodies to cross reactive against SARS-CoVs spike protein attachment. Protein-protein interactions also show that the mutants Lys467Val, Lys417Phe, Lys417Pro, Lys444Thr and Lys444Thr in which the basic group is removed, support the importance of the binding of the carboxyl group of the Asp35, Glu35 and Asp44 for the proper positioning of the RBD on the receptor surface exhibiting a very low affinity with ACE2. The results indicate that these positive and negative charges of all the nine subdomains expect hCoV-RBD subdomain Ala315-Tyr675 are directly involved in the formation of salt bridges in stabilizing ACE2-spike protein interactions which is higher in SARS-CoV-2 compared to other viral species used in this study. This may be the reason why only SARS-CoV-2 and SARS-CoV RBDs were recognized by SARS-CoV RBD-specific, but not MCoV RBD-specific, polyclonal antibodies, whereas only MCoV RBD was recognized by MCoV RBD-immunized polyclonal antibodies, suggesting the cross-reactivity of SARS-CoV-RBD-specific antibodies with SARS-CoV-2 RBD protein [42]. On the other side both full length and hCoV-RBD subdomain Ala315-Tyr675, binds away from the N-terminal helices of ACE2 which is in correlation with previous study demonstrating that hCoV uses certain types of O-acetylated sialic acid residues on glycoproteins to initiate the infection of host cells. Studies also reveal that HKU1 is only one of the six hCoVs identified with an unidentified cellular receptor [43]. The same way, MCoV also shows S-mediated attachment to sialo-
sides and entry into human airway epithelial cells [44]. In addition, studies also shown that coronavirus that belong to group 1 namely, human coronavirus-229E (HCoV-229E), feline infectious peritonitis virus (FIPV), canine coronavirus (CoCV), transmissible gastroenteritis virus (TGEV), and porcine epidemic diarrheaa virus (PEDV), are known to commonly use the aminopeptidase N (APN) of their natural host species as a functional receptor for virus entry [45-49]. This shows that these four viral species hCoV, MCoV, ACoV and PEDV-CoV spike proteins may not prefer ACE2 receptor as a viral host attachment. However, these subdomains hCoV RBD subdomain Ala575-Tyr675, MCoV RBD Gly372-Val616, ACoV RBD Asp250-Gln489 and PEDV-CoV RBD Ala575-Tyr675 may also show cross neutralization against SARS-CoV 2 viral infection since they bind on the surface of N-terminal alpha helices of ACE2 receptor. This may be supported by the previous data showing SARS-CoV RBD and MERS-CoV RBD efficiently induce production of neutralizing antibodies [50,51].

3.4. Sequence analysis of epitopic regions

Further sequence analysis of all epitopic sequences shows that Leu352 of SARS-CoV 2 which is highly conserved in BtRsRatTg13-CoV, BtRsBeta-CoV, BtRsCoV-related and hCoV is replaced with phenylalanine in PEDV1-CoV and isoleucine in both, MCoV and SARS-CoV. The residue Gln564 of SARS-CoV 2 which is also highly conserved in BtRsRatTg13-CoV, BtRsBeta-CoV, BtRsCoV-related and PEDV1-CoV is replaced by glutamic acid, serine, and proline in ACoV, hCoV and SARS-CoV. Another critical residue Phe665 which is highly conserved in BtRsRatTg13-CoV, BtRsBeta-CoV, BtRsCoV-related, ACoV, PEDV1-CoV and SARS-CoV is replaced by tyrosine and leucine in hCoV and MCoV. The residue Thr573 of SARS-CoV 2 which is conserved in BtRsRatTg13-CoV, BtRsBeta-CoV, BtRsCoV-related, hCoV and SARS-CoV is replaced with serine, proline, and glycine in ACoV, PEDV1-CoV and MCoV. Finally, the residue Val576 of SARS-CoV 2 which is conserved in BtRsRatTg13-CoV, BtRsBeta-CoV, BtRsCoV-related, ACoV is replaced with phenylalanine in hCoV, leucine in PEDV1-CoV, glycine in MCoV and alanine in SARS-CoV respectively (Table. 4). These conserved mutations along with variable amino acids may show impact on the cross neutralization of antibodies against SARS-CoV 2 showing unique structural features of the spike glycoprotein RBD of SARS-CoV 2 that confers potentially higher affinity binding for its receptor than found with other CoV viral species. These results show that the epitopic region from Gln319-Ser589 of PCoV RBD Gln319-Ser589 is only five residues with only Pro239 conserved with respect to SARS-CoV 2 RBD Pro232-Thr581 epitope. On the other hand, SARS-CoV RBD Pro232-Thr581 epitope shows high variations with other four epitopes of hCoV RBD subdomain Ala575-Tyr675 from Gly662-Gly708, MCoV RBD Gly372-Val616 from Asp222-Val245, ACoV RBD Asp250-Gln489 from Asp214-Ala425, and PEDV-CoV RBD Ala575-Tyr675 from Gly617-Cys327 used in the study (Fig. 5B). This predicts that the epitopes of these four subdomains along the epitope of PCoV RBD Gln319-Ser589 might be effective in a wide range in inducing antibodies for cross neutralization against SARS-CoV 2 spike protein attachment with its receptor ACE2. Previous data also confirms two-way antigenic cross reactivity between SARS-CoV and porcine group 1 CoVs through group 1 CoV N proteins and not the 5 protein [52]. More importantly, the epitopic sequences of six subdomains from SARS-CoV 2 RBD Pro232-Thr581, BtRsRatTg13-CoV RBD Thr322-Thr583, SARS-CoV RBD Pro239-Pro575, BtRsBeta-CoV RBD Ser311-Thr568 and BtRsCoV-related Arg306-Pro575 may play an important role as antigenic determinants against SARS-CoV 2 viral infection and cross neutralization.

The results show that the epitopic regions from Gly545-Thr581 of SARS-CoV 2 and BtRsRatTg13-CoV RBD Thr322-Thr583 are highly conserved with four variations in Leu560, Ala570, Ala575 and Gln580 in comparison to other two epitopic regions from Gly532-Thr568 in BtRsBeta-CoV RBD Ser311-Thr568 and from Gly531-Thr567 in BtRsCoV-related Arg306-Pro575 respectively. Expect Pro569 that is highly conserved, the epitopic region between Asp454-Pro577 that was predicted in SARS-CoV RBD Pro309-Pro575 show high variations with other five epitopic regions that were predicted in other five subdomains (Fig. 5A). Also, the epitopic region of SARS-CoV RBD Pro309-Pro575 show a total of five prolines at Pro459, Pro462, Pro466, Pro469, Pro470 and Pro477 which might be responsible for increasing the flexibility of SARS-CoV RBM. This allows lesser binding interaction than SARS-CoV 2 with hACE2 and shows distinct epitopic features in cross neutralization studies. This may be the reason why HEK293T cells when transfected with pCAGGS plasmids containing Flag-tagged SARS-CoV S or SARS-CoV 2 S show difference in the electrostatic surface potential maps leading to different immunogenic properties of the RBD subdomains [53]. Previous results also demonstrate that most SARS-CoV RBD-specific antibodies could cross-neutralize SARS like-CoV strain WIV1 from Bat [54]. Studies also identified human mAb S309 with broad neutralizing activity binding to N343-glycan (N330 in SARS-CoV S) epitope in the RBD domain which is in correlation with our SARS-CoV 2 RBD subdomain from Pro322-Thr581 [55]. This shows that the predicted subdomains might induce antibodies that binds to spike epitopes and shows cross neutralization. This is also supported by previous findings showing cross neutralization of antibody binding to the epitopes of SARS-CoV 2 spike protein 10-fold greater that was isolated from hyperimmune horse anti-SARS-CoV serum. Even SARS patient sera or rabbit hyperimmune sera also show cross neutralization on SARS-CoV 2 pseudo virus carrying spike protein in a limited level [10,56-58]. Recent cross-neutralizing data have also indicated that only one out of 15 SARS-CoV 2-infected patients was able to show cross reactive response weekly between SARS-CoV 2 and SARS-CoV viruses [59]. These results based on both computational and experimental clearly indicate that these five-proline shown above play an important role in both binding affinity and immunogenic properties in cross neutralization studies between SARS-CoV and SARS-CoV2.
4. Conclusion

There is a health and medical emergency to control the rapid and global ever-growing SARS-CoV-2 transmission and infection. Since we are at the beginning of understanding the immune responses to the virus and due to lack of knowledge, we may need to use our previous experiences with coronaviruses along with in silico approaches to design vaccines as the ultimate way to protect healthy individuals. In this study, we have comprehensively compared the sequences of spike protein from 10 different coronaviruses in the context of their interaction with ACE2 to identify the best subdomain of spike protein to be used for vaccine development. Although a full-length S protein may be a better candidate to induce immunity, a more focused immune induction based on an immunogenic part of S protein may warrant a stronger and more efficient vaccination outcome, while it significantly reduces the chance of development of antibody-dependent enhancement. In addition, industrial concerns support that the use of a shorter version of target antigen may be easier, faster, and more cost-efficient to be manufactured at the speed and large scale that is urgently required for the present pandemic. Although several vaccines have already been developed based on a full-length spike protein, this study suggests a shorter version of spike protein as a vaccine candidate with the same or even better immunogenicity because of its shorter length. In fact, vaccines that are designed based on shorter peptides have several advantages over longer peptides. First, a focused immune response against an essential component of a virus is much more favorable since it reduces the diversion or extension of the immune response toward less immunodominant segment of a target protein. Second, shorter peptides may reduce the chance of producing non-neutralizing or weakly-neutralizing antibodies, which can potentially facilitate viral entry through cellular FC receptor, even in cells without ACE2. This could result in a serious vaccine side effect, antibody dependent enhancement, which has been reported for respiratory syncytial virus in 1960s [60]. Third, shorter peptide can be easily scaled up and are less costly to manufacture compared to longer peptides. This is a critical industrial concern when large quantities of vaccine doses are required as such in the current SARS-CoV-2 pandemic.

This is the first in silico study that comprehensively compares the RBD subdomain of spike protein from ten closely related coronaviruses and their interaction with ACE2. Our protein-protein docking study identifies a short RBD subdomain of SARS-CoV-2 spike protein from Pro322 to Thr581 as the main binding site, interacting with ACE2. The current results in comparison to previous studies also indicate that SARS-CoV-2 RBD amino acids both in the full-length and subdomain Arg601, Glu606, Lys417, Lys444, Tyr463, Gln474, Gin508, Thr561, Asn568, and Tyr590 from SARS-CoV spike and Gln44, Asp59, Glu53, His54, Tyr61, Asn68 and Lys63 from ACE2 acts as common pharmacophores with stronger hydrogen bonds [61]. This 280aa peptide has very high potential to be used as an efficient vaccine candidate for SARS-CoV-2. Our study demonstrates that both RBD subdomain and full-length spike protein of SARS-CoV-2 binds to ACE2 with a similar but higher affinity in comparison to that of other coronaviruses including BtRsT4G13-CoV, BtRsBeta-CoV, PCoV, MCoV, ACoV, and PEDV1-CoV. This suggests that we might be able to design a universal vaccine that could induce cross-reactive neutralizing antibodies, which are capable of inhibiting entry of several closely related coronaviruses. These antibodies can also be produced ex vivo to be used as therapeutics in coronavirus infection such as COVID19. In addition, such a detailed study empowers us for an efficient and quick design or re-design of vaccine candidates to prevent future pandemic that might be caused by emerging or re-emerging coronaviruses infection. Taken together, this study provides an essential foundation for the design and development of SARS-CoV-2 RBD Pro322-Thr581-based vaccines and therapeutics while it may also be beneficial for infections caused by other coronaviruses.

Author’s contributions

Nataraj Sekhar Pagadala performed the complete study, processed information, interpreted results and written the manuscript. Dr. Amir Landi interpreted the results and written the manuscript. Dr. Paramahamsa Maturu interpreted the results and written the manuscript. Prof. Jack Tuszynski interpreted the results and written the manuscript.

Declaration of Competing Interest

No potential conflict of interest was reported by the authors.

References

[1] J. Cui, F. Li, Z.L. Shi, Origin and evolution of pathogenic coronaviruses, Nat. Rev. Microbiol. 17 (3) (2019) 181–192.
[2] P.C. Woo, Y. Huang, S.K. Lau, K.Y. Yuen, Coronavirus genomics and bioinformatics analysis, Viruses 2 (8) (2010) 1804–1820.
Coronavirus disease (COVID-19) continues to pose a significant global health threat. The novel coronavirus, SARS-CoV-2, has been shown to share several structural similarities with SARS-CoV and MERS-CoV, the causative agents of severe acute respiratory syndrome (SARS) and Middle East respiratory syndrome (MERS), respectively. These similarities include the spike protein, which is the primary receptor-binding protein of SARS-CoV-2, and the ACE2 receptor, which is also a common receptor for all three coronaviruses. The binding of the spike protein to ACE2 is crucial for viral entry into host cells, and antibodies targeting this interaction have shown promise as potential therapeutic agents.

Furthermore, the spike protein of SARS-CoV-2 has been studied extensively. It contains two subunits, S1 and S2, with S1 responsible for binding to ACE2 and S2 for fusion with the host cell membrane. Recent studies have focused on the development of vaccines and antibodies targeting the spike protein to prevent SARS-CoV-2 infection. These efforts have been complemented by research on the ACE2 receptor, which is overexpressed in certain cell types and may be a target for therapeutic intervention.

Overall, the understanding of SARS-CoV-2 and its interaction with the host cell continues to evolve. The development of effective vaccines and therapies requires a comprehensive understanding of the virus and its receptors. Further research is needed to optimize the efficacy and safety of these interventions and to address potential drug resistance and viral mutation.
J.B. Case, R.E. Chen, C. Havenar-Daughton, G. Snell, A. Telenti, H.W. Virgin, A. Lanzavecchia, M.S. Diamond, K. Fink, D. Veesler, D. Corti, Cross-neutralization of SARS-CoV-2 by a human monoclonal SARS-CoV antibody, Nature (2020).

[56] M. Hoffmann, H. Kleine-Weber, S. Schroeder, N. Kruger, T. Herrler, S. Erichsen, T.S. Schiergens, G. Herrler, N.H. Wu, A. Nitsche, M.A. Muller, C. Drosten, S. Pohlmann, SARS-CoV-2 cell entry depends on ACE2 andTMPRSS2 and is blocked by a clinically proven protease inhibitor, Cell 181 (2) (2020) 271–280 e8.

[57] M. Yu, V. Stevens, J.D. Berry, G. Cramer, J. McEachern, C. Tu, Z. Shi, G. Liang, H. Weingartl, J. Cardosa, B.T. Eaton, L.F. Wang, Determination and application of immunodominant regions of SARS coronavirus spike and nucleocapsid proteins recognized by sera from different animal species, J. Immunol. Methods 331 (1–2) (2008) 1–12.

[58] D.E. Anderson, C.W. Tan, W.N. Chia, B.E. Young, M. Linster, J.H. Low, Y.J. Tan, M.I. Chen, G.J.D. Smith, Y.S. Leo, D.C. Lye, L.F. Wang, Lack of cross-neutralization by SARS patient sera towards SARS-CoV-2, Emerg. Microbes Infect. 9 (1) (2020) 909–902.

[59] H. Lv, N.C. Wu, D.T. Tsang, M. Yuan, R. 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.

[60] M.F. Delgado, S. Coiello, A.C. Monsalvo, G.A. Melendi, J.Z. Hernandez, J.F. Bataille, L. Díaz, A. Trento, H.Y. Chang, W. Mitzner, J. Ravetch, J.A. Meler, P.M. Irusta, F.P. Polack, Lack of antibody affinity maturation due to poor Toll-like receptor stimulation leads to enhanced respiratory syncytial virus disease, Nat. Med. 15 (1) (2009) 34–41.

[61] Y. Wan, J. Shang, R. Graham, R.S. Baric, F. Li, Receptor recognition by the novel coronavirus from Wuhan: an analysis based on decade-long structural studies of SARS coronavirus, J. Virol. 94 (7) (2020) e00127-20.