Journal of High School Science

COVID-19 Variants: A Structural and Functional Analysis of Spike Proteins

Patel H.P[1], Zheng M[1], Agarwal R[2*], Makar N[1], Donnelly J[1], Armani M[1]

Abstract

The SARS-CoV-2 virus has led to a human pandemic of COVID-19. An infected person shows symptoms from mild to severe. Recently, more infectious and potentially more lethal COVID-19 variants have been identified. Four variants have been identified as "Variants of Concern": namely B.1.1.7 or Alpha variant (UK variant), B.1.351 or Beta variant (S. Africa variant), B.1.617.2 or Delta variant (India variant), and P.1 or Gamma variant (Brazil variant). The virus entry into the host cell is initiated by the interaction of its spike protein (S-protein) with the human ACE-2 receptors which plays a key role in the cell membrane fusion process and receptor recognition. In this study, a structure-function relationship of various mutations in S-proteins from the "Variants of Concern" was reported in order to establish a correlation with transmission and infectivity. Nearly 39 mutations were identified in the S-protein, of which, mutations: K417N (Beta variant), K417T (Gamma Variant), L452R and T478K (Delta variant), E484K (Beta & Gamma), and N501Y (Alpha, Beta & Gamma) were found in its Receptor Binding Domain (RBD). The residues at this location are critical because they are involved in significant binding interactions with the ACE-2 receptor. The structural and functional relationship suggests that the mutants L452R and T478K along with its native K417 on the RBD region of the Delta variant are involved in increasing the magnitude of electrostatic interactions with the ACE-2 receptor residues which in turn, may lead to its higher transmission rate and infectivity. These results may provide a better understanding of the mechanism of this host-pathogen interaction and enable better designed therapeutics for SARS-CoV-2.

Keywords

Structural analysis; Functional analysis; Sequence alignment; COVID-19, Spike protein; ACE-2 receptor; Variants of concern.

[2*]Corresponding Author: Dr. Rakhi Agarwal, Assistant Professor, Chemistry Department, Nassau Community College, Garden City, New York, Email: rakhi.agarwal@ncc.edu
[1]Jose Marti STEM Academy, Union City, New Jersey.

Introduction

SARS-CoV-2 also known as Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) is a single-stranded mRNA-enveloped virus that is a member of the coronaviridae family (SARS-CoV, MERS-CoV) (1). Each coronavirus contains nearly 30,000 nucleotides of RNA (2). The virus uses its genetic information and host cell machinery to replicate and make new virus copies; however, as the infected cell builds new virus copies, it periodically produces tiny copying errors called genetic mutations leading to a new strain. COVID-19 is caused by a strain of coronaviruses known as SARS-CoV-2 (2).

Throughout the pandemic, several variants of SARS-CoV-2 have emerged based on geographical locations that create further spread and threat to the human population. Four variants have been identified as "Variants of Concern" (VOCs): namely B.1.1.7 or Alpha variant (UK variant), B.1.351 or Beta variant (S. Africa variant), B.1.617.2 or Delta variant (India variant), and P.1 or Gamma variant (Brazil variant) (3,4). These VOCs contain specific mutations related to changing amino acid sequences on S-protein that interact with the human Angiotensin Converting Enzyme -2 (ACE-2) receptor to initiate the infection. For example, all the VOCs share one of the common D614G mutations. It is an alteration in a single DNA base pair that causes the substitution of aspartic acid (single-letter code: D) with glycine (single-letter code: G) in the protein. It is a dominant mutation of the coronavirus. (2,5,6). It is believed that the virus entry into the host cell is initiated by the interaction of its S-protein to the human ACE-2 receptors. Recently, the newly designed and FDA-approved mRNA vaccines from pharmaceutical companies including Pfizer-BioNTech and Moderna have specifically chosen the mRNA of the S-protein as a successful vaccine candidate (7).

Spike-protein and its Function: The genome of the SARS-CoV-2 virus encodes for 4 major structural proteins called the spike (S), envelope (E), membrane (M), and nucleocapsid (NSP) proteins (8). The Spike protein is a trimeric protein (Figure 1) and is crucial since it is responsible for host receptor binding and cell fusion, leading to infection and immune response in the host. The VOCs have shown significant mutations in their S-Protein leading to increased transmissibility and severity by impacting the immunity (3,4). The S-protein is 180-200 kDa in size and made up of a total of 1273 amino acids (1). It is made of two major subunits: S1 and S2 (9) (Figure 2). The S1 subunit is responsible for initiating host cell entry by receptor recognition and viral attachment. It is composed of 672 amino acids (14–685) (9) including two main subdomains: the N-terminal domain (NTD) and the C-

terminal domain (CTD). The RBD is the key element of the CTD in SARS-CoV-2 which consists of two motifs including an extended loop of the receptor-binding motif (RBM) (9). The function of the receptor-binding motif (RBM) of RBD is to bind and interact with ACE-2 receptors (9). They are present in the S1B region of the structure (Figure 2). The NTD is mainly responsible for determining the host range (8).

The S2 subunit is composed of 588 amino acids (686–1273) (10). The function of the S2 subunit is to mediate host-virus membrane fusion for host cell entry. The S2 subunit includes the N-terminal hydrophobic fusion peptide (FP) (943–982), two heptad-repeat regions (HR1 and HR2) (984–1104/1246–1295), a transmembrane domain (TM) (1296–1317), and a cytoplasmic tail (CT) (1318–1353) (9). The FP (Figure 2, highlighted in blue) of S2 becomes exposed upon cleavage at S2' site, facilitating its insertion to the host membrane for the viral fusion (Figure 2, highlighted in red). Mutations that occur in this region have been shown to block cell fusion for many viruses (9). The S2' site (Figure 2, highlighted in red), which is highly conserved in all sequenced SARS-CoV-2 and is cleaved by host proteases Transmembrane Protease Serine 2 (TMPRSS2), cathepsin L and furin, hence, facilitating the viral activation and representing SARS-CoV-2 pathogenicity (10). TMPRSS2 is also essential for SARS-CoV-2 to enter human lung cells because their cell lines are also susceptible to SARS-CoV-2. A notable feature of the S-protein in SARS-CoV-2 is the presence of a furin-like cleavage site (FCS) located at the boundary between the S1 and S2 subunits (Figure 2, highlighted in cyan). It increases the rate of its proteolytic processing by host-furin protease during SARS-CoV-2 infection. This enhances its infectivity and makes it more contagious (10). On the other hand, the HR1 and HR2 of S2, also known as the "fusion core region," mediates membrane fusion between the virus and a target cell by forming a 6-helix bundle (6-HB) (9). The transmembrane domain (TM) is essential for S-protein trimerization and membrane fusion. It is proposed that the TM interacts with the FP to facilitate the formation of the fusion pore (9). Lastly, the CD tail of the S2 subunit located at the C-terminal of the S-protein is involved in viral assembly, intracellular transport, and cell-cell fusion (9).

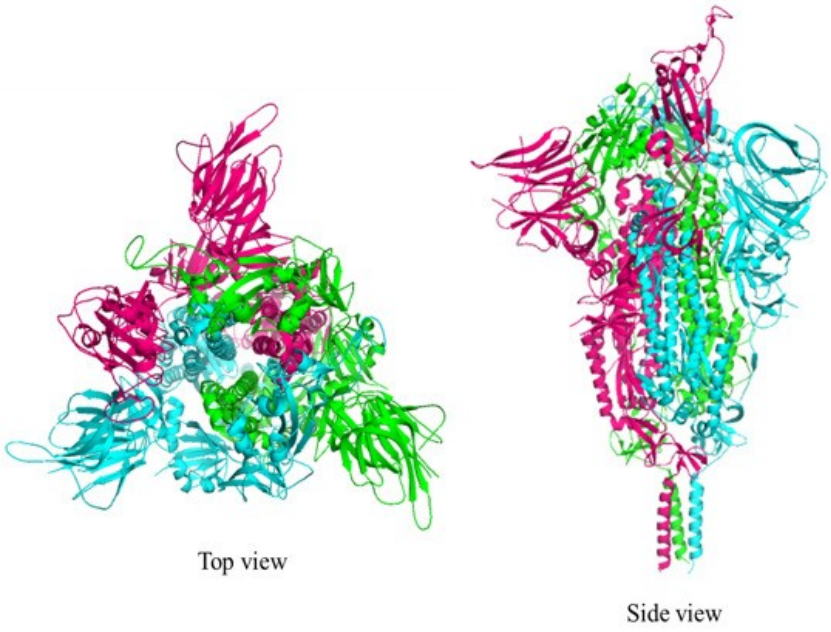


Figure 1: The structure of trimeric S-protein (top and side view). All the three chains have been colored pink, cyan, and green. (PDB ID: 7N1Q). The figure was made using the PYMOL program.

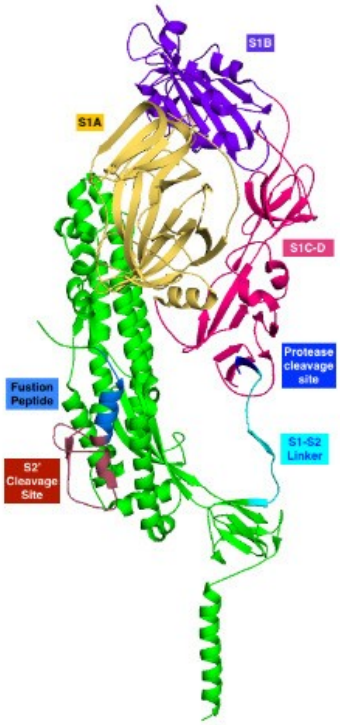


Figure 2: The ribbon structure of a single C-chain of S-protein (PDB ID 7N1Q). The S1 and S2 regions are shown in different colors. Other regions of the spike protein are shown, for example, protease cleavage site, S1-S2 linker, S2' cleavage, and fusion peptide (FP). The figure was made using the PYMOL program.

ACE-2 Receptor and its Function: ACE-2 receptor is an enzyme that fragments angiotensinogen (a larger protein) into small proteins and regulates cell functions (11). The normal function of the ACE-2 receptor is to control the increase in blood pressure, promote tissue regeneration and reduce inflammation through the renin-angiotensin-aldosterone system (RAAS) pathway by suppressing the activity of angiotensin II protein (ANG-II) (11). However, when the S-protein of the SARS-CoV-2 virus binds to the ACE-2 receptor, it fails to perform its normal function, as a result, an increase in ANG-II leads to a disruption in these pathways, including tissue injury in COVID-19 patients. The ACE-2 receptors are present in all individuals with varied counts among individuals, different tissues, and cells (11). Evidence also suggest that ACE-2 count may be higher in patients with conditions like hypertension, diabetes, and coronary heart disease (11).

In summary, the S-protein of the SARS-CoV-2 virus is critical because it interacts with the human ACE-2 receptor, thereby initiating the infection. Hence, the mutations at the residues located at interface S-protein interacting with the ACE-2 receptor and other crucial locations probably decide and differentiate the transmission and infectivity of a VOC. Therefore, determining specific S-protein mutations from the VOCs helps to determine its relatively higher transmission and infectivity. To achieve this objective, we analyzed the amino acid sequences by sequence alignment and successfully identified specific mutations among VOCs. The structure modeling of the mutations at the interface of S-protein and ACE-2 protein complex along with electrostatic map calculations deciphered the critical residues and mutations in VOCs especially those in the Delta variant. This helped us to establish a structure and function correlation in the S-protein given its function to enable higher transmission and infectivity rates in the Delta variant.

Methodology

1. Mutational Analysis of S-protein in Variants of Concern (VOCs):

To establish the structure and function relationship between S-protein and the VOCs (Alpha, Beta, Gamma, and Delta variants), the S-protein mutations were identified concerning common, shared, and exclusive mutations in each variant. (Table 1).

A. Clustal Omega Multiple Alignment: First, the FASTA sequence that represents the amino acid sequences and the Protein Data Bank ID (PDB ID) for each VOC was recorded from the NCBI website. Then, the amino acid sequences of all were compared to each other in the Clustal Omega Multiple Alignment program. The multiple amino acid sequence alignment of

each VOC showed major sequence differences among them and identified additional mutations that were unique to each variant (Figure 3). Finally, the common, shared, and exclusive S-protein mutations from VOCs were tabulated in Table 1 along with the amino acid Sequence IDs of each VOC.

2. Structural Modelling of Mutant Residues in S-protein and ACE-2 complex structures: To identify the role of each mutation, we performed modeling of the mutant residues in the S-protein and in the ACE-2 receptor-S-protein complex to establish structure and function correlation.

A. Structure Modeling of S-Protein and S-Protein-ACE-2 complex: After identifying the mutations on the S-protein for all VOCs, the mutant residues were modeled on the crystal structure of the S-protein using the PYMOL program using PDB ID: 7N1Q as a template (Figure 4a & Table 1, The color coding of mutations is consistent in both places). The mutations were then modeled in the RBD region of S-protein in complex with the human ACE-2 receptor using PDB ID: 6LZG as a template (Figure 4b). A conformational change in RBD of the S-protein was identified using the superimposition of two chains of trimer S-protein using PDB ID: 7N1Q as a template (Figure 4c.)

B. Electrostatic Potential Map Calculations of S-protein of Delta variant in complex with ACE-2 receptor: The mutant residues of Delta variant were modeled to show interactions with ACE-2 residues using PDB ID: 6LZG as a template by PYMOL program (Table II, Figure 5a). The other residues that are conserved between Alpha and Delta were modeled as well (Figure 5a-b, Table 2). An electrostatic potential map at the interfaces was calculated to compare unmutated and mutated sites to show the differences in charges. (Figure 5b-d).

Results and Discussion:

Typing S-protein Mutations in Variants of Concern (VOCs): Clustal Omega Multiple Alignment of VOCs S-protein sequences identified more than 39 mutations in total with at least 7-12 of them unique to each of the VOC (Figure 3). All mutations were further categorized as common, shared, and exclusive mutations in each variant to correlate them with their functional differences (Table 1). Furthermore, the Clustal Omega Multiple Alignment also revealed unique properties of different mutations based on the forms of conservative, semi-conservative, non-conservative, or deletions to compare and establish functional relevance to infectivity and transmission in VOCs (Table 1, Figure 4a). These mutations were located to different sites of the S-protein that were shown to be playing various roles in causing infections (Figure 4a). Most importantly, this analysis suggested that mutations K417N (Beta variant), K417T

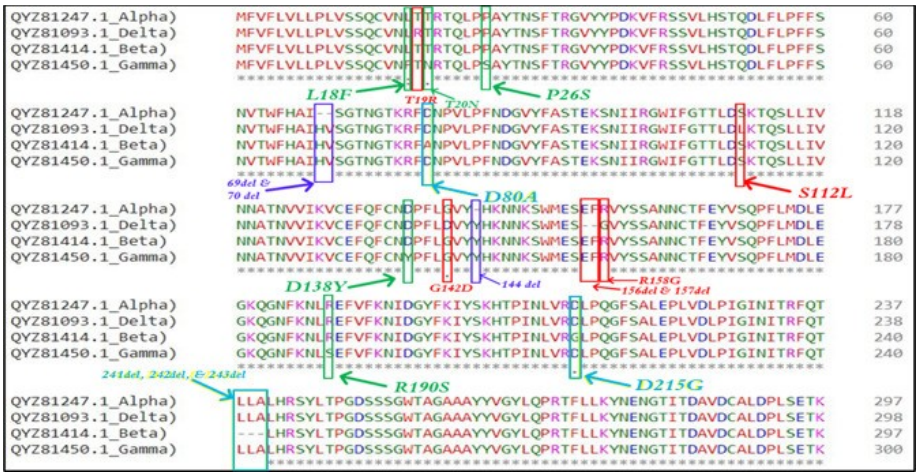
(Gamma variant), L452R (Delta variant), T478K (Delta variant), E484K (Beta & Gamma variants), N501Y (Alpha, Beta & Gamma variants) are the key non-conservative mutations located on the RBD of S1 subunit of the S-protein that interacts with host ACE-2 receptor (Figure 4 a & 4b). These significant mutations probably relate to the differences in infectivity and transmission among them. A list of mutations on each variant of SARS-CoV2 and the protein models showing the mutation sites are available for reference (3,4,12).

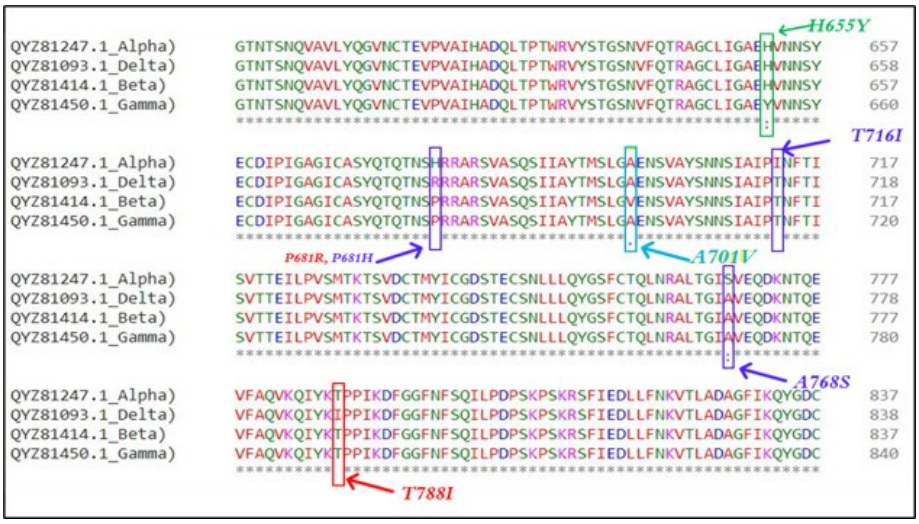
Table 1: COVID-19 variants of concern (VOCs) with a list of common, shared, and exclusive mutations.

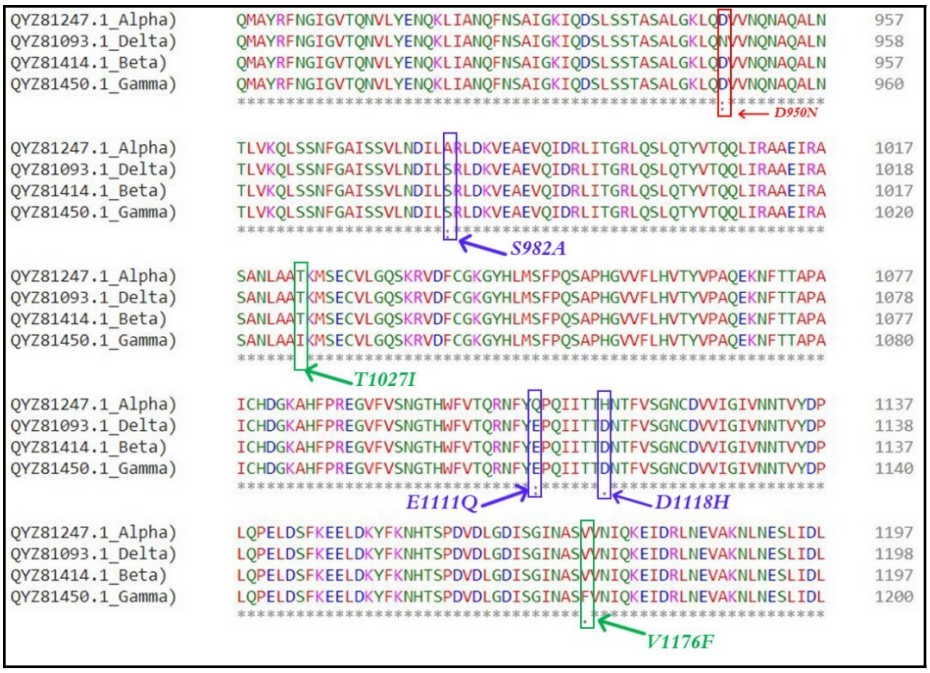
Exclusive mutations of Alpha, Beta, Gamma, and Delta variants are shown in purple, cyan, green, & red respectively. Shared and common mutations are shown in hot pink and light purple color respectively. The protein and nucleotide IDs for all VOCs are tabulated in last two rows.

Mutations	Alpha	Beta	Gamma	Delta
Common	D614G			
Shared	N/A	E484K		N/A
	N501Y			N/A
Exclusive	69del, 70del, 144del, (S494P*), A570D, P681H, T716I, S982A, D1118H (K1191N*), A768S, E1111Q	D80A, D215G, 241del, 242del, 243del, K417N, A701V	L18F, T20N, P26S, D138Y, R190S, K417T, H655Y, T1027I, V1176F	T19R, (G142D), 156del, 157del, R158G, L452R, T478K, P681R, D950N, S112L, T788I
Protein ID	QYZ81247	QYZ81414	QYZ81450	QYZ81093
Nucleotide ID	MZ815065	MZ815080	MZ815083	MZ815051

Figure 3: The Clustal Omega Multiple Alignment of variants of concern (Alpha, Beta, Gamma and Delta). The color coding of variant and mutations are consistent as in Table 1. [The symbols: (*) indicate conservative, (:): semi-conservative, (.): conservation between amino acid groups of weakly similar properties, and (): non-conservative mutations. Therefore, the hierarchy of conservation using these symbols is * (identical) >: (colon) >. (period).] (Note: Amino Acid sequences from 301-360, 840-900, and 1201-1273 are not shown in the images below as no significant change was detected among the variants of concern.)







S-protein Receptor Binding Domain and ACE-2 receptor Interface in VOCs: The S-protein adopts the pre-fusion and fusion conformations (10). In the pre-fusion conformation, the receptor-binding domains (RBDs) exist in either “closed” or “open” forms (70–80-degree rotation) (Figure 4c). In the “closed” form, all the three RBDs are in the downward position hiding the receptor-binding region. Due to the flexibility, the RBDs undergo hinge-like movements, to partially “open” the receptor-binding region, and then its receptor binding motif (RBM) of RBD interacts with the human ACE-2 receptor to initiate the attachment and infection (Figure 4b & c). This dynamic position is stabilized by additional interactions between interfaces. Ultimately, the proteolytic cleavage transforms the S-protein from pre-fusion to fusion conformation. These different conformations of RDB produce distinct antibodies which make it difficult for an individual's immune system to combat pathogens (10). Several mutations like T478K, E484K, S494P, N501Y, L452R, and K417N were found in the RBD of the S-protein in VOCs (Figure 4a, 4b & 5a). The residues at positions; 417, 452, 478, 484, 501 are crucial since they are involved in making interactions to the ACE-2 receptor (Figure 4a, & 4b). These RBD residues can affect the charge and shape of the protein near the S-protein-receptor electrostatic interaction interface.

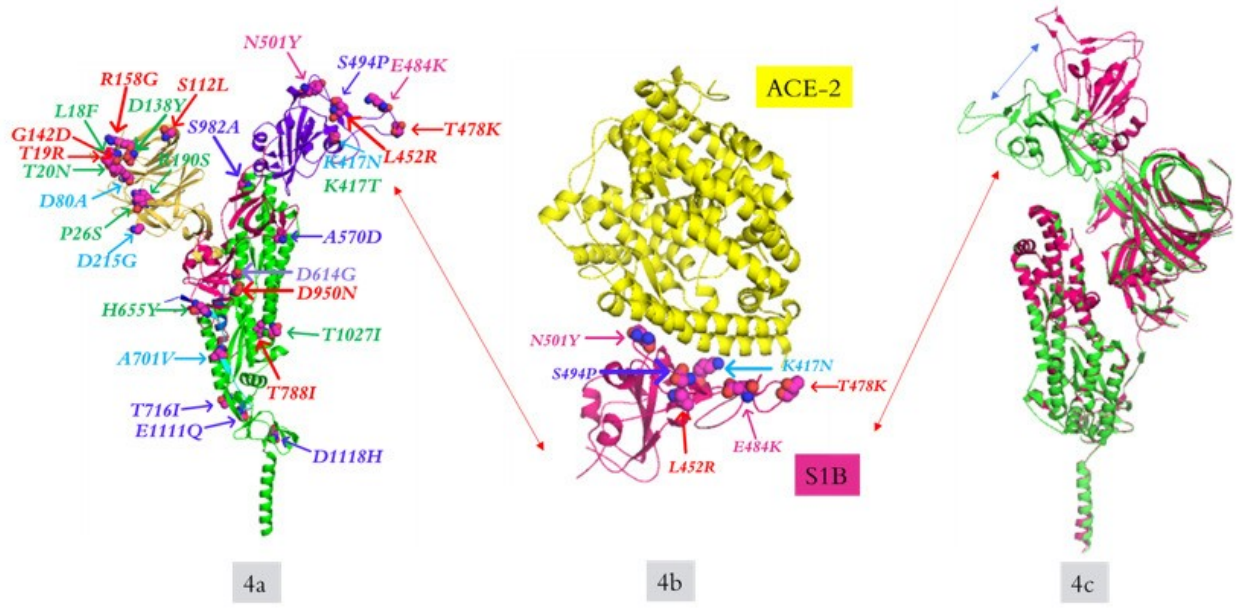


Figure 4a: The ribbon structure of a single chain (Chain C) of S- protein using PDB ID 7N1Q. The S1 and S2 regions are shown in different colors. The common, shared, and exclusive mutations are shown with the color coding consistent as in Table 1.

Figure 4b: The ribbon structure of ACE-2 (Yellow color) is complexed with S1B domain of S-protein (hot pink color) along with its S-protein mutations (PDB ID 6LZG). The color coding is consistent with Table 1 and Figure 4a.

Figure 4c: Superimposition of (S) protein chain A (hot pink color) on chain C (green color) showing the conformational change in S1B domain mimicking the binding to ACE-2 protein as in figure 2 (PDB ID 7N1Q). All figures (4a-c) were generated using the PYMOL program.

Interactions of Delta variant S-protein Residues with ACE-2 Receptor: The modeled mutant residues of Delta variant L452R showed significant electrostatic interactions with E35 and D38, and T478K showed interactions with Q24 and Y83 residues of ACE-2 receptor (Table 2, Figure 5a-d). The other two residues E484 and K417 that are conserved between Alpha, and Delta variants also had significant electrostatic interactions. The E484 conserved with wild type (WT) had electrostatic salt bridge interactions with K31 in Delta variant RBD. However, K384K change was shown to favor the stable complex formation (13). This residue was mutated to E484K in Beta and Gamma variants and was also identified in an Eta variant of interest (VOI) (14). On the other hand, K417 that had salt bridge interactions with D30 of the ACE-2 receptor protein (Figure 5a-d, Table 2) in the Delta variant has shown to be conducive to a stable protein-protein complex (13,15). The residue K417 was mutated to K417N in Beta and K417T in the Gamma variant. The

mutations K417N/T are found to be counterproductive in making stable RBD-ACE-2 complexes (13). The N501 residue that is conserved only in Delta was found to be making H-bond interactions with K353 and Y41 residues of ACE-2. The mutation N501Y that was detected in the other three VOCs (Alpha, Beta, and Gamma) was shown to increase the hydrophobic interactions and higher binding affinity to the ACE-2 receptor (13, 15,16). The RBD and ACE-2 interactions at the interface are most significant not only in viral attachment, transmission, and infectivity but also in how drugs/antibodies interactions with the S-protein (13, 15-18). Higher transmissibility and infectivity behavior of the Delta variant agrees with the lowest value of binding free energy calculations of its S-protein-ACE-2 complex using molecular mechanics Poisson Boltzmann surface area (MMPBSA) suggesting stronger binding of its RBD to ACE-2 in comparison to other VOCs (15). Interestingly, the mutation L452R in the Delta variant has been shown to reduce the binding affinity of many antibodies to the S-protein (18). Other mutations, N501Y, E484K, K417N, and L452R also found to influence the binding affinity of the antibodies at variable degrees (16).

Table 2: Interactions of major Spike protein residues of RBD with ACE-2 residues in Delta Variant.

RBD of S-protein Residues	WT	RBD of Delta Variant	ACE-2 protein Residues	Relevance changes in other VOCs
417	K	K	D30	K: in Alpha and Delta Variant N: in Beta and, T: in Gamma
452	L	R	E35, D38	New L452R mutation in Delta only
478	T	K	Y83, Q24	New T478K mutation in Delta only
484	E	E	K31	E: in Alpha and Delta variant and K: in Beta and Gamma
501	N	N	K353, Y41	Y: in Alpha, Beta, and Gamma

These results highlight the significance of stronger electrostatic interaction of the Delta variant via its signature key residues R452, K478, and K417 with the ACE-2 protein potentially leading to its higher transmission and infectivity relative to other VOCs (Figure 5a-d). These sites can be important targeted sites for variant-specific drug design because they are crucial in the interaction between the virus and the host cell. Also, it was reported that mutations on S-protein contribute to different antigenicity that affects antibody interactions (12-19). Therefore, it will be important in the future to consider further analyzing the changes in S-protein due to mutations, especially near the protein-receptor binding regions for the development of better therapeutics (antibodies, vaccines, and drugs).

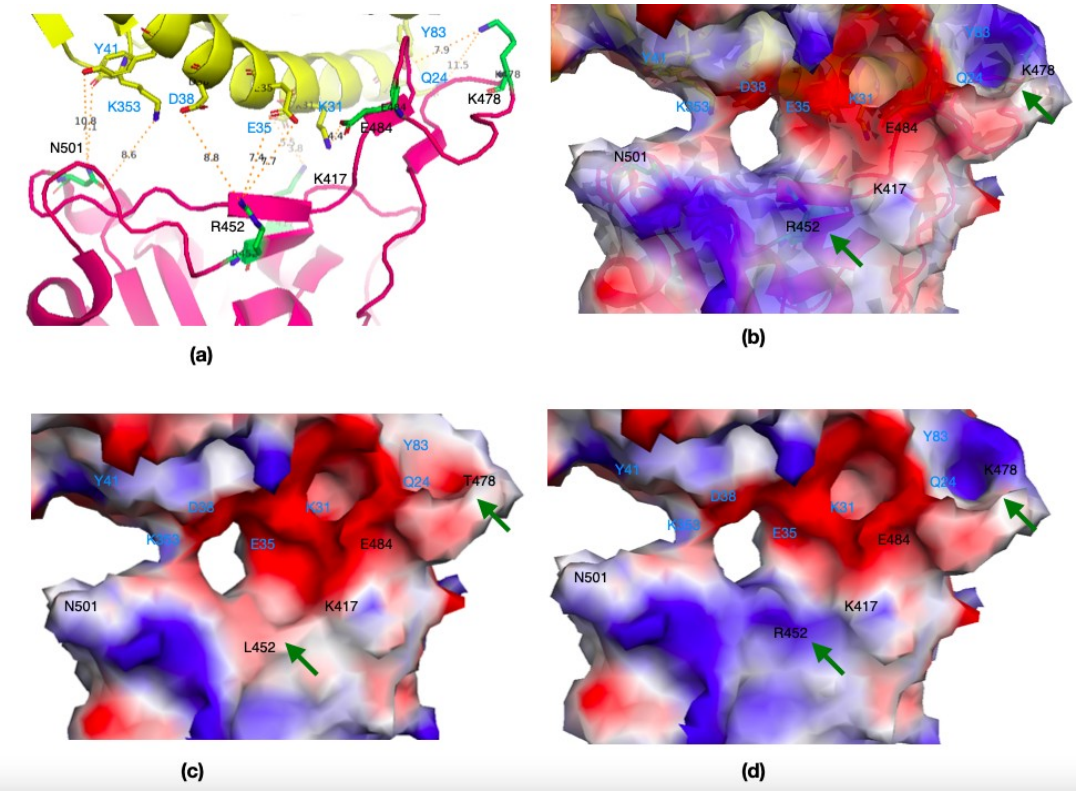


Figure 5: The interface of RDB (S-protein) and host ACE-2 receptor.

a. The RBD of S- protein (cartoon model in hotpink color) with residues K417, mutant L452R, mutant T478K, E484, and N501 of Delta variant are shown in green stick model (residues labelled in black) that are in interaction with ACE-2 (cartoon model in yellow color) residues (labeled in blue). Detailed list of interactions is listed on Table 2.
b. The superposition of figure (a) on its electrostatic potential map to correlate both to each other to aid in visual comparison.
c. The electrostatic potential map of S-protein-ACE-2 interface with wild type (WT) L452 and T478 residues of S-protein. (The residues are indicated with green arrows)
d. The electrostatic potential map of S-protein-ACE-2 interface with mutant L452R and T478K residues of S-protein in Delta Variant. (The residues are indicated with green arrows)

The positively and negatively charged surfaces are shown in blue and red respectively in the electrostatic potential map. All figures and models were generated using the PYMOL program with PDB ID 6LZG.

By accumulating information on SARS-CoV-2 variants and their S-protein mutations, appropriate vaccines can be developed targeting specific S-protein nucleotides in the vaccine. The above S-protein mutations [K417N (Beta variant), K417T (Gamma Variant), L452R and T478K (Delta variant), E484K (Beta & Gamma), and N501Y (Alpha, Beta & Gamma)] probably also alter the shape of the S-protein. As a result, this change may help the viral spikes to evade some types of coronavirus antibodies. The structure and function correlation of this study will help to better understand the mechanisms of transmission and infection so that strategies and variant-specific therapeutics may be developed to control the pandemic. In the future, vaccines can also be developed using mutant S-protein rather than native S-protein sequences. At the same time, it will also be critical to identify what causes new mutations to accumulate. For example, the temperature, environment, ethnicity, blood group type, population, and the replication rate are a few factors which can influence the emergence of new mutations. As a result, more studies need to be instituted to establish these correlations that may identify the next variant. This information may not only be used to stop the virus from mutating but can also be used to deliberately cause mutations that are less transmissible and less lethal to become dominant variants. This study underscores the importance of vaccinating the population; the more people that are vaccinated, the fewer the probability of the virus replicating and transmitting.

Conclusion

The detrimental mutations of S-protein are distinctive in different geographic regions around the globe which have undergone several waves of SARS-CoV-2 infections. Persistent infections cause the accumulation of adaptive mutations. The S-protein is the major determinant of the tropism of coronaviruses, and the modification of the S-protein can alter cell and tissue tropism and, in some cases, may lead to a change in virus pathogenicity. Hence, examining and comparing the S-protein mutations based on the amino acid sequences of the VOCs determined which variants caused the highest infectivity and transmission in humans. The multiple sequence alignment, performed in the Clustal Omega Multiple Alignment, identified unique S-protein mutations in Alpha, Beta, Gamma, and Delta variants of SARS-CoV-2. It also identified common and exclusive mutations in each variant. Through structure modeling, T478K, E484K, N501Y, L452R, and K417N mutations in RBM of RBD were found to be involved in interacting with the ACE-2 receptor. Additionally, conformational changes in the S-protein were identified that enabled it to bind with the ACE-2 receptor. Specifically, the

Delta variant mutations L452R and T478K combined with its native K417 residue are significant because they increase the positive charge surface on its S1B region of the RBD to increase the magnitude of electrostatic interaction with the negatively charged surface of ACE-2, which enhances the binding avidity of the virus to the ACE-2 receptor. Such mutations provide the basis of differences in transmission and infectivity among the various VOC.

References

1. Huang, Y., Yang, C., Xu, X., Xu, W., & Liu, S. (2020). Structural and functional properties of SARS-CoV-2 spike protein: potential antivirus drug development for COVID-19. *Acta Pharmacologica Sinica*, *41*(9), 1141–1149. https://doi.org/10.1038/s41401-020-0485-4

2. *D614G Mutation of COVID-19: What Does it Mean?* (2020, December 1). News Medical.net. https://www.news-medical.net/whitepaper/20201201/D614G-Mutation-of-COVID-19-What-Does-it-Mean.aspx

3. CDC. (2020, March 28). *COVID Data Tracker*. Centers for Disease Control and Prevention. https://covid.cdc.gov/covid-data-tracker/#variant-proportions

4. CDC. (2020, February 11). *Coronavirus Disease 2019 (COVID-19)*. Centers for Disease Control and Prevention. https://www.cdc.gov/coronavirus/2019-ncov/variants/variant.html

5. D614G Mutation in SARS-CoV-2 Spike Protein. (2020, July 30). News-Medical.net. https://www.news-medical.net/health/D614G-Mutation.aspx#:~:text=D614G%20mutation%20is%20a%20missense%20mutation%20wherein%20anhttps://www.news-medical.net/health/D614G-Mutation.aspx#:~:text=D614G%20mutation%20is%20a%20missense%20mutation%20wherein%20an

6. Zhang, L., Jackson, C. B., Mou, H., Ojha, A., Peng, H., Quinlan, B. D., Rangarajan, E. S., Pan, A., Vanderheiden, A., Suthar, M. S., Li, W., Izard, T., Rader, C., Farzan, M., & Choe, H. (2020). SARS-CoV-2 spike-protein D614G mutation increases virion spike density and infectivity. *Nature Communications*, *11*(1). https://doi.org/10.1038/s41467-020-19808-4

7. *Comparing the Pfizer, Moderna, and Johnson & Johnson COVID-19 Vaccines*. (n.d.). GoodRx. https://www.goodrx.com/health-topic/vaccines/comparing-covid-19-vaccines

8. Sironi, M., Hasnain, S. E., Rosenthal, B., Phan, T., Luciani, F., Shaw, M.-A., Sallum, M. A., Mirhashemi, M. E., Morand, S., & González-Candelas, F. (2020). SARS-CoV-2 and COVID-19: A genetic, epidemiological, and evolutionary perspective. *Infection, Genetics and Evolution*, *84*, 104384. https://doi.org/10.1016/j.meegid.2020.104384

9. Zhu, C., He, G., Yin, Q., Zeng, L., Ye, X., Shi, Y., & Xu, W. (2021). Molecular biology of the SARs-CoV-2 spike protein: A review of current knowledge. *Journal of Medical Virology*, *93*(10), 5729–5741. https://doi.org/10.1002/jmv.27132

10. Almehdi, A. M., Khoder, G., Alchakee, A. S., Alsayyid, A. T., Sarg, N. H., & Soliman, S. S. M. (2021). SARS-CoV-2 spike protein: pathogenesis, vaccines, and potential therapies. *Infection*, 1–22. https://doi.org/10.1007/s15010-021-01677-8.

11. Sriram, K., Insel, P., & Loomba, R. (2020, May 14). What is the ACE2 receptor, how is it connected to coronavirus and why might it be key to treating COVID-19? The experts explain. The Conversation. https://theconversation.com/what-is-the-ace2-receptor-how-is-it-connected-to-coronavirus-and-why-might-it-be-key-to-treating-covid-19-the-experts-explain-136928

12. GISAID: https://www.gisaid.org/collaborations/enabled-by-hcov-19-data-from-gisaid/

13. Shahhosseini, N.; Babuadze, G.; Wong, G.; Kobinger, G.P. Mutation Signatures and In Silico Docking of Novel SARS-CoV-2 Variants of Concern. *Microorganisms* **2021**, *9*, 926. https://doi.org/10.3390/microorganisms9050926

14. La Rosa, G., Brandtner, D., Mancini, P., Veneri, C., Bonanno Ferraro, G., Bonadonna, L., Lucentini, L., & Suffredini, E. (2021). Key SARS-CoV-2 Mutations of Alpha, Gamma, and Eta Variants Detected in Urban Wastewaters in Italy by Long-Read Amplicon Sequencing Based on Nanopore Technology. *Water*, *13*(18), 2503. https://doi.org/10.3390/w13182503

15. Celik, I.; Yadav, R.; Duzgun, Z.; Albogami, S.; El-Shehawi, A.M.; Fatimawali; Idroes, R.; Tallei, T.E.; Emran, T.B. Interactions of the Receptor Binding Domain of SARS-CoV-2 Variants with hACE2: Insights from Molecular Docking Analysis and Molecular Dynamic Simulation. *Biology* **2021**, *10*, 880. https://doi.org/10.3390/biology10090880

16. Li, J.; Hou, C.; Wang, M.; Liao, C.; Guo, S.; Shi, L., Ma, X.; Zhang, H.; Jiang, S.; Zheng, H.; Jiang, B.; Ye, L.; Yang, L; He, X. Hydrophobic interaction determines docking affinity of SARS CoV 2 variants with antibodies (2021). https://arxiv.org/abs/2103.00399**.**

17. Sridhar, S., Saha, G., Lata, S. *et al.* Molecular docking studies of Indian variants of pathophysiological proteins of SARS-CoV-2 with selected drug candidates. *J Genet* **100,** 64 (2021). https://doi.org/10.1007/s12041-021-01313-2

18. Yang, L.; Li, J.; Guo, S.; Hou, C.; Liao, C.; Shi, L.; Ma, X.; Jiang, S.; Zheng, B.; Fang, Y.; Ye, L.; He, X. SARS-CoV-2 Variants, RBD Mutations, Binding Affinity, and Antibody Escape. *Int. J. Mol. Sci.* **2021**, *22*, 12114. https://doi.org/10.3390/ijms222212114.

19. Harvey, W. T., Carabelli, A. M., Jackson, B., Gupta, R. K., Thomson, E. C., Harrison, E. M., Ludden, C., Reeve, R., Rambaut, A., Peacock, S. J., & Robertson, D. L. (2021). SARS-CoV-2 variants, spike mutations and immune escape. *Nature Reviews Microbiology*, *19*(7), 409–424. https://doi.org/10.1038/s41579-021-00573-0.