Title:

Spike protein of SARS-CoV-2: Impact of single amino acid mutation and effect of drug binding to the variant-*in silico* analysis.

Author: Pratibha Manickavasagam, Indian Institute of Technology Hyderabad.

Email: m.pratibha1996@gmail.com

Abstract:

Novel SARS-CoV-2, a bat based virus originated in Wuhan, China that caused a global pandemic in December, 2019 belongs to the Betacorona virus family and contains single stranded genome of ~29Kbp. The host cell invasion of SARS-CoV-2 is facilitated by interaction of C-Terminal Domain (CTD) of Spike (S) protein of virus and host ACE2 receptor in the presence of TMPRSS seine protease secreted by the host cell. In this study the mutation hotspots of S-protein will be identified and the impact of such mutation in the binding affinity will be studied. Additionally, the lead molecule which can bind to the mutated protein also will be identified. Multiple sequence alignment of the spike protein sequence of SARS-CoV-2 shows the number of single amino acid mutation hotspots such as L5F, R214L, R408I, G476S, V483A, H519Q, A520S, T572I, D614G and H655Y. Among these mutations D614G has 57.5% occurrence and G476S, V483A has 7.5% occurrence. The mutated proteins were modelled based on wild type homolog and docked to ACE2 receptor. When the mutated S protein is docked, the ΔG (binding free energy) value is very minimal in mutated protein showed the stability of variants. By the drug repurposing method, 1000 FDA approved drugs were virtually screened for its binding to RBD of S1 domain. Among these drugs Digitoxin, Gliquidone and Zorubicin Hcl binds to spike proteins with higher docking score (lesser than -8.5 Kcal/mol) to both wild type and mutants.

Key Words:

Spike Protein, SARS-CoV-2, mutation, Drug Repurposing, Digitoxin.

1. Introduction:

A novel corona virus (nCoV) that originated in Wuhan provinces of China in December 2019, and became a pandemic, has affected more than 10 million people worldwide with a death toll of >500 thousand people within six months of its origin (5). The nCoV belongs to Severe Acute Respiratory Syndrome (SARS) family; hence named SARS-CoV-2 affects the human respiratory system. Phylogenetic analysis of the genome of this SARS-CoV-2 shows more than 80% sequence identity with SARS-CoV and 50% to Middle East Respiratory Syndrome coronavirus (MERS-CoV) [1]. The isolates of the novel beta coronavirus shows 88% sequence identity to bat-derived SARS like coronavirus (2,3). Coronavirus is an enveloped virus with a single stranded positive ribonucleicacid (RNA) as genetic material inside it. Size of virus is around 80-120nm and the genome size varies from 26Kbp to 30Kbp.

SARS-CoV-2 is made up of four structural proteins and other non-structural proteins (NSPs) [4].The membrane of the coronavirus contains three main structural proteins: the spike protein (S), envelope protein (E) and membrane protein (M). The S-protein is a crown like homotrimer type I glycoprotein. E protein has a short ectodomain, a transmembrane domain and a cytoplasmic tail. M protein spans the membrane with a short N-terminal ectodomain and a cytoplasmic tail. Hemagglutinin esterase, an additional membrane protein is a non-essential protein which helps in the viral entry and/or pathogenesis in vivo. There are additional internal proteins encoded within the open reading frames (ORFs) [3]. The ORF 1 a/b is the largest gene comprising two-third of the viral RNA genome and codes for two large polypeptides pp1a and pp1b. These polypeptides further processed as 16 non-structural proteins. SARS-CoV-2 also express other polyproteins, nucleoproteins and membrane proteins like RNA polymerase, chymotrypsin or papine-like proteases, helicase, glycoprotein and accessory proteins which is the other one third of the genome [2,4].

As the SARS-CoV-2 shares more similarity to the SARS-CoV, the previously used animal models were used to study the infectious pathogenicity of CoV-2 [4]. Group 2 corona viruses like SARS-CoV enters into the cell through the interaction of spike protein with angiotensin converting enzyme 2 (ACE2) receptor. After the interaction, virus gets internalized by endocytosis and uncoating of the genome occurs by fusion with lysosome vesicles, which contains cathepsins, and the sense strand of the genome is translated to create replicase proteins from ORF1a/b. These proteins use the genome as a template to generate full-length anti-sense RNAs, which subsequently serve as templates in generating additional full-length genomes [9, 11].Simultaneously the mRNAs are translated into proteins from its 5' end. Both sense and anti-sense strands of RNA synthesized undergoes a unique discontinuous transcription mechanism that is not completely understood. After translation, protein forms double layered vesicle and also undergoes post-translational modification in the rough endoplasmic reticulum. Finally, the M and E membrane proteins along with other accessory proteins are localized to the Golgi intracellular membranes and get budded off and released out the cell by exocytosis [12].

The extremely stable triple-stranded coiled coil Spike protein (S; surface projection; 175 kDa) is a homotrimer glycoprotein containing two domains- S1 and S2 in equimolar

proportion [6,8]. The assembly of monomers to trimer is a rate limiting step. The S1 spike protein forms the head and S2 forms the stalk. Receptor Binding Domain (RBD) of the S1 specifically recognizes the ACE 2 receptor present on the epithelial cells [11]. The CoV spike protein mediates receptor binding via S1 domain and membrane fusion via S2 domain [13]. S1 domain contains two sub domains, a N-terminal domain (NTD) and a C-Terminal domain (CTD). S1 is very diverse among species. S2 domain is most conserved domain in the protein. The viral fusion protein is present in the S2 domain [14].For the priming of S protein to the ACE 2 receptor and for effective infection and pathogenesis; the host cellular serine protease TMPRSS2 activation is very significant. Priming of S protein by host cell protease is very essential for viral entry and encompasses S protein cleavage at the S1/S2 and S2' site. The S1/S2 cleavage site of SARS-2-S harbours several arginine residues (multi-basic), which indicates high cleavability [10].After dissociation the S2 domain undergoes a dramatic structural change. SARS-CoV-2 contains a proprotein convertase (PPCs) motifs at the boundary of S1/S2 [11].
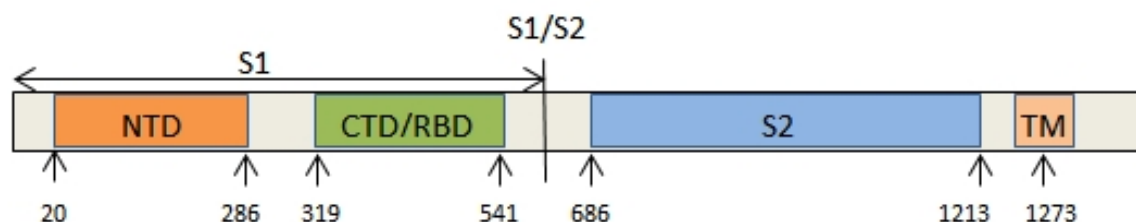

Fig.1 Schematic representation of spike protein- N Terminal Domain (NTD) and C Terminal Domain(CTD) or Receptor Binding Domain (RBD) belongs to S1 Domain. S2 domain is followed by the S1 domain of Spike protein in SARS-CoV-2.

Enveloped S protein is essential for the viral entry into the host, belongs to class I fusion proteins. Fusion proteins are conserved among species and are made up of 15-25 non-polar amino acids [13]. S2 domain has two heptad repeats (HR1 and HR2) regions, which forms a coiled coil fusion-active conformations. This hairpin conformation aids in the membrane fusion and subsequent viral entry. Drugs aiming at inhibiting the hairpin formation helped in the successful development of human immune deficiency virus-1 inhibitors [7].

To date, there is no proper antiviral drug available to cure SARS-CoV-2. Treatment is based on the symptoms like decreasing the cytokine storm to basal level [15]. It is essential to find a better effective drug to treat SARS-CoV-2. Currently drugs like Camostat, a TMPRSS2 inhibitor; Chloroquine and hydroxychloroquine, the anti-malarial drugs that affect endosomal fusions; Azithromycin, a broad spectrum antibiotic are being administered [16]. During this emergency state, utilizing available small molecules which are anti-infective agents might help us in combating the disease. Drug repurposing studies using computer aided drug interaction study helps us to fasten up in screening a large number of molecules.

Viral entry into the cell happens only when the spike protein interacts with the ACE2 receptor of the cell. If the binding itself is inhibited, further pathogenesis can be prevented. In this study, mutations in the spike protein were identified using sequence alignment and the

impact of mutation in the binding to receptor was analyzed by comparing the binding energies of wild type and mutants. Finally, all available FDA approved small molecules in the PubChem database will be virtually screened to find out the best binding small molecule to both wild type and mutant.

2. Methodology:

2.1 Data collection:

The amino acid sequences (N=80) of Spike protein were retrieved from the CDC translation of nucleic acid, NCBI Viruses. The sequences obtained are from different geographical locations where the samples are isolated from oral swabs or from blood. The Accession number of the collected samples was tabulated in Supplementary (Table 1). Conserved domains in S1 and S2 domains were analyzed using NCBI conserved domain analysis search tool [21].

2.2 Comparative Sequence Analysis:

Multiple sequence alignment (MSA) for all the protein sequences were collected to analyse the sequence/motif/residue involved in the substrate binding or function of protein using MEGA7 tool. MSA gives an insight into the conserved and non- conserved amino acids in the spike protein. All the sample sequence was aligned and phylogenetic tree was constructed using MEGA7 software [18]. The aligned sequence is converted to logo format using WebLogo online server for easy representation of mutations in sequences [19].

2.3 Modelling of Spike protein mutant using SWISS MODEL:

The most common mutants were identified using MSA. Comparative Homology modelling was carried out to model the structure of the mutants. The following criteria were used to choose the S-protein: the query should have >32% sequence identity and the query coverage should be as high as possible. The S- protein of the sequence (ID: NC_045512) obtained from Wuhan is assumed as wild type due to its origin and the D614G, G476S, V483A mutated samples were taken from India (ID: MT509504.1), USA (ID:MT252690),USA (ID: MT259236) respectively. The model was built with a template S Protein (PDB ID: 6vsb) which shares 99.59% similarity using SWISS-MODEL and 100% coverage [20]. ACE2 receptor was modelled using (PDB ID:1R42) template which is the human ACE2 receptor 'A' chain with 100% identity and 100% query coverage. The predicted models were assessed for its quality using ProSa web server [26]. Ramachandran plot constructed for validating the structure shows the amino acids of the modelled proteins mostly lies in the allowed regions (Supplementary figure 2). The best modelled proteins among several clusters were chosen based on the Z value. The Z score of the wild type, and mutants D614G,G476S and V483A are -12.59, -12.55,-12.65,-12.39 respectively. PyMOL Molecular Graphic System (Version 1.8.2.0 Schrodinger) was used for comparative structure analysis between the wild type and mutants.

2.4 Protein-protein interaction study:

The active and passive interacting amino acids of both Spike proteins and ACE2 receptors were analyzed using CPROT web server [22]. The protein-protein interaction was carried out between wild type/mutant spike proteins to human ACE2 receptor was studied using HADDOCK docking web server [23]. The binding energies and interacting amino acids of the docked proteins were analysed using the PRODIGY web server [28,29]

2.5 Molecular docking:

A library of FDA approved 1000 small molecules deposited in the PubChem [26] were screened against RBD domain of S1 protein via docking using Autodock Vina v.1.1.271 with the PyRx v.0.8 GUI [25]. The protein and the ligand were energy minimized before docking. The grid for docking was adjusted to cover only the RBD. Top 3 ranking molecules with highest binding energy were subjected to MD simulation.

3. Results:

3.1 Analysis of mutation in the sequence of Spike protein in SARS-CoV-2:

Comparative sequence analysis of the collected spike protein (length = 1273 amino acids) from NCBI viruses shows several mutations in the S1 domain of the S-protein. The negatively charged aspartate amino acid at the 614th position was found to be mutated to hydrophobic amino acid Glycine (D614G) in 46 (57.5%) sequences. The other two mutations with significantly higher percentages are at the RBD of S1 domain in the S-protein. One of the two mutations are at the 476th position where the hydrophobic amino acid glycine was mutated to hydrophilic amino acid serine (G476S) in 6 sequences with the mutation occurrence of 7.5%, and the other mutation was at the 483rd position the hydrophobic amino acid valine was mutated to hydrophobic amino acid alanine (V483A) in 6 sequences (7.5%). The other mutations which occur less frequently are L5F (3.75%), R214L (1.25%), R408I (1.25%), H519Q (1.25%), T572I (1.25%) (Table1). L5F and R214L mutations occur at NTD of S1 domain and rest of the other mutations occur at CTD of S1 domain. The major occurrence of mutations is D614G, G476S and V483A (Figure 1). Analysis of the phylogenetic tree constructed using neighbour joining method shows several clades (sub-types) in the S-protein. (Figure2). The sub-types arise because of the single amino acid mutations in the S-protein.

3.2 Conserved domain analysis in the S-protein across different species:

When different species like SARS, MERS, murine, bat and human corona viruses of NTD were aligned only very less number of amino acids were aligned (Supplementary Figure 1.1). The NTD of the S1 domain of S-protein is very less conserved among different species. The C-Terminal Domain (CTD) among different species like SARS, SARS-CoV-2, MERS, murine hepatitis virus, human coronavirus were aligned less than half of the amino acids were aligned (Supplementary Figure 1.2). Thus, the CTD of S1 domain of S-protein of coronavirus is less conserved across species. The S2 domain of the spike protein of different species like Feline infectious peritonitis virus, Human coronavirus, Avian infectious

bronchitis virus and SARS coronavirus shows higher similarity in the amino acid sequence when aligned using the Conserved Domain Search by using NCBI (Supplementary Figure 1.3). Thus, the S2 domain of S-protein of coronavirus is conserved across different species.
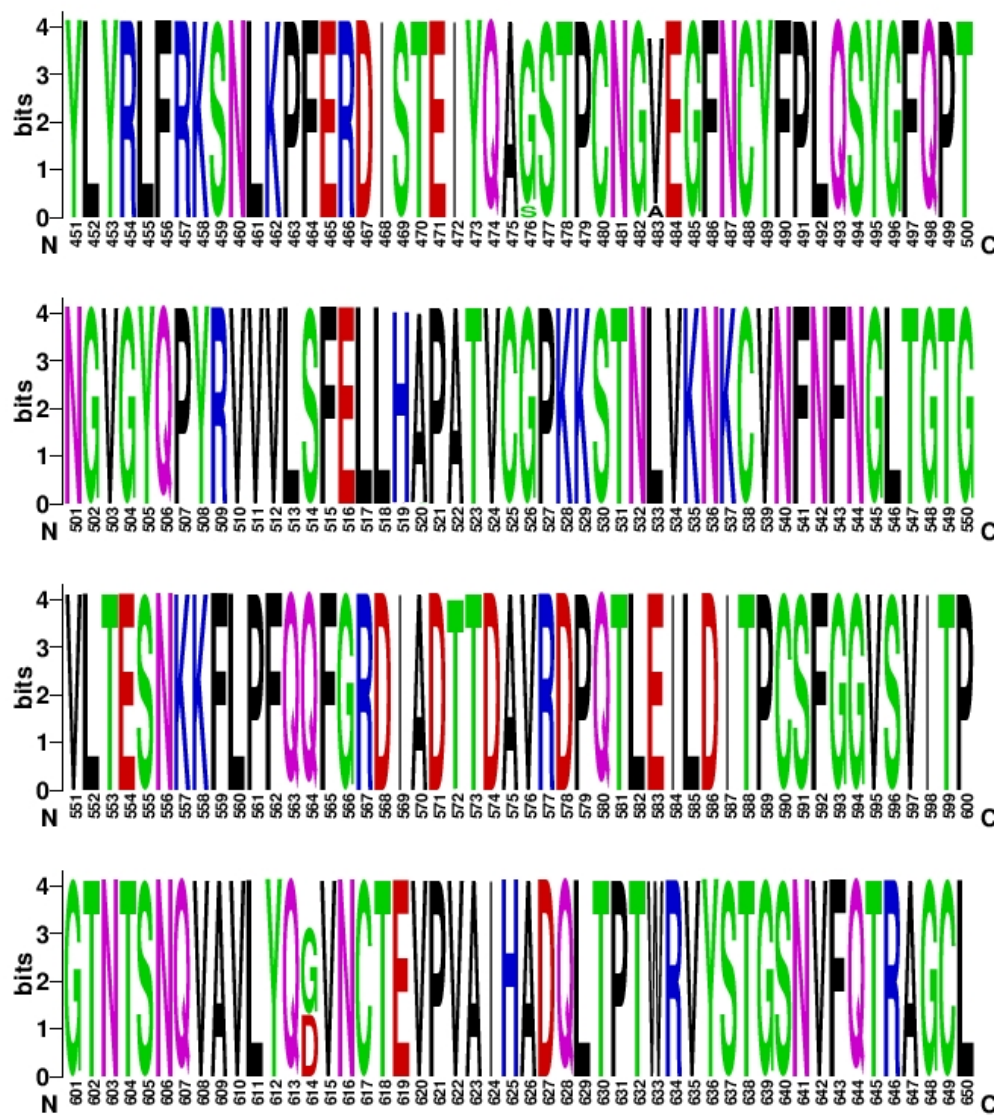

Figure 1: The sequence alignment is represented as a web logo. The prominent mutation such as D614G, V483A and G476S is shown as double letter.

Table 1: Mutations present in the SARS-CoV-2 spike protein at the S1 domain is listed with its percentage of mutation.

Spike Protein Domain	Mutation site	Percentage of mutation
S1 Domain (N-Terminal)	L5F	3.75%
	R214L	1.25%
S1 Domain (C-Terminal/Receptor Binding Domain)	R408I	1.25%
	G476S	7.5%
	V483A	7.5%
	H519Q	1.25%

	A520S	1.25%
	T572I	1.25%
	D614G	57.5%
	H655Y	1.25%
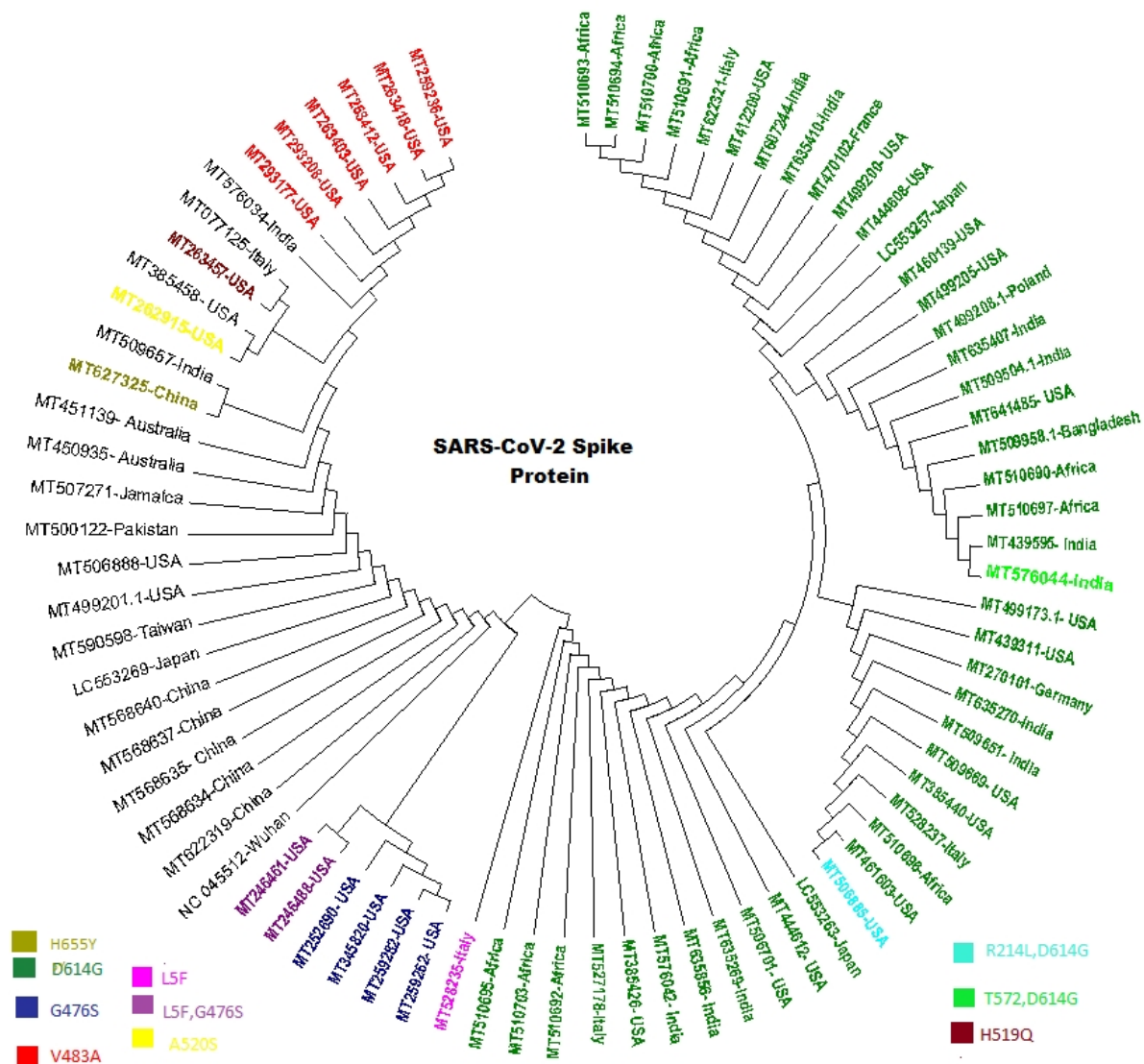
Figure 2: Phylogenetic tree constructed using neighbour joining method is displayed. The tree contains the NCBI code with its geographical location. The colour code represents each mutation and is as follows: D614G (dark green), G476S (ink blue, purple), V483A (red), L5F (pink, purple), A520S (yellow), R214L (sea blue), T572I (light green), H519Q (maroon), H655Y (olive green).

3.3 Protein structure analysis of wild type and mutants:

To gain deeper understanding of the domain architecture and the amino acids which are interacting with the ACE2 receptors in the wild type and the mutants, homology modelling was undertaken. The modelled structures were built using the template of S-protein crystal structure of SARS-CoV-2 isolated from infected individual. The structure of the homotrimer S-protein of the wild type showed a similar folding pattern except for the functional domain's

flexible loop region of the RBD of S1 domain and in the intermediate region between S1 and the S2 domains. The Root Mean Square Deviation (RMSD) value was calculated for its structure deviation from the wild type structure and the mutants. For RMSD value for wild type and the D614G, G476S and V483A mutants are 0.016, 0.043 and 0.053 Å respectively. In the D614G mutation, it was observed that the S1/S2 boundary loop got altered in the local region which is shown in Figure 3A and when G was mutated to S at 476th position, the loop region in the functional domain got altered at the site (Figure 3B). Similarly, when V got mutated to A at the 483rd position, the loop region at the functional domain of S1 protein (RBD) got altered in the local region (Figure 3C).
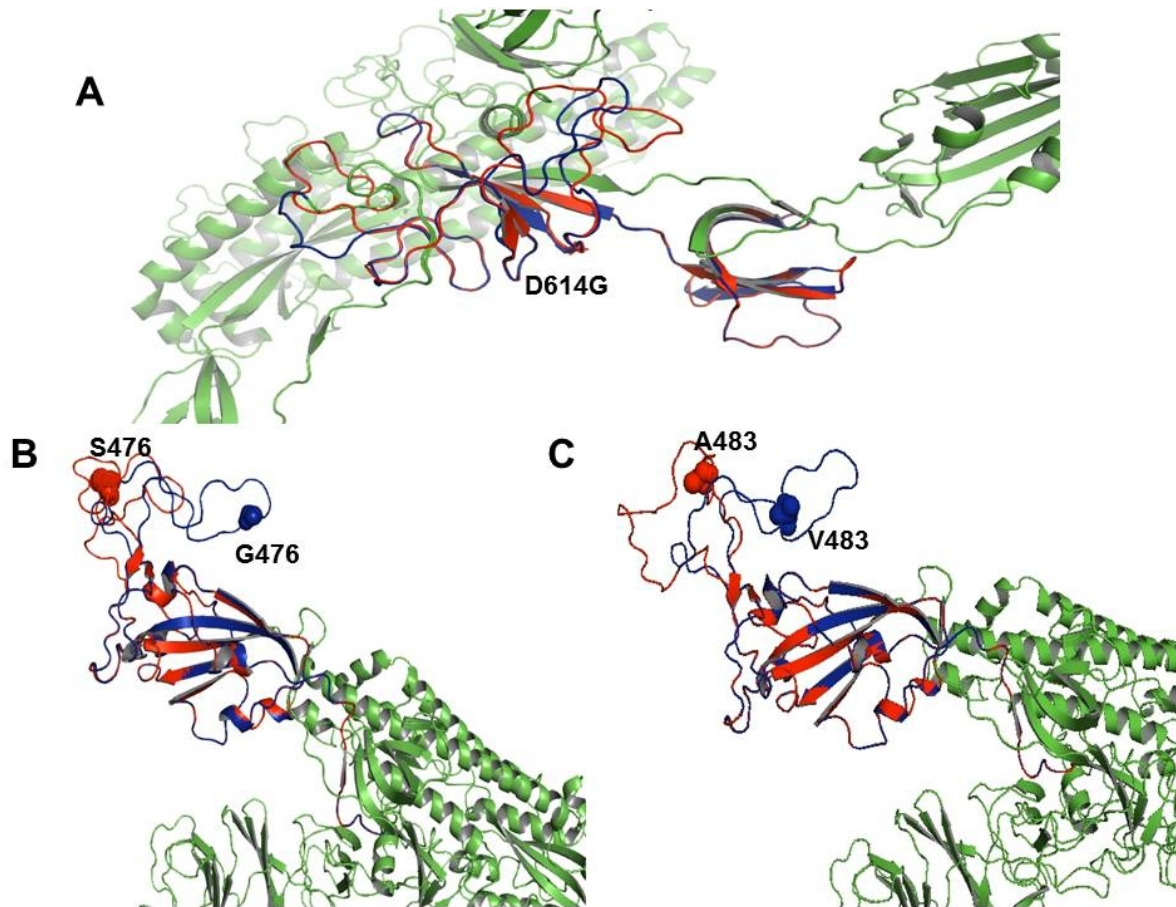
Figure 3: Analysis of the mutated S-protein by comparing with wild type. A) D614G mutant comparison with wild type. The cartoon represented in green shows the S1 (NTD, CTD) and S2 domains. The amino acid D and G are represented in blue and red spheres. B) G476S mutant compared with wild type. The red and blue colour represents the CTD/RBD of mutant and the wild type. The green colour represents the NTD and S2 domain. The amino acid G and S are represented in blue and red spheres C) V483A mutant compared with wild type. The red and blue colour represents the CTD/RBD of mutant and the wild type. The green colour represents the NTD and S2 domain. The amino acid V and A are represented in blue and red spheres.

3.4 Docking of S1-domain of S-protein of SARS-CoV-2 on human ACE2 receptor:

The modelled proteins were docked to the modelled human ACE2 receptor using HADDOCK flexible docking. The Active and passive interacting amino acids of both S-

protein and the ACE2 receptor were defined using the CPROT web server. The interacting amino acids of the ACE2 receptors in helix1 and 2 are: PHE10, LYS13, GLU17, ASP20, LEU21, TYR23, GLN24, LEU27, ASN31, GLU39, ASN43, ASN46, LYS50, PHE54, ALA53, GLU57, GLN58, LEU61, TYR65, ASN312, ASP337 and ARG339. But the interacting amino acids of the wild type and mutant S-protein are varying. Amino acids involved in the interactions are detailed in the supplementary table 2. The loop region of the functional domain RBD of S1domain of S-protein in wild type and mutant types are interacting with two helices of the A chain of ACE2 receptor. But the interactions of the loop of S-protein with the ACE2 receptor are differing based on the mutation involved (Figure 4).
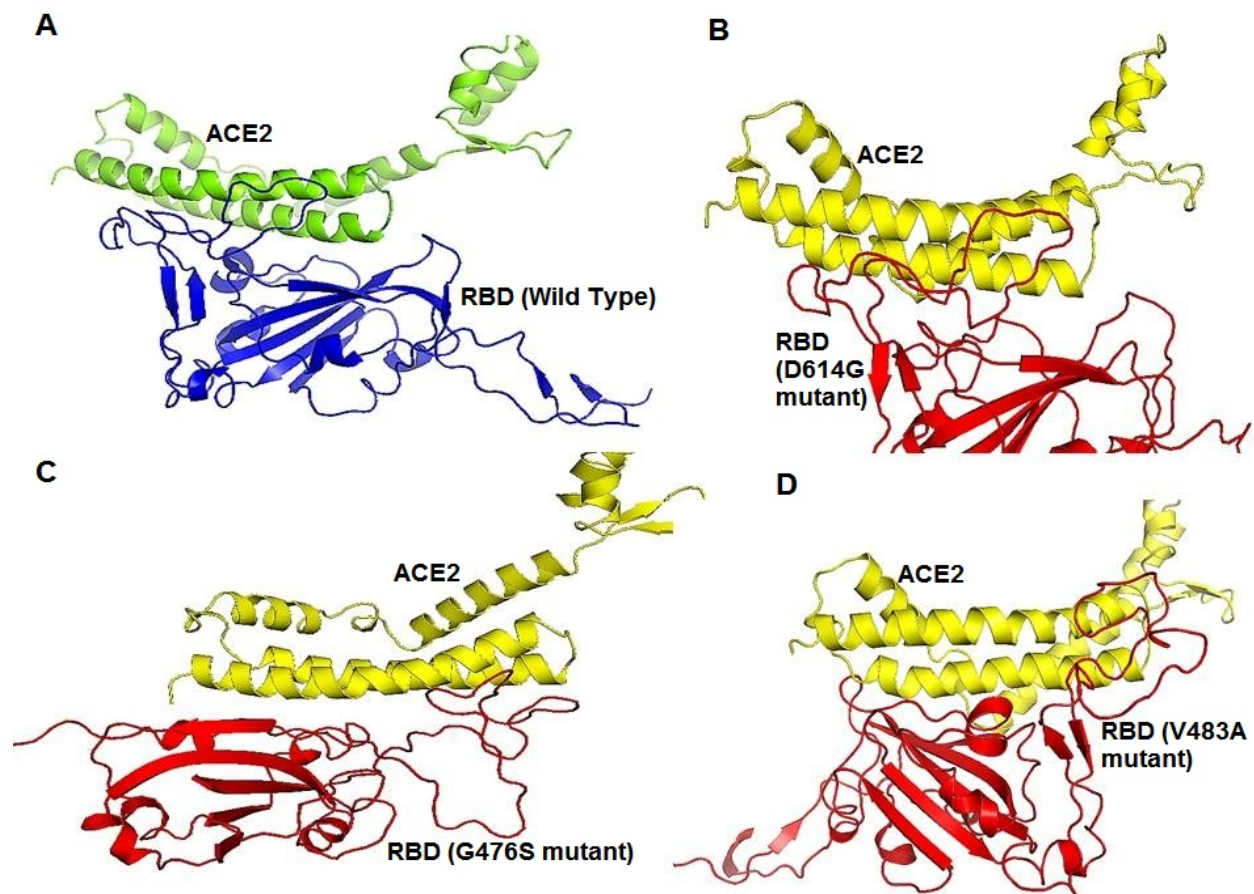
Figure 4. The docked poses of human ACE2 receptor helix 1 and 2 which interacts with the RBD of S-protein is shown in cartoon representation. a) The wild type RBD is shown in blue and the interacting helix of human ACE2 receptor A chain in green, b) The D614G mutant RBD is shown in red and ACE2 helix is in yellow, c) The G476S mutant RBD is shown in red and ACE2 helix in yellow, d) The V483A mutant RBD is shown in red and ACE2 helix in yellow.

The S-proteins were docked to the human ACE2 receptor and their binding energies were evaluated using PRODIGY web server. Due to the single amino acid mutations, the interacting amino acids are varying and as a result their binding energies are also varying. The list of mutations with their binding energies ΔG (Binding free energies in Kcal/mol) is listed in Table 2. These binding energies are calculated for room temperature (25°C). The difference in the binding energy is mainly because of the interacting amino acid side chain. Due to structure difference in the loop region of the RBD of S1 domain of S-protein, their

interacting amino acids are also altered. From the tabulated data, the mutated proteins have strong binding affinity towards the human ACE2 receptors when compared with the wild type. From the interacting amino acids of D614G mutant, it is observed that the number of tyrosine (TYR) interacting with human ACE2 receptor is higher than the other modelled proteins. Tyrosine is one of the amino acids which have the highest binding affinity. Also V483A mutant model has the highest number of tryptophan (TRP) than other S-protein models. Tryptophan is also one of the amino-acids which have the highest binding affinities.
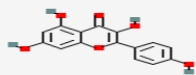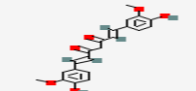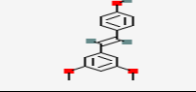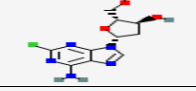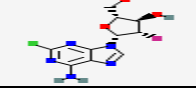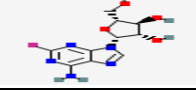
Table 2. List of ΔG (Binding energy) of the Spike protein interacting with hACE2 receptor is shown.
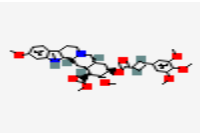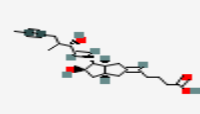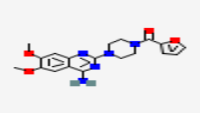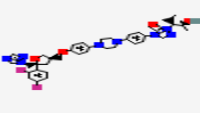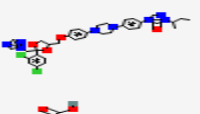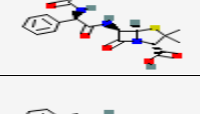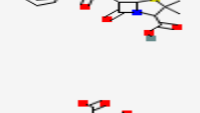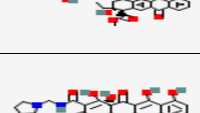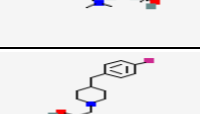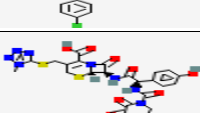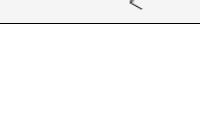
Mutations	Binding Energy ΔG (kcal/mol)
Wild Type	-14.1
G476S	-14.8
V483A	-15.2
D614G	-15.2
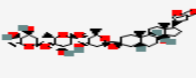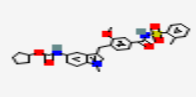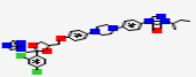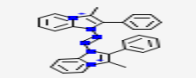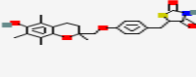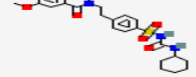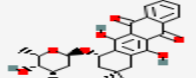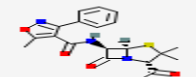
3.5 Identification of lead molecule by Virtual Screening:

Virtual screening was carried out for 1000 small molecules using PyRx Autodock Vina (1.8.2.0) targeting the RBD of the S1 domain of the S-protein wild type. The top 30 small molecules which have the least docking score were taken for further docking with the mutant type. Table 3 shows the docking score (free binding energy-ΔG) of all the mutant models and the wild type S-protein. Flexible docking method was carried out for this study.

Table 3: List of small molecules and their structures and binding energy to RBD of S1 domain is shown.

Drug Name	Structure	Wild Type	G476S	V483A	D614G
Flavanoids: 1. Kaempferol		-7.7	-6.3	-6.9	-7.7
2. Curcumin		-7.3	-6.1	-5.8	-5.5
3. Pterostilbene		-6.6	-5.8	-6.0	-6.2
4.Cladribine		-6.9	-5.3	-5.2	-5.8
5. Clofarabine		-6.7	-5.7	-6.1	-5.7
6.Fludarabine		-6.8	-7.5	-7.7	-8.5

Anti-hypertensive: 7. Rescinnamine		-8.3	-6.6	-6.9	-8.1
8. Iloprost		-7.0	-6.4	-6.1	-6.8
9. Prazosin		-7.3	-6.2	-6.5	-6.3
Anti-fungal: 10. Posaconazole		-8.0	-7.9	-8.0	-7.8
11. Itraconazole		-8.4	-7.2	-8.0	-8.4
Antibacterial 12. Sulfasalazine		-8.2	-7.2	-6.9	-7.9
13. Azlocillin		-7.2	-7.2	-7.2	-7.1
14. Penicillin		-6.2	-6.4	-6.1	-5.8
15.Cefsulodine		-7.7	-7.2	-7.1	-6.6
Anti-coagulant 16. Dabigatran etexilate mesylate		-7.3	-6.5	-6.9	-7.8
Anthracyclines 17. Zorubicin Hcl		-8.9	-9.0	-8.1	-8.9
18. Aclarubicin		-8.3	-8.2	-8.0	-8.8
Tetracyclin Derivative 19. Rolitetracycline		-7.2	-6.9	-7.1	-7.6
20. E155		-7.4	-6.7	-6.2	-6.1
Cephalosporin 21. Cefoperazone		-7.3	-7.0	-7.5	-7.5

22. Hesperidin		-7.9	-7.6	-7.0	-7.8
23. Digitoxin		-9.1	-10.0	-8.3	-9.9
24.Zafirlukast		-8.4	-8.0	-7.5	-8.4
25.Itraconazole		-8.4	-7.2	-8.0	-8.4
26. Fazadinium		-8.4	-7.5	-7.7	-8.5
27. Troglitazone		-8.7	-7.9	-8.0	-7.5
28. Gliquidone		-9.0	-8.2	-8.8	-8.9
29. Idarubicin		-7.7	-8.1	-7.7	-7.8
30. Oxacillin		-7.2	-6.4	-7.2	-7.5

From the top 30 small molecules which have the least binding energies, only top 3 drugs, Digitoxin, Gliquidone and Zorubicin Hcl were selected. These drugs can bind to the RBD of S1-domain of S-protein wild type and the mutants with the least binding energies and their RMDS score is zero. The amino acids which are involved in the interaction with Digitoxin are ARG355 (H-bonding), TRP353 (Pi-Sigma) and TYR473 (H bonding) is shown in Figure 6.1. For Gliquidone the interacting amino acids of S-protein are ALA348 (Pi-Alkyl), SER349 (Pi-Donor H bonding), ILE472 (Alkyl), LEU492 (Pi-CH), PHE486 (Pi-CH) , ILE468 (Pi-CH), ALA352 (Pi-CH), TYR351 (Pi-CH bonding) is shown in Figure 6.2. For Zorubicin HCl the interacting amino acids are TRP353 (H-bonding), TYR351 (H-bonding), TYR421 (H-bonding), ARG454 (H-bonding), GLU484 (H-bonding, Pi-Anion), GLN474 (Amide Pi stacked) is shown in Figure 6.3. The binding energies of Digitoxin, Gliquidone and Zorubicin Hcl to wild type are -9.1 Kcal/mol, -9.0 Kcal/mol and -8.9 Kcal/mol respectively.

Figure 6.1 Interaction of S-protein RBD S1 domain with the lead molecule Digitoxin. (a) Schematic representation of interacting amino acid of S-protein with the lead molecule (b) Ball and stick representation of the drug (in blue) interacting with the respective amino acids (c) ball and stick representation of small molecule bound to the RBD and the S-protein is shown as surface representation.

Figure 6.2 Interaction of S-protein RBD S1 domain with the lead molecule Gliquidone. (a) Schematic representation of interacting amino acid of S-protein with the lead molecule (b) Ball and stick representation of the drug (in blue) interacting with the respective amino acids (c) ball and stick representation of small molecule bound to the RBD and the S-protein is shown as surface representation.
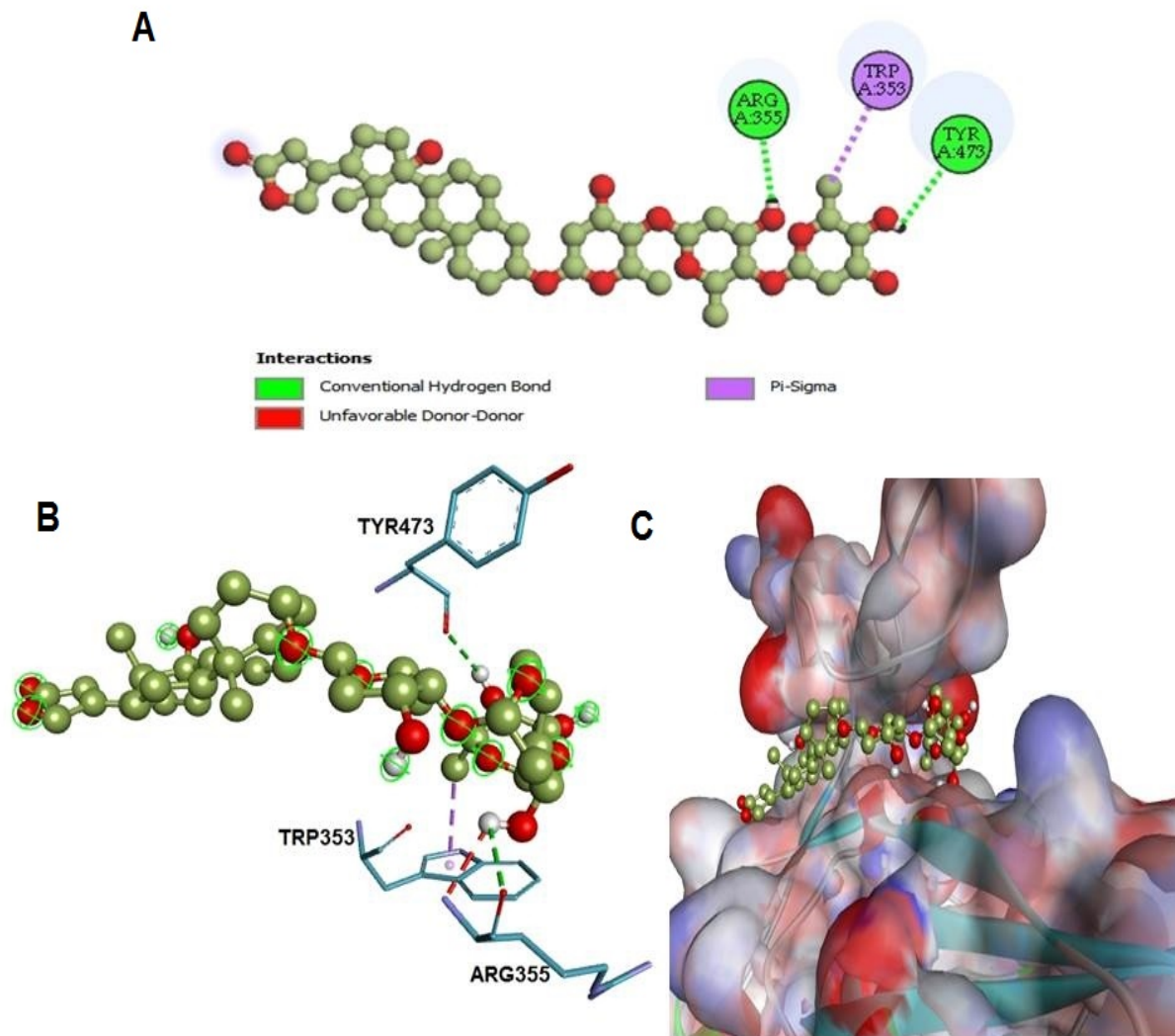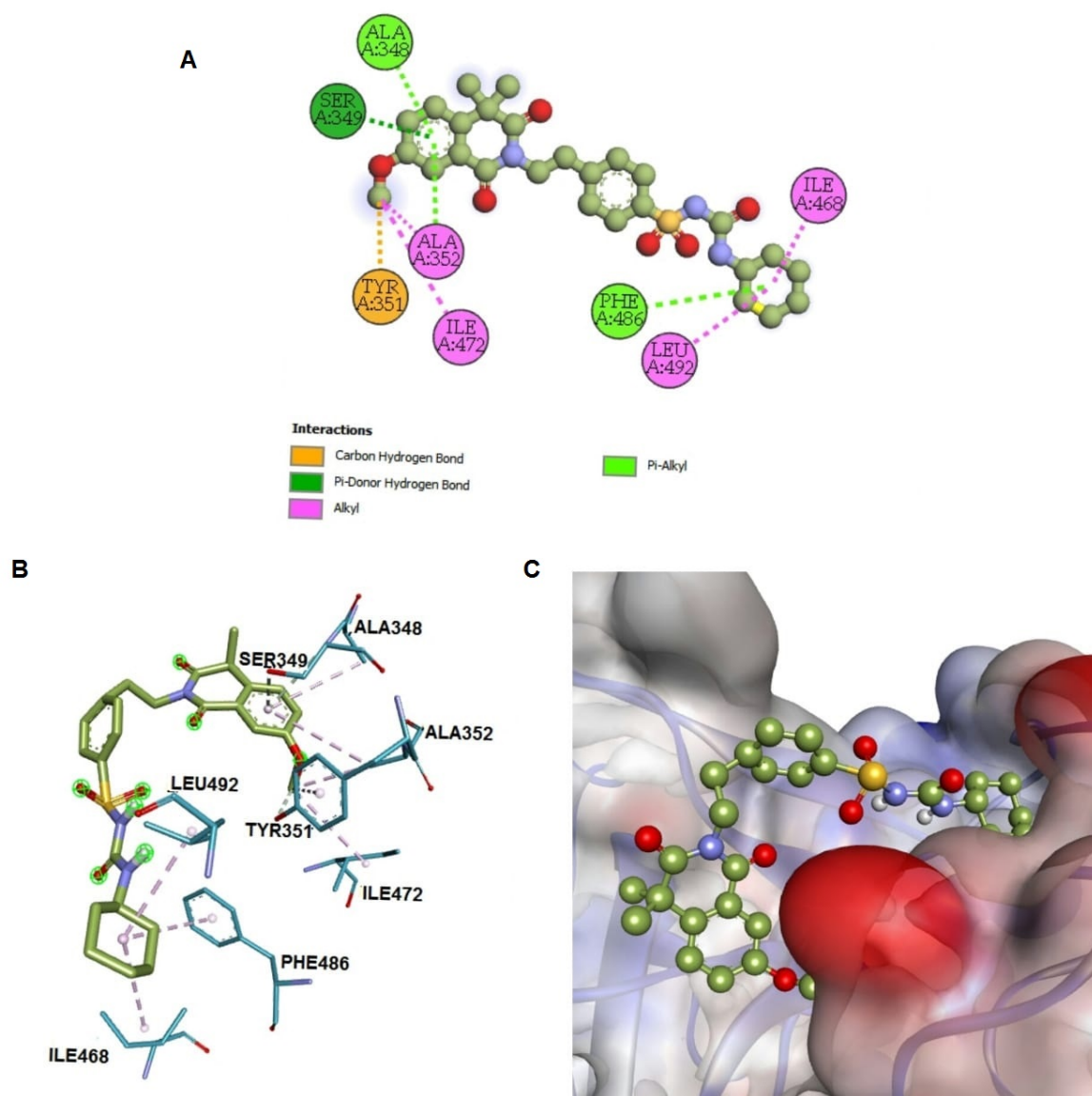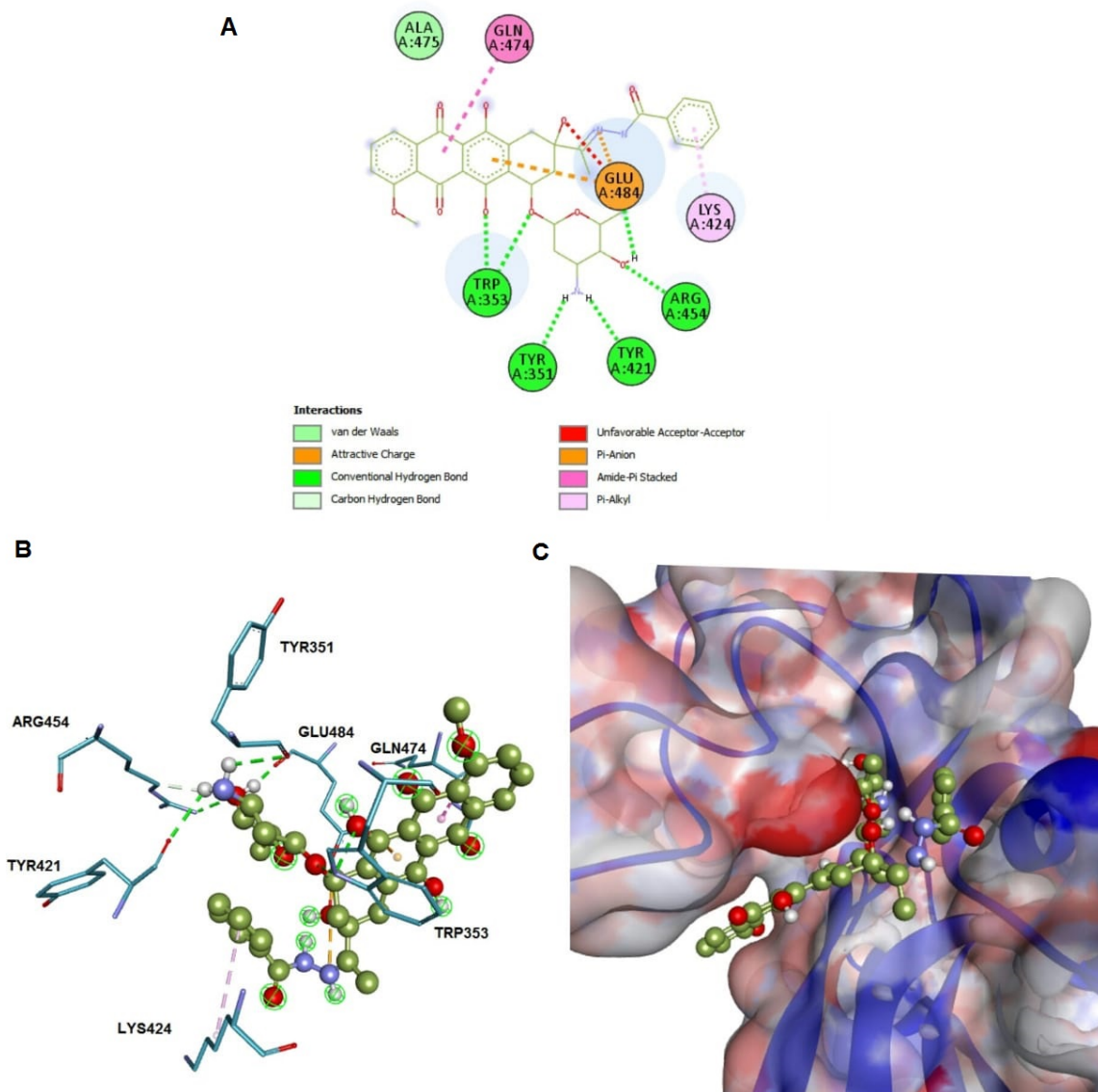
Figure 6.3 Interaction of S-protein RBD S1 domain with the lead molecule Zorubicin HCl. (a) Schematic representation of interacting amino acid of S-protein with the lead molecule (b) Ball and stick representation of the drug (in blue) interacting with the respective amino acids (c) ball and stick representation of small molecule bound to the RBD and the S-protein is shown as surface representation.

4. Discussion:

The RBD- S1 spike protein of SARS-CoV-2 is involved in the interaction with the human ACE2 receptor. The sequence alignment of S-protein shows mutation hotspots. Thus, the virus is undergoing a huge number of mutations in the recent days. Within six months of time since the pandemic occurred there are several mutations observed. This shows that the virus is getting evolved. D614G mutation is seen more predominantly in the Indian and American population. This mutation has occurred in the due course of pandemic which implies that this change might enhance the viral transmission. The mutation in the RBD is showing least binding energy when docked with hACE2 receptor. The enhancement in the binding may be indirectly co-related with the death rates. Because the USA has undergone a huge number of

mutation on RBD and the death rate is high in the country. This study reveals that the spike protein is becoming the hotspot for mutation and therefore the mutation may help either in the transmission or enhanced binding due to the alterations in the loop of RBD which is involved in the binding. Based on drug repurposing, library of 1000 small molecules were taken for study and from then top 30 drugs were analysed for their binding efficiency with mutants. Top 3 drugs were identified as best candidates which can bind specifically to RBD and can inhibit the binding of S-protein, which is the initial step for viral entry into the cell. These drugs can serve as a lead molecule for inhibiting the viral attachment to the receptor. Digitoxin is a cardiac drug, Gliquidone is ATP kinase 'K' inhibitor used for Type 2 diabetes and Zorubicin HCl is an anti-cancer drug. Future aspect of this study will be testing these drugs in cell lines and animal models.

5. Reference:

1. Rothan, H. A., & Byrareddy, S. N. (2020). The epidemiology and pathogenesis of coronavirus disease (COVID-19) outbreak. Journal of Autoimmunity, 109, 102433. https://doi.org/10.1016/j.jaut.2020.102433

2. Li, X., Geng, M., Peng, Y., Meng, L., & Lu, S. (2020). Molecular immune pathogenesis and diagnosis of COVID-19. Journal of Pharmaceutical Analysis, 10(2), 102–108. https://doi.org/10.1016/j.jpha.2020.03.001

3. Weiss, S. R., & Navas-Martin, S. (2005). Coronavirus Pathogenesis and the Emerging Pathogen Severe Acute Respiratory Syndrome Coronavirus. Microbiology and Molecular Biology Reviews, 69(4), 635–664. https://doi.org/10.1128/mmbr.69.4.635-664.2005

4. Shereen, M. A., Khan, S., Kazmi, A., Bashir, N., & Siddique, R. (2020). COVID-19 infection: Origin, transmission, and characteristics of human coronaviruses. Journal of Advanced Research, 24, 91–98. https://doi.org/10.1016/j.jare.2020.03.005

5. World Health Organization (WHO),2020. Coronavirus disease 2019 (COVID-19): situation report, 140.

6. Cavanagh, D. (1983). Coronavirus IBV: Structural Characterization of the Spike Protein. Journal of General Virology, 64(12), 2577–2583. https://doi.org/10.1099/0022-1317-64-12-2577

7. Xu, Y., Lou, Z., Liu, Y., Pang, H., Tien, P., Gao, G. F., & Rao, Z. (2004). Crystal Structure of Severe Acute Respiratory Syndrome Coronavirus Spike Protein Fusion Core. Journal of Biological Chemistry, 279(47), 49414–49419. https://doi.org/10.1074/jbc.m408782200

8. Delmas, B., & Laude, H. (1990). Assembly of coronavirus spike protein into trimers and its role in epitope expression. Journal of virology, 64(11), 5367–5375.

9. Perlman, S., & Netland, J. (2009). Coronaviruses post-SARS: update on replication and pathogenesis. Nature Reviews Microbiology, 7(6), 439–450. https://doi.org/10.1038/nrmicro2147

10. Hoffmann, M., Kleine-Weber, H., Schroeder, S., Krüger, N., Herrler, T., Erichsen, S., Schiergens, T. S., Herrler, G., Wu, N.-H., Nitsche, A., Müller, M. A., Drosten, C., & Pöhlmann, S. (2020). SARS-CoV-2 Cell Entry Depends on ACE2 and TMPRSS2 and Is Blocked by a Clinically Proven Protease Inhibitor. Cell, 181(2), 271–280.e8. https://doi.org/10.1016/j.cell.2020.02.052

11. Shang, J., Wan, Y., Luo, C., Ye, G., Geng, Q., Auerbach, A., & Li, F. (2020). Cell entry mechanisms of SARS-CoV-2. Proceedings of the National Academy of Sciences, 117(21), 11727–11734. https://doi.org/10.1073/pnas.2003138117

12. Weiss, S. R., & Navas-Martin, S. (2005). Coronavirus Pathogenesis and the Emerging Pathogen Severe Acute Respiratory Syndrome Coronavirus. Microbiology and Molecular Biology Reviews, 69(4), 635–664. https://doi.org/10.1128/mmbr.69.4.635-664.2005

13. Madu, I. G., Roth, S. L., Belouzard, S., & Whittaker, G. R. (2009). Characterization of a Highly Conserved Domain within the Severe Acute Respiratory Syndrome Coronavirus Spike Protein S2 Domain with Characteristics of a Viral Fusion Peptide. Journal of Virology, 83(15), 7411–7421. https://doi.org/10.1128/jvi.00079-09

14. Heald-Sargent, T., & Gallagher, T. (2012). Ready, Set, Fuse! The Coronavirus Spike Protein and Acquisition of Fusion Competence. Viruses, 4(4), 557–580. https://doi.org/10.3390/v4040557

15. Min, C.-K., Cheon, S., Ha, N.-Y., Sohn, K. M., Kim, Y., Aigerim, A., Shin, H. M., Choi, J.-Y., Inn, K.-S., Kim, J.-H., Moon, J. Y., Choi, M.-S., Cho, N.-H., & Kim, Y.-S. (2016). Comparative and kinetic analysis of viral shedding and immunological responses in MERS patients representing a broad spectrum of disease severity. Scientific Reports, 6(1). https://doi.org/10.1038/srep25359

16. Guy, R. K., DiPaola, R. S., Romanelli, F., & Dutch, R. E. (2020). Rapid repurposing of drugs for COVID-19. Science, 368(6493), 829–830. https://doi.org/10.1126/science.abb9332

17. Elmezayen, A. D., Al-Obaidi, A., Şahin, A. T., & Yelekçi, K. (2020). Drug repurposing for coronavirus (COVID-19): in silico screening of known drugs against coronavirus 3CL hydrolase and protease enzymes. Journal of Biomolecular Structure and Dynamics, 1–13. https://doi.org/10.1080/07391102.2020.1758791

18. Kumar, S., Stecher, G., & Tamura, K. (2016). MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. Molecular Biology and Evolution, 33(7), 1870–1874. https://doi.org/10.1093/molbev/msw054

19. Crooks, G. E. (2004). WebLogo: A Sequence Logo Generator. Genome Research, 14(6), 1188–1190. https://doi.org/10.1101/gr.849004

20. Biasini, M., Bienert, S., Waterhouse, A., Arnold, K., Studer, G., Schmidt, T., … Schwede, T. (2014). SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information. Nucleic Acids Research, 42(W1), W252–W258. https://doi.org/10.1093/nar/gku340

21. Marchler-Bauer, A. (2002). CDD: a database of conserved domain alignments with links to domain three-dimensional structure. Nucleic Acids Research, 30(1), 281–283. https://doi.org/10.1093/nar/30.1.281

22. De Vries, S. J., & Bonvin, A. M. J. J. (2011). CPORT: A Consensus Interface Predictor and Its Performance in Prediction-Driven Docking with HADDOCK. PLoS ONE, 6(3), e17695. https://doi.org/10.1371/journal.pone.0017695

23. Van Zundert, G. C. P., Rodrigues, J. P. G. L. M., Trellet, M., Schmitz, C., Kastritis, P. L., Karaca, E., Melquiond, A. S. J., van Dijk, M., de Vries, S. J., & Bonvin, A. M. J. J. (2016). The HADDOCK2.2 Web Server: User-Friendly Integrative Modeling of Biomolecular Complexes. Journal of Molecular Biology, 428(4), 720–725. https://doi.org/10.1016/j.jmb.2015.09.014

24. Wassenaar et al., WeNMR: Structural Biology on the Grid. J. Grid. Comp., 10, 743-767 (2012).

25. Dallakyan S, Olson AJ. Small-molecule library screening by docking with PyRx. Methods Mol Biol. 2015 doi: 10.1007/978-1-4939-22697_19.

26. Kim, S., Chen, J., Cheng, T., Gindulyte, A., He, J., He, S., Li, Q., Shoemaker, B. A., Thiessen, P. A., Yu, B., Zaslavsky, L., Zhang, J., & Bolton, E. E. (2019). PubChem 2019 update: improved access to chemical data. Nucleic acids research, 47(D1), D1102–D1109. https://doi.org/10.1093/nar/gky1033

27. Wiederstein M, Sippl MJ. ProSA-web: interactive web service for the recognition of errors in three dimensional structures of proteins. Nucleic Acids Res. 2007; 35:W407–10.

28. Vangone A. and Bonvin A.M.J.J. "Contact-based prediction of binding affinity in protein-protein complexes", *eLife*, 4, e07454 (2015).

29. Xue L., Rodrigues J., Kastritis P., Bonvin A.M.J.J.*, Vangone A.*, "PRODIGY: a web-server for predicting the binding affinity in protein-protein complexes", *Bioinformatics*, doi:10.1093/bioinformatics/btw514 (2016).