Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 7 September 2020 doi:10.20944/preprints202004.0281.v2

Mutation hot spots in Spike protein of SARS-CoV-2 virus

Feroza Begum[1,2], *Arup Kumar Banerjee[3], *Upasana Ray[1,2]

[1]Infectious Biology and Immunology Division, CSIR-Indian Institute of Chemical Biology, 4, Raja S.C., Mullick Road, Jadavpur, Kolkata-700032, West Bengal, India.
[2]Academy of Scientific and Innovative Research (AcSIR), Ghaziabad- 201002, India
[3]Department of Biochemistry, North Bengal Medical College and Hospital, Sushrutanagar, Siliguri-734012, West Bengal, India

**** Corresponding authors***

Correspondance:

ray.upasana@gmail.com

upasana.ray@iicb.res.in

Keywords: COVID-19, SARS-CoV-2, S protein, RBD, S1, S2

Abstract

Spike (S) protein of Coronaviruses help in receptor attachment and virus entry into the host cells. While S protein is required for virus entry, it is also important as an immunogen as it is the most accessible part of the virus architecture. S protein form knob like structures (viral spikes) protruding outwards in the form of homotrimers containing an S1 and S2 as monomers. Mutations in structural proteins of virus play crucial role in determining virulence and also in many instances influencing emergence of antibody escape variants and cellular tropism. In this paper we have performed in depth analyses of spike protein sequences from various parts of the world and tried to correlate the data with possible functional relevance of such mutations.

Introduction

In recent times novel coronavirus 2019/ nCoV-19/ COVID 19/ SARS CoV2 infection has become a pandemic and matter of concern worldwide. As per the World Health Organization (https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports/), globally, there have been millions of confirmed COVID 19 cases thousands have died.

The global sequencing effort have led to significant numbers of SARS-CoV-2 sequence information available now in various sequence databases specifically dedicated for COVID-19 research. Although number of sequences are increasing, the representation of various countries varies notably with more representations from few and very little from many others. Thus, this is a limiting factor in the field of sequence-based studies.

Spike protein is one of the most important structural protein of SARS-CoV-2 that plays the major role in virus attachment to its receptor followed by cellular entry [1-3]. It is one of the major structural proteins, 1273aa long, with two major sub domains, S1 and S2 (Figure 1). While S1 harbours the receptor binding domain or RBD and mediates virus attachment to its ACE2 receptor, S2 carries out the function of fusion to enable successful entry.

© 2020 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 7 September 2020 doi:10.20944/preprints202004.0281.v2

Viruses evolve by continuously mutating its gene sequences and the frequency of mutation is also attributed to the lack of proof-reading activity of the viral replication machinery. In case of SARS-CoV-2, the virus had proof-reading properties. Yet, mutations appear albeit at a slowly. The correlation between mutations and function/s of viral protein/s becomes important while designing drug/ vaccine candidates.

Here, we have studied the profile of mutations occurring in the spike protein world-wide and also laid more in-depth analyses of the sequences available from Indian patients.

Methods

Sequence information

A total of 11,571 available full length annotated sequences of SARS-CoV-2 spike protein (1-1273) of different geographical origins belonging to Asia (1017 sequences), Africa (103 sequences), Europe (441 sequences), Oceania (1426 sequences) (Australia: 1423 sequences and Guam 3 sequences), North America (8564 sequences) and South America (20 sequences) were downloaded from NCBI virus database for sequence analyses.

Additionally, 1648 whole genome sequences of SARS-CoV-2 deposited from Indian patients were also downloaded from GISAID database for detailed spike protein mutation analysis in Indian context. Although around 2000 sequences from Indian origin have been deposited in the GISAID database, not all are well sequenced with respect to the protein sequence of interest. Thus, only 1648 sequences with good translatable sequence information were used for current sequence analyses.

Accession YP_009724390.1 was used as the reference sequence.

Sequence Analyses

Multiple Sequence Alignment (MSA) was done using Clustal Omega and MAFFT V7 tools.

Surface glycoprotein coding regions from GISAID database were translated using ExPaSy protein translation tool (https://web.expasy.org/translate/). The translated sequences were further aligned using Clustal Omega tool (https://www.ebi.ac.uk/Tools/msa/clustalo/).

Prediction of protein stability change

Since D614G is one of the most dominant mutation, the effect of this mutation on spike protein stability and flexibility was studied using DynaMut prediction that uses Normal Mode Analysis (NMA). the SARS-CoV-2 spike ectodomain structure (open state) (PDB ID: 6VYB) was downloaded from RCSB PDB, uploaded on DynaMut software (University of Melbourne, Australia) and changes in vibrational entropy; the atomic fluctuations and deformation energies were determined for mutation D614G . For atomic fluctuation and deformation energy calculations, calculations were performed by the software over first ten non-trivial modes of the molecule.

Visualization of sites of RBD mutation with respect to ACE receptor interaction face

To visualise the mutated amino acid positions in SARS-CoV2 Spike RBD region with respect to ACE receptor the crystal structure of spike RBD with ACE2 was downloaded

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 7 September 2020 doi:10.20944/preprints202004.0281.v2

from PDB (PDB ID: 6LZG). The structure was viewed using PyMoL software and the two proteins were differently coloured along with the amino acids that were found to have mutated.

Results and Discussion

Multiple sequence alignment of COVID-19 spike protein sequences from various geographical locations world-wide revealed the types of mutations occurring since the outbreak as the pandemic progressed. Although, there were many mutations dispersed at various sites in the spike protein sequence, few mutations occurred more frequently (Figure 2A) i.e. in multiple isolates/ sequences.

Spike protein mutations world-wide

Till date there are several thousands of SARS-CoV-2 sequencing data available for analyses on different databases. To understand the mutation profile of spike protein on world-wide level representing multiple continents or geographical sources, we used all the available annotated full-length sequences of the spike protein from all the major continents from NCBI virus database. 11,571 sequences in total were taken into consideration as shown in Table 1.

While there were commonly occurring random mutations, as many as 15 mutations were detected to be occurring in a country specific manner and represented to be present in different domains of the SARS-CoV-2 spike protein. Out of 15 mutations detected, 8 of them lie in S1 domain, 6 of them lie in S2 domain and 1 of them lie in signal peptide. The signal peptide residue mutation, L5F is seen mainly in the USA sequences. Mutation in the signal peptide might modulate in the signalling events in the intracellular secretory pathway although the exact relevance can only be ascertained by cell culture-based assays on virus-host interactions. In the S1 domain, 8 mutations were detected in different continents out of which 4 were detected in Asia. India alone represented three of these mutations of the in the S1 domain namely L54F, R78M and E583D (Table 1). Further analysis of other Asian isolates revealed mutations at other positions, namely T22I (Iran), G1219V (Japan), T791I (Taiwan) and A829T (Thailand). Although more sequence analyses alter current judgements made, the emergence of mutations signify the evolving nature of the spike protein albeit at various levels. Three of the mutations S477N, G485R and N501Y lying in the receptor binding domain (RBD) might alter the virus-ACE2 receptor interactions or could potentially lead to involvement of other receptors or co-receptors. Alteration of receptor interactions does not necessarily indicate effect on virulence and the hypothesis by itself requires time-testing i.e. monitoring positive selection of the emerging mutations and also wet lab validations.

S2 domain helps in the fusion process. In S2 domain, 6 mutations were detected out of which 3 were detected in Asian isolates. The S2 mutations were more prevalent as compared to S1 especially T791I which occurred in 22.6% of sequence analysed from Taiwan and A829T which neared 60.3% in those from Thailand. Other 3 mutations were detected in France (A845S), Turkey (M900I) and USA (P1263L). These mutations have a frequency of 8.2%, 41.8% and 0.7% respectively.

Overall, there were 18 major mutations noted and each one emerged at different time points on the pandemic time scale (Table 2, Figure 2A). D614G and A829T mutations were the first major mutations to emerge as early as in the month of January albeit only in few places during that time point. Table 2 enlists the accession numbers of the

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 7 September 2020 doi:10.20944/preprints202004.0281.v2

sequences where each of the major mutations primarily appeared. M900I appear to be one of the recent ones with the sample collection date being in July.

Mutation 614G

Mutation at 614 from Aspartic acid to glycine is one of the mutations that appeared in more than 70% of the sequences in all the continents except in Oceania (major geographical location is Australia) where it appeared in 66.8% of the sequences (Table 3, Figure 7). Mutation from Aspartic acid to glycine is a potentially crucial change in a protein sequence as Aspartic acid is a big negatively charged, acidic amino acid whereas Glycine is a small neutral amino acid and thus a change from D to G might lead to electrostatic alterations. In a study by Feng et al [7], it was shown that when there was a mutation from G to D in the RNA dependent RNA polymerase PB1 of H5N1 virus, the binding of PB1 to viral RNA was hampered and H5N1 was attenuated. In case of COVID 19 spike protein since, residue 614 is positioned between S1 and S2 and a D to G change might affect the binding of spike with its receptor. In this study we used *in silico* platform of DynaMut prediction (which uses Normal Mode Analysis/ NMA) to check the effect of D614G mutation on the structural stability, flexibility, atomic fluctuations (amplitude of absolute atomic motions) and deformation energies (change in local flexibilities) (Figure 8). This mutation was seen to stabilize the structure and decrease molecular flexibility. Since Aspartic acid to Glycine is a major electrostatic change, this might also lead to formation of an antibody escape mutant if position 614 is a part of immunogenic epitope. In case this position is a part of an epitope, this mutation might help the virus to escape the immune system and proliferate into a new more adapted cluster. Currently, D614G mutation has been a topic of debate and functional relevance of this mutation with respect to virulence has not been established. This assumption can be validated by incorporating this mutation in a laboratory isolate followed by testing cellular tropism and affinity for neutralizing antibodies.

Mutation at Receptor Binding Domain

Tables 4-6 and Figure 3 show a compilation of all the RBD related mutations. We have shown this analysis separately as the receptor binding domain has been the focus of drug targeting and vaccine development because of the role in receptor binding. We observed that RBD has seen three major mutations showing percentage varying from 1%-4% of the total sequences of each geographical location. All these mutations were from Oceania (Table 4). In one of our previous studies published as a pre-print we had analysed mainly North American sequences as the study was performed at an earlier stage of the current pandemic and the availability of sequence was a limitation. In case of North America, we previously noted three mutations in the RBD domain, A348T, G476S and V483A. However, the frequency of mutations these are currently <1%, hence these mutations are not propagating further or propagating at a very slow pace. Similarly, in the current analyses which involved multiple continents, nearly 25 minor mutations were observed in the RBD region (Table 6) which have very low frequency of occurrence in the total sequence studied. Some of these mutations like N439K was detected in samples collected in the month of July, 2020 and thus more sequencing results that eventually gets deposited over next few months might shed light on the possibility of selection of such mutations and role in virus evolution.

To check the molecular location of the mutation sites of RBD with respect to the ACE2 receptor binding interface, we checked eleven of the RBD mutation sites on the RBD-

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 7 September 2020 doi:10.20944/preprints202004.0281.v2

ACE2 complex structure from PDB (Figure 4-6). Most of the major RBD mutations (except L518I) lie in the receptor binding motif (RBM) of the spike protein, the residues of which directly interact with the ACE2 receptor. As per the crystal structure used, amino acids of SARS-CoV2 RBD region observed to be interacting with ACE2 receptor were A475, Y489, F486, N487, E484, Y453, K417, Y449, G496, Q498, G446, G502 and T500. Mutation site 484 with respect to the E484Q mutation in the Indian isolates and other mutations localize around the ACE2 interacting amino acids. This imply possible effect of these mutations on receptor attachment.

Landscape of spike protein mutations in Indian patients

To correlate our basic overall understanding of the mutation hot-spots in the spike protein from other parts of the world with Indian variations, we tried to study almost all available translatable good quality sequences with respect to the spike protein from the GISAID database. Such sequences represented multiple Indian states and widely distributed sample collection time points. Multiple new mutations were detected in the spike glycoprotein sequence of the Indian isolates that were not detected in other countries (Table 7, Figure 2B-C). One particular mutation V367F lying in the RBD region of the protein was initially detected in a Chinese isolate in the month of January (Table 6) and our analysis detected this mutation only in Indian isolates collected in the month of May, 2020 with a frequency of 0.12%. This mutation was not detected in any other countries as per sequences analysed thus suggesting either minor independent mutation that might have emerged in India after the outbreak. Since the international flights were restricted inflow of infected individuals carrying this mutation from China is less likely. Since the percentage of individuals having this mutation was less and that the mutation emerged in the month of May, not detected in later months, it might not have spread very efficiently as of sequence information available currently. However, since the mutation lies in the RBD region, monitoring possible future expansion and selection of such mutations might add on to the therapeutics field where RBD is been targeted.

The major mutations that were detected in the RBD region of the spike protein in other countries were not detected in the Indian isolates except G476S. Rather, multiple independent new mutations emerged in the spike protein of Indian SARS-CoV-2 isolates as highlighted in Table 7. The new mutations which appeared in the samples collected in the months of June and July, 2020, were distributed in both the S1 and S2 domains out of which some were also from the RBD region of the protein.

Recent studies have identified two protease cleavage sites in the Spike glycoprotein, S1/S2 cleavage site (681st -684th) and S2' cleavage site (811th-815th). These two sites have been shown to be important for the proteolytic processing of the protein which increases its efficiency to interact with the host cell receptors and mediate cellular transduction process [4-6]. Our mutational analysis reveals multiple mutations near the S1/S2 cleavage site (A672V, Q675R, Q675H, Q677H, Q677R, N679T, S698L) and one in the S2' cleavage site (P812L) that emerged recently in the Indian isolates. These mutations might influence the efficacy of activity of the enzymes with respect to the cleavage sites although this hypothesis is subject to experimental validations. Nonetheless, considering importance of these proteolytic cleavage events in the virus entry process, mutations might alter the tropism and transduction mechanism if such mutations succeed to get selected in the virus evolutionary process.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 7 September 2020 doi:10.20944/preprints202004.0281.v2

The data presented here is based on the currently analyzed sequences. Further sequencing and mutation analyses would shed more light on the nature of this virus and might alternatively influence the mutational profile and inference drawn.

Continents (Total number of sequences analysed)	Counties (Number of sequences used)	Specific/Exclusive mutations observed in different country specific spike protein sequences (Reference sequence: YP_009724390.1)	Percentage of sequences showing the mutation
Asia (1017)	1. Bahrain (10) 2. Bangladesh(171) 3. China (121) 4. Georgia (9) 5. India(455)	A) L54F(372L, 83F) B) R78M(442R,13M) C) E583D (441E,14D)	18.2% 2.9% 3.1%
	6. Iran (6) 7. Japan (68) 8. Kazakhstan (4) 9. Malaysia(4) 10. Pakistan (3) 11. Saudi Arabia (58) 12. South Korea(6) 13. Sri Lanka (4) 14. Taiwan (31) 15. Thailand (63) 16. Viet Nam (2) 17. Israel (2)	A) T22I (3T,3I) B) G1219V (64G,4V) A) T791I (24T,7I) A) A829T(25A,38T)	50% 5.9% 22.6% 60.3%
Africa (103)	1. Egypt (83) 2. Morocco (10) 3. Tunisia (8) 4. Kenya (2)	A) Q677H(74Q,8H,1X)	9.6%
Europe (441)	1. Czech Republic(23) 2. France (85) 3. Germany (79) 4. Greece (98) 5. Italy (14) 6. Poland (31) 7. Russia (8) 8. Serbia (11) 9. Spain (37) 10. Turkey (55)	A) A845S (78A,7S) A) M900I (31M,23I,1X)	8.2% 41.8%
Oceania (1426)	1. Australia (1423)	A) S477N(1084S,59N,280X) B) G485R(1086G,31R,306X) C) N501Y (1096N,14Y,313X)	4.1% 2.2% 1.0%

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 7 September 2020 doi:10.20944/preprints202004.0281.v2

	2. Guam (3)		
North America (8564)	1. USA (8544) 2. Puerto Rico (12) 3. Jamaica (8)	A) L5F (8434L,97F,13X) B) P1263L (8476P,59L,9X)	1.1%Oce 0.7%
South America (20)	1. Brazil (6) 2. Chile (11) 3. Colombia (3)		

Table 1: Mutational analysis of Spike protein of SARS-CoV2 isolated from different continents. Mutations in the S1 domain (residues 13-685) and S2 domain (686-1273) are highlighted in red and blue respectively. Mutations lying in the Receptor binding domain (RBD) is highlighted in green and are also part of the S1 subunit..

Mutations	Date of it first occurrence in the year 2020 (day-month-year)	Accession ID of the sequence where the mutation was detected (NCBI)
1. D614G	00-01-2020	QJW69187.1
2. A829T	23-01-2020	QJQ84712.1
3. G485R	06-02-2020	QJG65949.1
4. T791I	26-02-2020	QJD20632.1
5. V483A	29-02-2020	QKS90227.1
6. A845S	00-03-2020	QJT72086.1
7. G476S	02-03-2020	QKS90479.1
8. L5F	09-03-2020	QIZ15981.1
9. P1263L	14-03-2020	QLC92372.1
10. S477N	19-03-2020	QLG76889.1
11. T22I	26-03-2020	QKV67308.1
12. R78M	11-04-2020	QJY40517.1
13. E583D	11-04-2020	QJY40565.1
14. G1219V	00-05-2020	BCI50555.1
15. L54F	04-05-2020	QJX44562.1
16. Q677H	02-06-2020	QKX65100.1
17. N501Y	03-06-2020	QLG75761.1
18. M900I	15-07-2020	QLK97783.1

Table 2: Major mutations detected in SARS-CoV2 Spike protein and its probable period of primary appearance (based on the collection date of the sample). The NCBI accession ID has been mentioned for the sequence where the mutation was detected.

Continents (Total number of sequences used)	Countries (Total number of sequences used)	Percentage of sequences showing D/G at the 614th position
Asia (1017)	1. Bahrain (10) 2. Bangladesh(171) 3. China (121) 4. Georgia (9) 5. India(455) 6. Iran (6)	9G,1D 168G,3D 3G, 118D 6G, 3D 414G,41D 6D

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 7 September 2020 doi:10.20944/preprints202004.0281.v2

Total: 1017	7. Japan (68) 8. Kazakhstan (4) 9. Malaysia(4) 10. Pakistan (3) 11. Saudi Arabia (58) 12. South Korea(6) 13. Sri Lanka (4) 14. Taiwan (31) 15. Thailand (63) 16. Viet Nam (2) 17. Israel (2)	59G,9D 2G, 2D 4D 3D 56G,2D 6D 2G, 2D 13G, 18D 11G, 52D 2D 1G,1D **744G (73.2%)** **273D (26.8%)**
Africa (103) Total: 103	1. Egypt (83) 2. Morocco (10) 3. Tunisia (8) 4. Kenya (2)	79G,2D,2X 10G 5G,3D 1G,1D **95G (92.2%)** **6D (5.8%)**
Europe (441) Total: 441	1. Czech Republic(23) 2. France (85) 3. Germany (79) 4. Greece (98) 5. Italy (14) 6. Poland (31) 7. Russia (8) 8. Serbia (11) 9. Spain (37) 10. Turkey (55)	23G 65G,20D 55G, 22D,2X 77G,21D 11G,3D 30G,1D 7G,1D 11G 16G,21D 54G,1D **349G (79.1%)** **90D (20.4%)**
Oceania (1426) Total: 1426	1. Australia (1423) **2.** Guam (3)	953G,459D,11X 3D **953G (66.8%)** **462D (32.4%)**
North America (8564) Total: 8564	1. USA (8544) 2. Puerto Rico (12) 3. Jamaica (8)	6313G,2200D,31X 9G,3D 3G ,5D **6325G (73.9%)** **2208D (25.8%)**
South America (20) Total: 20	1. Brazil (6) 2. Chile (11) **3.** Colombia (3)	4G, 2D 9G,2D 1G, 2D **14G (70%)** **6D (30%)**

Table 3: Mutational analysis of 614th a.a. of SARS-CoV2 Spike protein:

Major Mutations in Receptor binding region (RBD) of Spike	Continents (Total Number of sequences)	Percentage of sequences showing the mutation
S477N	1. Asia (1017)	1015S (99.8%),1X,1G

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 7 September 2020 doi:10.20944/preprints202004.0281.v2

	2. Africa (103) 3. Europe (441) 4. North America (8564) 5. South America (20) 6. Oceania (1426)	103S (100%) 422S(95.7%),19X 8493S (99.2%),3N (0.04%),68X 20S (100%) 1087S (76.2%),59N (4.1%),280X
G485R	1. Asia (1017) 2. Africa (103) 3. Europe (441) 4. North America (8564) 5. South America (20) 6. Oceania (1426)	1016G (99.9%), 1R(0.01%) 103G (100%) 423G (95.9%),18X 8514G (99,6%),50X 20G (100%) 1086G (76.2%),31R (2.2%),306X
N501Y	1. Asia (1017) 2. Africa (103) 3. Europe (441) 4. North America (8564) 5. South America (20) 6. Oceania (1426)	1017N (100%) 103N (100%) 423N (95.9%),18X 8499N (99.5%),65X 20N (100%) 1096N (76.9%),14Y (1.0%),313X

Table 4: Mutational analysis of Receptor binding domain (RBD) of SARS-CoV2 Spike protein:

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 7 September 2020 doi:10.20944/preprints202004.0281.v2

Mutations in the RBD domain (Major) (NCBI ID: YP_009724390.1)	NCBI Accession ID of the Spike protein sequence where the mutation is detected	Collection date of the patient samples used for sequencing (year-month-day) (as per availability of data)
1. G476S	1. QJC20487.1 2. QKS90479.1 3. QKS90179.1 4. QJD48075.1 5. QJA16664.1 6. QIS30625.1 7. QIS30425.1 8. QIZ13371.1 9. QIQ50152.1 10. QIQ49882.1	1. 2020-03-27 2. 2020-03-02 3. 2020-03-02 4. 2020-03-10 5. 2020-03-25 6. 2020-03-15 7. 2020-03-14 8. 2020-03-21 9. 2020-03-14 10. 2020-03-13
2. S477N **A. Australia**	1. QLG75977.1 2. QLG76673.1 3. QLG75581.1 4. QLG75569.1 5. QLG76949.1 6. QLG76409.1 7. QLG76925.1 8. QLG76901.1 9. QLG76889.1 10. QLG76877.1 11. QLG76841.1 12. QLG76829.1 13. QLG76781.1 14. QLG76769.1 15. QLG76757.1 16. QLG76733.1 17. QLG76721.1 18. QLG76661.1 19. QLG76649.1 20. QLG76637.1 21. QLG76625.1 22. QLG76613.1 23. QLG76601.1 24. QLG76493.1 25. QLG76481.1 26. QLG76445.1 27. QLG76433.1 28. QLG76421.1 29. QLG76349.1 30. QLG76325.1 31. QLG76301.1 32. QLG76229.1 33. QLG76217.1 34. QLG76205.1 35. QLG76193.1 36. QLG76169.1 37. QLG76061.1	1. 2020-06-11 2. 2020-06-13 3. 2020-05-30 4. 2020-05-30 5. 2020-06-18 6. 2020-06-19 7. 2020 8. 2020-06-19 9. 2020-03-19 10. 2020-06-19 11. 2020-06-15 12. 2020-06-16 13. 2020-06-15 14. 2020-06-15 15. 2020-06-14 16. 2020-06-13 17. 2020-06-14 18. 2020-06-13 19. 2020-06-13 20. 2020-06-12 21. 2020-06-10 22. 2020-06-09 23. 2020-06-08 24. 2020-06-20 25. 2020-06-19 26. 2020-06-19 27. 2020-06-19 28. 2020-06-19 29. 2020-06-18 30. 2020-06-17 31. 2020-06-15 32. 2020-06-15 33. 2020-06-15 34. 2020-06-15 35. 2020-06-15 36. 2020-06-15 37. 2020-06-13

	38. QLG76001.1 39. QLG75989.1 40. QLG75881.1 41. QLG75797.1 42. QLG75713.1 43. QLG75557.1 44. QLG76865.1 45. QLG76073.1 46. QLG75593.1 47. QKR86413.1 48. QKR85873.1 49. QKR86845.1 50. QKR86809.1 51. QLG75641.1 52. QLG75605.1 53. QLG75449.1 54. QLG75437.1 55. QLG75389.1 56. QKR86857.1 57. QKR86821.1 58. QKR87133.1 59. QKR86425.1	38. 2020-06-12 39. 2020-06-11 40. 2020-06-09 41. 2020-06-05 42. 2020-05-27 43. 2020-05-30 44. 2020-06-17 45. 2020-06-12 46. 2020-05-29 47. 2020-05-16 48. 2020-05-13 49. 2020-05-19 50. 2020-05-19 51. 2020-05-30 52. 2020-06-01 53. 2020-05-28 54. 2020-05-29 55. 2020-05-28 56. 2020-05-19 57. 2020-05-19 58. 2020-05-26 59. 2020-05-16
B. USA	1. QLC46901.1 2. QJZ32245.1 3. QKG90206.1	1. 2020-04-21 2. 2020-04-12 3. 2020-03-28
3. V483A	1. QIZ13131.1 2. QKV06535.1 3. QKU54117.1 4. QKU54081.1 5. QKU53649.1 6. QKS91079.1 7. QKS90659.1 8. QKS90239.1 9. QKS90227.1 10. QKS90095.1 11. QKS90059.1 12. QKS90023.1 13. QKS89879.1 14. QJD49095.1 15. QJD48987.1 16. QIZ14939.1 17. QIU81549.1 18. QIU81177.1 19. QIS60954.1 20. QIS60882.1 21. QIS60774.1 22. QIS30565.1 23. QIS30165.1 24. QJS57147.1 QKU53733.1	1. 2020-03-23 2. 2020-03-24 3. 2020-03-12 4. 2020-03-12 5. 2020-03-17 6. 2020-03-02 7. 2020-03-02 8. 2020-03-01 9. 2020-02-29 10. 2020-03-05 11. 2020-03-02 12. 2020-03-02 13. 2020-02-29 14. 2020-03-05 15. 2020-03-09 16. 2020-03-16 17. 2020-03-19 18. 2020-03-16 19. 2020-03-16 20. 2020-03-16 21. 2020-03-16 22. 2020-03-15 23. 2020-03-16 24. 2020-04-05 2020-03-17
4. G485R **A. Australia**	1. QKR85573.1 2. QKR86245.1 3. QKR86185.1 4. QKR85609.1	2. 2020-05-05 3. 2020-05-12 4. 2020-05-12 5. 2020-05-05

	5. QKR86449.1 6. QLG75965.1 7. QKR86221.1 8. QLG75509.1 9. QKR86785.1 10. QKR86881.1 11. QKR86233.1 12. QKR86353.1 13. QKR86197.1 14. QKR86089.1 15. QKR85825.1 16. QLG75773.1 17. QKR85945.1 18. QKR87169.1 19. QKR86389.1 20. QLG75461.1 21. QKR86497.1 22. QKR86365.1 23. QKR86341.1 24. QKR86209.1 25. QKR85909.1 26. QKR85597.1 27. QKR85513.1 28. QKR85381.1 29. QKR85765.1 30. QLG75497.1 31. QKR85453.1	6. 2020-05-17 7. 2020-06-10 8. 2020-05-12 9. 2020-05-28 10. 2020-05-18 11. 2020-05-17 12. 2020-05-12 13. 2020-05-09 14. 2020-05-12 15. 2020-05-09 16. 2020-05-05 17. 2020-06-03 18. 2020-05-05 19. 2020-05-24 20. 2020-05-14 21. 2020-05-27 22. 2020-05-17 23. 2020-05-12 24. 2020-05-12 25. 2020-05-12 26. 2020-05-14 27. 2020 28. 2020-05-05 29. 2020-05-05 30. 2020-05-09 31. 2020-05-27 32. 2020
B. China	1. QJG65949.1	1. 2020-02-06
5. N501Y	1. QLG76745.1 2. QLG76805.1 3. QLG76793.1 4. QLG76709.1 5. QLG76697.1 6. QLG76685.1 7. QLG76469.1 8. QLG76397.1 9. QLG76277.1 10. QLG76181.1 11. QLG76097.1 12. QLG76085.1 13. QLG75761.1 14. QLG76817.1	1. 2020-06-16 2. 2020-06-18 3. 2020-06-17 4. 2020-06-12 5. 2020-06-12 6. 2020-06-15 7. 2020-06-18 8. 2020-06-19 9. 2020-06-17 10. 2020-06-15 11. 2020-06-14 12. 2020-06-12 13. 2020-06-03 14. 2020-06-17

Table 5: Major mutations in receptor binding domain (RBD) of the spike protein based on the relative number of times the mutation appeared in each country (as shown in table 4) and the collection dates of each sample. 3 mutations that were reported earlier have also been included here although the percentage of occurrence is low.

RBD Mutations (Minor)	Country of origin (percentage of sequences showing the mutation)	NCBI Accession ID of the sequence showing the mutation	Collection date of the patient samples where mutation is detected (year-month-day)
1. T323I	USA (8396T, 3I, 145X) (0.04%)	a. QLK98059.1 b. QKV06079.1 c. QJD25193.1	2020-04-10 2020-03-27 2020-03-21
2. P330S	USA (8515P, 2S, 27X) (0.02%)	a. QKV37091.1 b. QKV37031.1	2020-04-17 2020-04-17
3. A348S	USA (8518A,1S,25X) (0.01%)	a. QKU32465.1	2020-05-02
4. A348T	USA (8538A 1T ,25X) (0.01%)	a. QIS30335.1	2020-03-16
5. V367F	China (120V,1F)(0.8%)	a. QJC20993.1	2020-01-22
6. C379F	France (84C,1F) (1.2%)	a. QJT72386.1	2020-03
7. V382L	USA (8509V, 1L, 34X) (0.01%)	a. QLG98191.1	2020-06-03
8. V382E	France (84V,1E) (1.2%)	a. QJT72806.1	2020-03
9. P384L	USA (8504P,2L,39X) (0.02%)	a. QKO24139.1 b. QKV08239.1	2020-05-04 2020-04-05
10. T393P	Greece (97T,1P) (1%)	a. QJS54106.1	2020-03-29
11. V395I	USA (8501V,2I, 41X) (0.02%)	a. QMI95674.1 b. QMI90807.1	2020-03-15 2020-03-15
12. R403K	A. USA (8503R, 9K,32X) (0.1%) B. Australia (1420R,1K,2X) (0.07%)	a. QLA47931.1 b. QLA47919.1 c. QLA47907.1 d. QLA47895.1 e. QLA47883.1 f. QLA47799.1 g. QLA47787.1 h. QLA47775.1 i. QLL36145.1 a. QKR87157.1	2020-05 2020-05 2020-05 2020-05 2020-05 2020-05 2020-05 2020-05 2020-04 2020-05-27
13. R408I	A. Egypt (79R,2I,2X)	a. QKT21014.1 b. QJY78056.1	2020-06-02 2020-05-02

	(2.4%) B. India (454R,1I) (0.2%)	a. QHS34546.1	2020-01-27
14. A411S	USA (San Diego, California) (8497A,6S,41X) (0.07%)	a. QMI96187.1 b. QMI93578.1 c. QMI93554.1 d. QMI93410.1 e. QMI93230.1 f. QMI93158.1	2020-04-29 2020-05-03 2020-05-03 2020-05-06 2020-05-05 2020-05-06
15. N439K	USA (8472N,1K,71X) (0.01%)	a. QLM06051.1	2020-07-07
16. I468T	Australia (1146I,1T,276) (0.07%)	a. QJR86937.1	2020-03-18
17. E471Q	India (454E,1Q) (0.2%)	a. QJX44562.1	2020-05-04
18. E484K	USA (8490E,1K,53X) (0.01%)	a. QLI51445.1	2020-05-22
19. S494P	USA (8474S,3P,67X) (0.04%)	a. QJF75359.1 b. QJF75347.1 c. QJF75335.1	2020-03-20 2020-03-20 2020-03-20
20. L518I	Bangladesh (168L,3I) (1.8%)	a. QLE10873.1 b. QLE10825.1 c. QLE10753.1	2020-06-24 2020-06-16 2020-06-18
21. H519Q	A. Australia (1031H,2Q,390X) (0.14%) B. Malaysia (3H,1Q) (25%)	a. QKR85081.1 b. QKR86401.1 a. QJD23249.1	2020-04-27 2020-05-17 2020-03-20
22. A520S	USA (8473A,6S,65X) (0.07%)	a. QJA41737.1 b. QIS60489.1 c. QKU53949.1 d. QKU53925.1 e. QKU53817.1 f. QJU70545.1	2020-03-19 2020-03-13 2020-03-19 2020-03-13 2020-03-19 2020-04-26
23. A522V	USA (8490A,1V,53X) (0.01%)	a. QKU28906.1	2020-04-12
24. A522S	Australia (1052A,1S,370X) (0.01%)	a. QJR91593.1	2020-03-27
25. K529E	USA (8532K,1E,11X) (0.01%)	a. QJC20043.1	2020-03-26

Table 6: Tabulation of all the minor mutations that are detected in the Receptor binding domain (RBD) of the Spike protein.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 7 September 2020 doi:10.20944/preprints202004.0281.v2

Mutations in SARS-CoV2 Surface glycoprotein in Indian isolates	Percentage of sequences showing the mutation (Total number of sequences =1648)	GISAID Accession ID of the respective sequence showing the mutation	Collection date (year-month-day)
1. L5F	1639L 9F (0.55%)	1. EPI_ISL_435108 2. EPI_ISL_463033 3. EPI_ISL_455756 4. EPI_ISL_451150 5. EPI_ISL_451151 6. EPI_ISL_458103 7. EPI_ISL_495290 8. EPI_ISL_495291 9. EPI_ISL_495297	1. 2020-03-15 2. 2020-05-10 3. 2020-05-07 4. 2020-05-06 5. 2020-05-05 6. 2020-05-05 7. 2020-05-13 8. 2020-05-13 9. 2020-05-15
2. S12F	1645S 1F (0.06%) 2X	EPI_ISL_508290	2020-06-21
3. T22I	1639T 7I (0.42%) 2X	1. EPI_ISL_508169 2. EPI_ISL_454558 3. EPI_ISL_479515 4. EPI_ISL_479516 5. EPI_ISL_452195 6. EPI_ISL_452196 7. EPI_ISL_496587	1. 2020-05-21 2. 2020-04-13 3. 2020-05-05 4. 2020-05-05 5. 2020-04-19 6. 2020-04-19 7. 2020-05-08
4. T29I	1630T 15I (0.91%) 3X	1. EPI_ISL_508211 2. EPI_ISL_508217 3. EPI_ISL_508225 4. EPI_ISL_508228 5. EPI_ISL_508233 6. EPI_ISL_508235 7. EPI_ISL_508272 8. EPI_ISL_508273 9. EPI_ISL_508281 10. EPI_ISL_508415 11. EPI_ISL_508224 12. EPI_ISL_508263 13. EPI_ISL_508265 14. EPI_ISL_508285 15. EPI_ISL_495203	1. 2020-06-17 2. 2020-06-17 3. 2020-06-20 4. 2020-06-20 5. 2020-06-20 6. 2020-06-20 7. 2020-06-18 8. 2020-06-18 9. 2020-06-18 10. 2020-06-16 11. 2020-06-17 12. 2020-06-18 13. 2020-06-18 14. 2020-06-17 15. 2020-06-09
5. L54F	1598L 49F (3%) 1X	1. EPI_ISL_508267 2. EPI_ISL_481148 3. EPI_ISL_512069 4. EPI_ISL_447046 5. EPI_ISL_476880 6. EPI_ISL_447032 7. EPI_ISL_447033 8. EPI_ISL_451153 9. EPI_ISL_458100 10. EPI_ISL_458101	1. 2020-06-18 2. 2020-05-26 3. 2020-07-27 4. 2020-05-04 5. 2020-06-12 6. 2020-05-03 7. 2020-05-03 8. 2020-05-10 9. 2020-05-24 10. 2020-05-24

		11. EPI_ISL_458112 12. EPI_ISL_458113 13. EPI_ISL_475027 14. EPI_ISL_475029 15. EPI_ISL_475030 16. EPI_ISL_475031 17. EPI_ISL_475033 18. EPI_ISL_475034 19. EPI_ISL_475037 20. EPI_ISL_475039 21. EPI_ISL_475046 22. EPI_ISL_475047 23. EPI_ISL_475048 24. EPI_ISL_475050 25. EPI_ISL_475051 26. EPI_ISL_475052 27. EPI_ISL_475053 28. EPI_ISL_475054 29. EPI_ISL_475055 30. EPI_ISL_475056 31. EPI_ISL_475057 32. EPI_ISL_476855 33. EPI_ISL_476856 34. EPI_ISL_476857 35. EPI_ISL_476858 36. EPI_ISL_476859 37. EPI_ISL_476860 38. EPI_ISL_476861 39. EPI_ISL_476862 40. EPI_ISL_476868 41. EPI_ISL_476870 42. EPI_ISL_476871 43. EPI_ISL_476872 44. EPI_ISL_476873 45. EPI_ISL_476876 46. EPI_ISL_476882 47. EPI_ISL_495021 48. EPI_ISL_495022 49. EPI_ISL_495033	11. 2020-05-18 12. 2020-05-18 13. 2020-06-07 14. 2020-06-07 15. 2020-06-11 16. 2020-06-11 17. 2020-06-11 18. 2020-06-11 19. 2020-06-11 20. 2020-06-11 21. 2020-06-11 22. 2020-06-05 23. 2020-06-05 24. 2020-06-08 25. 2020-06-08 26. 2020-06-08 27. 2020-06-08 28. 2020-06-08 29. 2020-06-08 30. 2020-06-08 31. 2020-06-08 32. 2020-06-08 33. 2020-06-08 34. 2020-06-08 35. 2020-06-08 36. 2020-06-08 37. 2020-06-08 38. 2020-06-08 39. 2020-06-08 40. 2020-06-12 41. 2020-06-11 42. 2020-06-11 43. 2020-06-11 44. 2020-06-11 45. 2020-06-12 46. 2020-06-12 47. 2020-06-30 48. 2020-06-30 49. 2020-06-15
6. K77M	1627K 15M (0.91%) 6X	1. EPI_ISL_436419 2. EPI_ISL_486389 3. EPI_ISL_486408 4. EPI_ISL_486409 5. EPI_ISL_481127 6. EPI_ISL_436450 7. EPI_ISL_447571 8. EPI_ISL_447556 9. EPI_ISL_447566 10. EPI_ISL_447850 11. EPI_ISL_447856 12. EPI_ISL_447858	1. 2020-03-31 2. 2020-05-10 3. 2020-05-10 4. 2020-05-10 5. 2020-05-22 6. 2020-04-13 7. 2020-04-01 8. 2020-03-30 9. 2020-04-01 10. 2020-04-01 11. 2020-04-02 12. 2020-04-06

		13. EPI_ISL_450330 14. EPI_ISL_458072 15. EPI_ISL_471585	13. 2020-04-02 14. 2020-04-08 15. 2020-04-08
7. R78M	1643R 5M (0.30%)	1. EPI_ISL_512075 2. EPI_ISL_512077 3. EPI_ISL_514435 4. EPI_ISL_514436 5. EPI_ISL_514437	1. 2020-07-09 2. 2020-07-09 3. 2020-07-09 4. 2020-07-09 5. 2020-07-09
8. D80Y	1645D 3Y (0.18%)	1. EPI_ISL_481163 2. EPI_ISL_481164 3. EPI_ISL_481165	1. 2020-06-04 2. 2020-06-04 3. 2020-06-04
9. D88H	1643D 1H (0.06%) 4X	EPI_ISL_486409	2020-05-10
10. E96D	1645E 1D (0.06%) 2X	EPI_ISL_486836	2020
11. S98F	1647S 1F (0.06%)	EPI_ISL_477208	2020-06-01
12. D138Y	1639D 9Y (0.55%)	1. EPI_ISL_477233 2. EPI_ISL_508200 3. EPI_ISL_508231 4. EPI_ISL_508269 5. EPI_ISL_508270 6. EPI_ISL_508271 7. EPI_ISL_482641 8. EPI_ISL_479554 9. EPI_ISL_508224	1. 2020-06-16 2. 2020-06-09 3. 2020-06-20 4. 2020-06-18 5. 2020-06-18 6. 2020-06-18 7. 2020-05-27 8. 2020-06-15 9. 2020-06-17
13. D178N	1646D 1N (0.06%) 1X	EPI_ISL_477239	2020-06-17
14. G181A	1643G 4A (0.24%) 1X	1. EPI_ISL_508211 2. EPI_ISL_508263 3. EPI_ISL_508265 4. EPI_ISL_508285	1. 2020-06-17 2. 2020-06-18 3. 2020-06-18 4. 2020-06-17
15. V213A	1638V 9A (0.55%) 1X	1. EPI_ISL_471590 2. EPI_ISL_471591 3. EPI_ISL_471592 4. EPI_ISL_471593 5. EPI_ISL_471596 6. EPI_ISL_471605 7. EPI_ISL_471606 8. EPI_ISL_471607 9. EPI_ISL_471611	1. 2020-05-25 2. 2020-05-25 3. 2020-05-25 4. 2020-05-25 5. 2020-05-25 6. 2020-05-25 7. 2020-05-25 8. 2020-05-25 9. 2020-05-25
16. A243S **A243V**	1642A 4S (0.24%) 1V (0.06%) 1X	1. EPI_ISL_436449 (S) 2. EPI_ISL_496569 (S) 3. EPI_ISL_496535 (S) 4. EPI_ISL_436426 (S) EPI_ISL_477262 (V)	1. 2020-04-12 2. 2020-05-21 3. 2020-05-21 4. 2020-04-05 2020-05-20

17. G261S	1643G 2S (0.12%) 3X	1. EPI_ISL_455666 2. EPI_ISL_463070	1. 2020-05-01 2. 2020-05-14
18. T274I	1644T 1I (0.06%) 3X	EPI_ISL_436416	2020-03-31
19. T286I	1646T 1I (0.06%) 1X	EPI_ISL_508182	2020-06-08
20. T323I	1644T 2I (0.12%) 2X	1. EPI_ISL_455655 2. EPI_ISL_455663	1. 2020-04-30 2. 2020-05-02
21. A344S	1647A 1S (0.06%)	EPI_ISL_477247	2020
22. V367F	1646V 2F (0.12%)	1. EPI_ISL_463064 2. EPI_ISL_463065	1. 2020-05-16 2. 2020-05-14
23. I402L	1647I 1L (0.06%)	EPI_ISL_455773	2020-04-04
24. R408I	1645R 1I (0.06%) 2X	EPI_ISL_413522	2020-01-27
25. T470N	1647T 1N (0.06%)	EPI_ISL_482563	2020-05-19
26. E471D **E471Q**	1646E 1D (0.06%) 1Q (0.06%)	EPI_ISL_508226 (D) EPI_ISL_447046 (Q)	2020-06-20 2020-05-04
27. G476S	1647G 1S (0.06%)	EPI_ISL_479493	2020-04-15
28. N481D	1646N 1D (0.06%) 1X	EPI_ISL_477238	2020-06-16
29. G482S	1644G 3S (0.18%) 1X	1. EPI_ISL_495222 2. EPI_ISL_495273 3. EPI_ISL_495264	1. 2020-06-11 2. 2020-06-13 3. 2020-06-13
30. E484Q	1645E 3Q (0.18%)	1. EPI_ISL_454530 2. EPI_ISL_495014 3. EPI_ISL_495015	1. 2020-03-22 2. 2020-06-18 3. 2020-06-18
31. S494P	1647S 1P (0.06%)	EPI_ISL_436419	2020-03-31
32. A520S	1627A 8S (0.49%) 13X	1. EPI_ISL_458045 2. EPI_ISL_458046 3. EPI_ISL_458047 4. EPI_ISL_458048 5. EPI_ISL_458049 6. EPI_ISL_458050 7. EPI_ISL_495161 8. EPI_ISL_495162	1. 2020-05-11 2. 2020-05-11 3. 2020-05-11 4. 2020-05-11 5. 2020-05-11 6. 2020-05-11 7. 2020-05-11 8. 2020-05-11
33. E554Q	1647E 1Q (0.06%)	EPI_ISL_508156	2020-06-01
34. K558N	1645K 2N (0.12%)	1. EPI_ISL_458035 2. EPI_ISL_452210	1. 2020-04-29 2. 2020-04-06

	1X		
35. F565L	1647F 1L (0.06%)	EPI_ISL_508320	2020-06-04
36. T572I	1642T 5I (0.30%) 1X	1. EPI_ISL_458103 2. EPI_ISL_447035 3. EPI_ISL_458090 4. EPI_ISL_458091 5. EPI_ISL_461484	1. 2020-05-05 2. 2020-05-03 3. 2020-05-24 4. 2020-05-24 5. 2020-05-27
37. E583D	1639E 8D (0.49%) 1X	1. EPI_ISL_476854 2. EPI_ISL_496576 3. EPI_ISL_512066 4. EPI_ISL_461485 5. EPI_ISL_461495 6. EPI_ISL_495271 7. EPI_ISL_467050 8. EPI_ISL_469028	1. 2020-05-11 2. 2020-05-21 3. 2020-07-27 4. 2020-05-27 5. 2020-05-27 6. 2020-06-13 7. 2020-06-03 8. 2020-06-03
38. D614G	499D (30.3%) 1122G (68.1%) 27X (1.64%)	Many	Many
39. V622F	1642V 6F (0.36%)	1. EPI_ISL_481163 2. EPI_ISL_481164 3. EPI_ISL_481165 4. EPI_ISL_481166 5. EPI_ISL_495195 6. EPI_ISL_495239	1. 2020-06-04 2. 2020-06-04 3. 2020-06-04 4. 2020-06-04 5. 2020-06-08 6. 2020-06-12
40. A653V	1646A 1X 1V (0.06%)	EPI_ISL_486837	2020
41. A672V	1647A 1V (0.06%)	EPI_ISL_508203	2020-06-08
42. Q675R	1642Q 1R (0.06%) 5H (0.30%)	EPI_ISL_508181 (R)	2020-06-02
Q675H		1. EPI_ISL_508231 (H) 2. EPI_ISL_508269 (H) 3. EPI_ISL_508270 (H) 4. EPI_ISL_508271 (H) 5. EPI_ISL_476864 (H)	1. 2020-06-20 2. 2020-06-18 3. 2020-06-18 4. 2020-06-18 5. 2020-06-11
43. Q677H	1632Q 12H (0.73%) 2R (0.12%) 2X	1. EPI_ISL_477216 (H) 2. EPI_ISL_508336 (H) 3. EPI_ISL_508419 (H) 4. EPI_ISL_452215 (H) 5. EPI_ISL_454550 (H) 6. EPI_ISL_452210 (H) 7. EPI_ISL_452197 (H) 8. EPI_ISL_508420 (H) 9. EPI_ISL_508263 (H) 10. EPI_ISL_508265 (H) 11. EPI_ISL_508285 (H) 12. EPI_ISL_476875 (H)	1. 2020-06-09 2. 2020-06-20 3. 2020-06-27 4. 2020-04-22 5. 2020-04-06 6. 2020-04-06 7. 2020-04-19 8. 2020-06-18 9. 2020-06-18 10. 2020-06-18 11. 2020-06-17 12. 2020-06-12

Q677R		1. EPI_ISL_508189 (R) 2. EPI_ISL_508190 (R)	13. 2020-05-23 14. 2020-05-23
44. N679T	1647N 1T (0.06%)	EPI_ISL_508181	2020-06-02
45. S698L	1647S 1L (0.06%)	EPI_ISL_508169	2020-05-21
46. T723I	1647T 1I (0.06%)	EPI_ISL_430466	2020-03-26
47. R765L	1646R 2L (0.12%)	1. EPI_ISL_477240 2. EPI_ISL_479744	1. 2020-06-17 2. 2020-06-17
48. P812L	1634P 4L (0.24%) 10X	1. EPI_ISL_508198 2. EPI_ISL_512068 3. EPI_ISL_512067 4. EPI_ISL_512076	1. 2020-06-09 2. 2020-07-27 3. 2020-07-27 4. 2020-07-09
49. D839Y	1645D 3Y (0.18%)	1. EPI_ISL_477222 2. EPI_ISL_486838 3. EPI_ISL_508428	1. 2020-06-11 2. 2020 3. 2020-05-17
50. A852T	1647A 1T (0.06%)	EPI_ISL_508198	2020-06-09
51. L858F	1647L 1F (0.06%)	EPI_ISL_508297	2020-06-17
52. S884F	1647S 1F (0.06%)	EPI_ISL_476894	2020-05-18
53. S929T **S929I**	1646S 1T (0.06%) 1I (0.06%)	EPI_ISL_476883 (T) EPI_ISL_508234 (I)	2020-05-13 2020-06-17
54. A930V **A930T**	1646A 1V (0.06%) 1T (0.06%)	EPI_ISL_413523 (V) EPI_ISL_455761 (T)	2020-01-31 2020-05-07
55. T941K	1647T 1K (0.06%)	EPI_ISL_455640	2020-03-27
56. S967G	1647S 1G (0.06%)	EPI_ISL_477243	2020-06-16
57. L981P	1647L 1P (0.06%)	EPI_ISL_508332	2020-05-27
58. T1027I	1647T 1I (0.06%)	EPI_ISL_454831	2020-04-29
59. S1030L	1647S 1L (0.06%)	EPI_ISL_486667	2020
60. K1045N	1646K 1N (0.06%) 1X	EPI_ISL_508330	2020-06-03
61. A1078S	1640A 7S (0.42%) 1X	1. EPI_ISL_496526 2. EPI_ISL_496531 3. EPI_ISL_496532 4. EPI_ISL_496573 5. EPI_ISL_496574 6. EPI_ISL_496575 7. EPI_ISL_496577	1. 2020-05-21 2. 2020-05-21 3. 2020-05-22 4. 2020-05-21 5. 2020-05-21 6. 2020-05-21 7. 2020-05-22
62. D1084Y	1645D 3Y (0.18%)	1. EPI_ISL_486389 2. EPI_ISL_486408	1. 2020-05-10 2. 2020-05-10

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 7 September 2020 doi:10.20944/preprints202004.0281.v2

		3. EPI_ISL_486409	3. 2020-05-10
63. P1162S	1647P 1S (0.06%)	EPI_ISL_477261	2020-05-18
64. I1179V **I1179N**	1645I 1V (0.06%) 2N (0.12%)	EPI_ISL_486836 1. EPI_ISL_454563 2. EPI_ISL_454528	2020 1. 2020-04-10 2. 2020-03-17
65. C1243F	1644C 4F (0.24%)	1. EPI_ISL_452792 2. EPI_ISL_481199 3. EPI_ISL_437445 4. EPI_ISL_437446	1. 2020-04-30 2. 2020-06-10 3. 2020-04-26 4. 2020-04-26
66. C1250F	1646C 1F (0.06%) 1X	EPI_ISL_428482	2020-04-08

Table 7: Mutational analysis of the SARS-CoV2 Surface glycoproteins from Indian isolates. Mutation highlighted in yellow is the one detected in China as well in the month of January. Mutations highlighted in green are new mutations that were detected in the patient samples collected in the month of June and July 2020. Mutations highlighted in blue are the ones lying close to the S1/S2 (furin) cleavage site (681st -684th) and S2' cleavage site (811th-815th)

Figure Legends

Figure 1: SARS-CoV-2 viral proteins

A. Schematic Representation of SARS-CoV2 virus

B. Schematic Representation of SARS-CoV-2 genome and different proteins encoded (NCBI Ref ID NC_045512.2)

C. Schematic showing structural organization of COVID 19 spike protein. Spike protein is 1273aa long with two main domains, S1 and S2. 13-685aa form the S1 domain and rest of the S protein comprises the S2 domain. The receptor binding domain (RBD) falls within S1 domain

Figure 2: Mutations of spike protein

A. The different major mutations are shown in terms of its first occurrence (i.e. with respect to the collection time) at different months of the year 2020.

B. Mutations in S1 domain of SARS-CoV-2 Spike protein in Indian Isolates. Red dots indicate mutations in RBD region. Blue dots signify mutations near the S1S2 cleavage site.

C. Mutations in S2 domain of SARS-CoV-2 Spike protein in Indian Isolates. Blue dots indicate mutations in the S2' cleavage site.

Figure 3: Positions of mutations in receptor binding domain (RBD).

Mutations in RBD region of Spike protein worldwide. Red dots signify mutations in the Receptor binding motif RBM (438-506 aa.)

Figure 4: RBD mutations.

PDB ID:6lzg has been used to indicate the position of the mutations (A-B: position 382; C-D: position 408; E-F: position 439; G-H: position 471; I-J: position 477). Structure in maroon colour indicates ACE receptor and that in green colour represents RBD. Mutation points are shown in blue.

Figure 5: RBD mutations

PDB ID:6lzg has been used to indicate the position of the mutations (A-B: position 481; C-D: position 482; E-F: position 484; G-H: position 485). Structure in maroon colour indicates ACE receptor and that in green colour represents RBD. Mutation points are shown in blue.

Figure 6: RBD mutations

PDB ID:6lzg has been used to indicate the position of the mutations (A-B: position 501; C-D: position 518). Structure in maroon colour indicates ACE receptor and that in green colour represents RBD. Mutation points are shown in blue.

Figure 7: Distribution pattern of mutation D614G across different continents

Figure 8: DynaMut prediction of effect of D614G mutation on Spike protein. A. Entropy and free energy changes as assessed by the DynaMut prediction for D614G mutation B. Inermolecular interactions in wild-type spike protein at position 614. C. Inermolecular interactions in mutant spike protein at position 614. D-E Visual analysis of atomic fluctuations. F-G. Inermolecular interactions in mutant spike protein at position 614. D-E Visual analysis of deformation energies.

Acknowledgements

CSIR is acknowledged for research support and fellowship to FB. North Bengal Medical College and Hospital is also deeply acknowledged. We thank AcSIR for academic support.

Author Contributions

FB did sequence alignments. AKB analysed biochemistry of amino acid changes and proof-read the manuscript. UR analysed the results and wrote the manuscript.

Funding

There was no funding funding for this work.

Conflict of Interest

Authors declare no conflict of interests.

References

1. Tai W, He L, Zhang X, Pu J, Voronin D, Jiang S et al. Characterization of the receptor-binding domain (RBD) of 2019 novel coronavirus: implication for development of RBD protein as a viral attachment inhibitor and vaccine. 2020.
2. Lan J, Ge J, Yu J, et al. Structure of the SARS-CoV-2 spike receptor-binding domain bound to the ACE2 receptor. Nature. 2020;581(7807):215-220. doi:10.1038/s41586-020-2180-5
3. Hoffmann M, Kleine-Weber H, Pöhlmann S. A Multibasic Cleavage Site in the Spike Protein of SARS-CoV-2 Is Essential for Infection of Human Lung Cells. Mol Cell. 2020;78(4):779-784.e5. doi:10.1016/j.molcel.2020.04.022

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 7 September 2020 doi:10.20944/preprints202004.0281.v2

4. Hoffmann M, Kleine-Weber H, Schroeder S, et al. SARS-CoV-2 Cell Entry Depends on ACE2 and TMPRSS2 and Is Blocked by a Clinically Proven Protease Inhibitor. Cell. 2020;181(2):271-280.e8. doi:10.1016/j.cell.2020.02.052
5. Walls AC, Park YJ, Tortorici MA, Wall A, McGuire AT, Veesler D. Structure, Function, and Antigenicity of the SARS-CoV-2 Spike Glycoprotein. Cell. 2020;181(2):281-292.e6. doi:10.1016/j.cell.2020.02.058
6. Bestle et al 2020. TMPRSS2 and furin are both essential for proteolytic activation and spread of SARS2 CoV-2 in human airway epithelial cells and provide promising drug targets..https://doi.org/10.1101/2020.04.15.042085.
7. Feng X, Wang Z, Shi J, Deng G, Kong H et cl. Glycine at Position 622 in PB1 Contributes to the Virulence of H5N1 Avian Influenza Virus in Mice. Journal of Virology. 2016;90(4):1872-1879.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 7 September 2020 doi:10.20944/preprints202004.0281.v2

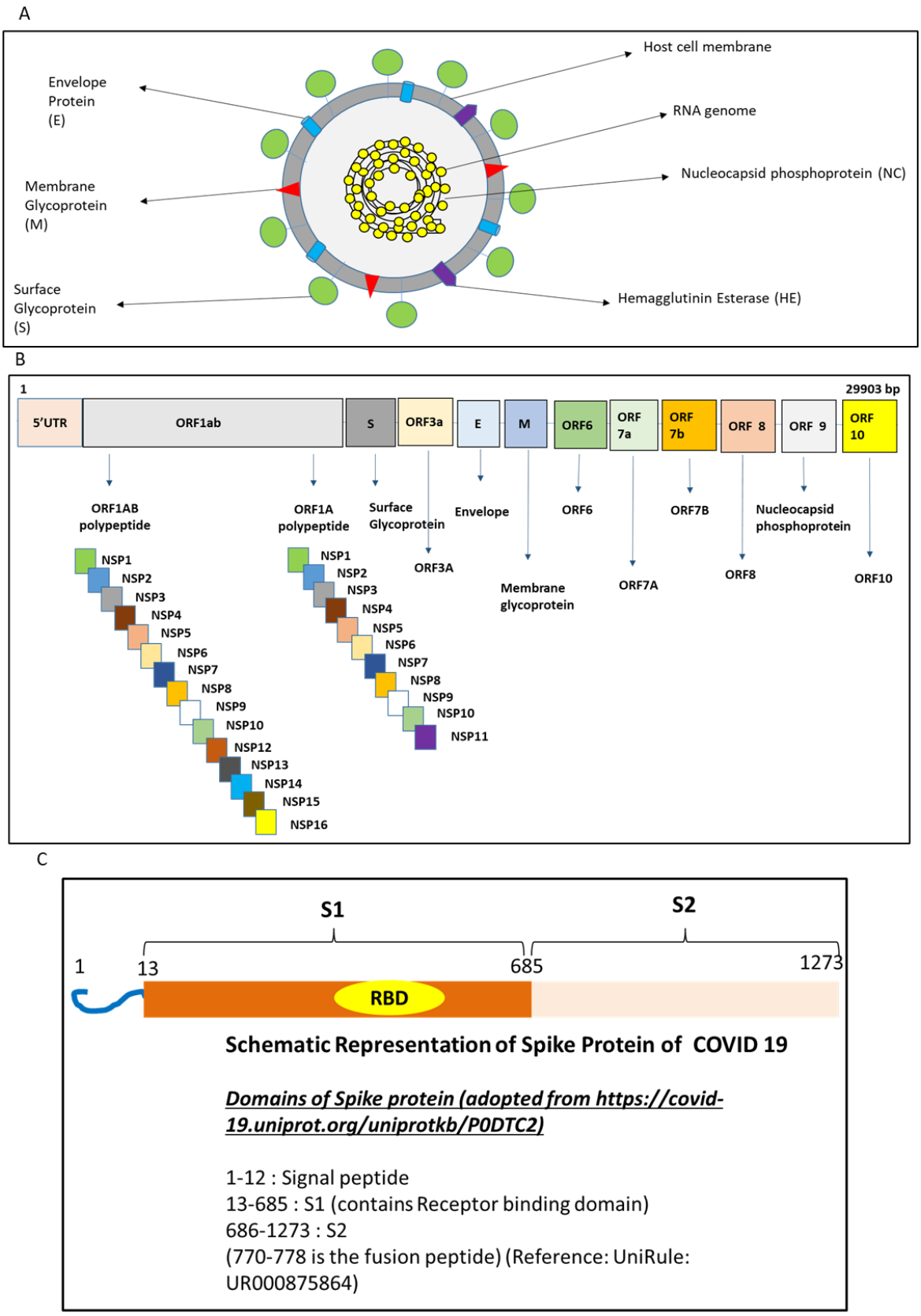

Figure 1: **SARS-CoV-2 viral proteins**

A. Schematic Representation of SARS-CoV2 virus

B. Schematic Representation of SARS-CoV-2 genome and different proteins encoded (NCBI Ref ID NC_045512.2)

C. Schematic showing structural organization of COVID 19 spike protein: Spike protein is 1273aa long with two main domains, S1 and S2. 13-685aa form the S1 domain and rest of the S protein comprises the S2 domain. The receptor binding domain (RBD) falls within S1 domain

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 7 September 2020 doi:10.20944/preprints202004.0281.v2

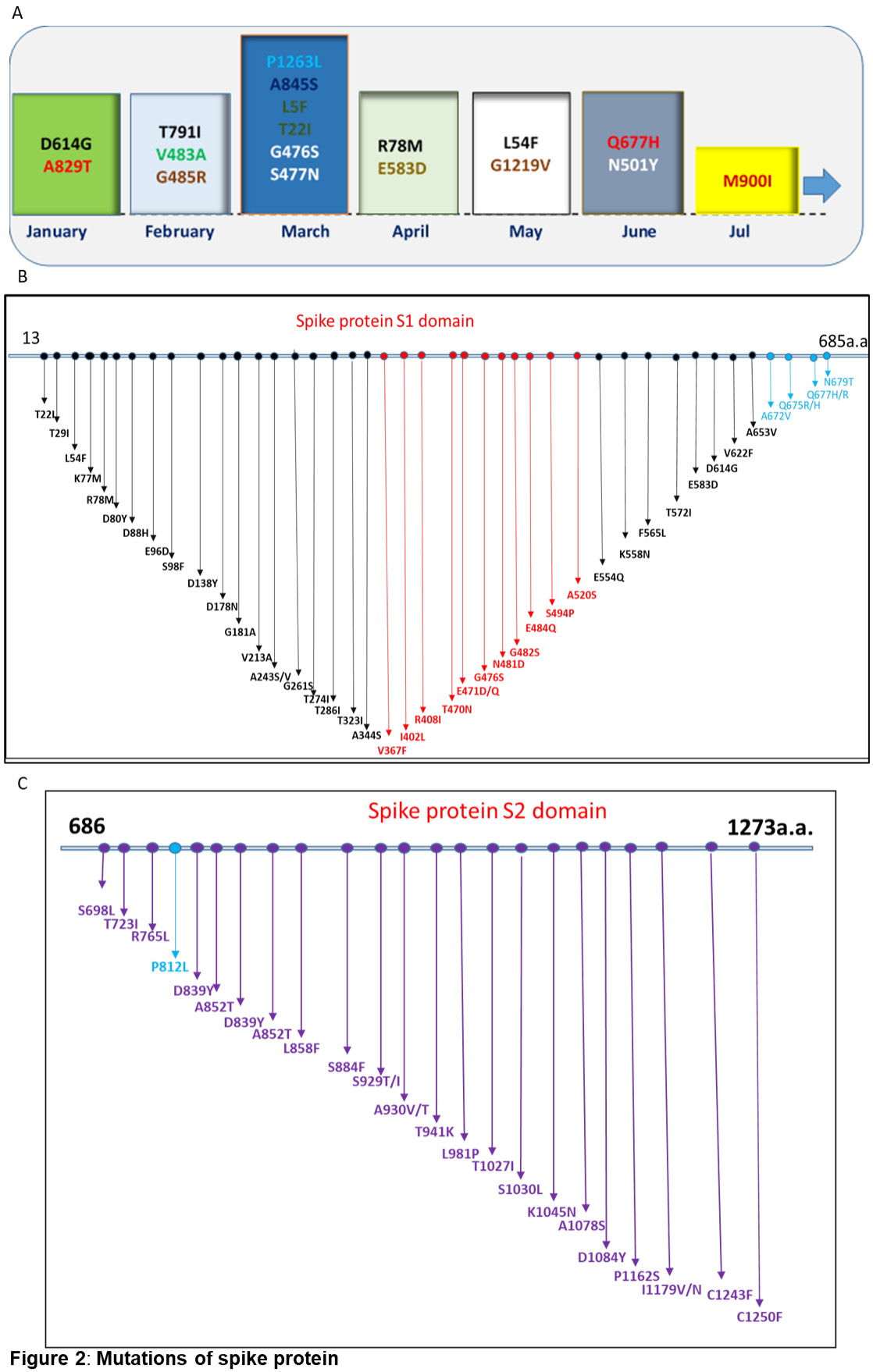

Figure 2: **Mutations of spike protein**

A. The different major mutations are shown in terms of its first occurrence (i.e. with respect to the collection time) at different months of the year 2020.

B. Mutations in S1 domain of SARS-CoV2 Spike protein in Indian Isolates. Red dots indicate mutations in RBD region. Blue dots signify mutations near the S1S2 cleavage site

C. Mutations in S2 domain of SARS-CoV2 Spike protein in Indian Isolates. Blue dots indicate mutations in the S2' cleavage site

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 7 September 2020 doi:10.20944/preprints202004.0281.v2

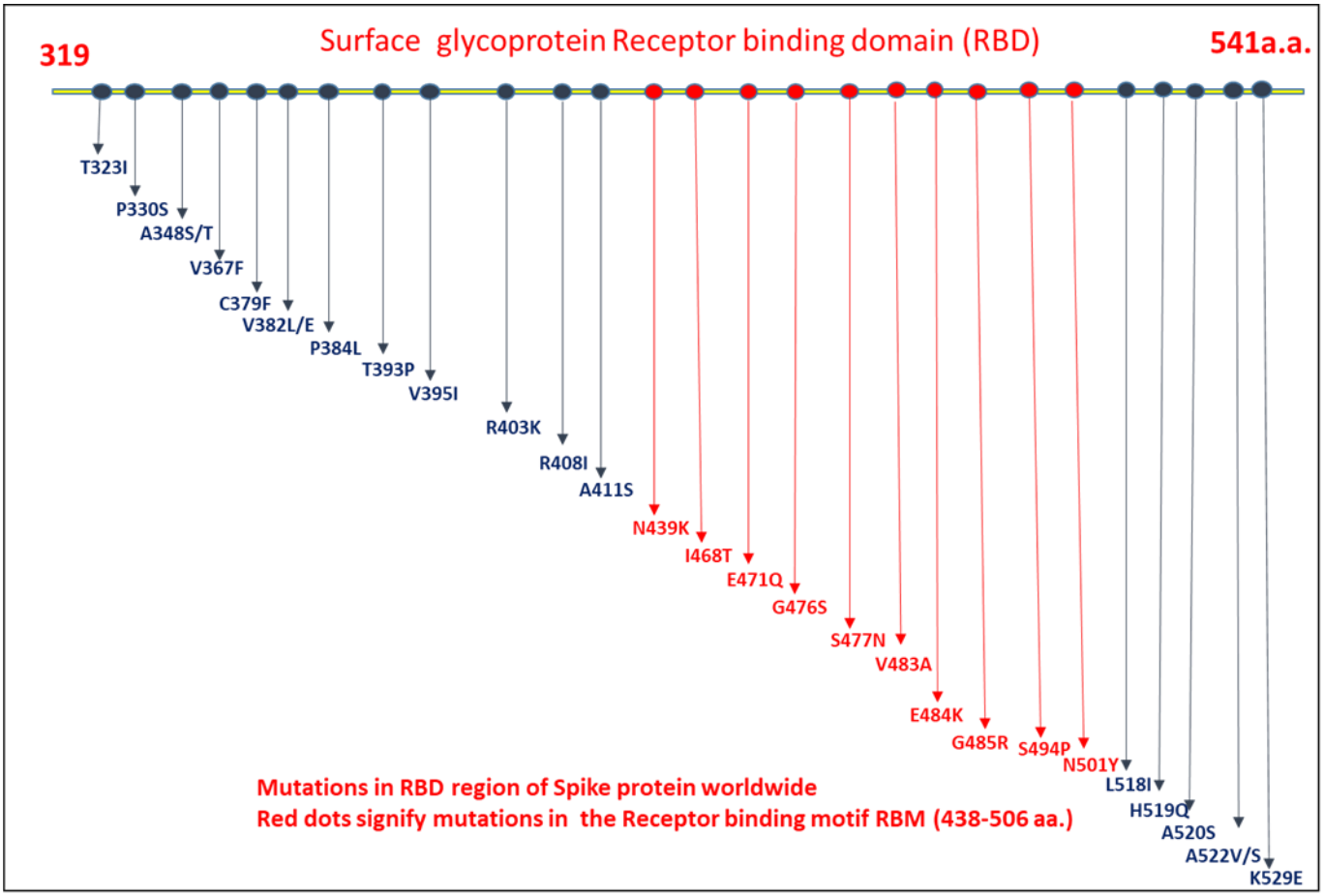

Figure 3: Positions of mutations in receptor binding domain (RBD). Mutations in RBD region of Spike protein worldwide. Red dots signify mutations in the Receptor binding motif RBM (438-506 aa.)

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 7 September 2020 doi:10.20944/preprints202004.0281.v2

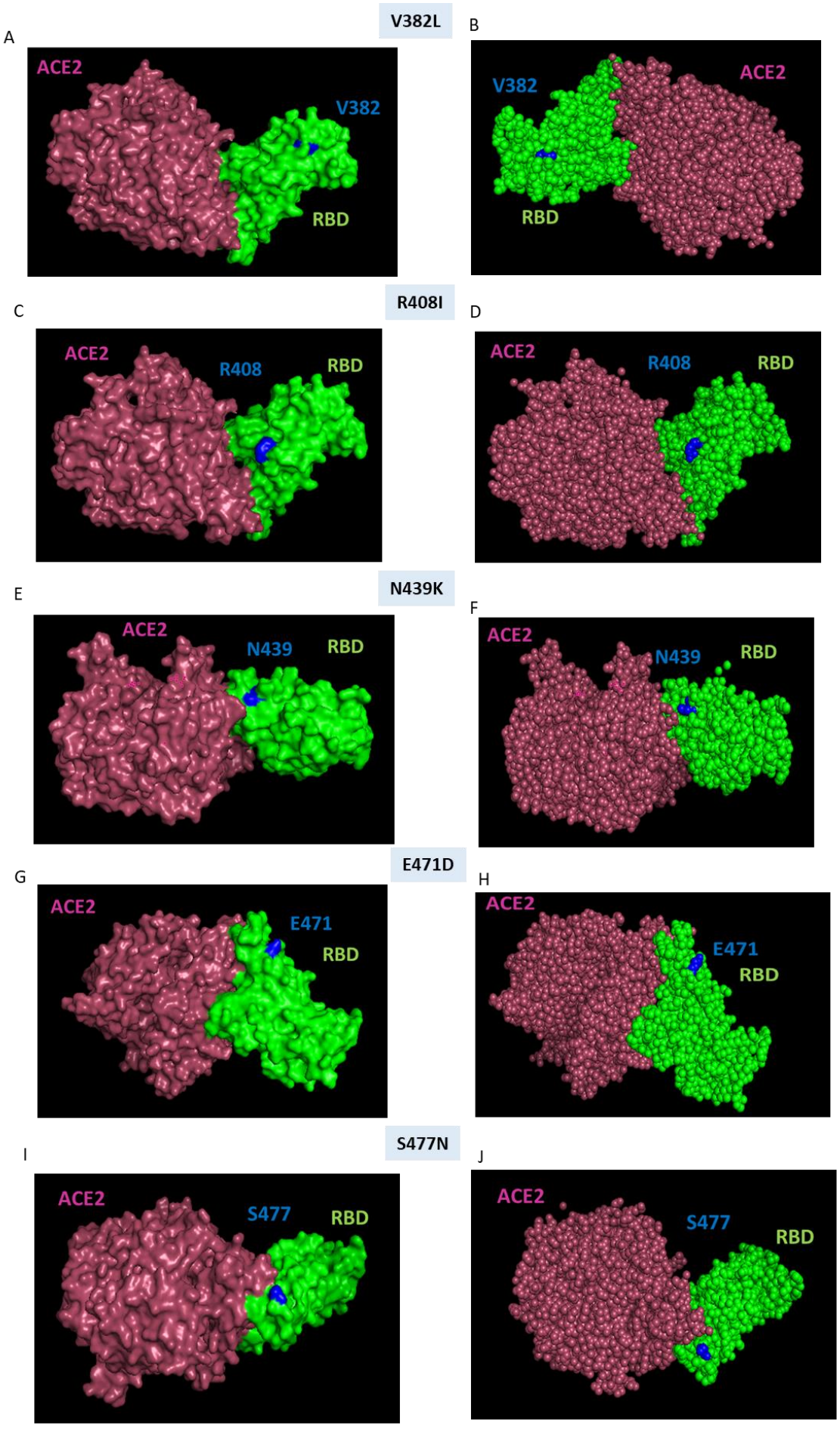

Figure 4:**RBD mutations**. PDB ID:6lzg has been used to indicate the position of the mutations (A-B: position 382; C-D: position 408; E-F: position 439; G-H: position 471; I-J: position 477). Structure in maroon colour indicates ACE receptor and that in green colour represents RBD. Mutation points are shown in blue

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 7 September 2020 doi:10.20944/preprints202004.0281.v2

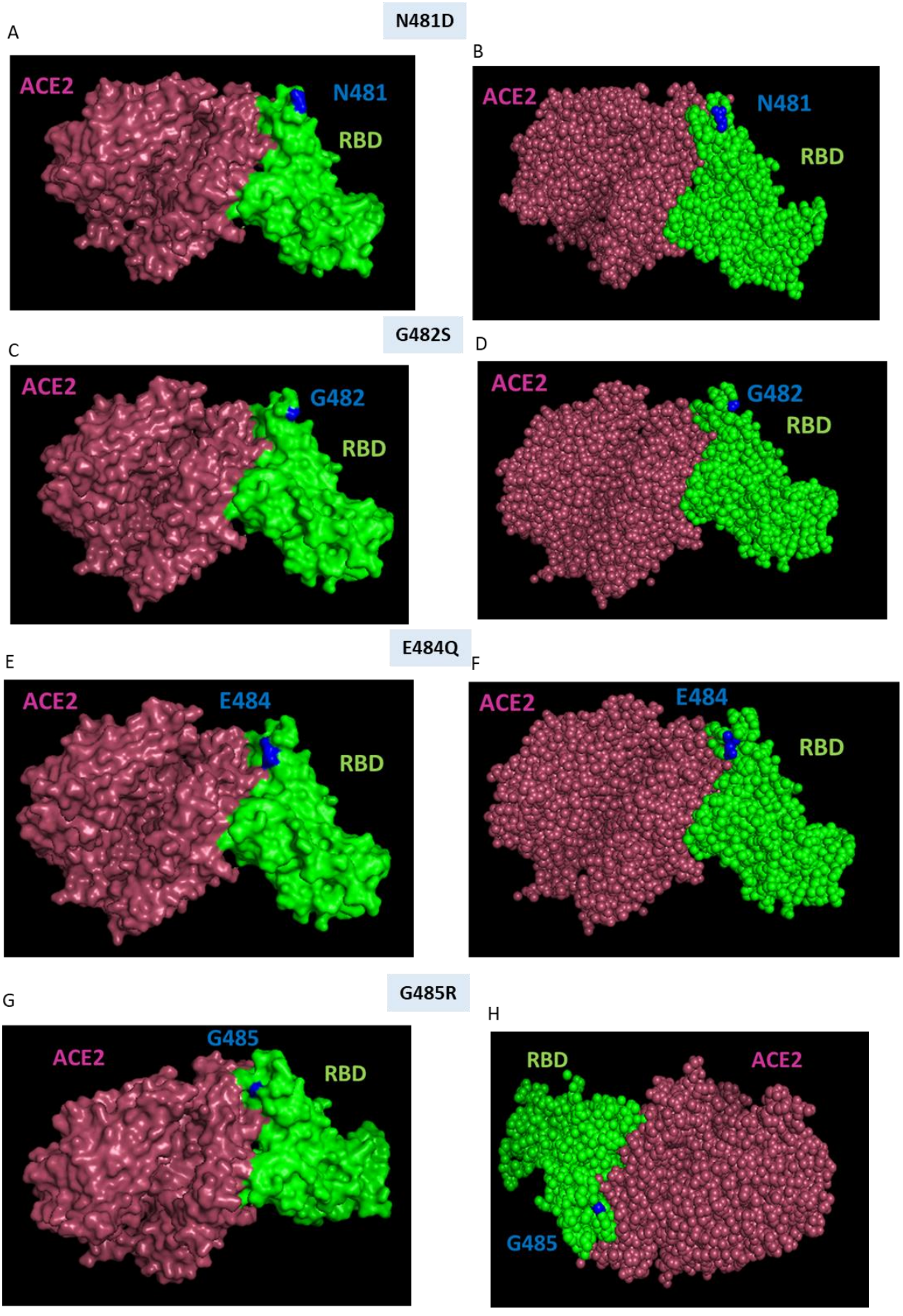

Figure 5:**RBD mutations**. PDB ID:6lzg has been used to indicate the position of the mutations (A-B: position 481; C-D: position 482; E-F: position 484; G-H: position 485). Structure in maroon colour indicates ACE receptor and that in green colour represents RBD. Mutation points are shown in blue

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 7 September 2020 doi:10.20944/preprints202004.0281.v2

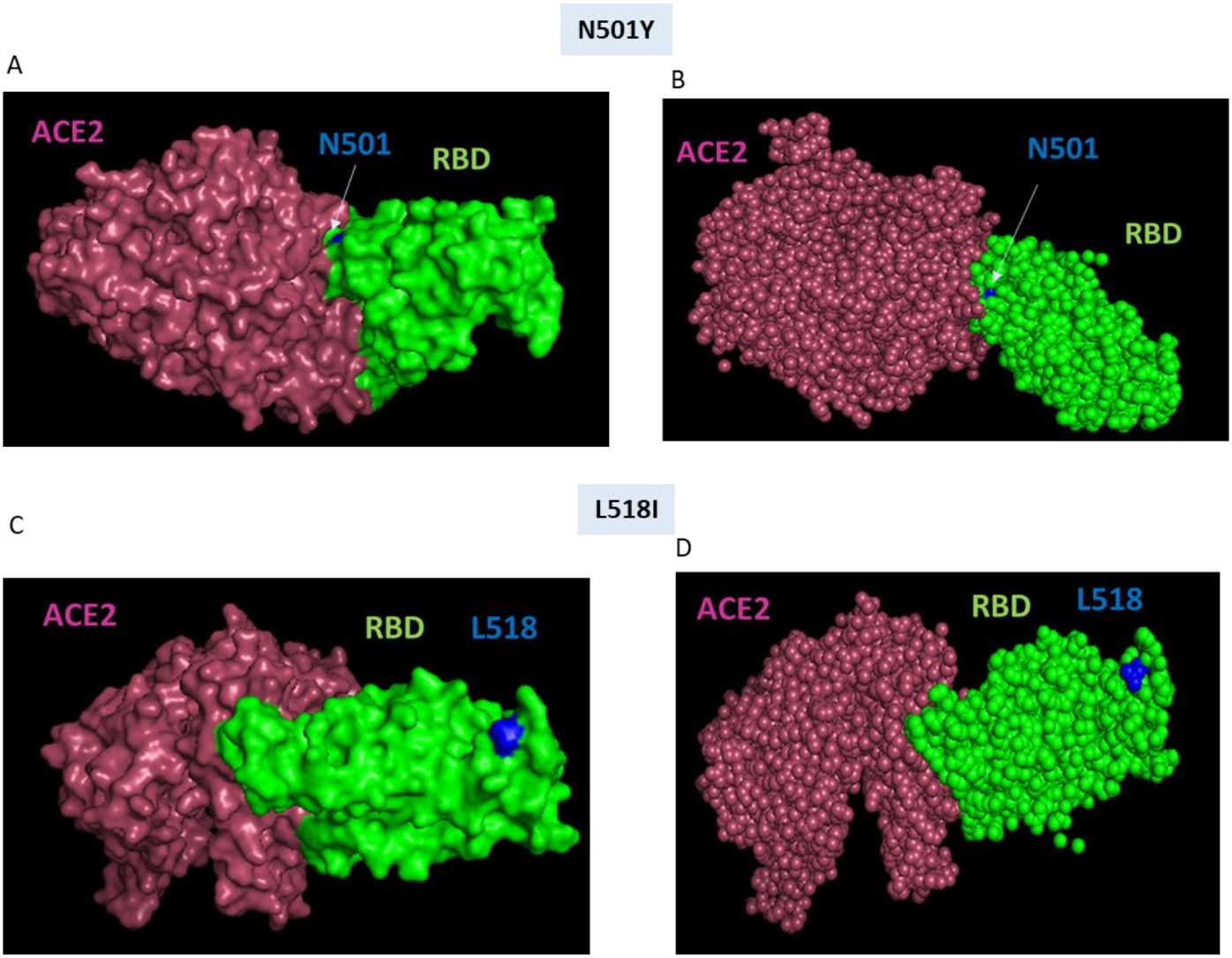

Figure 6:**RBD mutations**. PDB ID:6lzg has been used to indicate the position of the mutations (A-B: position 501; C-D: position 518). Structure in maroon colour indicates ACE receptor and that in green colour represents RBD. Mutation points are shown in blue

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 7 September 2020 doi:10.20944/preprints202004.0281.v2

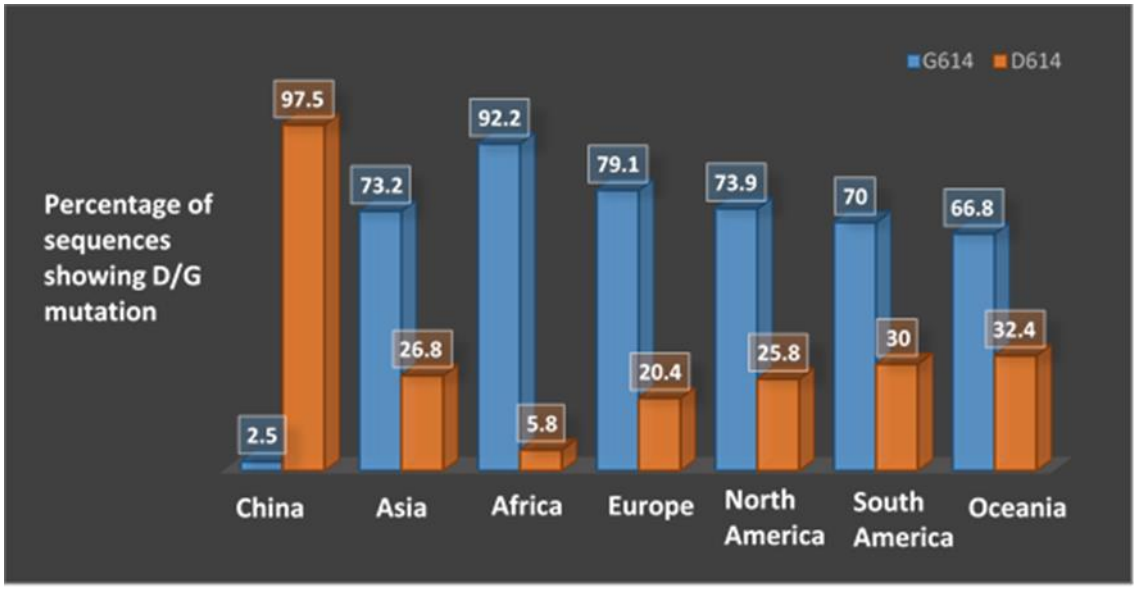

Figure 7: Distribution pattern of mutation D614G across different continents

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 7 September 2020 doi:10.20944/preprints202004.0281.v2

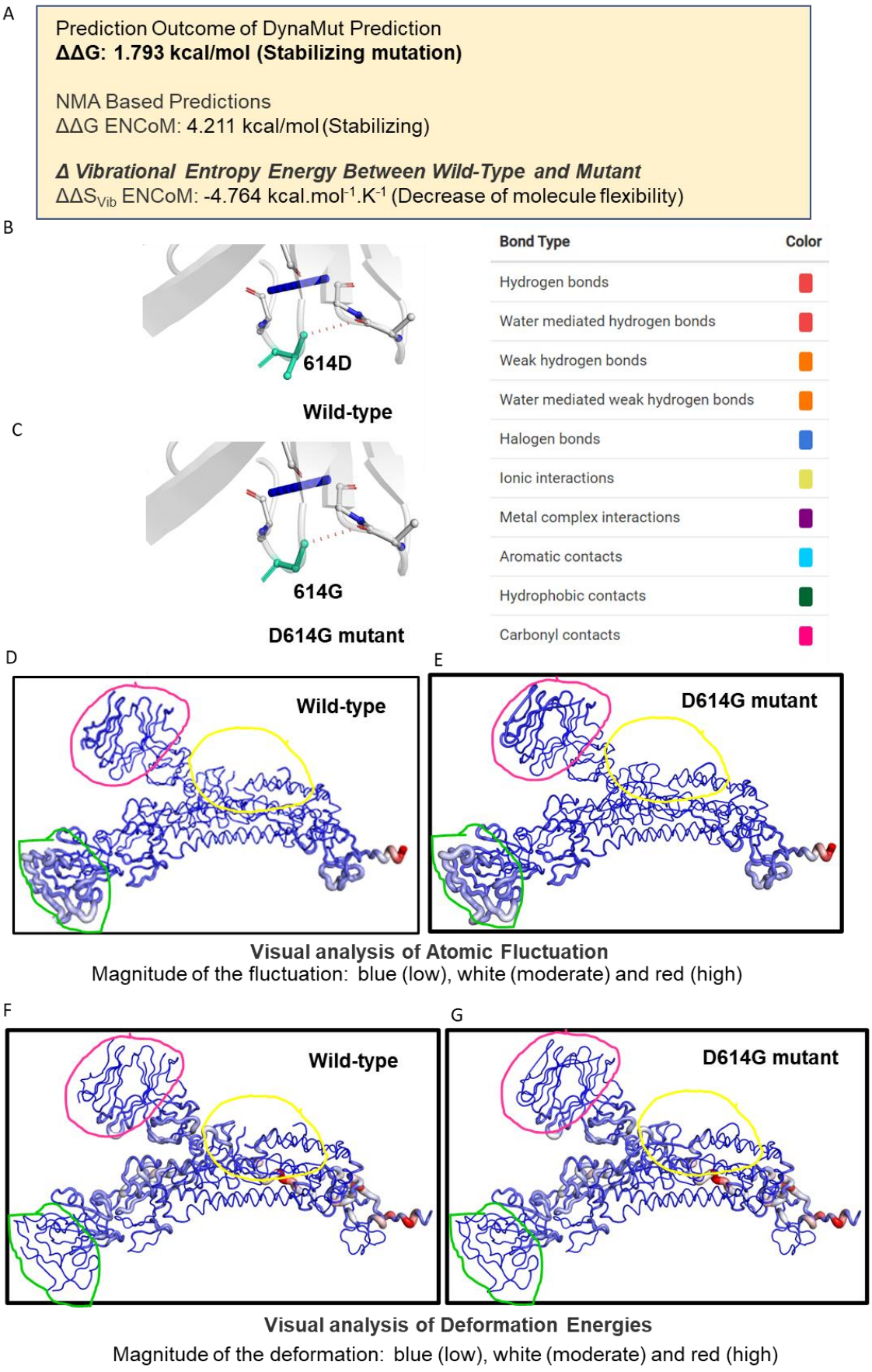

Figure 8: DynaMut prediction of effect of D614G mutation on Spike protein. A. Entropy and free energy changes as assessed by the DynaMut prediction for D614G mutation B. Inermolecular interactions in wild-type spike protein at position 614. C. Inermolecular interactions in mutant spike protein at position 614. D-E Visual analysis of atomic fluctuations. F-G. Inermolecular interactions in mutant spike protein at position 614. D-E Visual analysis of deformation energies.