

Non-Synonymous Mutation analysis in SARS-CoV-2 variants isolated from Humans and Prediction of Conserved Linear Antibody Epitopes for Use in Country-wise Epitope-based Vaccine Development

Authors: Kunal Sharma^{1,2}, Naima Noor Khan^{1,2}, Hamza Bin Sana^{1,2}, Mohammad Ali Khan^{1,2}, Pankaj Kumar^{1,#}

¹Department of Biochemistry, School of Chemical and Life Sciences, Jamia Hamdard University, New Delhi, 110062

²All authors contributed equally

#Correspondence to:

Dr. Pankaj Kumar

Department of Biochemistry,

School of Chemical and Life Sciences

Jamia Hamdard University

New Delhi 110062

pankajkumar@jamiahamdard.ac.in

Abstract:

COVID-19 pandemic has caused a large-scale havoc in almost every country across the globe, putting major challenges for the healthcare system in many parts of the world. Several of the laboratories are running in the race with undying efforts for developing potential vaccine, drugs or therapeutics to treat or prevent the infection. However, with the limited time window and high rate of infection, the task is very big for humanity to find a cure. With hundreds of genomic data of SARS-CoV-2 virus isolates from humans are being submitted almost every day, it is coming into knowledge that virus is mutating, slower in countries with sporadic cases, but higher in countries experiencing large outbreak. These types of mutations in virus may bring challenges in vaccine or therapeutic development for use in each and every country, as each hotspot region may have their own pattern of mutations in virus with ongoing outbreak. In our current study, we retrieved non-synonymous mutation data of around 12,225 SARS-CoV-2 virus samples isolated from humans globally, and discovered all mutations that are collectively happening in antibody epitope regions of the virus country-wise. We found a few numbers of epitope regions in SARS-CoV-2 that are highly conserved collectively in all variants and may be used for epitope-based vaccine development for whole world. We also found epitope regions that are conserved collectively in SARS-CoV-2 variants country-wise and can be used for customized epitope-based vaccine development in each different country.

Introduction:

Since the first report of a novel coronavirus SARS-CoV-2 in late December 2020 from Wuhan province of China, millions of cases have been reported worldwide, affecting most of the countries [1, 2]. After China shared first genetic sequence of this novel human coronavirus, based on the genome sequence of SARS-CoV-2, the virus belongs to B-lineage of beta-coronavirus family of the beta-coronaviruses [1, 3]. Virus is 96% identical to bat coronavirus at whole genome level [4]. SARS-CoV-2 genome encodes four major structural genes, namely, nucleocapsid protein (N), spike protein (S), membrane glycoprotein (M) and additional membrane glycoprotein (HE). Virus genome also code for a very long unstructured polyprotein, Orf1ab, which can yield nsp1-nsp16. Several of important enzymes from SARS-CoV-2 including helicase, 3'-5'-exonuclease, Endo-RNase, 2'-O-Ribose Methyltransferase and RNA dependent RNA polymerase (RdRp) are yielded from Orf1ab polyprotein (Figure 1).

Based on the whole genome sequencing data, several mutation hotspots with non-synchronous mutations have been identified in SARS-CoV-2 [5-7]. It is observed that mutations become unprecedentedly high with increase in burden of infections in various geographical regions. However, there has been no single study on finding relationship between viral mutations and human immune response from various geographical locations. Virus may have to evolve under the pressure of diverse human immune response to gain thermodynamic fitness. A study finds that acute immune response to SARS-CoV-2 is very dynamic in humans and this should be taken into consideration for pathogenesis studies in COVID-19 [8]. More studies are required to gain better understanding of host-viral interactions, host immune response and pathogen immune evasion and discover if mutations in SARS-CoV-2 virus have any role in them. A phylogenetic analysis based study on S gene from 144 sequences of SARS-CoV-2 surmised

that virus evolves to evade the host immune system with non-synonymous mutation as part of the positive selection and more evolved virus may have greater fitness to cause more outbreak [9]. It is believed that amino acid substitutions may alter the immunogenic determinants of the virus and consequently reducing the immunogenicity by hampering the immune cell recognition of SARS-CoV-2virus [9]. A USA strains is found to have a non-synonymous mutation in S protein, D614G, that shows less immunogenic response possibly because of alteration in S-protein epitopes [9].

There have been several attempts on prediction of epitope-based subunit vaccine candidates through in-silico approaches [10-14] or also development of monoclonal antibodies [15]. Highly conserved epitopes in Receptor Binding Domain of S-protein have been reported that can bind a neutralizing antibody CR3002 that was actually developed for SARS virus [16]. However, if the SARS-CoV-2virus is mutating to evolve more, and the mutations are in B or T-cell epitopes, this can bring more challenges in developing a vaccine. Also, mutations in SARS-CoV-2virus develops variants that may show different response to inhibitory drugs [5].

There is a need to understand the mutations in SARS-CoV2 that are happening at large scale globally, especially in epitope region of the virus. If we can tabulate the portfolio of mutations happening in epitope regions of the SARS-CoV-2 virus globally, it may be possible to discover epitopes that are conserved in a country or region and can be good candidate for customized epitope-based vaccine development. However, we may have to keep an eye on genomic data that are being submitted almost every day to track the mutations in virus epitope regions in each and every country for customized the epitope-based vaccine. In our current study, we have retrieved the mutation data of around 12,225 SARS-CoV-2 virus isolates from humans deposited in National Genomics Data Center, China National Center for Bioinformation / Beijing Institute of Genomics, Chinese Academy of Sciences. We tabulated and highlighted all of the non-synchronous mutations

that have happened collectively in predicted antibody epitope regions of SARS-CoV-2 proteins country-wise. We further listed most conserved epitopes regions in all virus variants globally. This knowledge can be very useful in designing customized epitope-based vaccine country-wise.

Material and Method:

1. Homology modeling of proteins

Crystal structure or homology model of SARS-CoV-2 proteins were downloaded from RCSB and I-TASSER [17]. NSP1 (QHD43415_1.pdb); NSP2 (QHD43415_2.pdb); NSP3 (QHD43415_3.pdb); NSP4 (QHD43415_4.pdb); NSP5 (PDB ID: 6LU7); NSP6 (QHD43415_6.pdb); NSP7 (QHD43415_7.pdb); NSP8 (QHD43415_8.pdb); NSP9 (QHD43415_9.pdb); NSP10 (QHD43415_10.pdb); 2'O-methyltransferase (QHD4315_15.pdb); Uridylate-specific endoribonuclease (QHD43415_14.pdb); ExoN (QHD43415_13.pdb); Helicase (QHD43415_12.pdb); RdRp (QHD43415_11.pdb); Spike (QHD43416.pdb); Orf7a (QHD43421.pdb); Orf8 (QHD43422); Orf9 (QHD43423); Orf10 (QHI43199)

2. Prediction of antibody epitopes in SARS-CoV-2

ElliPro was used for prediction of antibody epitopes based on a protein 3-dimensional structure (PMID 19055730). We only searched for linear antibody epitopes due to very high number of mutations in SARS-CoV-2 proteins from across the globe from several countries. ElliPro predictions with a score above 0.5 were only selected in the study.

3. Retrieval of global mutation data of SARS-CoV-2 from human isolates

Mutation data for 12,225 genome sequence of SARS-CoV-2 virus from human isolates was retrieved on 24th April 2020 from China National Centre for Bioinformation (CNCB), (https://bigd.big.ac.cn/ncov/release_genome). CNCB has genomic data source from NGDC, NMDC, GISAID, GenBank and Genome Warehouse. Data was tabulated country-wise for every single mutation in SARS-CoV-2 from human isolates. After gathering predicted epitopes from ElliPro, further emphasis was laid on retrieval of country wise mutated epitope data.

4. Highlight of mutant regions

In the 3-dimensional structure of various proteins from SARS-CoV-2, mutant residues were highlighted in red color in PyMOL, Schrödinger LLC.

Result and Discussion:

All types of collective mutations in SARS-CoV-2 (from 12,225 genomic data) in non-structural polyprotein (Orf1ab), which yield nsp1-nsp16, have been highlighted in red color in three-dimensional structural model of the proteins (Figure 2-4). All predicted epitopes in different proteins (nsp1-nsp16) of Orf1ab have been tabulated (Supplementary table 1-16) and red color highlighted residues in predicted epitopes have been found mutated in several countries listed in the table. Most conserved antibody epitopes regions were also found in Orf1ab proteins that were not mutated in any country (Table 1). For example, in NSP3, six antibody epitopes region were found that were not mutated in any country in the amino acid sequence 1585-1589, 1752-1755, 178-1792, 1933-1950 and 2143-2154. Conserved epitopes were also found in NSP4 protein in amino acid sequence 2813-2817, 2856-2865, 2958-2975 and 3176-3182. Conserved epitopes in NSP6 were found at positions 3628-3633 and 3741-3749. In NSP8, a single sequence epitope region 4133-4140

was conserved. Two epitope regions 4231-4212 and 4250-4253 were found conserved in NSP9. NSP10 has one epitope from 4268-4275 that showed no mutation in any country. A list of all antibody epitope regions in Orf1ab that have been mutated or conserved in many countries can be found in supplementary table 1-16. No conserved epitope was found in other orf1ab proteins, for example, nsp1, nsp2, nsp5, nsp7, ExoRNAase, EndoRNAase and Methyltransferase.

Only single epitope region was found conserved in Membrane glycoprotein, Spike protein and N-protein, residues 145-148 in membrane glycoprotein, 371-376 in spike protein and 265-268 in N-protein. Not a single conserved epitope was found in Orf3a, E-protein, Orf6, Orf7a, Orf8 and Orf10 proteins. A list of all epitopes that have mutated in SARS-CoV-2 virus country-wise in Spike protein, Orf3a, E-protein, M-glycoprotein, Orf6, Orf7a, Orf8, N-protein and Orf10 proteins can be found in supplementary table 17-25.

We further looked in SARS-CoV-2 epitopes that have very high tendency of mutations in maximum number of countries (in at-least more than 30 countries). We find that nsp2 protein epitope regions 183-268, 306-407, 716-747 and 760-778 are highly mutated in large number of countries (Figure 2 and Supplementary table 2). In epitope region 183-268, there are two amino acids R207 and T265 that are largely mutated in SAR-CoV-2 virus from several countries. Residue R207 is found mutated in several countries, namely, DNK, GBR, USA, TUR, ARE, IND, KAZ, AUS, CHN, CHE, BEL, UGA, LBN, CAN, ISR, IND, NOR, SVK, KWT and residue T265 is found mutated in countries: USA, SWE, CHN, ESP, GBR, DNK, FRA, BEL, CHE, AUS, JAM, POL, NOR, COL, ISR, AUT, DEU, TWN, ARE, RUS, HUN, CZE, LUX, FIN, CAN. Abbreviations of countries are given in legend of table 1. In another epitope region of nsp2, 306-407, two amino acid residues V378 and G392 are found mutated in large number of countries, residue V378 in AUS, CHN, UGA, IND, BLR, LBN, TWN, THA, IRN, TUR, ARE, KAZ, IND, SWE, DNK, DEU, CHN, IND, HKG, CAN, NZL, CHN and residue G392 in USA, ISL, BEL, GBR, GBR, ISR, DEU, SWE,

AUT, SGP, GRC, DEU, CAN, HUN, TWN, NLD, BEL, CHN, IND, PRT, LVA, BRA and POL.

In nsp3 protein, very high number of epitope regions are mutated in large number of countries and these are 1224-1258, 2465-2566, 1880-1928, 1969-2070, 1309-1389, 920-968, 1041-1130, 864-917, 1458-1502, 2178-2205 (Figure 2 and Supplementary table 3). In nsp3 epitope region 864-917, one amino acid residue A876 is mutated in DEU, GBR, DNK, BEL, USA, NOR, ISR, GBR, NLD, SWE, GBR, GRC, HUN, ISL, TWN, LUX, LVA, PRT, AUS, POL, BRA and CAN. In nsp3 epitope 1224-1258, one amino acid residue T1246 is mutated in large number of countries, DNK, BEL, POL, CHE, HRV, GBR, CHL, AUT, DNK, ZAF, VNM, HUN, CRI, ARG, SWE, RUS, AUS, USA, COD and ISL). In an another nsp3 epitope region 1969-2070, an amino acid residue T2016 is found mutated in SARS-CoV-2 virus from large number of countries, namely, IND, USA, THA, IND, JAM, GBR, TWN, IND, FRA, SGP, GRC, BRN, PHL, CHN, CAN, BRA, SVN, BEL, AUS, MYS and SAU.

In nsp4, there are two epitope regions, 2764-2808 and 3084-3151, that are mutated in more than 30 countries are (Figure 2 and Supplementary table 4). A residue M2796 in epitope region 2764-2808 is found mutated in virus from large number of countries, namely, UGA, USA, TUR, ARE, KAZ, GBR, CHN, AUS, DNK, NOR, IND, KWT, CAN, LBN, IND and FRA.

In nsp5 protein, three epitope regions 3306-3346, 3352-3364 and 3474-3549 are abundantly mutated in large number of countries (Figure 3 and Supplementary table 5). An amino acid residue G3334 from nsp5 epitope region 3306-3346 is found mutated in virus from various countries USA, CHE, TWN, CHL, SRB, DNK, SRB, BRA, AUS and DEU. An another residue K3353 from 3352-3364 epitope region is found mutated in BGD, DNK, USA, ESP, GBR, GBR, AUS, ISL, GBR, FRA and CHN.

In nsp6, a single epitope region from 3596-3612 residues is mutated at large scale in more than 30 countries (Figure 3 and Supplementary table 6). A single amino acid residue L3606 is found mutated in virus from large number of countries, namely, USA, IND, CHE, ITA, CHN, AUS, UGA, BEL, CAN, JAM, GBR, THA, JOR, POL, LBN, ISR, TWN, DNK, CHL, AUT, FRA, SGP, IRN, TUR, COD, DEU, RUS, ,CZE, QAT, BRA, KOR, HKG, SVN, NLD, CHN, NOR, ESP, ISR, MYS, JPN, GEO, GRC, ESP, BRN, ARE, VNM, IDN, PHL, ARG and LUX.

In nsp8 protein, one epitope region, 3955-3990 is mutated in large number of countries and a residue S3983 in this epitope is found mutated from many countries, namely, SWE, GBR;SCOTLAND, GBR;ENGLAND, GRC, ARG, CHN and USA (Figure 3 and Supplementary table 8). We also find that many cities in GBR country has SARS-CoV-2 virus with mutations in nsp9 epitope region 4214-4225 at residue P4220 and in nsp10 epitope region 4309-4317 at D4317 residue (Figure 4 and Supplementary table 9 and 10).

Two epitopes 4399-4514 and 5210-5324 from RdRp protein are found mutated in large number of countries (Figure 4 and Supplementary table 12). A residue T4418 is mutated in SARS-CoV-2 virus from countries like GBR;ENLAND, USA, GBR;SCOTLAND, DNK, FRA, LUX, TWN, SWE and BEL. Another residue in the same epitope, 4489, is mutated in virus from very large number of countries, namely, IND, CHN, TWN, THA, GBR;ENGLAND, COL, FRA, SGP, BRN, PHL, CAN, MYS, AUS, SAU and GMB. One more residue S5039 in epitope region 5035-5042 is found mutated in virus from many countries, namely, BEL, CAN, NOR, DNK, FRA, AUT, GBR;SCOTLAND, GBR;WALES, GBR;ENGLAND, CRI, DEU and ISL.

One epitope region 5349-5428 in helicase protein have many mutations in SARS-CoV-2 virus from large number of countries (Figure 5 and Supplementary table 13). An amino acid residue 5621 from epitope region 5618-5628 is found mutated in virus in many countries, namely, USA, NLD, GBR;ENGLAND, DNK, GBR;SCOTLAND, ESP,

GBR;WALES, ESP and BEL. In 3'-5' Exonuclease, epitopes 5934-5978, 6368-6418 and 6050-6086 are found highly mutated in virus from many countries (Figure 5 and Supplementary table 14). Two epitope region 6467-6498 and 6663-6686 endo-RNAase are highly mutated in many countries (Figure 5 and Supplementary table 15). In 2'-O-Ribose methyltransferase, 7031-8096 and 6809-6839 are highly mutated in large number of countries (Figure 5 and Supplementary table 16).

There are many epitopes regions in S-protein, 1067-1134, 432-536, 239-264, 1-26, 108-188 and 1233-1273 that has most abundant number of mutations across the globe (Figure 6 and Supplementary table 17). An amino acid residue L5 from epitope 1-26 is found mutated in several countries, namely, USA, IND, LVA, GBR;ENGLAND, GBR;SCOTLAND, GBR;WALES, BEL, DNK, FRA, ISL, TWN, PRT, CAN and JPN. An another amino acid P1263 in spike protein epitope region 1233-1273 is found mutated in virus from several countries, GBR;ENGLAND, USA, FRA, ISR, CHL, GBR;NORTHERN, IRELAND, GRC, IND, GBR;WALES, GBR;SCOTLAND, AUS and ISL. We did not find a single amino acid residue in epitopes of receptor binding domain (RBD) that has mutation with high abundancy country-wise. This also suggest that RBD domain may not be important determinant of immunogenicity.

Two epitope region 13-41, 56-64, 179-197 and 250-275 are found highly mutated in ORF3a proteins in SARS-CoV-2 virus from several countries (Figure 7 and Supplementary table 18). In 56-64 epitope region, a residue Q57 is found highly mutated in virus from many countries, namely, AUS, BEL, USA, SWE, IND, JAM, LBN, POL, NOR, BGD, CAN, GBR;ENGLAND(LONDON), GBR;SCOTLAND, VNM, SVN, TWN, COL, ISR, GEO, ROU, GBR;WALES, ESP, CHL, FRA, AUT, SAU, DNK, LVA, DEU, BRA, LUX, CHE, THA CZE, RUS, SEN, NOR, CAN and CAN. One more residue G196 in epitope region 179-197 is found mutated in virus from countries: AUS, COL, CHL, DNK, ESP, POL, GBR;ENGLAND, GBR;SCOTLAND, GRC, ESP, USA, KAZ, IND, BRA, MEX, NLD, LUX, ESP, VASCO, GBR;WALES, ESP, CHN, ESP, GEO and FRA.

In M-glycoprotein, one epitope region at N-terminal of protein from 1-22 residues is highly mutated in many countries (Figure 7 and Supplementary table 20). One residue, D3, is found mutated in virus in several countries, namely, GBR;ENGLAND, BEL, USA, CAN, NOR, ISR, COD, DNK, AUT, FRA, ISL, IND, GBR;SCOTLAND, LVA, RUS, LUX, SWE, CHE, CHN, KAZ, THA, ARE, LTU, ITA, SVK, PRT, AUS and FIN. One epitope region 42-57 residues in ORF6, 23-38 epitope residues in ORF7a and 62-69 epitope region in ORF8 are found to have mutations in virus from many countries (Figure 7-8 and Supplementary table 21-23). In ORF8 protein, an amino acid residue A65 is mutated in large number of countries, GBR;ENGLAND, BGD, CHL, GBR;NORTHERN, IRELAND, SWE, ESP, NLD, GBR;SCOTLAND, GRC, USA, AUS and FRA.

In N-protein, high number of epitope regions are found mutated in virus from many countries, and these are 113-144, 231-250, 411-419, 19-43, 361-388. One epitope region in N1-protein, 184-208, is found mutated in virus from almost every country (Figure 8 and Supplementary table 24). Several amino acid residues in 184-208 residue, namely, S188, R191, S193, S194, S197, S202, R203, G204 and T205 are found highly mutated almost across the globe. From our overall mutation analysis study in SARS-CoV-2 virus, we find that N-protein epitope region 184-208 show maximum number of mutations in amino acids in virus from across the globe which suggests that N-protein can be the important determinant of immunogenicity.

Conclusion:

This is the first study of its own kind where all the mutations that have happened collectively in thousands of SARS-CoV-2 virus isolates from humans from various countries have been highlighted in predicted epitope regions of viral proteins. The goal of this study is to discover antibody epitope regions that are most conserved collectively in all SARS-CoV-2 variants isolated from humans and can be used for designing a common epitope-based

vaccine (or multiepitope-based vaccine) for whole world. We found around 25 epitope regions that are conserved so far collectively in 12,225 number of SARS-CoV-2 virus isolates from human from various countries. Also, we have highlighted all mutations that have happened in epitope regions of SARS-CoV-2 virus country-wise. We find that there is different portfolio of mutated epitopes in different countries, for example, in nsp1 protein, almost all of the nsp1 epitopes have been found mutated in virus isolates from United States, but this has not happened in virus isolates from China, and the possible reason may be very high outbreak of COVID-19 in United States as compared to China. This also suggests that high number of COVID-19 outbreak may lead to high number of mutations in virus. Countries like United States with high abundance of mutations in epitope regions may find challenges in finding a successful vaccine for treatment of COVID-19. However, on a positive note, in our current study, we analyzed and highlighted mutations collectively happening across all epitope regions of SARS-CoV-2 virus variants country-wise and find that there are epitopes which are still conserved in countries with highest burden of COVID-19. For example, we found nsp2 epitope region 806- 818 that is conserved in SARS-CoV-2 isolates from United States, the epitope is found mutated in other countries like Great Britain, Australia, Turkey and Sweden.

Overall, a new approach is needed to develop vaccine against COVID-19 that also considers mutagenesis happening globally in SARS-CoV-2 virus. We suggest to look for conserved epitopes in SARS-CoV-2 variants country-wise and develop a customized multiepitope-based vaccine.

Declaration of Competing interest: The authors declare no competing interests.

References:

1. Zhu, N., et al., *A Novel Coronavirus from Patients with Pneumonia in China*, 2019. N Engl J Med, 2020. **382**(8): p. 727-733.
2. Wang, D., et al., *Clinical Characteristics of 138 Hospitalized Patients With 2019 Novel Coronavirus-Infected Pneumonia in Wuhan, China*. JAMA, 2020.
3. Velavan, T.P. and C.G. Meyer, *The COVID-19 epidemic*. Trop Med Int Health, 2020. **25**(3): p. 278-280.
4. Zhou, P., et al., *A pneumonia outbreak associated with a new coronavirus of probable bat origin*. Nature, 2020. **579**(7798): p. 270-273.
5. Pachetti, M., et al., *Emerging SARS-CoV-2 mutation hot spots include a novel RNA-dependent-RNA polymerase variant*. J Transl Med, 2020. **18**(1): p. 179.
6. Tiwari, M. and D. Mishra, *Investigating the genomic landscape of novel coronavirus (2019-nCoV) to identify non-synonymous mutations for use in diagnosis and drug design*. J Clin Virol, 2020. **128**: p. 104441.
7. Chang, T.J., et al., *Genomic Analysis and Comparative Multiple Sequence of SARS-CoV2*. J Chin Med Assoc, 2020.
8. Ong, E.Z., et al., *A Dynamic Immune Response Shapes COVID-19 Progression*. Cell Host Microbe, 2020.
9. Kim, S.J., et al., *A Novel Synonymous Mutation of SARS-CoV-2: Is This Possible to Affect Their Antigenicity and Immunogenicity?* Vaccines (Basel), 2020. **8**(2).
10. Peele, K.A., et al., *Molecular docking and dynamic simulations for antiviral compounds against SARS-CoV-2: A computational study*. Inform Med Unlocked, 2020: p. 100345.
11. Ojha, R., et al., *High throughput and comprehensive approach to develop multi-epitope vaccine against minacious COVID-19*. Eur J Pharm Sci, 2020. **151**: p. 105375.
12. Yi, C., et al., *Key residues of the receptor binding motif in the spike protein of SARS-CoV-2 that interact with ACE2 and neutralizing antibodies*. Cell Mol Immunol, 2020.
13. Lee, C.H. and H. Koohy, *In silico identification of vaccine targets for 2019-nCoV*. F1000Res, 2020. **9**: p. 145.
14. Kalita, P., et al., *Design of a peptide-based subunit vaccine against novel coronavirus SARS-CoV-2*. Microb Pathog, 2020. **145**: p. 104236.
15. Wang, C., et al., *A human monoclonal antibody blocking SARS-CoV-2 infection*. Nat Commun, 2020. **11**(1): p. 2251.
16. Yuan, M., et al., *A highly conserved cryptic epitope in the receptor binding domains of SARS-CoV-2 and SARS-CoV*. Science, 2020. **368**(6491): p. 630-633.
17. Roy, A., A. Kucukural, and Y. Zhang, *I-TASSER: a unified platform for automated protein structure and function prediction*. Nat Protoc, 2010. **5**(4): p. 725-38.

TABLES AND FIGURES

TABLE 1. List of most conserved epitopes collectively from 12,225 isolates of SARS-CoV-2 from humans across the globe.

SARS-CoV-2 protein	Conserved Antibody Epitope
NSP3	1933 DNIKFADDLNQLTGYKKP 1950
	2143 KPFLNKVVSTTT 2154
	1585 DMSMT 1589
	1788 TCGKQ 1792
	1752 CKTC 1755
	2436 VYANGGKGF 2444
NSP4	2856 AVITREVGfV2865
	2958 DGSIIQFPNTYLEGsvrv 2975
	2856 AVITREVGfV2865
	3176 EEaALCT 3182
	2813 DGGVT 2817
NSP6	3628 FVKHKH 3633
	3741 TSNYSGVVT 3749
NSP8	4133 ANSAVKLQ 4140
NSP9	4250 VRLQ 4253
	4231 IKGLNNL 4212
NSP10	4268 SFCAFAVD 4275
RdRp	4999-SDVENP-5004
	4720-PLVRKIFVDGVPfVVSTG-4737
	4586-DAMRNAGIVG-4595
Helicase	5331- LCNSQTSLR- 5339
	5325- AVGA-5328
M glycoprotein	145 LRGH 148
Spike	371-SASFST-376
N-protein	265-TKAY-268

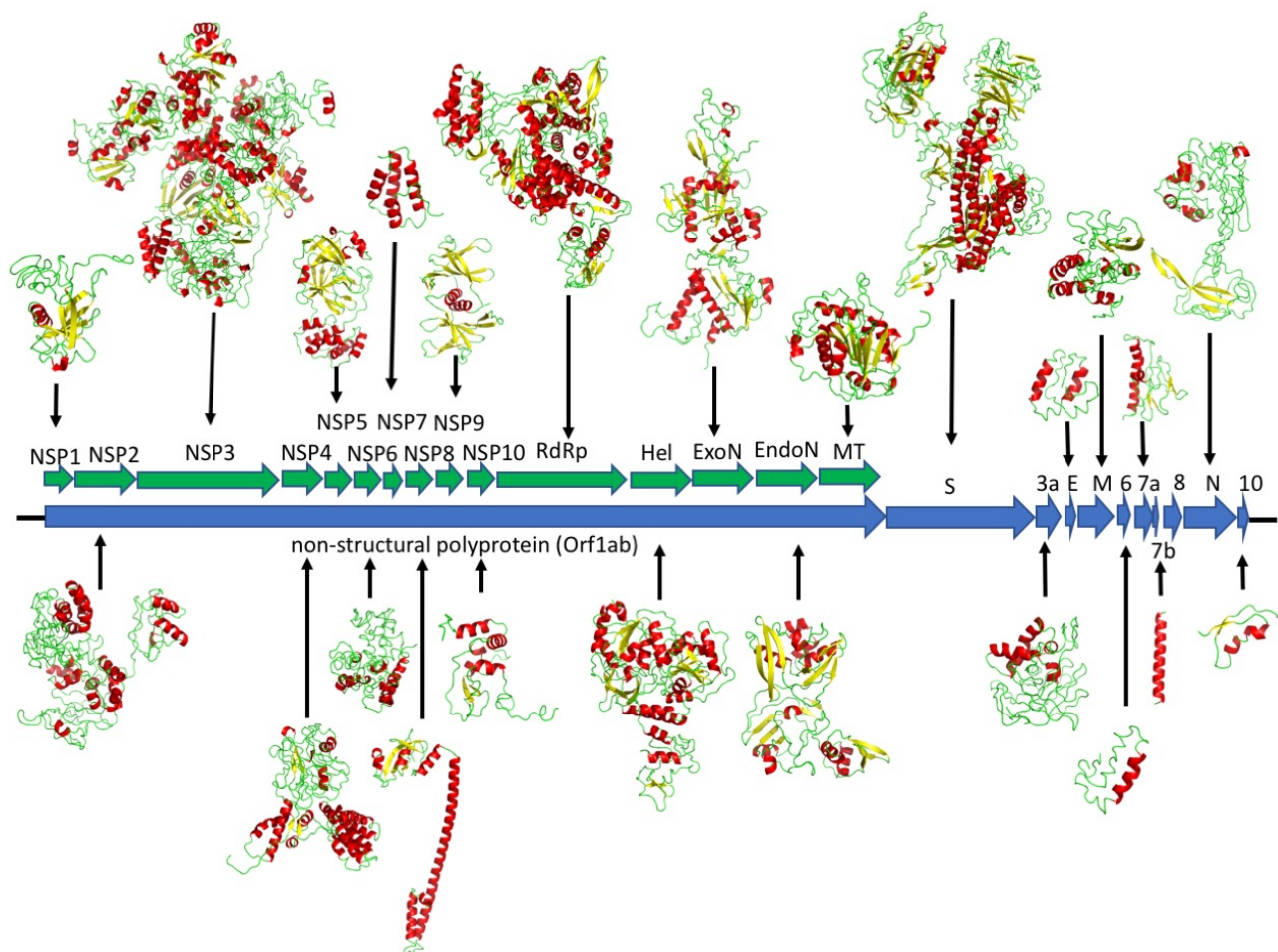


FIGURE 1. Structure of all proteins encoded in SARS-CoV-2 genome. (RdRp: RNA dependent RNA Polymerase , Hel: Helicase, ExoN: 3'-5' exonuclease, Endo: Endo-RNAase, MT: 2'-O-Ribose Methyltransferase, S: Spike protein, 3a: ORF3a ,E: Envelop protein, M: Membrane Glycoprotein 6: ORF6, 7a: ORF7a, 7b: ORF7b, N: Nucleocapsid protein, 10: ORF10)

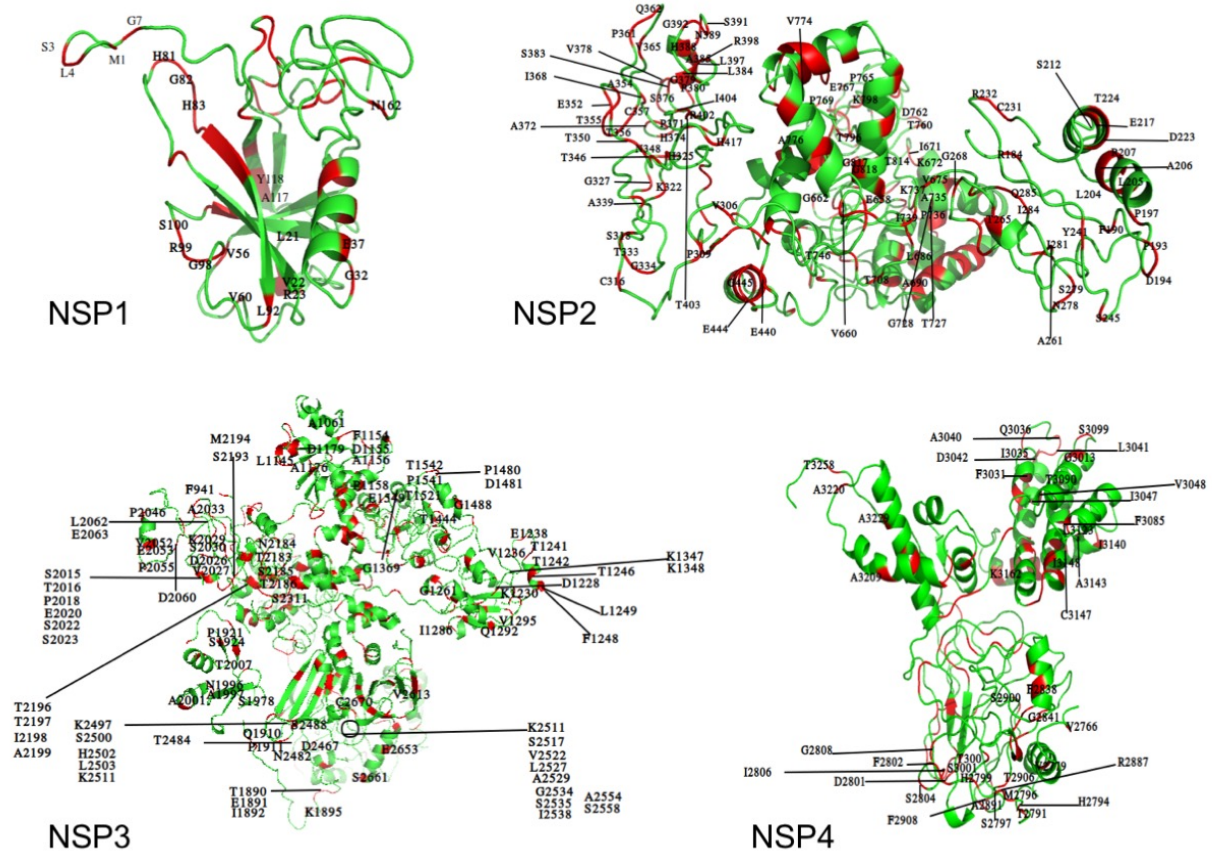


FIGURE 2. 3-dimensional structures of SARS-CoV-2 proteins, NSP1, NSP2, NSP3 and NSP4. Region highlighted in red colour are mutations collectively from 12,225 isolates of SARS-CoV-2 from humans across the globe. Only mutations in epitope regions were labelled.

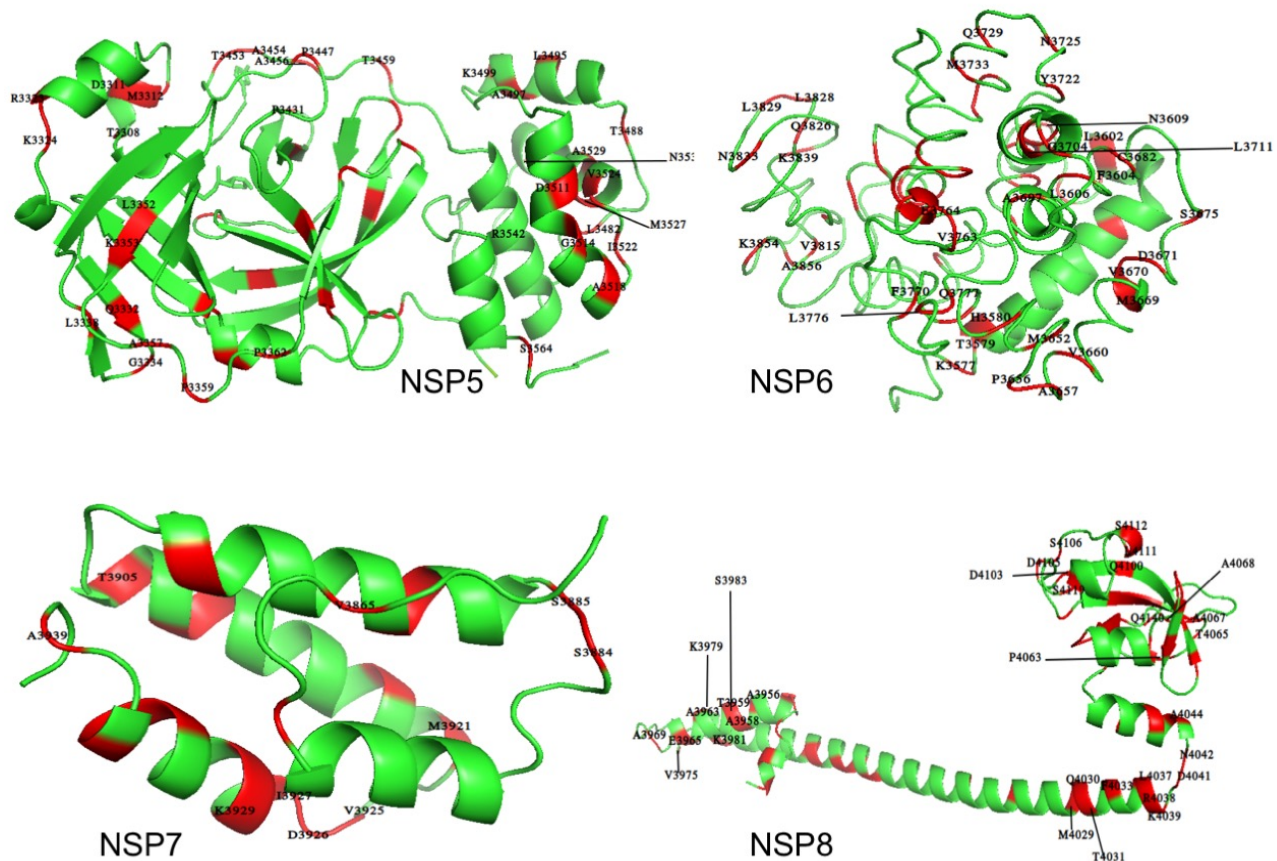


FIGURE 3. 3-dimenstional structures of SARS-CoV-2 proteins, NSP5, NSP6, NSP7 and NSP8. Region highlighted in red colour are mutations collectively from 12,225 isolates of SARS-CoV-2 from humans across the globe. Only mutations in epitope regions were labelled.

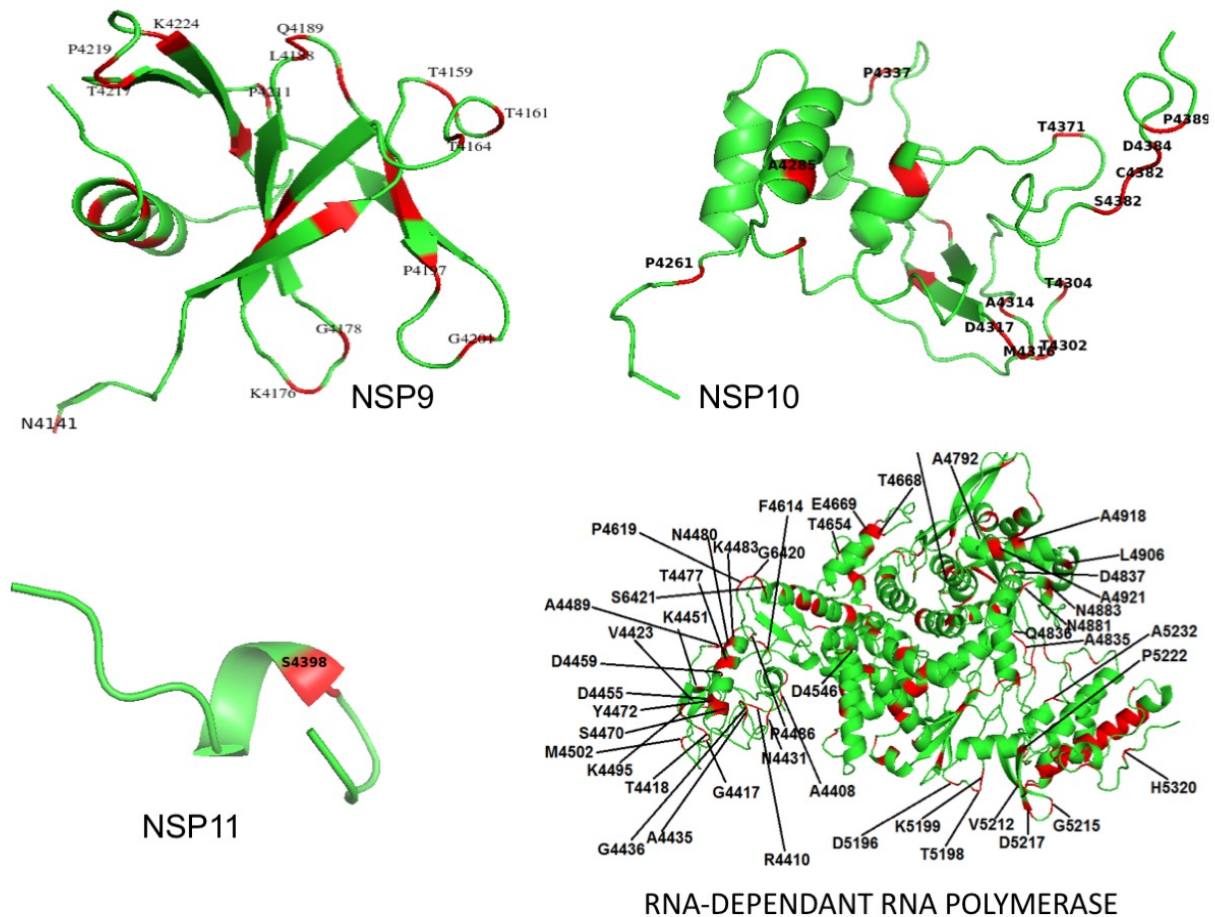


FIGURE 4. 3-dimentional structures of SARS-CoV-2 proteins, NSP9, NSP10, NSP11 and RNA-dependent RNA polymerase (RdRp). Region highlighted in red colour are mutations collectively from 12,225 isolates of SARS-CoV-2 from humans across the globe. Only mutations in epitope regions were labelled.



FIGURE 5. 3-dimensional structures of SARS-CoV-2 proteins, Helicase, 3'-5' exonuclease, Endo-RNAase and 2'-O-Ribose Methyltransferase. Region highlighted in red colour are mutations collectively from 12,225 isolates of SARS-CoV-2 from humans across the globe. Only mutations in epitope regions were labelled.



FIGURE 6. 3-dimentional structure of SARS-CoV-2 spike protein. Region highlighted in red colour are mutations collectively from 12,225 isolates of SARS-CoV-2 from humans across the globe. Only mutations in epitope regions were labelled.

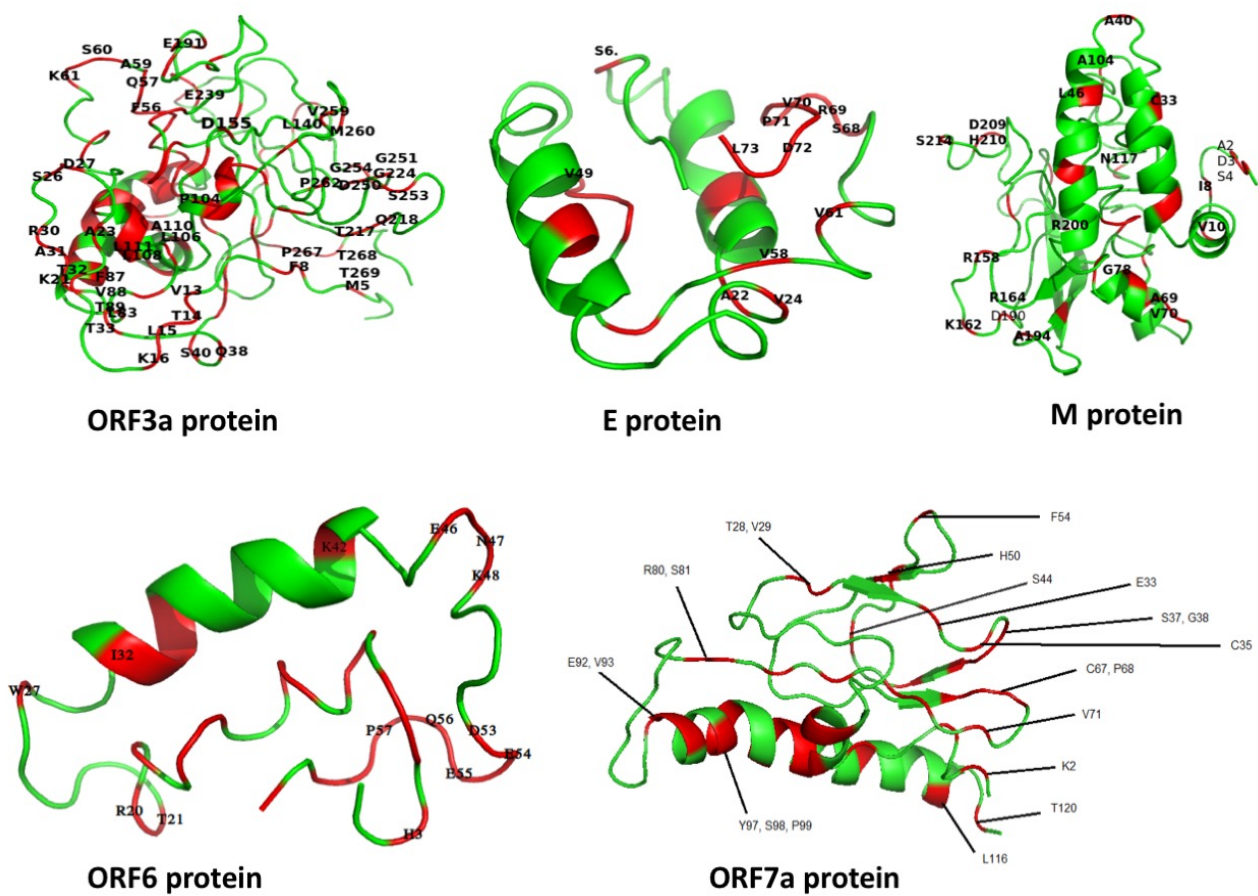


FIGURE 7. 3-dimenstional structure of SARS-CoV-2 proteins ORF3, E-protein, M-glycoproteins, ORF6 and ORF7a. Region highlighted in red colour are mutations collectively from 12,225 isolates of SARS-CoV-2 from humans across the globe. Only mutations in epitope regions were labelled.

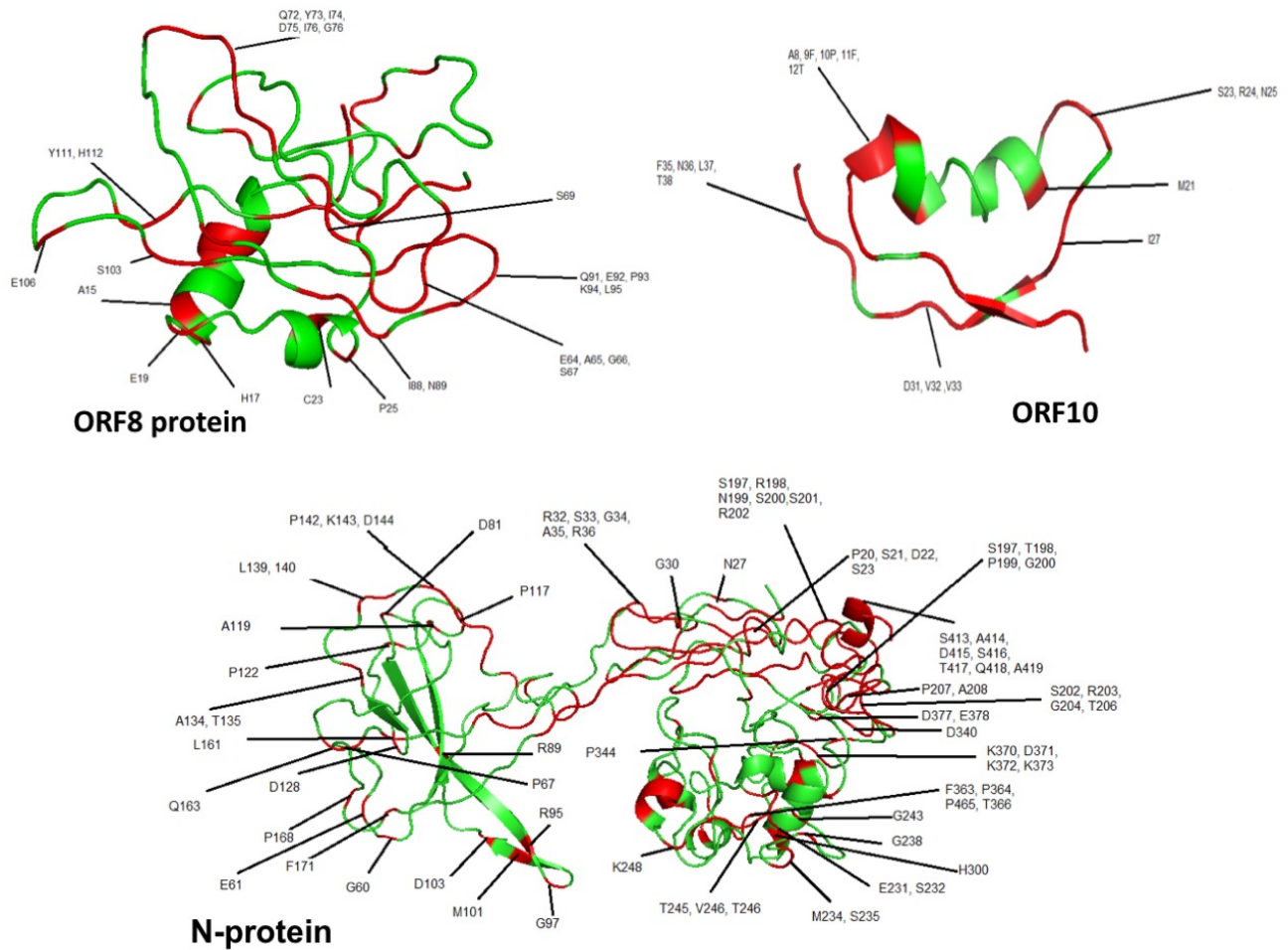


FIGURE 8. 3-dimentional structure of SARS-CoV-2 proteins, ORF8, N-proteins and ORF10. Region highlighted in red colour are mutations collectively from 12,225 isolates of SARS-CoV-2 from humans across the globe. Only mutations in epitope regions were labelled.