

Mutation sites are distant from Remdesivir binding site in human SARS-CoV-2 RNA-dependent RNA polymerase

Dr. Kunchur Guruprasad, Ph.D

ABREAST™, Plot Nos.14/A & 15, Sitaramnagar, Safilguda, Hyderabad-500056, India

E.mail: abreastkgp@gmail.com, kunchur.guruprasad@gmail.com

Website: <https://www.abreast.in>

Abstract

The comparison of 10,929 human SARS-CoV-2 RdRp protein sequences representing six geographical locations with the reference protein sequence in human SARS-CoV-2 genome isolate from Wuhan, China, identified 222 distinct mutation sites in the RdRp protein. The NiRAN and interface domains, Fingers, Palm and Thumb sub-domains were each associated with ~20% or more mutations compared to mutations in N-terminal, beta-hairpin or C-terminal regions of the protein. The Pro4715Leu mutation was predominantly observed in RdRp proteins from all six geographical locations; Africa, Asia, Europe, North America, Oceania and South America. None of the mutation site residues were within 3.2 Å interacting distance from remdesivir as observed in the three-dimensional cryo-electron microscopy structures of RdRp protein complexes available in the Protein Data Bank. Therefore, the mutations in human SARS-CoV-2 RdRp proteins, described in the present work, are not likely to cause resistance to remdesivir binding. Further, the mutations were also not associated with functionally important residues that would affect the enzyme's function.

Keywords: human SARS-CoV-2; RNA dependent RNA polymerase; geographical locations; mutation mapping, RdRp sequence, RdRp three-dimensional structure; remdesivir

Introduction

The current coronavirus pandemic coronavirus disease 2019 (Covid-19), caused due to the severe acute respiratory syndrome coronavirus-2 (SARS-CoV-2) has resulted in 85,159,304 coronavirus cases and 1,846,242 deaths worldwide (<https://www.worldometers.info/coronavirus/>). The first complete genome sequence of human

SARS-CoV-2 from infected individual in the city of Wuhan, Hubei-1 province, China was reported in January 2020 (Wu et al., 2020). The SARS-CoV-2 is a positive sense single stranded RNA genome containing 29,903 base pairs (NCBI Accession code: NC_045512.2) (<https://www.ncbi.nlm.nih.gov/genome/>). The orf1AB gene comprising 7084-7096 amino acid residues represents fifteen polyproteins; leader protein, (non-structural proteins) nsp2, nsp3, nsp4, nsp6, nsp7, nsp8, nsp9, nsp10 and the 3C-like proteinase, RNA dependent RNA polymerase (RdRp), helicase, 3' -to-5' exonuclease, endoRNase and 2' -O-ribose methyltransferase.

In earlier study, mutations relative to the reference polyprotein sequences in human SARS-CoV-2 orf1AB gene from infected individuals in four Indian states between January 2020 to mid-April 2020 were reported (Guruprasad, 2020a). The sequences were available in the publicly accessible NCBI database (<https://www.ncbi.nlm.nih.gov/labs/virus/vssi/#/>). Later, the mutations were mapped on to protein three-dimensional structures for the RdRp, helicase, endoRNase and spike proteins available in the Protein Data Bank (Berman et al., 2000). The protein secondary structure corresponding to the mutations and the proximity of the mutation sites to functionally important residues in the protein, and to the remdesivir and tipiracil drug binding sites in RdRp and endoRNase targets, respectively, were analysed (Guruprasad, 2020b). The study included 8 mutations for the RdRp proteins from the 22 human SARS-CoV-2 Indian isolates that were then mapped on to the three-dimensional structure. Subsequently, mutations relative to the reference polyproteins present in a much larger dataset comprising 10,929 orf1AB genes representing human SARS-CoV-2 genome isolates from six geographical locations; Africa, Asia, Europe, North America, Oceania and South America were reported (Guruprasad, 2020c).

The RdRp or (nsp12) protein is involved in replication and transcription cycle of the virus whereby it catalyzes synthesis of viral RNA with assistance of cofactors (Subissi et al., 2014). There are currently no drugs specifically developed for the treatment of Covid-19 patients.

A recent final report on remdesivir for the treatment of Covid-19 that was based on double-blind, randomized, placebo-controlled trial in adult patients hospitalized with Covid-19 and who had evidence of lower respiratory tract infection showed that remdesivir was superior to placebo in shortening the time to recovery (Beigel et al., 2020). In the report on the trial known as the Adaptive COVID-19 Treatment Trial (ACTT-1), funded by the National Institute of

Allergy and Infectious Diseases (NIAID), researchers showed that remdesivir provides moderate clinical benefits in the treatment of Covid-19 considered an important advance in developing effective therapies for SARS-CoV-2 infections (Dolin et al., 2020). In another recent and significant case study, researchers from a number of collaborating universities and institutions in the United Kingdom demonstrated the effectiveness of remdesivir in the treatment of Covid-19 infection (Buckland et al., 2020). The study of a 31-years old man who tested positive for SARS-CoV-2 and who had a rare genetic condition, X-linked agammaglobulinaemia (XLA), that affects the body's ability to produce antibodies to fight infection provided rare insights in to the effectiveness of remdesivir as a treatment for coronavirus infection. Remdesivir (GS-5734) is a monophosphoramidate prodrug of an adenosine analogue and has a broad antiviral spectrum including coronaviruses as described in the DrugBank (Wishart et al., 2018) (Accession Number:DB14761) and is the first approved treatment for severe Covid-19 (Malin et al., 2020).

In light of the recent reports, I intended to evaluate whether any mutations known, so far, relative to the reference protein sequence was likely to affect binding of remdesivir to the RdRp protein at its binding site and therefore possibly confer drug-resistance.

In this work, all mutations in the 10,929 human SARS-CoV-2 RdRp proteins representing six geographical locations; Africa, Asia, Europe, North America, Oceania and South America identified earlier (Guruprasad, 2020c) were mapped on to the protein sequence and the three-dimensional structure. The proximity of mutation sites within 3.2 Å interacting distance from remdesivir and to the functionally important residues, ligands, solvent defined in the cryo-electron microscopy RdRp protein structures complexes available in the Protein Data Bank were analysed.

Materials and Methods

Six representative RdRp protein sequences were considered, one for each geographical location, in order to perform a multiple sequence alignment. The protein sequences considered were identical to the reference human SARS-CoV-2 RdRp protein sequence from Wuhan, China (NCBI Accession code: YP_009725307.1) (<https://www.ncbi.nlm.nih.gov/protein/>) and were aligned using the CLUSTAL OMEGA software (Sievers et al., 2011) available at (<https://www.ebi.ac.uk>). The RdRp protein sequence in the cryo-electron microscopy structure

(PDB code: 7BV2) was also included in the alignment. The mutations in RdRp protein were derived from my earlier work on the mutation analyses of polyproteins in human SARS-CoV-2 orf1AB gene compared with the equivalent reference polyproteins in human SARS-CoV-2 orf1AB gene from Wuhan, China (NCBI Accession code: YP_009724389.1) (Guruprasad, 2020c).

Accordingly, the mutations in the RdRp proteins corresponding to each geographical location was highlighted (in bold). Mutations at a particular site occurring in RdRp proteins representing one or more geographical locations were highlighted (in yellow) along the RdRp sequence corresponding to (PDB code:7BV2) in the alignment. The different domains/regions in RdRp protein described in the cryo-electron microscopy structures (PDB codes: 6M71, 7BV2) were also highlighted in different colours. The mutation density was evaluated as the ratio of the number of mutations observed to the number of amino acid residues in a particular domain/region of the RdRp protein. The mutation propensities were calculated as the ratio of mutation frequency to the amino acid sequence length frequency corresponding to the different domain/regions. The superposition of cryo-electron microscopy three-dimensional structures corresponding to the nsp12-nsp7-nsp8 complex with ligands, solvent, bound template primer RNA and triphosphate form of remdesivir at 2.5Å resolution (PDB code: 7BV2) (Yin et al., 2020) on to the human SARS CoV-2 RdRp (nsp12) in complex with cofactors nsp7, nsp8 determined at 2.9Å resolution (PDB code: 6M71) (Gao et al., 2020) was used to locate the positions of the mutations in the structure. A distance cut-off value $\leq 3.2 \text{ \AA}$ was used to define interacting residues between the RdRp protein and remdesivir, ligands or solvent. The functionally important residues in RdRp protein were noted according to the description in the literature on the cryo-electron microscopy structures (PDB codes: 7BV2, 6M71). The PyMol software (DeLano et al., 2002) was used for the graphics visualization, structural superposition, modelling and generating the figures.

Results and discussion

The 10,929 RdRp protein sequences represented in human SARS-CoV-2 from different geographical locations were as follows; Africa (137), Asia (999), Europe (378), North America (8860), Oceania (526), South America (29). Accordingly, the number of mutations observed in the RdRp proteins were; Africa (12), Asia (56), Europe (15), North America (186), Oceania (19), South America (1).

Mutations in human SARS CoV-2 RdRp protein

The RdRp protein comprises 932 amino acid residues and a total number of 222 distinct mutation sites were observed in 10,929 RdRp proteins corresponding to ~23.8% of the protein. The mutation site residues are listed in **Table 1**. The mutations observed in RdRp proteins corresponding to the orf1AB gene polyproteins comprising 7096 amino acid residues and distributed according to the geographical locations are listed in **Table 2a**. Likewise, the mutations in RdRp proteins corresponding to the orf1AB gene polyproteins comprising 7084-7095 amino acid residues are listed in **Table 2b**. The Pro4715Leu was the only mutation observed to be common among the human SARS-CoV-2 RdRp proteins in all the six geographical locations. Besides, the Pro4715Leu mutation is also the predominant mutation compared to the other mutations in the RdRp protein within each geographical location.

The mutation sites corresponding to the different geographical locations are highlighted (bold) in Figure 1. These mutation sites represent amino acid positions at which mutations are known to occur in at least one of the several human SARS-CoV-2 RdRp protein sequences corresponding to the geographical location compared to the reference RdRp protein sequence. Such mutation sites were highlighted (yellow) along the RdRp protein sequence as in (PDB code:7BV2) for mapping on to the cryo-electron microscopy three-dimensional structure. The different domains/regions in RdRp protein are highlighted in different colours in Figure 1. The Pro4715Leu mutation is associated with the interface domain in the RdRp protein as shown in Figure 1.

The mutation percentages corresponding to the different domains/regions in the RdRp proteins is shown in Figure 2. The maximum number of mutations in the protein were associated with the NiRAN domain (28.37%), followed by mutations in the Fingers sub-domain (24.32%), palm sub-domain (18.46%), thumbs sub-domain (13.96%), interface domain (13.51%), β -hairpin domain (0.9%) and the N-terminal region (0.45%). Accordingly, the mutation propensities were; N-terminal region (1.39), NiRAN domain (1.18), β -hairpin domain (0.38), Interface domain (1.08), Fingers sub-domain (0.82), Palm sub-domain (0.98) and Thumbs sub-domain (1.11). The mutations in β -hairpin domain were significantly low. This may be due to importance of the domain in maintaining overall stability of the protein (Gao et al., 2020).

In certain regions of the RdRp protein, the mutation sites were concentrated at three or four continuous amino acid positions as can be seen from Figure 1. These regions where the virus has accommodated the mutations are; TGT (4416-4418), VTY (4539-4541), TVQ (4581-4583), TLTR (4638-4641), VDT (4649-4651), PLV (4720-4722), AQD (4835-4837), YSDV (4998-5001), MVM (5058-5060), VNE (5134-5136), TET (5193-5195), VYLP (5219-5222), LTK (5261-5263), NQE (5266-5268), ADV (5270-5272), EAM (5314-5316). These regions may correspond to the ‘mutation hot spots’ in the human SARS-CoV-2 RdRp protein.

Remdesivir, inhibits virus proliferation and has been suggested as a potential treatment for Covid-19 viral infections with clinical potential (Wang et al., 2020, Holshue et al., 2020). Recent reports indicate the utility of remdesivir in the treatment of Covid-19 patients (Beigel et al., 2020, Dolin et al., 2020, Buckland et al., 2020). In order to understand whether any of the known mutations are likely to confer resistance to remdesivir, interactions made by the RdRp protein with remdesivir defined by a distance cut-off value $\leq 3.2 \text{ \AA}$ were identified. The cryo-electron microscopy structure of the complex with template RNA and remdesivir was used for this purpose (PDB code: 7BV2). The structure defines a partial double-stranded RNA template in the central channel of RdRp with remdesivir covalently incorporated into the primer strand at the first replicated base pair terminating the chain elongation (Yin et al., 2020).

The amino acid residues in the RdRp protein structure complex (PDB code:7BV2) that are within interacting distance ($\leq 3.2 \text{ \AA}$) of remdesivir (F86), magnesium (MG) and zinc (ZN) ions, RNA, pyrophosphate 2^- (POP) and solvent (HOH) molecules included the following 33 residues; His295, Cys301, Cys306, Cys310, Cys487, Ser501, Asn507, Gln541, Arg553, Lys577, Gly590, Ser592, Tyr595, Trp617, Asp618, Tyr619, Asp623, His642, Cys645, Cys646, Ser682, Gly683, Asp684, Ser759, Asp760, Asp761, Trp800, Glu811, Ser814, Arg836, Leu854, Arg858, Tyr915. The interaction types along with interacting distances are listed in Table 3. Asp760 is the only residue at a distance $\leq 3.2 \text{ \AA}$ from remdesivir according to the three-dimensional structure and this residue is not known to be mutated. Further, none of the 194 residues among the 222 distinct mutation sites that were mapped on to the protein three-dimensional structure was within 3.2 \AA distance from remdesivir at the drug binding site in the RdRp complex structure (PDB code:7BV2) as shown in Figure 3. Of the remaining 28 distinct mutation sites that could not be mapped owing to missing residues in the three-dimensional structure, S4393 is in the N-terminal region, Q4397, S4398, A4408, L4411, P4413, C4414, T4416, G4417, T4418, T4420, K4451, D4454, D4455, S4460, H4467, T4468, K4495, R4497,

D4499, D4501, M4502, P4504, I4506 are associated with the NiRAN domain and D5293, E5314, A5315, M5316 are associated with the Thumb sub-domain as observed from the multiple sequence alignment in Figure 1. The corresponding residue numbering in the three-dimensional structure (PDB code: 7BV2) are; S(1) in the N-terminal region, Q(5), S(6), A(16), L(19), P(21), C(22), T(24), G(25), T(26), T(28), K(59), D(62), D(63), S(68), H(75), T(76), K(103), R(105), D(107), D(109), M(110), P(112), I(114) in the NiRAN domain and D(901), E(922), A(923), M(924) in Thumb sub-domain. Examination of three-dimensional structure of the protein (PDB code: 7BV2) suggests these 28 mutation sites too are distant (≥ 3.2 Å) from remdesivir at the binding site.

Functionally important residues in RdRp

The protein residue interactions within 3.2 Å with different ligands were as follows: Asp760, Asp761, Trp800 and Glu811 with magnesium ions (MG: 1004, 1005, 101). His295, Cys301, Cys306, Cys310, Cys487, His642, Cys645, Cys646 with zinc ions (ZN: 1001, 1002). Arg553, Asp623 with POP (1003). Ser501, Asn507, Gln541, Lys577, Gly590, Ser592, Tyr595, Ser682, Gly683, Asp684, Ser759, Asp760, Ser814, Arg836, Leu854, Arg858, Tyr915 with template RNA primer in the complex. Trp617, Asp618, Tyr619, Asp761 and Glu811 with the solvent molecules (HOH: 1102, 1103, 1104 and 1105). It is interesting to note that none of the above functionally important residues, except Trp617, is known to be mutated in the human SARS-CoV-2 RdRp protein. Trp617 is involved in interaction at a distance of 2.33 Å with solvent (HOH:1102) in the structure (PDB code:7BV2).

Further, the divalent-cation binding residue; Asp618, the catalytic site residues; Ser759, Asp760, Asp761 associated with the Palm sub-domain, and residues; Lys545, Arg553 and Arg555 that form the NTP entry channel associated with the Fingers sub-domain (see Figure 1) are also not known to be mutated in the human SARS-CoV-2 RdRp protein. The β -hairpin is stabilized by interactions with the NiRAN domain and Palm sub-domain and these interactions are important for stability of the overall protein structure (Gao et al., 2020). The β -hairpin residues; Arg33, Phe35, Asp36, Tyr38, Val42, Phe45 and Phe48 make interactions with residues; Lys121, Tyr122, Asp126, Tyr129, Thr206, Asp208 and Ser236 in the NiRAN and with residues; Asp711, Asn713, His725, Tyr728, Glu729 and Arg733 in Palm sub-domain, according to the three-dimensional structure (PDB code:6M71). Except Asn713 and His725, none of the residues mentioned above are known to be mutated in the RdRp protein (refer

Figure 1). The residue Asn713 (numbered 5105 in the alignment shown in Figure 1) and His725 (5117) are known to be associated with the following mutations; N5105S and H5117Y (observed in North America) and H5117Y (observed in Asia) (refer Table 2a). However, an examination of these two residues in the protein three-dimensional structure (PDB code:7BV2) showed they are near the protein surface as in Figure 4. Further, models of the mutants; N713S and H725Y suggested that the mutated residues are not likely to cause steric hindrance.

The polyproteins of human SARS-CoV-2 orf1AB gene comprising 7084-7095 amino acid residues, represented 158 RdRp proteins that comprised deletions and certain additional mutations. Their distribution according to the geographical locations were; Africa (2), Asia (9), Europe (7), North America (113), Oceania (26) and South America (1). The deletion mutations were associated with the 9-residues region; Y4738-V4746 (YHFRELGVV) in the RdRp protein from Africa. These deletion mutations are associated with the interface domain (refer Figure 1). The location of equivalent mutation sites (Tyr346-Val354) in three-dimensional structure (PDB code:7BV2) suggest they are distant ($\geq 3.2 \text{ \AA}$) from remdesivir binding site. The additional mutation sites; V4503L, A4591V (North America) are associated with NiRAN domain. The equivalent residues; Val111, A199 (according to numbering in PDB code:7BV2) are also, accordingly, distant from remdesivir.

In summary, of the 222 mutation sites known so far, in human SARS-CoV-2 RdRp proteins relative to the reference protein sequence and representing the six geographical locations, none were within 3.2 \AA interacting distance of remdesivir at the drug binding site as observed from the cryo-electron microscopy three-dimensional structure (PDB code:7BV2). Thereby, the known mutations in human SARS-CoV-2 RdRp protein are not likely to offer resistance to the binding of remdesivir. Also, none of the mutations were associated with the functionally important residues.

Conclusions

The human SARS-CoV-2 RdRp protein is known to be associated with at least 222 distinct mutation sites compared to the reference RdRp protein sequence in the first reported human SARS-CoV-2 genome isolate from Wuhan, China during January 2020. It is interesting to note from the cryo-electron microscopy three-dimensional structures for the protein complexes available in the Protein Data Bank, that none of the mutations are within 3.2 Å interacting distance from remdesivir at the drug binding site. This suggests that the mutations observed in human SARS-CoV-2 RdRp proteins are not likely to confer resistance to remdesivir. Further, none of the mutations are associated with the functionally important residues. Thereby, the capability of the novel human SARS-CoV-2 to survive and propagate itself in the human host may not be affected by mutations discussed in the present work.

References

- Beigel, J. H., Tomashek, K. M., Dodd, L. E., Mehta, A. K., Zingman, B. S., Kalil, A. C., ... & de Castilla, D. L. (2020). Remdesivir for the treatment of Covid-19. *New England Journal of Medicine*. 383 (19);1813-1826.
- Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N., Bourne, P. E. (2000). The Protein Data Bank *Nucleic Acids Research*, 28, 235-242.
- Buckland, M. S., Galloway, J. B., Fhogartaigh, C. N., Meredith, L., Provine, N. M., Bloor, S., ... & Mann, T. (2020). Treatment of COVID-19 with remdesivir in the absence of humoral immunity: a case report. *Nature communications*, 11(1), 1-11. Buckland et al., 2020
- DeLano, W. L. (2002). PyMOL: An open-source molecular graphics tool. *CCP4 Newsletter on protein crystallography*, 40, 44-53.
- Dolin, R., & Hirsch, M. S. (2020). Remdesivir—An Important First Step. 383(19);1886-1887.
- Gao, Y., Yan, L., Huang, Y., Liu, F., Zhao, Y., Cao, L., ... & Ge, J. (2020). Structure of the RNA-dependent RNA polymerase from COVID-19 virus. *Science*, 368(6492), 779-782.
- Guruprasad, K. (2020a). Amino Acid Mutations in the Protein Sequences of Human SARS CoV-2 Indian Isolates Compared to Wuhan-Hu-1 Reference Isolate from China. *ChemRxiv*. Preprint. <https://doi.org/10.26434/chemrxiv.12300860.v1>.
- Guruprasad, K. (2020b): Mapping Mutations in Proteins of SARS CoV-2 Indian Isolates on to the Three-Dimensional Structures. ChemRxiv. Preprint. <https://doi.org/10.26434/chemrxiv.12683771.v1>
- Guruprasad, K. (2020c): Geographical Distribution of Amino Acid Mutations in Human SARS-CoV-2 Orf1ab Poly-Proteins Compared to the Equivalent Reference Proteins from China. ChemRxiv. Preprint. <https://doi.org/10.26434/chemrxiv.12951617.v1>

Holshue, M. L., DeBolt, C., Lindquist, S., Lofy, K. H., Wiesman, J., Bruce, H., ... & Diaz, G. (2020). First case of 2019 novel coronavirus in the United States. *New England Journal of Medicine*, 382, 929-936.

Malin, J. J., Suárez, I., Priesner, V., Fätkenheuer, G., & Rybniker, J. (2020). Remdesivir against COVID-19 and other viral diseases. *Clinical microbiology reviews*, 34(1).

Sievers, F., Wilm, A., Dineen, D., Gibson, T. J., Karplus, K., Li, W., ... & Thompson, J. D. (2011). Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Molecular systems biology*, 7(1), 539.

Subissi, L., Posthuma, C. C., Collet, A., Zevenhoven-Dobbe, J. C., Gorbalenya, A. E., Decroly, E., ... & Imbert, I. (2014). One severe acute respiratory syndrome coronavirus protein complex integrates processive RNA polymerase and exonuclease activities. *Proceedings of the National Academy of Sciences*, 111(37), E3900-E3909.

Wang, M., Cao, R., Zhang, L., Yang, X., Liu, J., Xu, M., ... & Xiao, G. (2020). Remdesivir and chloroquine effectively inhibit the recently emerged novel coronavirus (2019-nCoV) in vitro. *Cell research*, 30(3), 269-271.

Wishart, D. S., Feunang, Y. D., Guo, A. C., Lo, E. J., Marcu, A., Grant, J. R., ... & Assempour, N. (2018). DrugBank 5.0: a major update to the DrugBank database for 2018. *Nucleic acids research*, 46(D1), D1074-D1082.

Wu, F., Zhao, S., Yu, B., Chen, Y. M., Wang, W., Song, Z. G., ... & Yuan, M. L. (2020). A new coronavirus associated with human respiratory disease in China. *Nature*, 579(7798), 265-269.

Yin, W., Mao, C., Luan, X., Shen, D. D., Shen, Q., Su, H., ... & Chang, S. (2020). Structural basis for inhibition of the RNA-dependent RNA polymerase from SARS-CoV-2 by remdesivir. *Science*.

Conflict of interest

The author declares no conflict of interest

Funding

None

Table 1. The 222 distinct mutation sites in human SARS-CoV-2 RdRp proteins.

S4393, Q4397, S4398, A4408, L4411, P4413, C4414, T4416, G4417, T4418, T4420, A4435, G4436, K4451, D4454, D4455, S4460, H4467, T4468, T4477, N4480, K4483, P4486, A4487, A4489, K4495, R4497, D4499, D4501, M4502, P4504, I4506, R4510, M4516, V4520, H4525, D4532, T4533, E4536, V4539, T4540, Y4541, D4546, D4556, I4563, L4564, G4571, R4573, A4577, T4581, V4582, Q4583, M4588, I4593, V4594, W4608, F4614, Q4616, P4619, S4621, P4624, Y4630, T4638, L4639, T4640, R4641, E4646, V4649, D4650, T4651, T4654, K4655, T4668, E4670, R4671, D4676, R4677, Q4684, T4685, N4689, V4691, D4696, V4707, T4711, F4713, P4715, P4720, L4721, V4722, I4725, G4729, T4736, Y4738, L4743, V4746, Q4749, E4762, L4764, M4772, A4774, A4792, T4794, V4797, A4798, V4802, N4806, D4810, A4815, K4818, K4822, V4827, E4828, A4835, Q4836, D4837, A4840, A4841, R4849, P4853, M4855, C4856, I4858, V4865, K4870, D4873, N4883, L4906, D4909, D4915, A4918, A4921, P4929, S4941, A4946, C4955, M4958, H4964, A4973, G4976, A4977, V4980, I4981, G4988, H4991, M4993, T4996, Y4998, S4999, D5000, V5001, H5005, M5007, W5009, M5021, I5024, L5028, L5030, A5031, T5035, T5036, S5039, L5040, A5048, M5058, V5059, M5060, A5091, N5105, K5110, H5117, N5126, D5128, V5129, T5131, V5134, N5135, E5136, M5148, L5150, V5155, A5163, G5166, L5167, A5169, K5172, F5185, T5193, E5194, T5195, T5198, H5202, Q5214, G5215, D5217, V5219, Y5220, L5221, P5222, V5240, L5261, T5262, K5263, N5266, Q5267, E5268, A5270, D5271, V5272, H5274, H5284, D5285, D5293, M5298, T5300, N5301, T5304, S5305, P5310, E5311, E5314, A5315, M5316

Table 2a. Geographical distribution of mutations in human SARS-CoV-2 RdRp proteins observed in polyproteins of orf1AB gene comprising 7096 amino acid residues.

AFRICA	ASIA	EUROPE	NORTH AMERICA	OCEANIA	SOUTH AMERICA
L4411F	Q4397P	T4418I	S4393L	G4436V	P4715L
A4489V	T4418I	N4480S	S4398L	D4455Y	
R4510C	K4451N	A4489V	A4408V	P4486L	
D4546Y	K4483R	T4638I	P4413L	A4489V	
V4649F	P4486L	T4668M	C4414F	I4563V	
P4715L	A4487V	P4715L	T4416I	P4619L	
K4870R	A4489V	M4772I	G4417Y	T4640I	
T5035I	R4510C	A4841V	T4418I	T4654A	
S5039I	M4516I	A4918V	T4420A	R4671S	
G5215S	D4532G	S5039I	A4435V	R4677H	
S5305L	E4536D	K5110R	G4436V	V4707F	
E5314D	T4540I	V5219L	K4451N	P4715L	
	Y4541C	L5261F	D4454Y	A4841V	
	L4564V	A5270V	D4455Y	K4870N	
	G4571S	S5305L	S4460F	A4921V	
	A4577V		H4467Y	H4964Y	
	M4588I		T4468I	I4981V	
	I4593L		T4477S	T5036M	
	W4608R		K4483R	V5240L	
	Q4616K		A4487T		
	P4619S		A4489V		
	L4639F		K4495R		
	E4646D		R4497I		
	E4670D		D4499N		
	P4715L		D4501G		
	L4721I		M4502V		
	V4746A		P4504S		
	A4798V		I4506M		
	N4806S		M4516I		
	A4815V		V4520I		
	M4855I		H4525Y		
	C4856F		D4532Y		
	L4906F		T4533I		
	D4909G		V4539I		
	A4921S		D4546Y		
	H4991Y		D4556N		
	Y4998C		I4563M		
	S4999I		R4573C		
	L5030F		A4577V		
	T5035I		T4581K		
	T5036M		V4582L		

M5060I A5091S H5117Y T5131I V5155A E5194A Q5214H L5221F V5240L T5262I V5272I D5285Y M5298I T5300I S5305L		Q4583L M4588I I4593V V4594F F4614L P4619L S4621N P4624S Y4630H T4638I T4640I R4641K D4650Y T4651I K4655N R4671S D4676Y Q4684H T4685I N4689S V4691F D4696Y T4711I F4713L P4715L P4720S V4722E I4725T G4729C T4736I Y4738H L4743P V4746L Q4749H E4762G L4764F M4772I A4774V A4792S T4794I V4797F V4802I D4810N A4815V K4818N K4822N V4827F E4828G A4835V Q4836H	
----------------------------------------------------------------------------------------------------------------------------------------------------	--	-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	--

			D4837G A4840V A4841V R4849C P4853T C4856F I4858V V4865F K4870N D4873G N4883S L4906F D4909G D4915N A4921V P4929S S4941N A4946V C4955F M4958I A4973T G4976V A4977T V4980L G4988S M4993I T4996I S4999I D5000G V5001I H5005Y M5007V W5009L M5021V I5024V L5028F L5030F A5031V T5035I T5036M S5039I L5040F A5048T M5058I V5059I M5060I A5091S N5105S K5110R H5117Y		
--	--	--	------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	--	--

			N5126T D5128N V5129F T5131I V5134M N5135S E5136D M5148I L5150I A5163V G5166S L5167V A5169T K5172T F5185L T5193I T5195A T5198I H5202Y Q5214H G5215C D5217Y V5219A Y5220H P5222S V5240L T5262I K5263R N5266D Q5267R E5268D D5271Y H5274Y H5284Y D5293Y M5298I T5300I N5301D T5304N S5305L P5310S E5311D E5314D A5315T M5316V		
--	--	--	----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	--	--

Table 2b. Geographical distribution of mutations in human SARS-CoV-2 RdRp proteins observed in polyproteins of orf1AB gene comprising 7084-7095 amino acid residues.

AFRICA	ASIA	EUROPE	NORTH AMERICA	OCEANIA	SOUTH AMERICA
P4715L Y4738- H4739- F4740- R4741- E4742- L4743- G4744- V4745- V4746-	D4676Y P4715L	P4715L	A4489V V4503L A4591V P4715L C4955F A4973G A4977S T4996I V5240L	P4715L K5110N	P4715L

Table 3. RdRp protein-ligand interactions ≤ 3.2 Å (in PDB code:7BV2)

Ligand No.	Ligand type	Residue No.	Residue Name	Distance (in Å)
102	F86	760	ASP	2.99
1102	HOH	617	TRP	2.33
1103	HOH	811	GLU	2.73
1103	HOH	811	GLU	2.62
1103	HOH	811	GLU	2.51
1104	HOH	619	TYR	2.63
1105	HOH	618	ASP	2.80
1105	HOH	761	ASP	2.94
1004	MG	760	ASP	3.17
1005	MG	800	TRP	3.18
1005	MG	811	GLU	3.09
101	MG	761	ASP	3.08
1001	ZN	295	HIS	2.41
1001	ZN	301	CYS	2.38
1001	ZN	306	CYS	3.18
1001	ZN	306	CYS	2.31
1001	ZN	310	CYS	2.29
1002	ZN	487	CYS	2.39
1002	ZN	642	HIS	3.07
1002	ZN	642	HIS	2.21
1002	ZN	645	CYS	2.63
1002	ZN	646	CYS	2.40
1003	POP	553	ARG	2.70
1003	POP	553	ARG	2.84
1003	POP	623	ASP	2.82
15	A	854	LEU	3.15
17	U	858	ARG	2.44
18	U	836	ARG	3.10
18	U	836	ARG	3.13
19	A	836	ARG	2.92
20	U	814	SER	2.70
20	U	760	ASP	3.09
20	U	759	SER	3.04
20	U	759	SER	2.81
8	U	507	ASN	2.87
8	U	541	GLN	3.18
9	U	501	SER	2.78
9	U	501	SER	3.06
10	U	683	GLY	3.19
10	U	682	SER	2.92
10	U	683	GLY	3.15
10	U	683	GLY	3.16
11	A	683	GLY	3.13
11	A	684	ASP	3.06
13	A	577	LYS	3.20
13	A	590	GLY	3.16
14	A	592	SER	2.60

16	U	595	TYR	2.88
16	U	915	TYR	3.08
16	U	915	TYR	3.02

RdRp_N.America	4861	LLFVVEVVDK Y FD C YDGGCINANQ V IVNNLDKSAGFPFNKWG K ARLY D SM S YEDQ D ALF	4920
RdRp_S.America		LLFVVEVVDK Y FD C YDGGCINANQ V IVNNLDKSAGFPFNKWG K ARLY D SM S YEDQ D ALF	4920
RdRp_Africa		LLFVVEVVDK Y FD C YDGGCINANQ V IVNNLDKSAGFPFNKWG K ARLY D SM S YEDQ D ALF	4920
RdRp_Asia		LLFVVEVVDK Y FD C YDGGCINANQ V IVNNLDKSAGFPFNKWG K ARLY D SM S YEDQ D ALF	4920
RdRp_Oceania		LLFVVEVVDK Y FD C YDGGCINANQ V IVNNLDKSAGFPFNKWG K ARLY D SM S YEDQ D ALF	4920
RdRp_Europe		LLFVVEVVDK Y FD C YDGGCINANQ V IVNNLDKSAGFPFNKWG K ARLY D SM S YEDQ D ALF	4920
7BV2_1 Chn_A		LLFVVEVVDK Y FD C YDGGCINANQ V IVNNLDKSAGFPFNKWG K ARLY D SM S YEDQ D ALF	528
RdRp_N.America	4921	AYTKRNVIP T ITQ M NLKYA I SAK N RARTVAGV S IC S T M TNRQ F HQ K LLK S IA A TRGAT V	4980
RdRp_S.America		AYTKRNVIP T ITQ M NLKYA I SAK N RARTVAGV S IC S T M TNRQ F HQ K LLK S IA A TRGAT V	4980
RdRp_Africa		AYTKRNVIP T ITQ M NLKYA I SAK N RARTVAGV S IC S T M TNRQ F HQ K LLK S IA A TRGAT V	4980
RdRp_Asia		AYTKRNVIP T ITQ M NLKYA I SAK N RARTVAGV S IC S T M TNRQ F HQ K LLK S IA A TRGAT V	4980
RdRp_Oceania		AYTKRNVIP T ITQ M NLKYA I SAK N RARTVAGV S IC S T M TNRQ F HQ K LLK S IA A TRGAT V	4980
RdRp_Europe		AYTKRNVIP T ITQ M NLKYA I SAK N RARTVAGV S IC S T M TNRQ F HQ K LLK S IA A TRGAT V	4980
7BV2_1 Chn_A		AYTKRNVIP T ITQ M NLKYA I SAK N RARTVAGV S IC S T M TNRQ F HQ K LLK S IA A TRGAT V	588
RdRp_N.America	4981	IGTSK F YGGW H N M LK T V S D V EN P H L M G W D Y P K C D R A M P N L R I M A S L V L A R K H T T C C S L	5040
RdRp_S.America		IGTSK F YGGW H N M LK T V S D V EN P H L M G W D Y P K C D R A M P N L R I M A S L V L A R K H T T C C S L	5040
RdRp_Africa		IGTSK F YGGW H N M LK T V S D V EN P H L M G W D Y P K C D R A M P N L R I M A S L V L A R K H T T C C S L	5040
RdRp_Asia		IGTSK F YGGW H N M LK T V S D V EN P H L M G W D Y P K C D R A M P N L R I M A S L V L A R K H T T C C S L	5040
RdRp_Oceania		IGTSK F YGGW H N M LK T V S D V EN P H L M G W D Y P K C D R A M P N L R I M A S L V L A R K H T T C C S L	5040
RdRp_Europe		IGTSK F YGGW H N M LK T V S D V EN P H L M G W D Y P K C D R A M P N L R I M A S L V L A R K H T T C C S L	5040
7BV2_1 Chn_A		IGTSK F YGGW H N M LK T V S D V EN P H L M G W D Y P K C D R A M P N L R I M A S L V L A R K H T T C C S L	648
RdRp_N.America	5041	SHRFYRLAN E CAQ V L S E M V M C G G S L V V K P G G T S S G D A T T A Y A N S V F N I C Q A V T A N V N A L L	5100
RdRp_S.America		SHRFYRLAN E CAQ V L S E M V M C G G S L V V K P G G T S S G D A T T A Y A N S V F N I C Q A V T A N V N A L L	5100
RdRp_Africa		SHRFYRLAN E CAQ V L S E M V M C G G S L V V K P G G T S S G D A T T A Y A N S V F N I C Q A V T A N V N A L L	5100
RdRp_Asia		SHRFYRLAN E CAQ V L S E M V M C G G S L V V K P G G T S S G D A T T A Y A N S V F N I C Q A V T A N V N A L L	5100
RdRp_Oceania		SHRFYRLAN E CAQ V L S E M V M C G G S L V V K P G G T S S G D A T T A Y A N S V F N I C Q A V T A N V N A L L	5100
RdRp_Europe		SHRFYRLAN E CAQ V L S E M V M C G G S L V V K P G G T S S G D A T T A Y A N S V F N I C Q A V T A N V N A L L	5100
7BV2_1 Chn_A		SHRFYRLAN E CAQ V L S E M V M C G G S L V V K P G G T S S G D A T T A Y A N S V F N I C Q A V T A N V N A L L	708
RdRp_N.America	5101	STDG N KI A D K Y V R N L Q H R L Y E C L Y R N R D V D T D F V N E F Y A L R K H F S M M I L S D D A V V C F N S	5160
RdRp_S.America		STDG N KI A D K Y V R N L Q H R L Y E C L Y R N R D V D T D F V N E F Y A L R K H F S M M I L S D D A V V C F N S	5160
RdRp_Africa		STDG N KI A D K Y V R N L Q H R L Y E C L Y R N R D V D T D F V N E F Y A L R K H F S M M I L S D D A V V C F N S	5160
RdRp_Asia		STDG N KI A D K Y V R N L Q H R L Y E C L Y R N R D V D T D F V N E F Y A L R K H F S M M I L S D D A V V C F N S	5160
RdRp_Oceania		STDG N KI A D K Y V R N L Q H R L Y E C L Y R N R D V D T D F V N E F Y A L R K H F S M M I L S D D A V V C F N S	5160
RdRp_Europe		STDG N KI A D K Y V R N L Q H R L Y E C L Y R N R D V D T D F V N E F Y A L R K H F S M M I L S D D A V V C F N S	5160
7BV2_1 Chn_A		STDG N KI A D K Y V R N L Q H R L Y E C L Y R N R D V D T D F V N E F Y A L R K H F S M M I L S D D A V V C F N S	768
RdRp_N.America	5161	TYAS Q GL V AS I K N F K SV L Y Y Q N N V F M SE A K C W T E T D L T K G P H E F C S Q H T M L V K Q G D D Y V	5220
RdRp_S.America		TYAS Q GL V AS I K N F K SV L Y Y Q N N V F M SE A K C W T E T D L T K G P H E F C S Q H T M L V K Q G D D Y V	5220
RdRp_Africa		TYAS Q GL V AS I K N F K SV L Y Y Q N N V F M SE A K C W T E T D L T K G P H E F C S Q H T M L V K Q G D D Y V	5220
RdRp_Asia		TYAS Q GL V AS I K N F K SV L Y Y Q N N V F M SE A K C W T E T D L T K G P H E F C S Q H T M L V K Q G D D Y V	5220
RdRp_Oceania		TYAS Q GL V AS I K N F K SV L Y Y Q N N V F M SE A K C W T E T D L T K G P H E F C S Q H T M L V K Q G D D Y V	5220
RdRp_Europe		TYAS Q GL V AS I K N F K SV L Y Y Q N N V F M SE A K C W T E T D L T K G P H E F C S Q H T M L V K Q G D D Y V	5220
7BV2_1 Chn_A		TYAS Q GL V AS I K N F K SV L Y Y Q N N V F M SE A K C W T E T D L T K G P H E F C S Q H T M L V K Q G D D Y V	828
RdRp_N.America	5221	LPYPDPSR I L G A C F V D D I V K T D G T L M I E R F V S L A I D A Y P L T K H P N Q E Y A D V F H L Y L Q Y I	5280
RdRp_S.America		LPYPDPSR I L G A C F V D D I V K T D G T L M I E R F V S L A I D A Y P L T K H P N Q E Y A D V F H L Y L Q Y I	5280
RdRp_Africa		LPYPDPSR I L G A C F V D D I V K T D G T L M I E R F V S L A I D A Y P L T K H P N Q E Y A D V F H L Y L Q Y I	5280
RdRp_Asia		LPYPDPSR I L G A C F V D D I V K T D G T L M I E R F V S L A I D A Y P L T K H P N Q E Y A D V F H L Y L Q Y I	5280
RdRp_Oceania		LPYPDPSR I L G A C F V D D I V K T D G T L M I E R F V S L A I D A Y P L T K H P N Q E Y A D V F H L Y L Q Y I	5280
RdRp_Europe		LPYPDPSR I L G A C F V D D I V K T D G T L M I E R F V S L A I D A Y P L T K H P N Q E Y A D V F H L Y L Q Y I	5280
7BV2_1 Chn_A		LPYPDPSR I L G A C F V D D I V K T D G T L M I E R F V S L A I D A Y P L T K H P N Q E Y A D V F H L Y L Q Y I	888
RdRp_N.America	5281	RKL H DEL T G H M L D M Y S V M L T N D N T S R Y W E P E F Y E A M Y T P H T V L Q	5324
RdRp_S.America		RKL H DEL T G H M L D M Y S V M L T N D N T S R Y W E P E F Y E A M Y T P H T V L Q	5324
RdRp_Africa		RKL H DEL T G H M L D M Y S V M L T N D N T S R Y W E P E F Y E A M Y T P H T V L Q	5324
RdRp_Asia		RKL H DEL T G H M L D M Y S V M L T N D N T S R Y W E P E F Y E A M Y T P H T V L Q	5324
RdRp_Oceania		RKL H DEL T G H M L D M Y S V M L T N D N T S R Y W E P E F Y E A M Y T P H T V L Q	5324
RdRp_Europe		RKL H DEL T G H M L D M Y S V M L T N D N T S R Y W E P E F Y E A M Y T P H T V L Q	5324
7BV2_1 Chn_A		RKL H DEL T G H M L D M Y S V M L T N D N T S R Y W E P E F Y E A M Y T P H T V L Q	932

Note: Mutations sites mapped on to the sequence of human SARS-CoV-2 RdRp proteins corresponding to the six geographical locations. The mutations are with reference to the human

SARS-CoV-2 RdRp protein sequence from Wuhan, China as in the orf1AB gene comprising 7096 amino acid residue polyproteins identified earlier (Guruprasad, 2020c). The RdRp protein sequence in polyproteins of the human SARS-CoV-2 orf1AB gene (NCBI Accession code: YP_009724389.1) comprises 932 amino acid residues and corresponds to the region 4393-5324. The RdRp protein sequence numbering according to the three-dimensional structure (as in PDB code:7BV2) is also included. The different domains/regions in RdRp protein are according to definition in the cryo-electron microscopy structure (PDB code:7BV2) (Yin et al., 2020). **Color code:** NiRAN domain (green), β -hairpin (cyan), Interface domain (gray), Fingers sub-domain (magenta), Palm sub-domain (red), Thumb sub-domain (blue). The representative mutation sites along the RdRp protein sequence are highlighted (yellow).

Figure 2. Mutation percentage corresponding to the different domains/regions in human SARS-CoV-2 RdRp proteins.

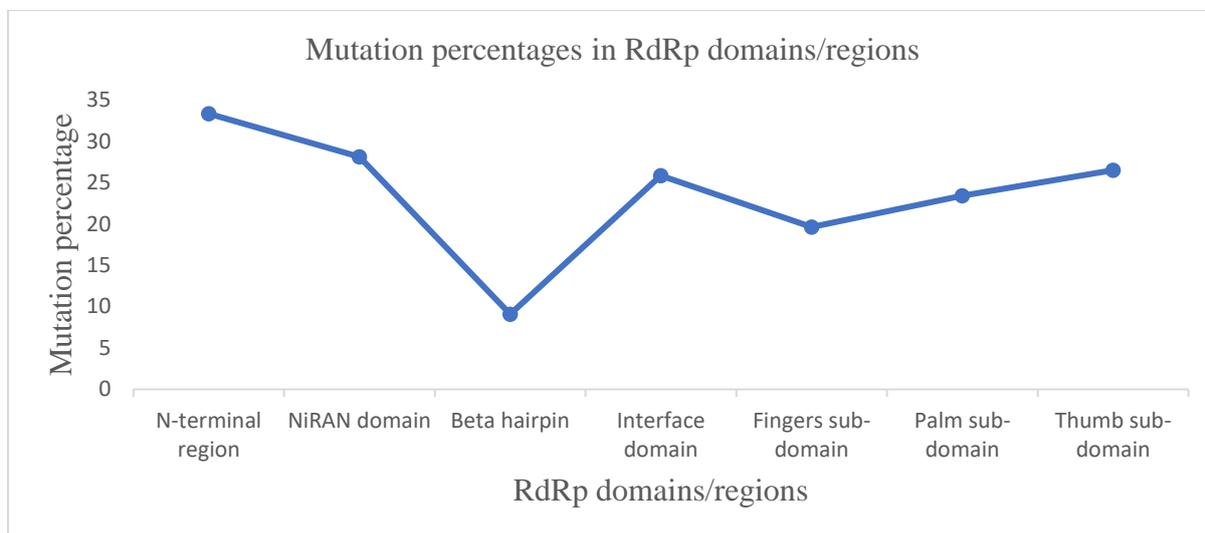
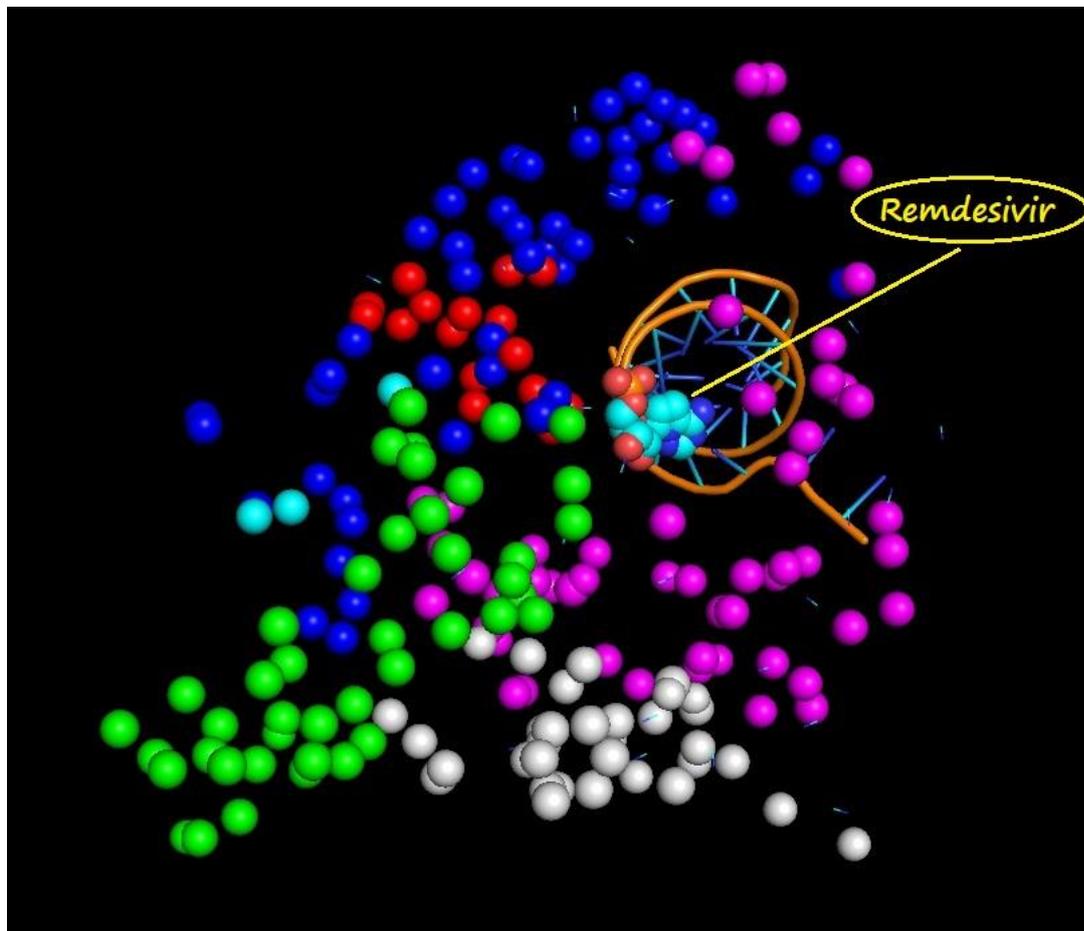
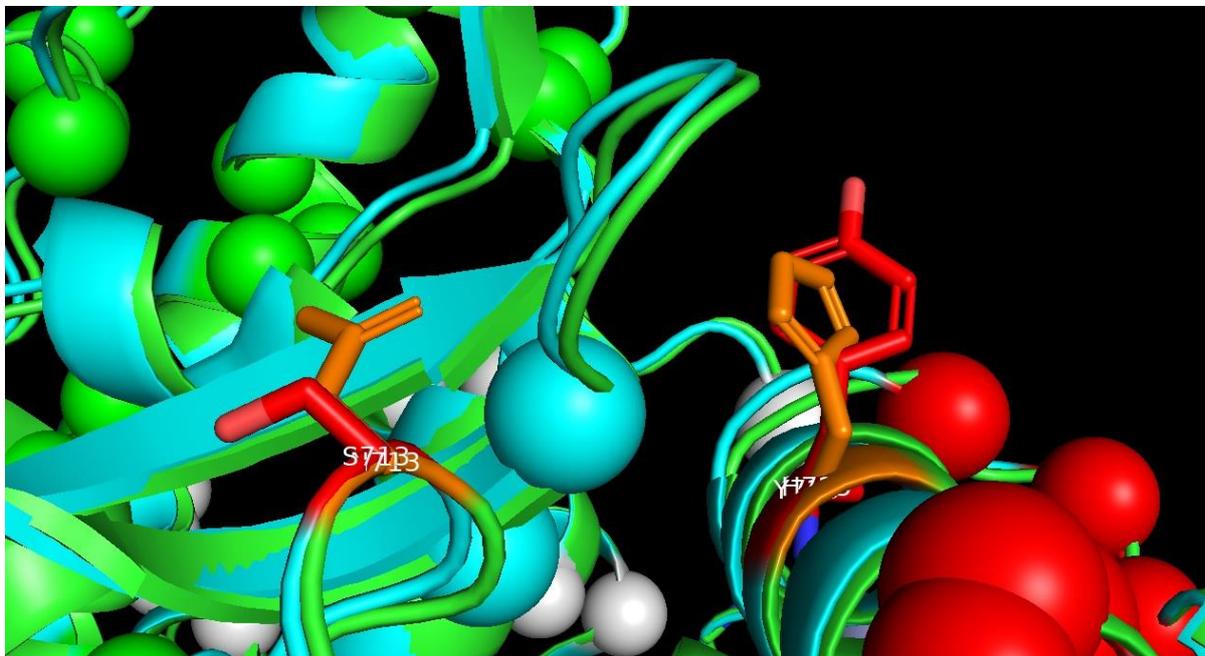


Figure 3. Mutation sites in human SARS-CoV-2 RdRp proteins mapped on to cartoon representation of the cryo-electron microscopy three-dimensional structure (PDB code:7BV2).



Note: The template primer RNA, triphosphate form of the bound remdesivir and solvent molecules are shown mapped on to the three-dimensional structure (PDB code:7BV2). The protein is excluded for the sake of clarity. The residues at the mutation sites corresponding to the different RdRp domains/regions are indicated as coloured spheres. The colour coding for the different domain/regions is same as described in Figure 1.

Figure 4. Models of the mutations; N713S and H725Y in Palm sub-domain close to the β -hairpin residues (Val31-Lys50) in (PDB code:7BV2).



Note: The interaction of the β -hairpin residues (Val31-Lys50) with residues in NiRAN domain and Palm sub-domain are important for stability of the human SARS-CoV-2 RdRp protein (Gao et al., 2020). Except, N5105 (or N713) and H5117 (or H725) in the Palm sub-domain, none of the other residues involved in the interactions described in the text are known to be mutated. Models of the mutations; N713S and H725Y (in PDB code:7BV2) suggest that the mutated residues are not likely to cause steric hindrance in the protein.