

Mapping mutations in proteins of SARS CoV-2 Indian isolates on to the three-dimensional structures

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

Mobile: +91 7337554324

Website: www.abreast.in

Abstract:

The amino acid residue mutations observed in SARS CoV-2 RNA dependent RNA polymerase, helicase, endoRNase and spike proteins from Indian isolates, relative to the reference SARS CoV-2 proteins from the Wuhan Hu-1 isolate from China, were mapped onto the protein three-dimensional structure templates available in the Protein Data Bank. The secondary structure conformations corresponding to the mutations, their locations and proximity to functionally important residues in the proteins and to the drug binding sites in RNA dependent RNA polymerase and endoRNase targets were analysed. Our analyses provide structural insights into the mutations in these SARS CoV-2 proteins.

Keywords:

SARS CoV-2 Indian isolate proteins, Amino acid mutations, RNA-dependent RNA polymerase, Helicase, EndoRNase, Spike glycoprotein, Secondary structure conformation, drug-binding sites

Introduction:

The current COVID-19 pandemic disease that has resulted in over 11 million coronavirus cases and above half a million deaths worldwide (<https://www.worldometers.info/coronavirus/>), seriously awaits intervention in the form of vaccine and drugs for the prevention and treatment of this deadly disease. Several organizations and research groups in the world are currently engaged in this endeavour to come up with a successful therapeutic treatment. A successful vaccine depends upon its utility

over populations worldwide. Mutations in SARS CoV-2 spike glycoproteins from several infected individuals representing different countries have been reported and the mutations in the spike protein have been mapped onto the protein structure (Korber et al., 2020). Also, two mutations in the spike protein of the first two SARS-CoV-2 sequences from India are mapped on to protein structure (Yadav et al., 2020). Based on the analysis of protein sequences from SARS CoV-2 representing 22 Indian isolates available in the publicly accessible NCBI databank (<http://www.ncbi.nlm.nih.gov>), we identified not only certain new mutations in the spike protein of SARS CoV-2 among the Indian isolates but also mutations in ten additional SARS CoV-2 proteins (Guruprasad, 2020).

In this work, we analyse the secondary structure conformations and map the mutation site residues for SARS CoV-2 proteins observed in the Indian isolates on to the protein three-dimensional structures of RdRp, helicase, endoRNase and spike proteins for which suitable templates are available in the Protein Data Bank (PDB). We further analyse the proximity of the mutation sites to functionally important residues in the protein three-dimensional structure and to the Remdesivir and Tipiracil drug binding sites in RdRp and endoRNase proteins, respectively.

Materials and Methods:

The structural templates for mapping the mutation sites of the SARS CoV2 proteins in the Indian isolates were obtained by searching the Protein Data Bank (PDB) (Berman et al., 2000) (<http://www.rcsb.org>) with the PSI-BLASTP program (Schäffer et al., 2001) available at the NCBI website (<http://www.ncbi.nlm.nih.gov>). The reference protein sequences of SARS CoV-2 Wuhan-Hu-1 isolate were used as the query. The secondary structure conformations were inferred from PDBsum (Laskowski et al., 2001) (<https://www.ebi.ac.uk/pdbsum>). The PyMol software (DeLano, 2002) was used for molecular visualization and structure superposition.

Results and Discussion:

Among the eleven proteins; Nsp2, Nsp3, Nsp4, Nsp6, RNA-dependent RNA polymerase (RdRp), helicase, endoRNase, spike glycoprotein, nucleocapsid phosphoprotein 'N', orf3a and orf8 proteins associated with mutations in SARS CoV-2 Indian isolates (Guruprasad,

2020), suitable structural templates to map the mutations were available for the SARS CoV-2 RdRp, endoRNase, spike and helicase proteins. The secondary structure conformations associated with the mutation site residues in the protein structures are shown in Table 1.

Mapping the mutation sites in RNA-dependent RNA polymerase (Nsp12)

The RdRp (or Nsp12) protein plays a central role in replication and transcription cycle of the virus by catalyzing the synthesis of viral RNA with assistance of cofactors (Subissi et al., 2014). The protein with PDB code: 6M71 was selected as template for RdRp that corresponds to the electron microscopy structure of human SARS CoV-2 RdRp in complex with cofactors; Nsp7, Nsp8 determined at 2.9Å resolution (Gao et al., 2020). The overall fold of the protein complex is shown in Figure 1 along with mutation site residues (shown as red spheres) and labelled for the A97V, A185V, I201L, P323L, L329I, A406V, T644M, V880I mutations. The A97, A185, I201 mutation sites are associated with nidovirus-specific N-terminal extension domain (cyan), P323 and L329 mutation sites with interface domain (yellow), A406 and T644 mutations with fingers subdomain (blue) and V880 mutation with thumb subdomain (green) in the polymerase domain. None of the mutation site residues are close to the divalent-cation-binding residue D618, or to the catalytic residues $S_{759}DD_{761}$ in the palm subdomain (orange), or to the NTP entry channel in RdRp formed by Lys545, Arg553 and Arg555 in the fingers domain (blue) as shown in Figure 2. The residues; Arg33, Phe35, Asp36, Tyr38, Val42, Phe45 and Phe48 in the β -hairpin (magenta surface) make contacts with residues; Lys121, Tyr122, Asp126, Tyr129, Thr206, Asp208 and Ser236 in NiRAN domain (cyan surface) and with residues; Asp711, Asn713, Tyr728, Arg733, His725 and Glu729 in palm sub-domain (orange surface) and these interactions are important for stability of the overall protein structure (Gao et al., 2020). It can be seen from Figure 3, that none of the residues at the mutation sites (red spheres) are close to the residues at interaction interface of β -hairpin with the NiRAN domain or β -hairpin with palm sub-domain that could affect overall stability. We infer from Figure 4, that mutation of leucine to isoleucine at position 329 would not affect the hydrogen bond involving the main-chain atoms of Leu329 and Thr344. Remdesivir drug inhibits virus proliferation and is suggested as a potential treatment for COVID-19 viral infections with clinical potential (Wang et al., 2020, Holshue et al., 2020). Based on the superposition of the electron microscopy structure of human SARS CoV-2 RdRp in complex with cofactors (PDB code: 6M71) (magenta) and the electron microscopy structure of the Nsp12-Nsp7-Nsp8 complex bound to the template primer RNA and triphosphate form of Remdesivir (PDB code:

7BV2) (green) (Yin et al., 2020), we note that the SARS CoV-2 Indian isolate mutation site residues are distant from the Remdesivir drug binding site in the RdRp protein as shown in Figure 5. The protein is excluded for the sake of clarity and the mutation site residues (red spheres) are shown in Figure 6, along with the zinc ions (green spheres) and magnesium ions (purple spheres) as in the crystal structure. The side-chains of residues in the polymerase domain important for the functional activity and residues that interact with the β -hairpin required for maintaining overall stability of the structure are also shown.

Ala97 which is in a loop conformation is present immediately after a type I beta-turn and is located near the protein surface. Ala185 is in the middle of helix. Ile201 is at the beginning of a beta-strand near the protein surface and its side-chain makes close contacts with side-chains of Ile86 and Leu90 on a neighbouring helix. Pro323 is at the start of a G-type helix near the protein surface and at the junction between helices; H13 and H15. Leu329 is associated with a twisted anti-parallel beta-strand and is near the protein surface with side-chain close to Cys114, Pro116, Gly345, His347 on neighbouring beta-strands. Ala406 is near the protein surface and present on a long loop between Ala400-Asn414. Thr644 is associated with $i+2^{\text{nd}}$ position or equivalent $i+1^{\text{th}}$ position of overlapping type IV beta-turns and its side-chain is exposed to the protein surface. Val880 is associated with helix (H39) located near the protein surface and its side-chain makes contacts with Pro834 and Leu838 on a neighbouring helix located within the interior of a core formed by four helices (H36, H37, H39 and H18).

In summary, the SARS CoV-2 mutation sites in the RdRp protein are associated with loops, helices, beta-strand, beta-turns and located near the protein surface. The mutation site residues are distant from the functionally important residues in the protein, and from the interface residues of β -hairpin interactions with the NiRAN and palm domains that is required for maintaining stability of the overall structure and also from the Remdesivir drug binding site. The main-chain hydrogen bond involving the residue at position 329 in the RdRp protein of SARS CoV-2 Indian isolate will not be affected due to the L329I mutation.

Mutation sites in helicase (Nsp13)

The helicase SARS CoV (Nsp13) protein catalyzes the unwinding of duplex oligonucleotides into single strands in an NTP-dependent manner and its activity is enhanced by the SARS-Nsp12 protein (Jia et al., 2019). The protein with PDB code: 6JYT was used as template for

mapping the mutations and represents the crystal structure of the full-length SARS-Nsp13 protein upon ATP hydrolysis determined at 2.8Å resolution with the overall fold in the shape of a pyramid (Jia et al., 2019). The SARS-Nsp13 protein fold is shown in Figure 7 and it comprises five domains; the 1A (magenta), 2A (cyan) domains, 1B domain (yellow) at the base, ZBD (green) at the apex and S domain (blue) connecting the ZBD domain to the base of the pyramid. The S166A and T214I mutation sites are associated with the 1B domain and P504L mutation is associated with the 2A domain and the side-chains of all three residues are exposed near the protein surface. Ser166 is located close to end of a beta-strand and T214 is at the $i+1^{\text{th}}$ position of a type IV beta-turn. P504 lies within a 14-residue loop (VREFLTRNPAWRKA) essentially comprising beta and gamma turns and is associated with the $i+1^{\text{th}}$ or equivalent i^{th} position of overlapping type IV beta-turns. None of the mutations are close to any of the functionally important residues; R337, R339, K345 and K347 in the $\beta 19$ - $\beta 20$ loop (tint) in 1A domain as shown in Figure 8, or to the six residues; Lys288, Ser289, Asp374, Glu375, Gln404, Arg567 in the 1A and 2A domains (shown in purple and cyan, respectively), in Figure 9, or to the ZBD domain in SARS-Nsp13 which interacts with SARS-Nsp12. The side-chain atom (NH1) of Arg337 (3.2Å) is involved in hydrogen bond with main-chain carbonyl oxygen atom (C=O) of T214 in the 1B domain and the T214I mutation will not affect this hydrogen bond interaction.

In summary, mutations in the SARS CoV-2 Indian isolate helicase protein are associated with beta-strand and beta-turns near the protein surface and distant from functionally important residues in the protein.

Mutation sites in the spike (S) glycoprotein

The trans-membrane spike (S) glycoprotein forms homotrimers protruding from the viral surface and mediates coronavirus entry into host cells (Tortorici and Veasler, 2019). Three identical protein chains labelled; 'A', 'B', and 'C' in the protein assemble to form the 'crown-like' structure or 'corona' on the virion surface. Each protein chain has 1273 amino acid residues that comprises two subunits; S1 and S2. The S1 subunit contains the receptor binding domain (RBD) involved in mediating the binding of the spike glycoprotein to the host cell human angiotensin converting enzyme 2 (ACE-2) receptor and an N-terminal domain (NTD) which is distant from the RBD. The S2 subunit comprises fusion of the viral and cellular membranes. Host proteases cleave the spike protein at specific sites and a concerted action of

the receptor-binding and proteolytic processing of the S-protein is required to promote virus-cell fusion and entry of the coronavirus into the host (Walls et al., 2020).

The spike glycoproteins structures available in the PDB include the electron microscopy structures of SARS CoV-2 spike protein representing the pre-fusion with a single receptor binding domain in the UP conformation (PDB code: 6VSB) (Wrapp et al., 2020), the SARS CoV-2 spike glycoprotein (closed state) (PDB code: 6VXX) and SARS CoV-2 spike ectodomain (open state) (PDB code: 6VYB) (Walls et al., 2020), crystal structures of the novel coronavirus spike receptor binding domain (RBD) complexed with ACE-2 receptor determined at 2.5 Å resolution (PDB code: 6LZG) (Wang, et al., 2020) and at 2.45 Å resolution (PDB code: 6M0J) (Lan J. et al., 2020). These structures were used to map the mutations in the SARS CoV-2 spike glycoprotein.

Seven mutations in the spike glycoproteins of the human SARS CoV-2 Indian isolates (Guruprasad, 2020) correspond to: Y28H, D614G, C1250F, Q271R, Y145-del, R408I, A930V in the protein structure (PDB code: 6VXX). Y28 is located at the start of a beta-strand near the N-terminus and is associated with the NTD in the S₁ subunit. The side-chain of Y28 shown in Figure 10A (red spheres) is exposed on the protein surface and is close to the glycosylation site NAG-1301. The Y145-del is associated with the NTD in S₁ subunit and is distant from RBD (Korber et al., 2020). In Figure 10B, the equivalent residue Lys142 according to the sequence alignment is shown corresponding to the Y145-del site for SARS CoV (PDB code: 6ACG) as the region corresponding to the mutation site is not available in the SARS CoV2 structure. A930 is in the S₂ subunit (Yadav et al., 2020) and is close to the mutational cluster that possibly impacts the fusion core of the first heptad repeat region HR1 (Korber et al., 2020). A930 is present in the middle of helix (H15) comprising 21 amino acid residues (Q₉₂₀KLIANQFNSAIGKIQDSLSS₉₄₀) (PDB code: 6VXX_A-chain, green) with side-chain pointing towards Val722 on a neighbouring twisted beta-strand S₇₁₁IAIPTNFTISVTTEILP₇₂₈ (blue) as shown in Figure 10C. The H15 helix immediately follows a short helix (H14); N₉₁₄VLYE₉₁₈ (orange) and Asn₉₁₉ is at the 'kink' region between the two helices close to the glycosylation site NAG-1312. Ala930Val mutation in human SARS CoV-2 is in a position to make favourable hydrophobic contacts with the side-chain of Val722. The D614G mutation is associated with the S₁ subunit of spike glycoprotein (Korber et al., 2020). D614 is the first residue in a loop immediately at end of a beta-strand associated with a motif comprising 5 anti-parallel beta-strands in the beta-sheet 'E' (PDB code: 6VXX) and is close to the

glycosylation binding motif N₆₁₆CT near NAG-1309 (Figure 10D). The Cys617 residue in the loop makes disulfide bridge with Cys649 on a beta-strand in the beta-sheet ('E') and the side-chain of Asp614 is oriented towards the protomer 'B' (shown in cyan) in the spike glycoprotein. The D614G mutation site is distant from the RBD. Cys1250 is associated with the cytoplasmic region of the spike glycoprotein and a suitable template was not available to map this region onto the protein structure. Q271 is associated with the S1 subunit and is located at the start of the 2nd beta-strand in a motif comprising 3 anti-parallel beta-strands in the beta-sheet 'C' (shown in orange) with its side-chain (shown as spheres) exposed on the protein surface (Figure 10E). R408 is associated with the RBD in S1 subunit (Yadav et al, 2020). In the crystal structure of the RBD of spike protein (cyan) bound to the ACE-2 receptor (green) (PDB code: 6LZG), a long curved helix comprising 32 amino acid residues; Ile21-Thr52 (orange) in the ACE2 receptor is mainly involved in interactions with the RBD of the human SARS CoV-2 spike glycoprotein as shown in Figure 10F. The RBD of spike protein Thr333-Pro527 (cyan) comprises five beta-strands in the beta-sheet 'D'; N354-I358, T376-Y380, N394-R403, G431-N437, P507-E516 and labelled from 1 to 5 as they occur along the protein sequence as shown in Figure 10G. The R408 (shown as red spheres) is present between the beta-strands 3 and 4 on a loop comprising two helices (yellow); a G-type helix G₄₀₄DEV**R**Q₄₀₉ and a H-type helix K₄₁₇IADY₄₂₁ and the arginine is associated with the G-type helix with side-chain exposed on the protein surface. The long loop between Ser438-Gln506 (shown in tint) that makes interactions with the ACE-2 receptor connects beta-strands 4 and 5. Arg408 does not interact with the ACE-2 receptor and its side-chain is observed to be flexible when the two crystal structures of SARS CoV-2 spike protein RBD complexed with ACE-2 receptor are compared (PDB codes: 6LZG, 6M0J). In (PDB code: 6LZG), the side-chain atom (NH1) of Arg408 is at a distance of 3.5Å to main-chain carbonyl atom C=O of Thr415, whereas, in the structure (PDB code: 6M0J), the NH1 atom in the side-chain of R408 is involved in hydrogen bond with the main-chain carbonyl oxygen of Asp405. Also, in the electron microscopy structures of SARS CoV spike glycoproteins bound to the ACE-2 receptor in three different conformations (PDB codes: 6ACG, 6ACJ, 6ACK) (Song et al., 2018) and in the SARS CoV spike protein where a different protein chain (B-chain) is involved in binding to the ACE-2 receptor (PDB code: 6CS2) (Kirchdoerfer, 2018), Arg408 (numbered R395) is not involved in interactions with the receptor.

In summary, the mutations observed in the spike glycoprotein of the SARS CoV-2 Indian isolates are associated with the NTD and RBD in the S1 subunit and membrane fusion domain

in the S2 subunit. The secondary structures corresponding to the amino acid residues at the mutation sites are associated with loops, helices or beta-strand. None of the mutations interact with ACE-2 receptor.

Mutation sites in the endoRNase (Nsp15)

The endoRNase protein (<https://www.uniprot.org/uniprot/P0DTD1>) is a Mn^{2+} -dependent uridylyate-specific enzyme and it leaves 2'-3'-cyclic phosphates 5' to the cleaved bond. High resolution crystal structures of Nsp15 endoribonuclease NendoU from SARS CoV-2 in the apo-form (PDB code: 6VWW) and citrate bound form (PDB code: 6W01) are available in the PDB (Kim et al., 2020a). In addition, three crystal structures that represent the Nsp15 from SARS CoV-2 complexed with uridine-5'-monophosphate (PDB code: 6WLC), a complex with product nucleotide GpU (PDB code: 6X1B) and another structure as complex with a potential refurbishing drug - Tipiracil (PDB code: 6WXC) are available in the Protein Data Bank (Kim et al., 2020b). The monomer chain in the crystal structure of the citrate bound form of SARS CoV-2 Nsp15 endoribonuclease NendoU (PDB code: 6W01_A-chain, cyan) comprising 348 amino acid residues is shown in Figure 11. The structure is associated with three domains (Kim et al., 2020a); the N-terminal oligomerization domain, middle domain and C-terminal catalytic NendoU domain. Three mutation sites of SARS CoV-2 endoRNase protein observed among the Indian isolates (Guruprasad, 2020) correspond to V149, D268, D273 in the crystal structure and are shown in Figure 11 (red spheres). V149 is associated with the middle domain, D268 and D273 are associated with the C-terminal catalytic domain. All three amino acid residues are associated with beta-turns. V149 is at i^{th} position of type II beta-turn and its side-chain makes van-der-Waals contacts with Val132 (green spheres) on a neighbouring helix (H5). D268 is at i^{th} position of type VIII beta-turn or the equivalent $i+1^{th}$ position of an overlapping gamma-turn and D273 is in a 2-residue loop between two type VIII beta-turns. The D268 and D273 mutation sites are in a loop connecting the two beta-strands; F264-E267 (magenta) and K277-D283 (blue) within the beta-sheet ('F') that comprises three anti-parallel beta-strands as shown in Figure 11. The loop is oriented within the protein cleft and the side-chains of the two aspartic acid residues are exposed to the solvent. The side-chains of six key residues that constitute the active site in the C-terminal NendoU domain of the Nsp15 of the SARS CoV-2 and that are important for the catalytic function; H235, H250, K290, T341, Y343, S294 (Kim et al., 2020a) are shown. None of the mutation site residues of Nsp15 in SARS CoV-2 Indian isolates are close to residues involved in the catalytic function. Further,

the SARS CoV-2 NendoU hexamer made up of dimer of trimers and that is essential for the enzymatic activity (Kim et al., 2020a) was generated using the PISA server available at https://www.ebi.ac.uk/msd-srv/prot_int/cgi-bin/piserver and the mutations were also mapped onto the hexamer. The active sites in the protein within the hexamer are located at the top and bottom of the assembly (Kim et al., 2020a) as shown in Figure 12 and the sites of the mutations are distant from the active sites also within the hexamer. The NendoU monomers interact extensively with all five subunits of the hexamer rendering it susceptible to mutations that affect the oligomeric assembly (Kim et al., 2020). In Figure 13, the model of the NendoU hexamer is shown with monomer (orange) making contacts with other five monomers (green, magenta, yellow, blue and cyan) and the three mutation sites; Val149, Asp268, Asp273 are also shown (red spheres) for the monomer (orange). The mutation sites are distant from the interface sites of inter-monomer interactions in the hexamer. Further, it was observed that the mutation site residues are distant from the site where Tipiracil drug (5-chloro-6-[(2-iminopyrrolidin-1-yl)methyl]-1H-pyrimidine-2,4-dione) binds to the protein as observed in the crystal structure complex of the Nsp15 SARS CoV-2 complex (PDB code: 6WXC_A-chain) (Kim et al., 2020b) and shown in Figure 14.

In summary, the residues at the three mutation sites in Nsp15 protein of human SARS CoV-2 Indian isolates are associated with beta-turns or loops in the protein structure and the side-chains are exposed to the protein surface. The mutation sites are distant from the functionally important residues, inter-monomer interaction interfaces in the hexamer and from the Tipiracil drug-binding site in the SARS CoV-2 Nsp15 protein.

We could not find suitable templates to map the mutations for the Nsp2, Nsp3, Nsp4, Nsp6, orf3a, orf8 proteins or the nucleocapsid phosphoprotein 'N' (Guruprasad, 2020).

Conclusions:

The RdRp (Nsp12) mutations in the SARS CoV-2 Indian isolates are associated with the nidovirus-specific N-terminal extension domain, the interface domain and the fingers and thumb subdomains within the polymerase domain. The mutation sites are distant from the divalent-cation-binding residue D618 and the catalytic residues $S_{759}DD_{761}$. The mutations in the helicase (Nsp13) are associated with the 1B and 2A domains on the protein surface and are distant from amino acid residues involved in the helicase activity. The T214I mutation will

not affect the hydrogen bond with Arg337 that is among the residues in the β 19- β 20 loop in the 1A domain and critical for helicase activity. The mutations in the spike glycoprotein are associated with NTD and RBD domains in the S1 subunit and with the membrane fusion domain in the S2 subunit and the mutations are associated with loops, helix or the strand conformations. None of the spike glycoprotein mutations are close to the ACE-2 receptor binding site. The mutations in endoRNase (Nsp15) are associated with beta-turns or loops and distant from residues involved in the catalytic function within the monomer and the monomer interaction interfaces within the hexamer. The mutation sites in the SARS CoV-2 Indian isolates of the RdRp and endoRNase proteins are also distant from the Remdesivir and Tipiracil drug binding sites, respectively.

In summary, mutations analysed for the RdRp, helicase, spike and endoRNase proteins among the twenty-two SARS CoV-2 Indian isolates are situated at locations that are generally distant or that do not affect the functionally important residues in the protein three-dimensional structure.

Acknowledgements:

The author sincerely acknowledges resources available at the PDB, EBI, NCBI, UniProt. The author particularly thanks research groups who have made available protein structure data in the PDB prior to publication.

Conflict of interest:

The author declares no conflict of interest

Funding:

None

References:

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.

DeLano, W. L. (2002). PyMOL.

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. (2020). 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>.

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.

Jia, Z., Yan, L., Ren, Z., Wu, L., Wang, J., Guo, J., ... & Rao, Z. (2019). Delicate structural coordination of the Severe Acute Respiratory Syndrome coronavirus Nsp13 upon ATP hydrolysis. *Nucleic Acids Research*, 47(12), 6538-6550.

Kim, Y., Jedrzejczak, R., Maltseva, N. I., Wilamowski, M., Endres, M., Godzik, A., ... & Joachimiak, A. (2020a). Crystal structure of Nsp15 endoribonuclease NendoU from SARS-CoV-2. *Protein Science*.

Kim, Y., Maltseva, N., Jedrzejczak, R., Endres, M., Chang, C., Michalska, K., Joachimiak, A., Center for Structural Genomics of Infectious Diseases (CSGID) (2020b). Crystal Structure of NSP15 Endoribonuclease from SARS CoV-2 in the Complex with potential repurposing drug Tipiracil doi: [10.2210/pdb6wxc/pdb](https://doi.org/10.2210/pdb6wxc/pdb)

Kirchdoerfer, R. N., Wang, N., Pallesen, J., Wrapp, D., Turner, H. L., Cottrell, C. A., ... & Ward, A. B. (2018). Stabilized coronavirus spikes are resistant to conformational changes induced by receptor recognition or proteolysis. *Scientific reports*, 8(1), 1-11.

Korber, B., Fischer, W., Gnanakaran, S. G., Yoon, H., Theiler, J., Abfalterer, W., ... & Partridge, D. G. (2020). Spike mutation pipeline reveals the emergence of a more transmissible form of SARS-CoV-2. *bioRxiv*.

Lan, J., Ge, J., Yu, J., Shan, S., Zhou, H., Fan, S., Zhang, Q., Shi, X., Wang, Q., Zhang, L., & Wang, X. (2020). Structure of the SARS-CoV-2 spike receptor-binding domain bound to the ACE2 receptor. *Nature*, 581(7807), 215–220. <https://doi.org/10.1038/s41586-020-2180-5>.

Laskowski, R. A. (2001). PDBsum: summaries and analyses of PDB structures. *Nucleic Acids Research*, 29(1), 221-222.

Schäffer, A. A., Aravind, L., Madden, T. L., Shavirin, S., Spouge, J. L., Wolf, Y. I., ... & Altschul, S. F. (2001). Improving the accuracy of PSI-BLAST protein database searches with composition-based statistics and other refinements. *Nucleic Acids Research*, 29(14), 2994-3005.

Song, W., Gui, M., Wang, X., & Xiang, Y. (2018). Cryo-EM structure of the SARS coronavirus spike glycoprotein in complex with its host cell receptor ACE2. *PLoS pathogens*, 14(8), e1007236.

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.

Tortorici, M. A., & Veessler, D. (2019). Structural insights into coronavirus entry. In *Advances in virus research* (Vol. 105, pp. 93-116). Academic Press.

Walls, A. C., Park, Y. J., Tortorici, M. A., Wall, A., McGuire, A. T., & Veessler, D. (2020). Structure, function, and antigenicity of the SARS-CoV-2 spike glycoprotein. *Cell*, 180, 281-292.

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.

Wang, Q., Zhang, Y., Wu, L., Niu, S., Song, C., Zhang, Z., Lu, G., Qiao, C., Hu, Y., Yuen, K. Y., Wang, Q., Zhou, H., Yan, J., & Qi, J. (2020). Structural and Functional Basis of SARS-CoV-2 Entry by Using Human ACE2. *Cell*, 181(4), 894–904.e9. <https://doi.org/10.1016/j.cell.2020.03.045>

Wrapp, D., Wang, N., Corbett, K. S., Goldsmith, J. A., Hsieh, C. L., Abiona, O., ... & McLellan, J. S. (2020). Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation. *Science*, 367(6483), 1260-1263.

Yadav, P. D., Potdar, V. A., Choudhary, M. L., Nyayanit, D. A., Agrawal, M., Jadhav, S. M., ... & Cherian, S. S. (2020). Full-genome sequences of the first two SARS-CoV-2 viruses from India. *The Indian journal of medical research*, 151(2-3), 200.

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*.

Table 1. Secondary structure conformations associated with the SARS CoV-2 protein mutation site residues.

S.No.	Protein	Mutation	Secondary structure conformation	PDB template
1	RdRp	A97V	Loop	6M71
2	RdRp	A185V	Helix	6M71
3	RdRp	I201L	Beta-strand	6M71
4	RdRp	P323L	Helix	6M71
5	RdRp	L329I	Beta-strand	6M71
6	RdRp	A406V	Loop	6M71
7	RdRp	T644M	Beta-turn	6M71
8	RdRp	V880I	Helix	6M71
9	Helicase	S166A	Beta-strand	6JYT
10	Helicase	T214I	Beta-turn	6JYT
11	Helicase	P504L	Beta-turn	6JYT
12	Spike	Y28H	Beta-strand	6VXX
13	Spike	D614G	Loop	6VXX
14	Spike	Q271R	Beta-strand	6VXX
15	Spike	A930V	Helix	6VXX
16	Spike	R408I	Helix	6LZG, 6M0J
17	endoRNase	V149A	Beta-turn	6W01
18	endoRNase	D268Y	Beta-turn	6W01
19	endoRNase	D273G	Loop	6W01

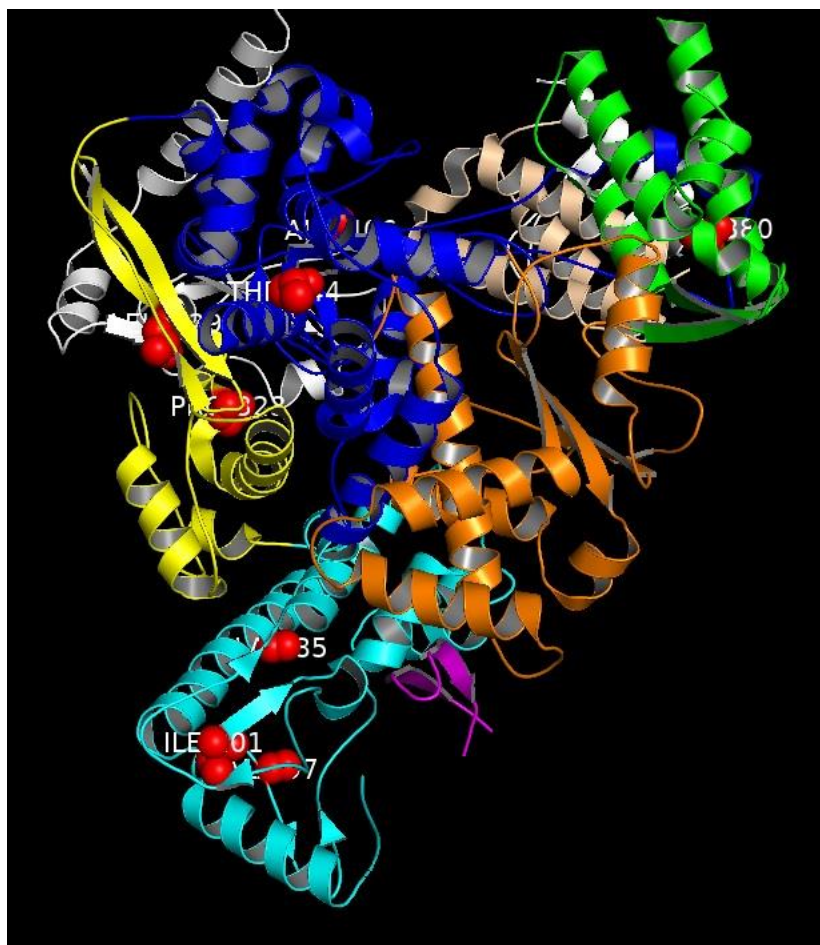


Figure 1. The overall fold of the RdRp protein complex (PDB code: 6M71) comprising a right-hand RdRp polymerase domain (blue, orange and green), a nidovirus-specific N-terminal extension domain (cyan) and an interface domain (yellow). The three subdomains of the RdRp domain; fingers subdomain (blue), palm subdomain (orange) and thumb subdomain (green). The RdRp is bound to cofactors Nsp7 (tint), Nsp8 (white) in the nsp12-nsp7-nsp8 complex. The N-terminal β -hairpin (magenta). The mutation site residues (red spheres) are labelled.

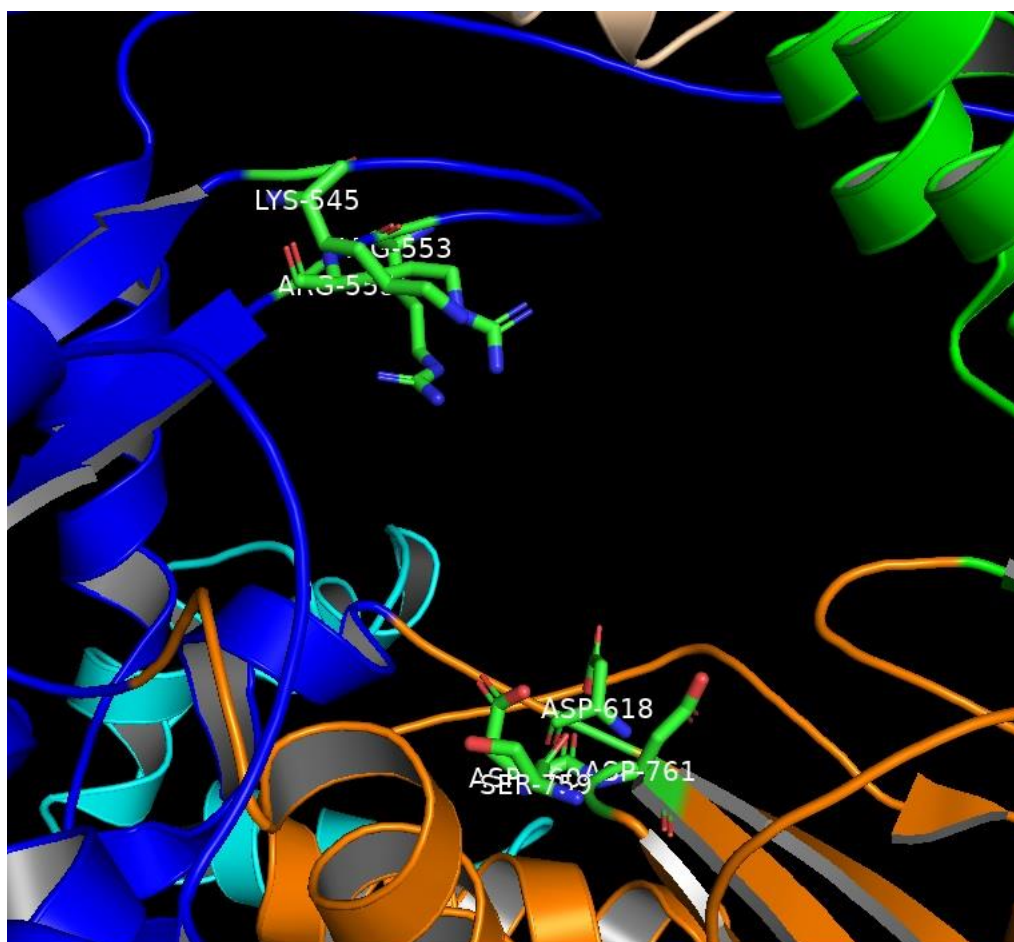


Figure 2. The catalytic site residues in palm subdomain (orange) and residues near the NTP entry channel in RdRp finger domain (blue).

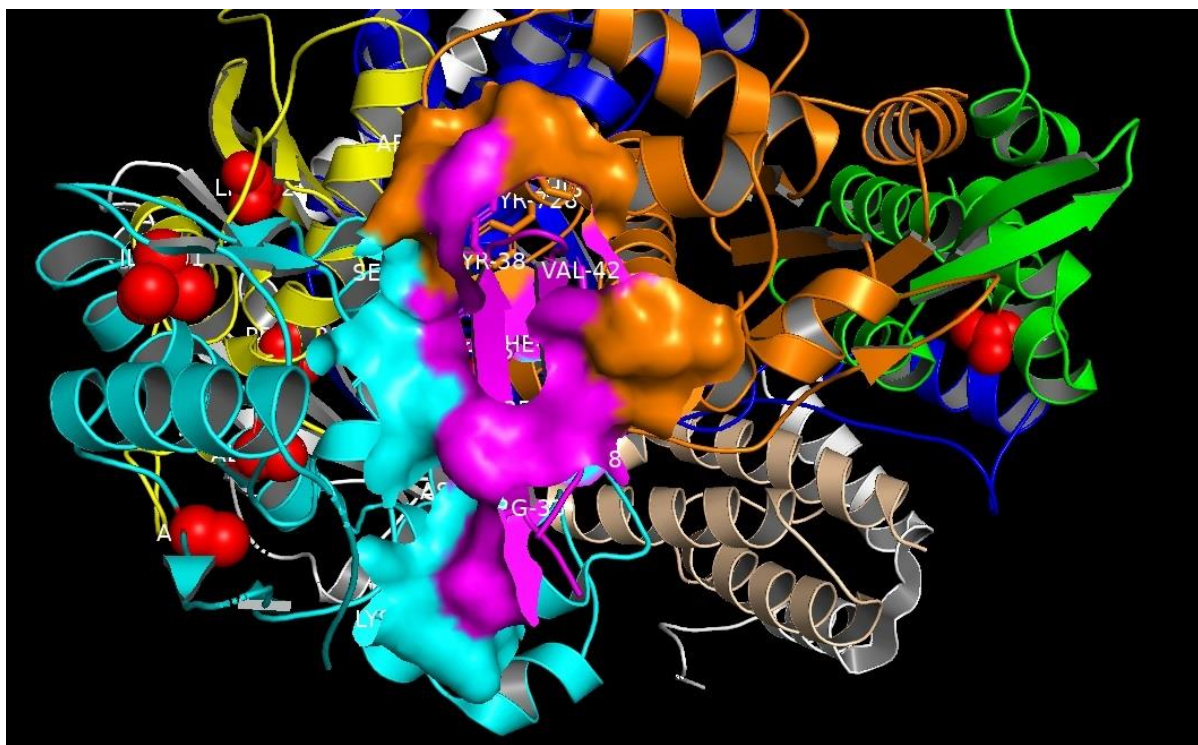


Figure 3. The surfaces of the residues in β -hairpin (magenta) that make contacts with the residues in NiRAN domain (cyan) and residues of the palm sub-domain (orange) in the RdRp protein (PDB code: 6M71). The sites of mutated residues (red spheres) in RdRp protein in the SARS CoV-2 Indian isolates are distant from the interface regions.

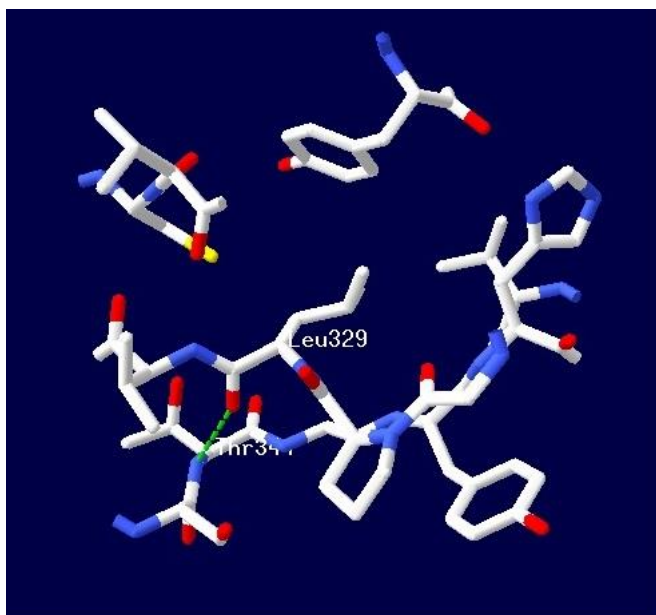


Figure 4. The Leu329-Thr344 main-chain hydrogen bond in the RdRp protein (PDB code: 6M71).

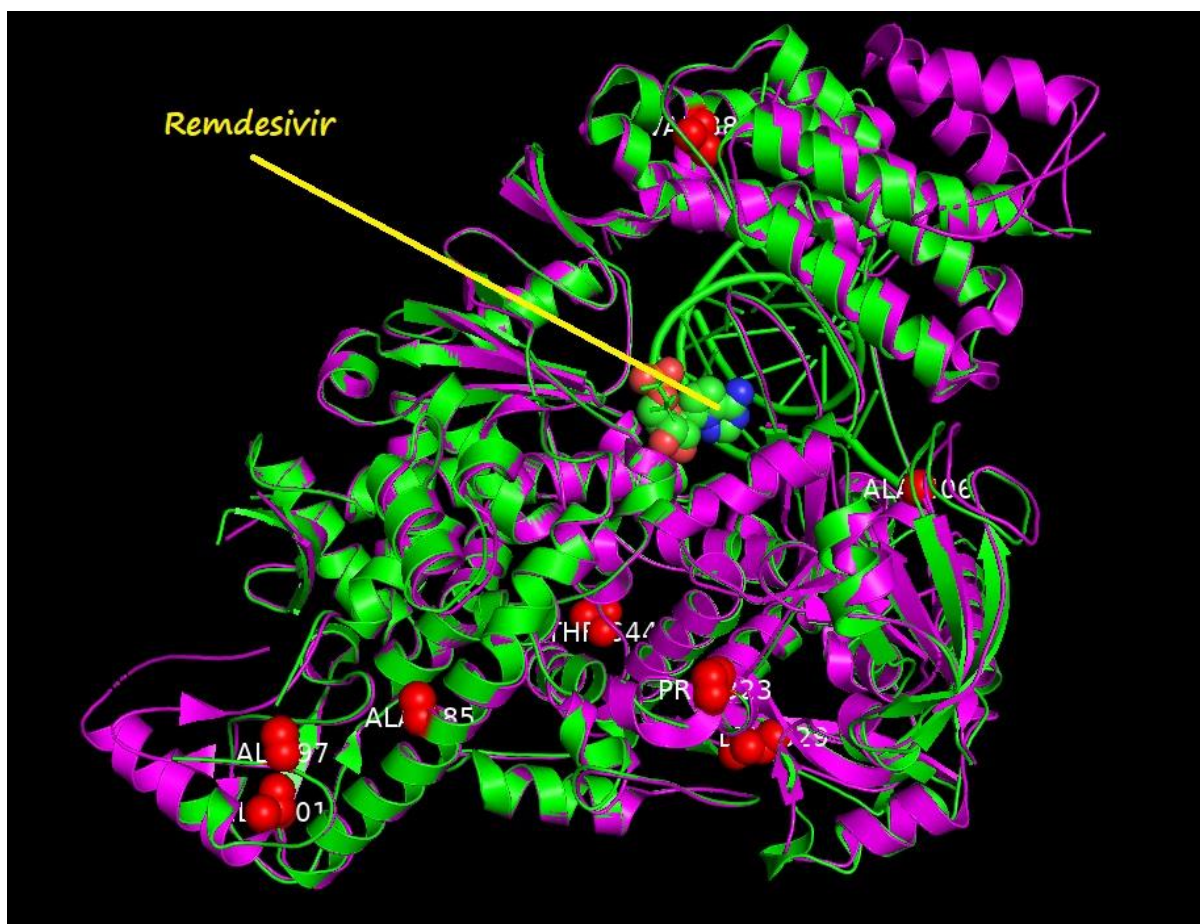


Figure 5. Structural overlay of the human SARS CoV-2 RdRp protein complex bound to cofactors (PDB code: 6M71) (magenta) and the SARS CoV-2 RdRp protein complex bound to the template-primer RNA and triphosphate form of Remdesivir (PDB code: 7BV2) (green) along with the eight RdRp mutation site residues of the SARS CoV-2 in Indian isolates (red spheres).

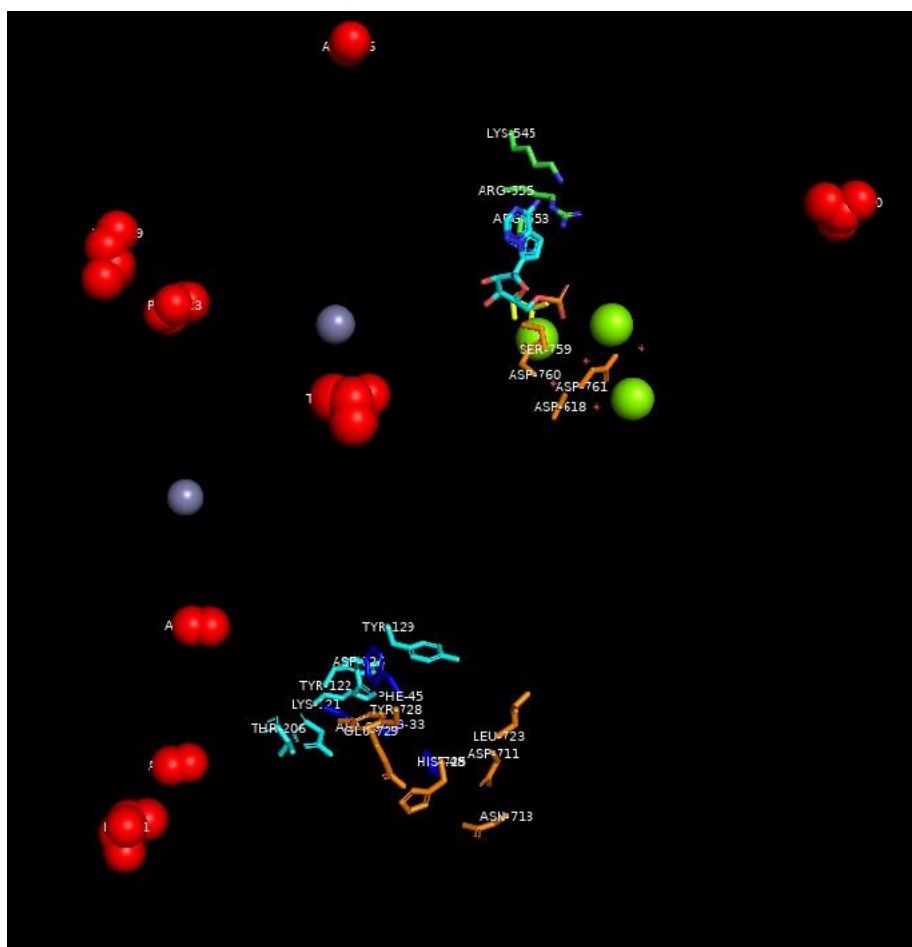


Figure 6. Mapping of the RdRp mutation site residues (red spheres) of SARS CoV-2 in Indian isolates. The protein is excluded, for clarity. The Remdesivir drug (stick representation) binding site and side-chains of residues important for the protein function in polymerase domain (labelled) near the top and side-chain of residues that interact with the β -hairpin (at the bottom). The zinc ions (green spheres) and magnesium ions (purple spheres).

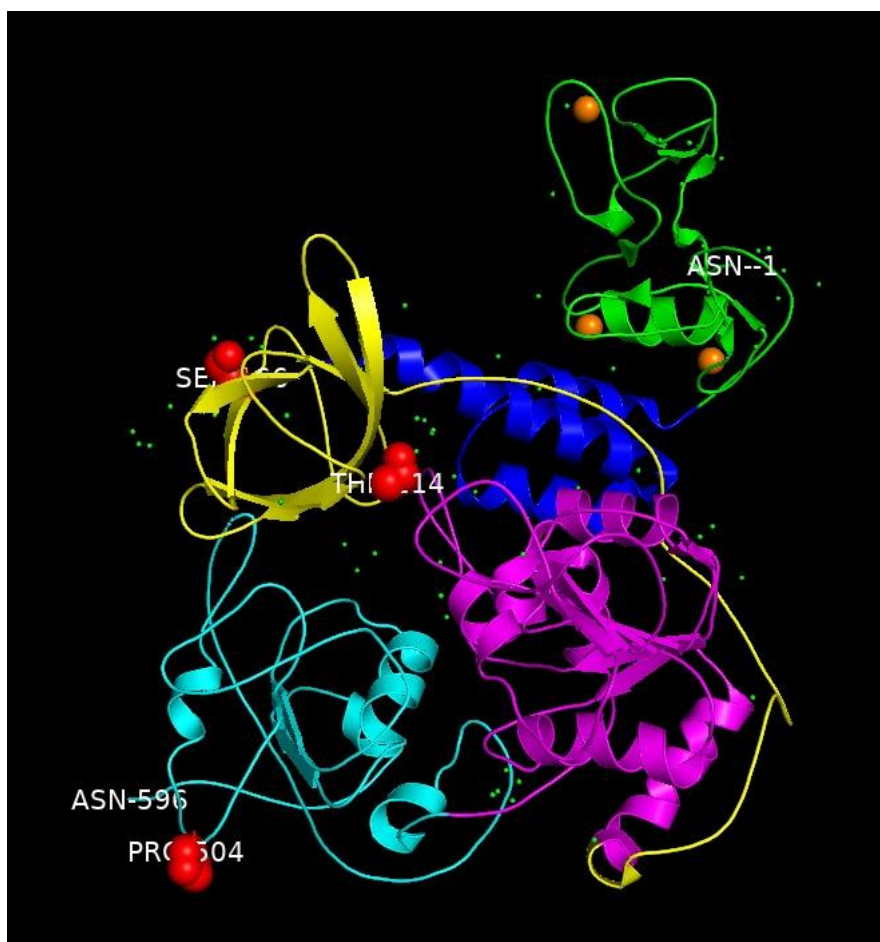


Figure 7. The overall fold of the helicase protein (Nsp13) comprising five domains; zinc-binding domain (green), stalk domain (blue), 1B domain (yellow), 1A domain (magenta), 2A domain (cyan) (PDB code: 6JYT). The mutation site residues (red spheres) in helicase of human SARS CoV-2 Indian isolates. The start and end amino acid residues are labelled, Asn1 and Asn596, respectively.



Figure 8. The β 19- β 20 loop (tint) in the 1A domain (magenta) of helicase in the crystal structure (PDB code: 6JYT) showing side-chains of residues important for the helicase activity. Mapping the site of the T214I mutation (red sphere) in human SARS CoV-2 Indian isolate within the 1B domain (yellow) of helicase. The main-chain carbonyl oxygen (C=O) atom of Thr214 makes hydrogen bond with side-chain (NH1) atom of Arg337.

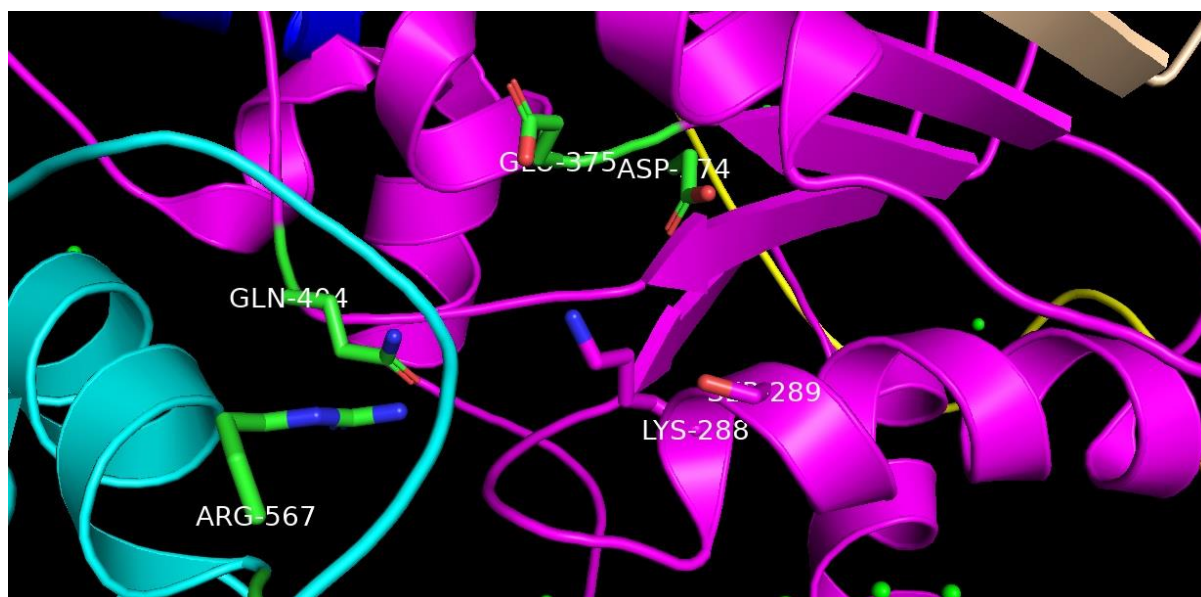
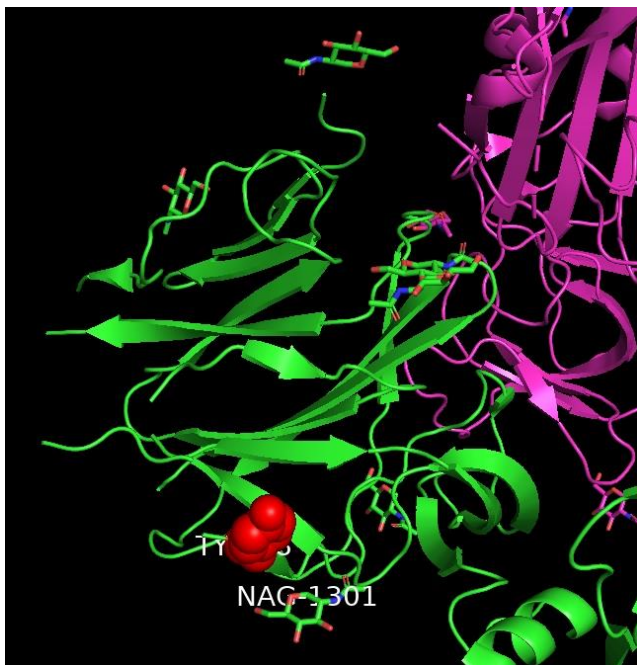
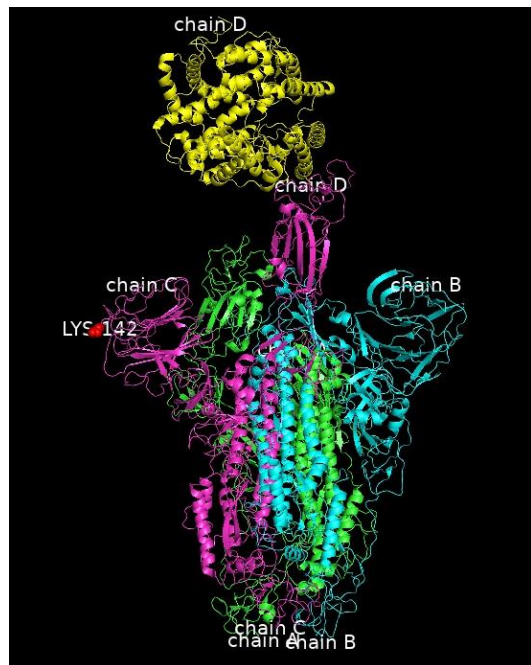


Figure 9. The amino acid residues important for helicase activity that lie between the 1A (magenta) and 2A (cyan) domains in helicase (PDB code: 6JYT).

A

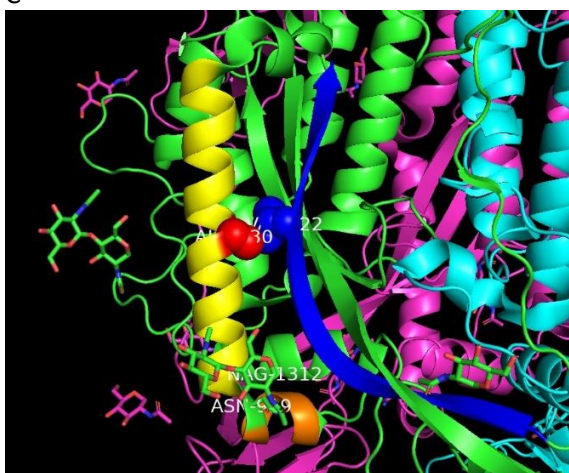


B

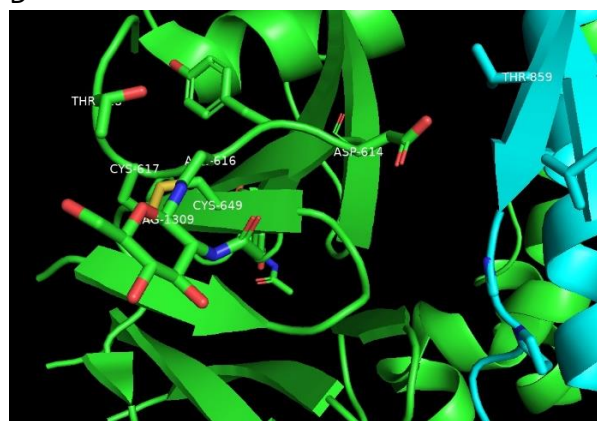


24

C



D



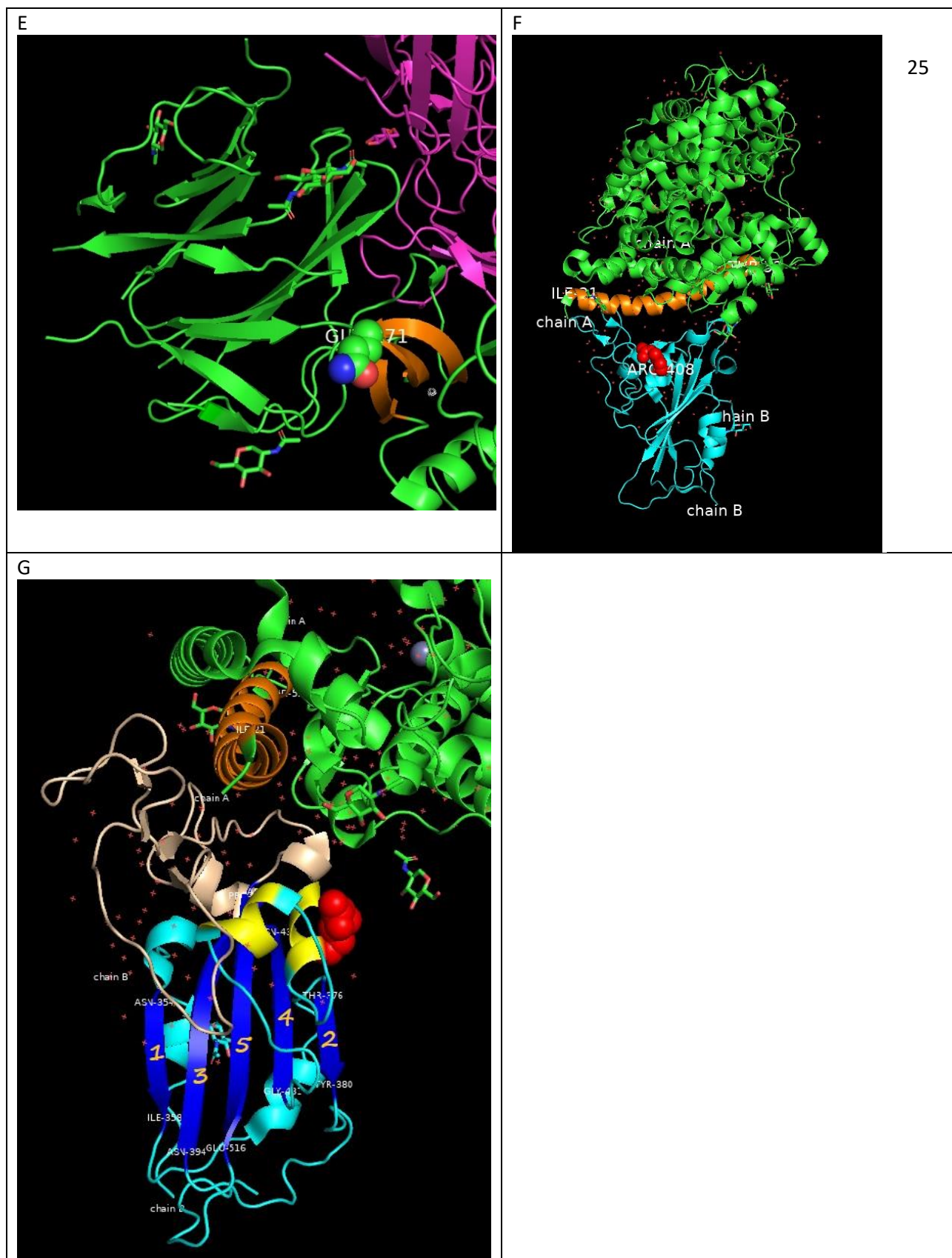


Figure 10. The mutation sites in the spike glycoprotein of the human SARS CoV-2 Indian isolates: (A) Y28, (B) Y145-del (Lys142, see text), (C) A930, (D) D614, (E) Q271, (F) R408, (G) The location of R408 (red sphere) in the RBD of spike protein comprising residues; Thr333-Pro527.

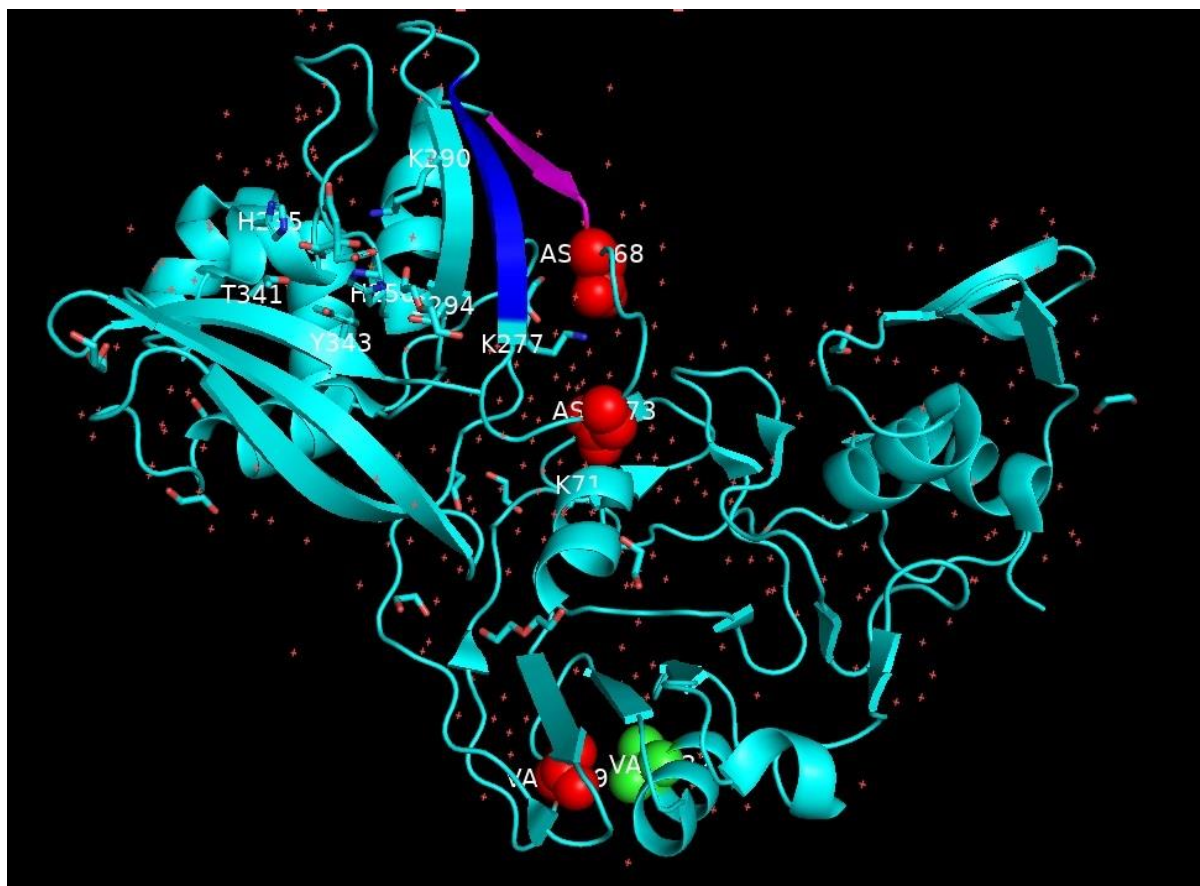


Figure 11. The overall fold corresponding to the monomer chain in the crystal structure of the citrate bound form of SARS CoV-2 Nsp15 endoribonuclease NendoU (PDB code: 6W01_A-chain, cyan). The mutation sites residues in human SARS CoV-2 Indian isolates are shown (red spheres). The functionally important residues are labelled. Hydrogen bonds are formed between D268-K277, D273-K71. Val149 (red spheres) makes hydrophobic contacts with Val132 (green spheres) on neighbouring helix.

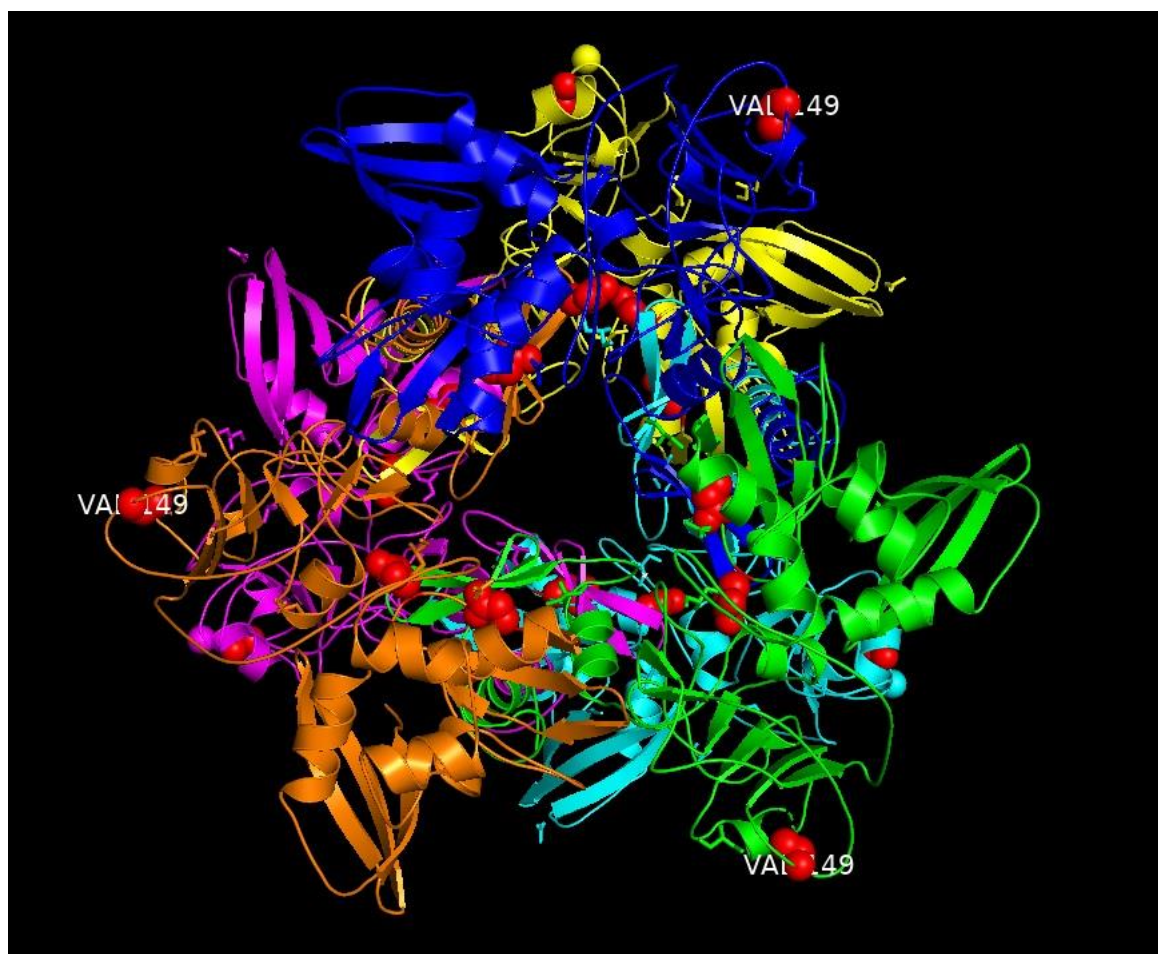


Figure 12. The SARS CoV-2 Nsp15 hexamer comprising the six monomers (magenta, blue, yellow, green, cyan, orange) showing the mutation site residues (red spheres) of Nsp15 SARS CoV-2 Indian isolates for each monomer within the hexamer.

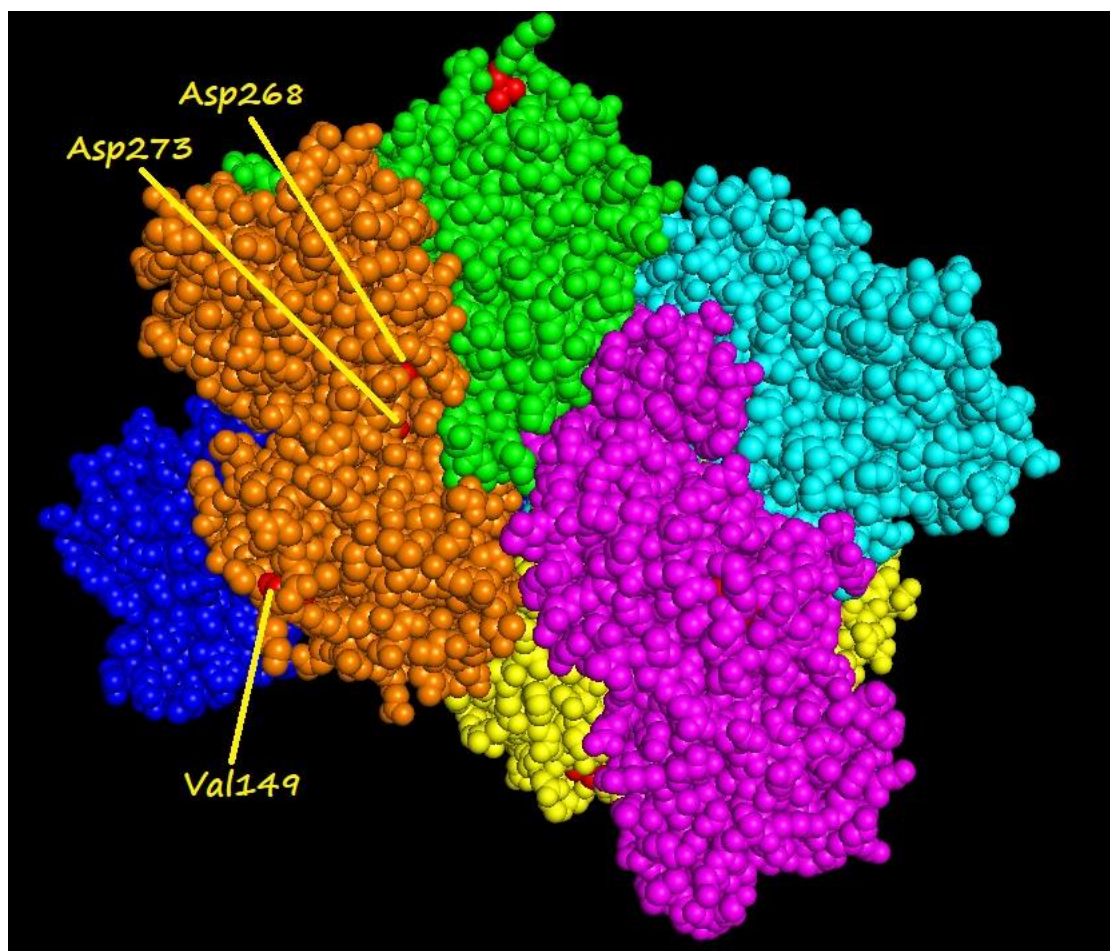


Figure 13. The oligomeric contacts made by the monomer (orange) with the other five monomers (green, magenta, yellow, blue and cyan) in the Nsp15 SARS CoV-2 hexamer showing the mutation sites; Val149, Asp268, Asp273 (red spheres) for the (orange) monomer in the complex. The mutation sites are distant from the inter-monomer contact interfaces in the hexamer.

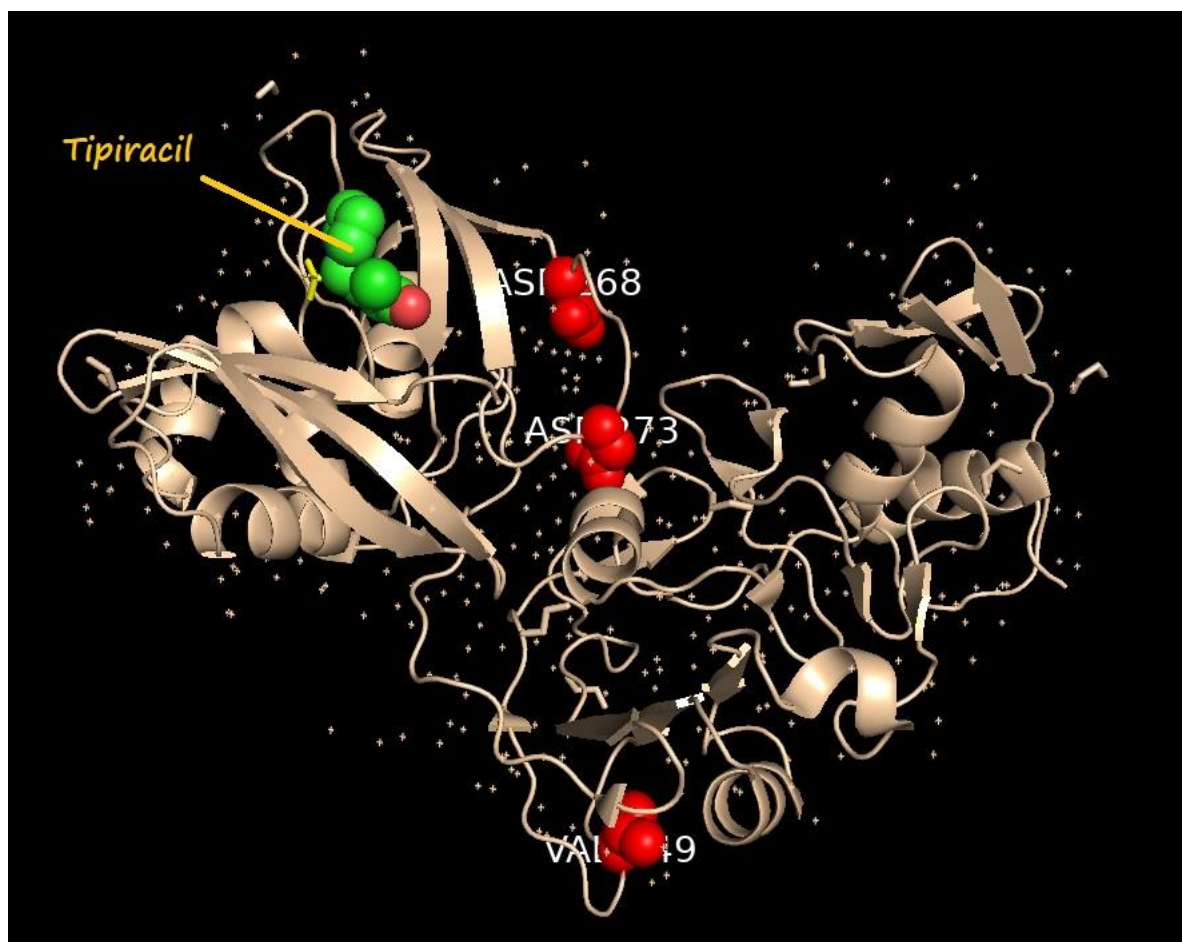


Figure 14. The Tipiracil drug binding site in SARS CoV-2 Nsp15 protein as in the crystal structure (PDB code: 6WXC_A-chain) with the mutation site residues (red spheres) of the human SARS CoV-2 Indian isolates mapped.