bioRxiv preprint doi: https://doi.org/10.1101/2021.03.26.437194; this version posted March 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

D155Y Substitution of SARS-CoV-2 ORF3a Weakens Binding with Caveolin-1: An *in silico* Study

Suchetana Gupta[a¥], Ditipriya Mallick[b¥], Kumarjeet Banerjee[b¥], Soumyadev Sarkar[c], Sonny T M Lee[c], Partha Basuchowdhuri[a#], Siddhartha Sankar Jana[b#]

[a]School of Mathematical and Computational Sciences, Indian Association for the Cultivation of Science, India

[b]School of Biological Sciences, Indian Association for the Cultivation of Science, India

[c]Division of Biology, Kansas State University, USA

#Correspondence: Siddhartha Sankar Jana, bcssj@iacs.res.in and Partha Basuchowdhuri, partha.basuchowdhuri@iacs.res.in

¥Authors contributed equally

Running Head: D155Y substitution of ORF3a weakens binding with Caveolin-1

Keywords:

ORF3a, SARS-CoV-2, Molecular Dynamics Simulation, Graph theory, Mutation, Caveolin-1

Word count for the abstract: 234

Word count for the text: 4272

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.26.437194; this version posted March 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Abstract

The clinical manifestation of the recent pandemic COVID-19, caused by novel SARS-CoV-2, varies from mild to severe respiratory illness. Although environmental, demographicand co-morbidity factors have an impact on the severity of the disease, the contribution of mutations in each of the viral genes towards the degree of severity needs to be elucidated for designing better therapeutic approach against COVID-19. Here, we studied the effect of two substitutions D155Y and S171L, of ORF3a protein, found in COVID-19 patients. Using computational simulations we discovered that the substitutions at 155th and 171st positions changed the amino acids involved in salt bridge formation, hydrogen-bond occupancy, interactome clusters, and the stability of the protein. Protein-protein docking using HADDOCK analysis revealed that out of the two observed substitutions, only the substitution of D155Y, weakened the binding affinity of ORF3a with caveolin-1. The increased fluctuation in the simulated ORF3a-caveolin-1 complex suggested a change in the virulence property of SARS-CoV-2.

Importance

The binding interaction of viral ORF3a protein to host caveolin-1 is essential for entry and endomembrane trafficking of SARS-CoV-2. The D155Y substitution in SARS-CoV-2 ORF3a is located near its caveolin-binding Domain IV and thus the substitution can interfere with the binding affinity of ORF3a to host caveolin-1. Our *in silico* study report decreased molecular stability of D155Y mutant of ORF3a and increased fluctuation of the simulated D155Y ORF3a-caveolin-1 complex. Thus, we hypothesize that the D155Y substitution could change the virulence property of SARS-CoV-2.

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.26.437194; this version posted March 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

1. Introduction:

The Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) is the causative agent of the novel Coronavirus Disease 2019 (COVID-19)[1].Till February 2021, 114.1 million cases have been reported worldwide spanning across 215 countries and territories, out of which 11.1 million people have been infected with SARS-CoV2 in India[2].Mortality rate across the world varies drastically from 9.1% (Mexico) to 0.9% (Turkey) [3].Although age, ethnicity and sex contribute to the demographic variation in the viral transmission and its case fatality rate, how mutation in viral genomecan change,in such variation forpathological manifestation needs to be explored.

The SARS-CoV-2 genome consists of approximately 30 kilobasesand shares about 82% sequence identity with both SARS-CoV and MERS-CoV. It also shares more than 90% sequence identity for essential enzymes and structural proteins [4].Despite the similarity, only SARS-CoV-2showssevere pathological manifestations in humans, suggesting theexistence of differential molecular interactions between viral proteins and host cell machinery.

The SARS-CoV-2 genome broadly consists of 14 open reading frames (ORF), which are generated from nested transcription of subgenomic RNAs.Interestingly, ORF1a and 1b encode for 16 non-structural proteins (nsp) known as replicase/transcriptase complex. The other ORFs code for 4 structural proteins and 8 accessory proteins [1,4].ORF3 is wedged between spike (S) and envelop (E) ORFs and encodes for a membrane-spanning, ion channel protein ORF3a. It is also known as the single largest accessory protein of 275 amino acids[5,6]. Ribosomal profiling has identified two putative overlapping genes, namely ORF3b and ORF3c, at the 3' end of ORF3 with an alternative reading frame to the canonical ORF3a[7-10], whose functional importance is not well understood.ORF3a can localise at plasma membrane and Golgi complex, and can exist in both glycosylated and non-glycosylated forms[11]. This viral protein has been shown to be highly immunogenic as antisera isolated from SARS-CoV-infected patients can detect ORF3a [10]. Yount*et.al*and others have shown that ORF3a has been co-evolved with Spike (S) protein, suggesting the possibility of direct or indirect interactions between ORF3a and S protein[10,12,13].Studies in SARS-CoV-infected Caco2 cells show that ORF3a can be efficiently released in detergent-resistant membrane structures and the diacidic motif, ExD, located within the domain VI, plays importantrole in membrane co-localisation[14].ORF3a has multi-functional roles including activating NLRP3 inflammasome and NFkB pathway, upregulating fibrinogen

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.26.437194; this version posted March 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

secretion, downregulating IFN Type Iand inducing ER stress and pro-apoptotic activity[5,15-17].Therefore, mutations in this protein warrant further study to understand its role in the virulence and immune evasive potential of the recent SARS-CoV-2.Several mutations have been reported in the ORF3a gene and have been classified in the form of clades and sub-clades.The mutation patterns of ORF3a gene have been characterized as largely non-synonymous(Q57H, H93Y, R126T, L127I, W128L, L129F, W131C, D155Y,S171L, D173Y, G196V, and G251V). G251V and Q57H exhibit severe virulence property[18-21].Interestingly, the 57th position in ORF3a of pangolin SARS-CoV is H. D155Y and S171L mutations were detectedin Indian patients in May 2020[22]. To understand the functional importance of these mutants, their characterization is needed.

Our study aims to understand the effect of these two substitutions (D155Y and S171L) in the structural stability of the ORF3a protein and its ability to form complex with caveolin-1. Using computational simulation, protein-protein docking we find that the amino acids involved inthe hydrophobic interactions,hydrogen bond formation,salt bridge formation and residue interaction patterns are different inwild type (WT), i.e., the original Wuhan sequencecompared with the two mutants having D155Y and S171L substitutions.

2. Methods:

2.1 Bioinformatic Methods

A total of26,656 sequences of ORF3a protein deposited in NCBI database as on Nov 17, 2020 were considered for the bioinformatics analysis. The keywords used for the search were “SARS-CoV-2”, “ORF3a protein”, and“complete structure”. These structures were aligned using the BLAST algorithm on the NCBI website. Some of the post-BLAST sequences were larger than 275 due to erroneous performance of the code. But such cases were very low in number. Subsequently, the erroneous sequences were manually cleaned to obtain the final alignments of the complete protein sequences (275 amino acids). The number of samples whose locations were geo-tagged to India was 614. These sequences were then compared to the Wuhan sequence (NCBI Accession No: YP_009724391.1[23]) and the amino acid positions were compared. The positions, where mismatches were observed with respect to the reference Wuhan sequence (WT), were considered as locationsof mutations. Clearly, lesser number of mutations denote a sequence more similar to the WT, whereas more number of mutations denote a more deviant mutant.Overall the sequences found from NCBI database

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.26.437194; this version posted March 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

were compared against the WT and the number of mutations for every position of ORF3awas stored. This essentially provides us with the frequency distribution of the mutations found at each position of ORF3a.We have used PROVEAN score to assess whether the effect of a mutation is deleterious or neutral. PROVEAN score of each mutation was determined using PROVEAN web server [24].

2.2 Preparation of structure of the ORF3a proteins

The cryo-EM structure of WT ORF3a protein of SARS-CoV-2 was obtained from PDB (PDB ID: 6XDC[25]).The symmetry information present in the PDB file was used to convert the structure into the functional dimeric form using PDBe PISA server online [26]. The residues on the second monomer have been numbered using " ' " throughout the manuscript.As no homologous structure of the protein was available, we considered the PDB structure of ORF3a, which has residues from 40th to 238th. Weintroduced the necessary mutations (D155Y and S171L) by modelling the residues in Swiss PDB Viewer[27].

2.3 Molecular dynamics simulation

We performed classical Molecular Dynamics (MD) simulation in AMBER20[28]using AMBER ff14sb force field[29]. The missing hydrogen atoms in the protein structure were added by the LEaP module of AMBER20 package. The protein was then subjected to energy minimisation for 2000 steps using steepest descent and conjugate gradient algorithms. We then solvated the energy minimised structures using rectangular water boxes comprising of TIP3P water molecules [30].Particle mesh Ewald was used to calculate the electrostatic interactions at a cut-off distance of 12Å. We performed initial minimisation and equilibration in order to avoid bad contacts. This was followed by equilibration using NVT ensemble at 300K for about 500ps. The systems were then equilibrated using NPT ensemble at 1 atm pressure for 1 ns. We considered 2fs as the time step throughout the minimisation-equilibration-production. After the energy values and the density values converged, the systems were subjected to 100ns production runs using NPT ensemble at 300K and 1 atm pressure. The coordinates were saved after intervals of 2ps. We performed the analyses CPPTRAJ module of AMBER[31] and visualizations were performed using VMD [32].The binding energies for the complexes werecalculated using MMGBSA[33]suite of AMBER.

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.26.437194; this version posted March 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

2.4 Graph theory

We used graph theory to decipher the composition of interactomes in terms of participating amino acids involved in pairwise interactions. Briefly,the graph structure G(V,E, W) is denoted by three sets. The first one is, the set of residues, denoted as the vertex set (V) of a graph. Each individual residue, here, is considered as an independent entity, formally termed as a vertex or a node. Say, we denote V as $\{v_1, v_2, v_3, ...,v_n\}$, where v_i is the ith residue. Therefore, |V| = n, where n is the number of nodes in the graph, otherwise also known as the order of the graph. The second set is the set of interactions between residues denoted as the edge set (E). The interaction between the ith and the jth residue may be represented as an edge $e(v_i, v_j)$ and the edge set E may be represented as $\{e_1, e_2, e_3, ..., e_m\}$. Therefore, |E| = m, where m is the number of edges in the graph, otherwise also known as the size of the graph. Please note that the edges have not been considered as directed because, there is no significance of the roles of the interacting residues in these interactions. Weinitially calculate the average energy values (calculated per unit time) over the time of observation for all ${}^{n}C_2$possible interactions. Some of them turn out to be high and are deemed insignificant. We use a threshold to remove the average energy values of those interactions. Here, m is the number of interactions with significant average energy values.These average energy values represent the importance of the interactions and may be denoted as the set of edge weights (W). For every edge, there is a corresponding edge weight, therefore it may be concluded that |W| = m.

In this work, we are interested in studying the interaction dynamics of the residues. Due to the difference in interaction energies, from observation, we could intuitively understand that a group of residues are more prone to interact among themselves than the other residues. But, to discover underlying densely interacting residue groups or clusters, we apply algorithms that could reveal the clusters accurately. The equivalent problem in residue-residue interaction graphs or networks is known as graph clustering. We have used one of the most popular community detection techniques, Louvain method, to find out the clusters in this residue interaction network[34]. Please note that we have used the terms cluster and community interchangeably. It provides us with a cover $C = \{c_1, c_2, c_3, ….., c_k\}$, where k is the number of communities and c_i is the ith cluster/community. Each vertex in V belongs to exactly one of the clusters. Therefore, union of the vertex sets of all the clusters would lead back to V.

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.26.437194; this version posted March 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

2.5 Modelling the protein-protein interaction complex

We used hierarchical approach to predict the structure of the protein in the absence of a suitable template structure for caveolin-1.I-TASSER server was used to generate five initial models[35-37]. One model was selected based on the C-score (confidence score). The model was then evaluated using the SAVES v5.0 server, where Ramachandran plot and ERRAT analyses were performed[38,39]. Model visualizations were done using Chimera[40]. This model was then simulated for 100ns to generate a more stable structure. The average structure was then considered as the initial structure for docking after proper structural evaluation by Ramachandran Plot and ERRAT analyses[38,39]. The two molecules of human caveolin-1 were docked to the WT ORF3a by using HADDOCK[41]. HADDOCK not only considers traditional energetics and shape complementarity, but also incorporates experimental data in terms of restraints to guide the docking of two proteins.The residues of domain IV on ORF3a and the residues Asp82 to Arg101 on human caveolin-1 were defined as active residues in docking based on the cryo-EM structure information[42]. On the basis of the most negative binding energy,we selected a starting structure for the WT ORF3a-caveolin-1 complex. Necessary mutations were introduced in these structures by modelling with the Swiss PDB Viewer. These structures were subjected to all atomic MD simulations.

3Results:

3.1 Worldwide prevalence of D155Y and S171L substitution of ORF3a

ORF3a protein is important for the viral infection, spreading and modulating the host immune system. To understand the role of mutations in the function of this protein, we first checked the prevalence of each mutationsfound in ORF3a. Fig. 1a shows the number of mutant samples at each position of the ORF3a protein in the global and Indian population from a total of 26,656 samples, deposited in NCBI dataset (dated November 17, 2020). From this figure, we observe that mutations occurat 258 positions of the protein for the global population. Whereas, for the Indian population, mutations were found only at 13 positions.Mutations at the 57th position were the highest in both the global and the Indian population. For the Indian population, the positions 155th and 171st showed two instances of mutations while the number of instanceswas 23 and 34 respectively in the global population. Fig. 1b shows the distribution of these mutations worldwide.The mutations in ORF3a protein

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.26.437194; this version posted March 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

for both world and Indian populations are distributed in the entire protein.Table S1 shows a list of the mutation counts we have found at all the positions of the ORF3a protein sequence.

3.2 Description of the protein systems

The SARS-CoV-2 ORF3a protein can form dimer[43].The monomeric ORF3a has been dividedinto six domains, each having its own functional importance[20]. Fig. 2 shows the locations of D155Y and D155'Y (red spheres) and S171L and S171'L (cyan spheres). The locations of these substitutions between domains IV and V, and in domain VI suggest their possible role in caveolin binding, intracellular protein sorting and intracellular membrane trafficking of ORF3a[20].

3.3Stability of the two ORF3a variants, D155Y and S171L

The cryo-EM structure of WT ORF3a protein (residues 40 to 238) was downloaded from the Protein Data Bank (PDB ID: 6XDC) and processed as discussed in the methodology section. The substitutions D155Y and S171L on each monomer were modelled on the WT structure separately using SwissPDBViewer[27]. Each of these structures was simulated in triplicate till 200ns. Fig. 3 shows thetime evolution of the root mean square deviation (RMSD) of the simulated structure with respect to the starting frame of simulation. Note that WT and both mutants have shown reasonable stability. WT and D155Y (black and red profiles in Fig. 3) showed lesser RMSD (the final RMSD being 2.25Å) and lesser fluctuation, whereas the S171L (green profile in Fig. 3) variant showed higher RMSD (the final RMSD being 2.75Å).The overall fluctuation in RMSD was also greater in S171L compared to the WT and D155Y. This indicates that S171L substitution causes more deviation and may interfere with its membrane localisation.However, the final RMSD values attained by the WT and the two mutants were comparable, indicating a similar final simulated structure.So it can be concluded that the substitutions at D155Y and S171L do not cause a major conformational change of ORF3a from the WT.

3.4 Differential behaviour of the constituent residues

While RMSDs are a measure of the overall stability of the biological systems under consideration, the B-factor values give an idea on the flexibility of the individual residues. The B-factor values were measured to determine the average flexibility of the protein residues around their mean position across all trajectories (Fig. 4). Fig. 4 shows thatthe

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.26.437194; this version posted March 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

residues in the WT ORF3a protein exhibit the least deviation from their mean position, whereas the residues in both mutants show moreflexibility. Interestingly,in the D155Y variant, the 155th residue showed higher flexibility compared to the WT(14.52 Å^2 for WT and 61.7 Å^{2}for D155Yand16.95 Å^2 for WT and 55.3 Å^{2}for the D155Yat positions 155 and 155' respectively). Similarly, for the S171L variant, the flexibility of the 171st residue in the mutant was higher than the WT or D155Y(13.69 Å^2 for WT and64.94 Å^{2}and 26.59 Å^2 for WT and 124.48 Å^{2}for S171Lat positions 171 and 171'respectively). The terminal residues are exposed to solvent and are more flexible, resulting in their high B-factor values as seen in Fig. 4 (as marked). Several other residues which are located both near the positions of mutations as well as distallyalso showed greater flexibility. We thus observe an effect of the mutation on the overall dynamics of the protein at distant locations. Thus, these mutations may have allostericeffects on the domain specific function of the ORF3a protein.

3.5 Differential contribution of stabilizing residues in the ORF3a proteins

The WT and the mutant ORF3a proteinswere analysed to understand the role of individual amino acids, hydrogen bond occupancies and salt bridges in their structural stability. The free energies of the three variants(WT and the two mutants) were calculated by using the MMGBSA module of Amber20 and are tabulated in Table 1, which shows the differences in stability among them.The S171L mutant with a free energy of -5376.51 (± 19.34) kcal/molwas the most stable, followed by WT (-5356.85± 12.95kcal/mol) and D155Y mutant (-5266.41± 12.56kcal/mol).This indicates that the mutants D155Y and S171L can also exist independently just like the WT ORF3a protein. To understand the contributing factors for these variations in stabilizing energy, we looked at the contributions of eachamino acidto the overall free energy and tabulated the top contributors for each variantin Table 2. While we observe that the group of residues contributing to the overall stability remains almost unchanged among the variants, their ranking differs. For instance, in WT, Arg68 plays the most important role, whereas in case of the mutant systems, it is Arg126' that has the most contribution. However, Arg68 features as the second most contributory residue in D155Y mutant, whereas in the S171L mutant, it has the fourth position. In this mutant, the Arg126 plays the second most important role, following its corresponding residue on the second monomer.

We also checked the hydrogen bond interactions in WT and the two mutant ORF3a proteins, and found that thetotal number of hydrogen bonds remain same (average number is 95, Fig. S1), inall the three variants.In contrast, the individual residuesthat have the most hydrogen

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.26.437194; this version posted March 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

bond occupancy vary among the ORF3a proteins. We found that the top three residue pairs involved in forming hydrogen bonds with the maximum occupancy are Tyr156'-Lys192', Arg134-Asp155 and Ser205'-Asn144'. In D155Y, the top three residue pairs forming the hydrogen bonds with maximum occupancy are Tyr212'-Thr164, Ser205'-Asn144' and Ser205-Asn144. In the mutant S171L, the top three residue pairs forming hydrogen bonds with maximum occupancy areLeu203-Asp210, Leu203'-Asp210' and Thr89-Leu85'. A detailed list is given in Table S2.We also calculated the salt bridge interactions for the WT and the two mutant proteins and tabulated the list of salt bridges in Table S3, which shows that D155Yforms lesser number (24) of salt bridges compared to the WT and the S171L (31 each). Interestingly, mutation at position 155, but not at 171, breaks the salt bridge formation between Asp155-Arg134. This residue pair is formed at the end of the alpha helix and the beginning of a beta sheet in the proximity of domains III and IV of ORF3a.Thus this loss of salt bridge interaction in D155Y is significant and may play a role in the binding affinity of the interacting partner of the ORF3a protein at this region.

3.6 Changes in interactome interactions – Agraph theoretic perspective

The variation in hydrogen base pairing and salt bridges prompted us to check the interactome interactions in ORF3a variants. We represented the interactions between the interactomes in terms of a network. The pairwise hydrophobic interaction energies of the residues in theWT and the two mutants of ORF3a were calculated using MMGBSA suite of Amber20. These hydrophobic interaction energies were considered for building a residue-residue interaction network for the WT and the two mutant proteins. Here, we have used graph data structures and relevant algorithms, to model the interactionsamong the residues.

The spatial orientation of the protein, adjacency of the residues and interactions among them play a role in finding the clusters or communities of interacting residues. In the visualization of the clusters, as seen in Fig. 5a, we see the whole interaction network and an overview of the clusters. In Fig. 5b, we zoom on one part of the graph and provide a closer view of the interactions. The node colours denote its affiliation to a certain cluster. The edge colours are determined by the colours of the nodes it is incident upon. The edge thickness denotes the strength of the interactions between the residues, i.e., the weight of the edge. In Fig. 5c, one residue has been selected to show the nodes adjacent to it (also known as its neighbourhood). Fig. 6 shows these clusters as can be seen in the actual protein. We observe that the membership of the residues in the clusters in each protein has shown substantial variation. A

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.26.437194; this version posted March 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

list of the clusters and their constituent residues has been provided for the WT and the two mutants (D155Y and S171L) in Table S4. Fig. 6 and Table S5 indicate that the residues of the functional domains have rearranged in different interacting clusters in WT and the two mutants. Domain III being the largest in size has split into the most number of clusters. However, the clusters are different in terms of the constituent residues for WT and the two mutants. Thus, we may conclude that the mutations have changed the interaction patterns of the interactomes present in the protein. Due to changes in residue interactions, the clusters have changed from WT to the other mutants. But it should be noted that the cluster membership for the nodes in the regions of mutations do not change. This indicates that the mutations may have distal effects too, which can be explored further for better understanding of the protein function.

3.7 Formation of complex with partner protein caveolin-1

We are interested to check if the substitution can change the binding interaction of ORF3a protein with host caveolin-1.Issa*et.al*.[20]have suggested that domain IV of ORF3a binds to caveolin-1 protein, which is required forviral uptake and regulation[44,45].We modelled caveolin-1 using a hierarchical approach to predict the structure of the protein. Five initial models were then generated using the I-TASSER server. Out of these, one model was selected based on the C-score (confidence score). This model was then evaluated using the SAVES v5.0 server, where Ramachandran plot analysis and ERRAT analysis were performed as shown in Fig. 7a-b. Ramachandran plot showed that 96.9% of the residues of caveolin-1 were within the favoured and allowed regions, while 3.1% of the residues were in the disallowed regions. On the other hand, ERRAT analysis had an overall quality factor of 89.412.

The modelled structure of human caveolin-1 was simulated for 100ns to generate a well equilibrated and stable structure. The stability of the simulation, as evident from the time evolution of the RMSD of the protein from its starting structure, has been shown in Fig. S2a. The average structure from this simulation (Fig. S2b) was considered as the starting structure of the ORF3a-caveolin-1 complexesafter proper structural evaluation (Fig. 7c-d). Ramachandran plot of the average simulated and stable structure showed that 98.1% of the residues were within the ranges of favourable and allowed regions, and only 1.9% of the residues were in the disallowed regions. ERRAT plot too showed an improvement with the overall quality factor increased to 94.304%.Thus, both Ramachandran plot and ERRAT

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.26.437194; this version posted March 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

analysis (Fig. 7c-d) indicated that thecaveolin-1 model was of acceptable quality, and could be used as the starting structure for docking.

We carried outour protein-protein docking using the HADDOCK webserver[41,46]. We consider binding domains on ORF3a and caveolin-1 as the interacting residues [42]. Our analysis generated twelve probable structures from three clusters as shown in Fig. S3. In each of these structures, we had two molecules of the human caveolin-1 interacting with the dimeric form of ORF3a protein. The top structure from the topmost cluster (left-most structure in first row of Fig. S3), having a HADDOCK score of -155.3 (±22.2) was considered as our starting structure. This structure showed a symmetrical nature. The buried surface area of this complex was found to be 3031.4 (± 181) $Å^2$, signifying a strong complex. The necessary mutations were introduced into the protein-protein complex by Swiss PDBViewer.

The starting structures for WT and the two mutants were simulated for 100ns. The stability for these structures was assessed by plotting the time evolution of their RMSD values with respect to the starting frame of simulation (as shown in Fig. 8). We observe that the WT and the S171L are stable having an average RMSD value around ~-9Å. Although the absolute value is high, yet the protein complexes reached stability and showed a plateau in the RMSD plot from ~40ns, again indicating a stable complex. However, for the D155Y system, the protein complex showed a lot more fluctuation and deviation from the starting structure. This indicates a not-so-stable complex structure, which is further supported by the lower PROVEAN score of D155Y (Table S6). Since the mutation is present in the vicinity of the caveolin binding domain in ORF3a, it can be said that the presence of the mutation leads to an unstable protein-protein complex formation. Thus, the D155Y substitution interfereswith the caveolin binding activity of ORF3a protein. We also calculated the free energy,corresponding to the binding of caveolin-1 to the ORF3a protein, in these three protein complex systems. The values for WT, D155Y and S171L were -37.6385 (±8.3248) kcal/mol, -11.5504 (±2.9333) kcal/moland -31.9254 (± 5.0812) kcal/mol, respectively. From these values, it is evident that the binding affinity of caveolin-1 is considerably less in D155Y mutant compared to WT and S171L. This is in corroboration with the unstable protein-protein complex in the D155Y system. The change in hydrogen bonding, salt bridge patternand hydrophobic interaction pattern associated with D155Y substitution may have contributed to the weakened interaction betweenD155YORF3a and caveolin-1.

4 Discussion:

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.26.437194; this version posted March 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

In this study, we have established that D155Y substitution changes the intramolecular hydrogen bonding, salt bridge formation, and disrupts the interaction between ORF3a and caveolin-1.

Several other mutations are present in ORF3a of SARS-CoV-2. Wu*et. al.* have shown that the incidence of mutation at position 57 is high, compared to the other positions[47]. In order to consider the effect of mutation at the 57th position in D155Y and S171L, we simulated Q57H, Q57H-D155Y and Q57H-S171L variants of ORF3a, for 200ns. Their structural stabilities were calculated with MMGBSA and have been tabulated in Table 3. We noted that the structural stabilities for Q57H and the Q57H-S171L variants were comparable to WT, D155Y and S171L (values listed in Table 1).The Q57H-D155Y variant had considerably lesser stability.Previously Hassan *et.al.* reported the presence of H at the 57th position in ORF3a protein of pangolin CoV[22].Thus, we may hypothesize that the presence of H may provide natural stability of ORF3a.We checked the structural stability of W131C, W131R, G172C and G172V, which were found in Indian patients. We simulated the variants W131C, W131R, G172C and G172V for 200ns. These four variants were stable and showed an average RMSD value of 2.5Å with respect to the starting structure as shown in Fig. S4. The overall binding free energies for the four variants were also calculated, and we noted that the stabilities for the systems W131C, G172C and G172V were similar to that of the WT ORF3a protein as listed in Table 1. The variant W131R, was more stable than WT ORF3a. This indicates that these four mutants are very stable and can have independent existence. Further study is needed to check the effects of these substitutions both *in silico* and *in vitro*. On the contrary, mutation at 155th position (D155Y) reduced the binding affinity of ORF3a to caveolin. The disrupted interaction can be indicative of improved viral fitness, wherein, the virion particles can continue to build the host intracellular viral load without inducing host cell apoptosis or promoting their egress thus lengthening the asymptomatic phase of the infection. Contrariwise, the ORF3a-caveolin-1 affinity change can also affect the virion internalisation into host cells, endomembrane sorting and assembly of the viral components.

Direct Coupling Analysis revealed eight genes involved in epistatic interactions at several polymorphic loci. The locusof ORF3a is involved in three out of eight potentially significant epistatic links with, namely, nsp2, nsp6 and nsp12. These intragenetic interactions open up the possibility of potential evolutionary links of the above described substitutions at D155Y and S171L with other positively selected loci in viral genes, which is reportedly subject to demographic variations[48].Moreover, the Neanderthal-derived COVID-19 risk haplotype is

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.26.437194; this version posted March 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

altogether positively selected in some populations and has 30% allele frequency thus introducing an evolutionary landscape to the current COVID-19 pandemic[49].

Our simulation studies and further analyses ofORF3a protein have shown that the presence of mutations affects the structural stability of the ORF3a protein. The residues involved in forming several stabilising interactions in these proteins also change with the presence of mutations. Although the overall stability of the protein structures in the WT and the two variants are not much different,these mutations may affect the binding affinity of ORF3a with its partner proteins. For instance, in this study we have shown the presence of mutation drastically reduces the binding affinity of ORF3a with caveolin-1.SARS-CoV-2 enters the host cell by both membrane fusion and by clathrin/caveolin-mediated endocytosis after binding to the ACE2 cell-surface receptors in the upper respiratory tract and alveolar epithelial cells[50-52].Caveolin/cholesterol mediated endocytosis has been previously implicated in SARS-CoV through an in silico study wherein several caveolin-1 binding domains (CBD) were found in SARS-CoV proteins and internalisation of virion was proposed to be facilitated in a caveolin-1 and lipid raft-dependent manner. However, the role of caveolin-1 was not limited to viral entry.Rather, it was associated with all stages of viral life cycle starting from virus binding to surface receptors, fusion and endomembrane trafficking of virus in caveosomes, sorting of viral components to endomembrane surfaces, replication, assembly and to subsequent egress. The host-derived lipid bilayer surrounding the enveloped viral nucleocapsid contains caveolin-1 incorporated during viral fission from the host membrane [53]. Thus, binding interactions of SARS-CoV-2 ORF3a WT and mutational variants with caveolin-1 provides a putative alternative route for viral pathogenesis in COVID-19. Change in interaction of D155Y ORF3a with caveolin-1 may provide an alternative route to exhibit SARS-CoV-2 virulence properties in COVID-19 patients.

Cryo-EM structural analysis of SARS-CoV-2 proteins shows that ORF3a can exist in dimeric and tetrameric complex arrangements with six functional domains (Domain I to VI) of each protomer. The protein comprises of 3 helices, spanning the transmembrane domain (Domains II and III) and a cytosolic domain with multiple beta-strands (Domain IV, V, VI) [20,54]. Each domain of SARS-CoV ORF3a interacts with different host proteins and modulateshost signalling pathways. Domain II has binding sites for TRAF3 and ASC, and interacts to activate NLRP3 inflammasome[5].Domain III has a conserved Cysteine residue at 133rd positionknown to stabilize ORF3a homodimer and homotetramers for its ion channel

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.26.437194; this version posted March 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

activity[43,54].Cytosolic domain IV has a conserved motif YDANYFVCW from amino acids 141-149 that binds with host caveolin-1[55].Finally, domains V and VI, comprising of the YXXφ and diacidicExD motifs respectively, are essential for intracellular viral protein sorting, trafficking and localization of ORF3a to the host membrane followed by its release into culture medium[14]. The various other host proteins as binding partners to ORF3a protein include components of the anti-inflammatory pathway HMOX1, innate immune signalling pathway, TRIM59, glycosylation pathway (ALG5) and nucleus-inner-membrane proteins (SUN2 and ARL6IP6)[56]. ORF3a also regulates Caspase 8-mediated extrinsic apoptotic pathway for its pro-apoptotic activity in HEK293T cells[17]. Thus, mutations in the binding regions or in its close proximity may interfere with the host protein-viral ORF3a interaction, which needs to be further validated. SARS-CoV-2 ORF3a was among the other candidate proteins that has been found to elicit significant CD4+ and CD8+ T-cell response and it has been suggested that an optimal vaccine should be inclusive of class I epitopes derived from M, nsp6 and ORF3a[57-59].Our *in silico* study provides support to carry out *in vivo* and *in vitro* studies for evaluating viral pathogenesis with mutant SARS-CoV-2.

Author contributions: PB and SSJ conceptualized the study. SG and PB performed experimental work. SG, PB, DM, KB, SS and SSJ contributed to analysis of the results and preparation of the figures. SG, DM and KB were involved in the bioinformatics analysis of the data. SG, DM and KB wrote the initial draft of the manuscript. SSJ and PB edited the manuscript with input from all the authors. All authors agreed to the submission of this work to the Journal of General Virology.

Conflicts of interest:

The authors declare that there are no conflicts of interest.

Funding Information: This study was funded by Technical Research Centre (TRC), Indian Association for the Cultivation of Science (IACS), Kolkata, India.

Acknowledgements

This work was supported by the Indian Association for the Cultivation of Science (IACS). We thank Google Cloud Research Credits (GCRC) program and the high-performance computing (HPC) support provided by Fluid Numerics. We also thank IACS for fellowship to

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.26.437194; this version posted March 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

SG and DM, DST-INSPIRE for fellowship to KB, and SERB-DIA (DIA/2018/000005) award to SSJ.

Abbreviations:

ARL6IP6: ADP Ribosylation Factor Like GTPase 6 interacting protein 6, ASC: Apoptosis-associated speck-like protein containing a caspase recruitment domain, BLAST: Basic Local Alignment Search Tool, CD4+: Cluster of Differentiation 4+, CD8+: Cluster of Differentiation 8+, COVID-19: Coronavirus Disease 2019, Cryo-EM: Cryo Electron Microscope, HMOX1: Heme Oxygenase 1, IFN: Interferon, MERS-CoV: Middle East respiratory syndrome coronavirus, MMGBSA: Molecular mechanics with generalized Born and surface area solvation, NCBI: National Centre for Biotechnology Information, NF-κB: Nuclear factorkappa light chain enhancer of activated B cells, NLRP3: Nucleotide-binding oligomerization domain, Leucine rich repeat and Pyrin domain containing, ORF: Open Reading Frame, PDB: Protein Data Bank, PISA: Protein Interfaces Surfaces and Assemblies, PROVEAN: Protein Variation Effect Analyzer, RMSD: Root Mean Square Deviation, SUN2: SUN domain-containing protein 2, TRIM59: Tripartite motif-containing protein 59.

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.26.437194; this version posted March 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Figure legends

Fig. 1 Distribution of mutations in ORF3a protein. The number of instances of the various mutations at different positions of ORF3a protein in the total number of samples considered in this study has been plotted in (a). The global distribution of the mutations at positions 155 and 171 have been shown in blue and red respectively in (b)

Fig. 2 Structure of ORF3a protein. The structure of WT ORF3a (PDB ID: 6XDC) marking the functional domains as known from literature has been shown. The positions of mutation at the 155th and 171st positions have been shown in orange and cyan spheres respectively.

Fig. 3 Stability of structure. The time evolution of the RMSD of the ORF3a proteins with respect to the starting structure. (a) Black: WT, (b) Red: D155Y and (c) Green: S171L

Fig. 4 Flexibility of residues. The B-Factor plot for ORF3a for the three systems. (a) Black: WT, (b) Red: D155Y and (c) Green: S171L

Fig. 5 Visualization of residue interaction network in SARS-CoV-2 ORF3a protein using Gephi[60]**.**(a) The whole residue interaction network showing the complete cover C, with nodes coloured with the membership colour of a particular cluster, (b) A magnified view of the residue interaction network, and (c) Shows one particular residue (here, GLY209) and the residues it is directly interacting with.

Fig. 6 Interactome clusters in SARS-CoV-2 ORF3a protein. The different interactomes are shown for (a) WT, (b) D155Y and (c) S171L

Fig. 7 Modelling the human caveolin-1 structure. (a) The ERRAT analysis of the modelled structure of caveolin-1. (b) The distribution of the residues of the modelled structure on the Ramachandran Plot.(c) The ERRAT analysis of the simulated structure of caveolin-1. (d) The distribution of the residues of the simulated structure on the Ramachandran Plot.

Fig. 8 Stability of the ORF3a-caveolin-1 complex. The time evolution of the RMSD of the ORF3a-caveolin-1 complex with respect to the starting structure. (a) Black: WT-caveolin-1, (b) Red: D155Y-caveolin-1 and (c) Green: S171L-caveolin-1

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.26.437194; this version posted March 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

References

1. **V'kovski P, Kratzel A, Steiner S, Stalder H, Thiel V**. Coronavirus biology and replication: implications for SARS-CoV-2. *Nature Reviews Microbiology* 2020;19:155–170.

2. **Asia S-WHO**. India Situation. 2020;10–11.

3. Mortality Analyses - Johns Hopkins Coronavirus Resource Center. https://coronavirus.jhu.edu/data/mortality (accessed 31 December 2020).

4. **Naqvi AAT, Fatima K, Mohammad T, Fatima U, Singh IK, *et al.*** Insights into SARS-CoV-2 genome, structure, evolution, pathogenesis and therapies: Structural genomics approach. *BBA - Mol Basis Dis* 2020;1–17.

5. **Siu KL, Yuen KS, Castano-Rodriguez C, Ye ZW, Yeung ML, *et al.*** Severe acute respiratory syndrome Coronavirus ORF3a protein activates the NLRP3 inflammasome by promoting TRAF3-dependent ubiquitination of ASC. *FASEB J* 2019;33:8865–8877.

6. **Tan Y-J, Teng E, Shen S, Tan THP, Goh P-Y, *et al.*** A Novel Severe Acute Respiratory Syndrome Coronavirus Protein, U274, Is Transported to the Cell Surface and Undergoes Endocytosis. *J Virol* 2004;78:6723–6734.

7. **Firth AE**. A putative new SARS-CoV protein, 3c, encoded in an ORF overlapping ORF3a. *J Gen Virol* 2020;101:1085–1089.

8. **Finkel Y, Mizrahi O, Nachshon A, Weingarten-Gabbay S, Morgenstern D, *et al.*** The coding capacity of SARS-CoV-2. *Nature*. Epub ahead of print 2020. DOI: 10.1038/s41586-020-2739-1.

9. **Konno Y, Kimura I, Uriu K, Fukushi M, Irie T, *et al.*** SARS-CoV-2 ORF3b Is a Potent Interferon Antagonist Whose Activity Is Increased by a Naturally Occurring Elongation Variant. *Cell Rep* 2020;32:108185.

10. **Yount B, Roberts RS, Sims AC, Deming D, Frieman MB, *et al.*** Severe Acute Respiratory Syndrome Coronavirus Group-Specific Open Reading Frames Encode Nonessential Functions for Replication in Cell Cultures and Mice. *J Virol* 2005;79:14909–14922.

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.26.437194; this version posted March 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

11. **Oostra M, de Haan CAM, de Groot RJ, Rottier PJM**. Glycosylation of the Severe Acute Respiratory Syndrome Coronavirus Triple-Spanning Membrane Proteins 3a and M. *J Virol* 2006;80:2326–2336.

12. **Tan YJ**. The Severe Acute Respiratory Syndrome (SARS)-coronavirus 3a protein may function as a modulator of the trafficking properties of the spike protein. *Virology Journal* 2005;2:5.

13. **Zeng R, Yang RF, Shi M De, Jiang MR, Xie YH, *et al.*** Characterization of the 3a protein of SARS-associated coronavirus in infected vero E6 cells and SARS patients. *J Mol Biol* 2004;341:271–279.

14. **Huang C, Narayanan K, Ito N, Peters CJ, Makino S**. Severe Acute Respiratory Syndrome Coronavirus 3a Protein Is Released in Membranous Structures from 3a Protein-Expressing Cells and Infected Cells. *J Virol* 2006;80:210–217.

15. **Tan Y-J, Tham P-Y, Chan DZL, Chou C-F, Shen S, *et al.*** The Severe Acute Respiratory Syndrome Coronavirus 3a Protein Up-Regulates Expression of Fibrinogen in Lung Epithelial Cells. *J Virol* 2005;79:10083–10087.

16. **Minakshi R, Padhan K, Rani M, Khan N, Ahmad F, *et al.*** The SARS coronavirus 3a protein causes endoplasmic reticulum stress and induces ligand-independent downregulation of the Type 1 interferon receptor. *PLoS One* 2009;4:e8342.

17. **Ren Y, Shu T, Wu D, Mu J, Wang C, *et al.*** The ORF3a protein of SARS-CoV-2 induces apoptosis in cells. *Cellular and Molecular Immunology* 2020;17:881–883.

18. **Roy C, Mandal SM, Mondal SK, Mukherjee S, Mapder T, *et al.*** Trends of mutation accumulation across global SARS-CoV-2 genomes: Implications for the evolution of the novel coronavirus. *Genomics* 2020;112:5331–5342.

19. **Velazquez-Salinas L, Zarate S, Eberl S, Gladue DP, Novella I, *et al.*** Positive Selection of ORF1ab, ORF3a, and ORF8 Genes Drives the Early Evolutionary Trends of SARS-CoV-2 During the 2020 COVID-19 Pandemic. *Front Microbiol* 2020;11:550674.

20. **Issa E, Merhi G, Panossian B, Salloum T, Tokajian S**. SARS-CoV-2 and ORF3a: Nonsynonymous Mutations, Functional Domains, and Viral Pathogenesis. *mSystems*;5. Epub ahead of print 5 May 2020. DOI: 10.1128/msystems.00266-20.

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.26.437194; this version posted March 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

21. **Gonzalez-Reiche AS, Hernandez MM, Sullivan MJ, Ciferri B, Alshammary H, *et al.*** Introductions and early spread of SARS-CoV-2 in the New York City area. *Science (80-)* 2020;369:297–301.

22. **Hassan SS, Choudhury PP, Basu P, Jana SS**. Molecular conservation and differential mutation on ORF3a gene in Indian SARS-CoV2 genomes. *Genomics* 2020;112:3226–3237.

23. **Wu F, Zhao S, Yu B, Chen YM, Wang W, *et al.*** A new coronavirus associated with human respiratory disease in China. *Nature* 2020;579:265–269.

24. **Choi Y, Chan AP**. PROVEAN web server: A tool to predict the functional effect of amino acid substitutions and indels. *Bioinformatics* 2015;31:2745–2747.

25. **Kern D, Sorum B, Hoel C, Sridharan S, Remis J, *et al.*** Cryo-EM structure of the SARS-CoV-2 3a ion channel in lipid nanodiscs. *bioRxiv Prepr Serv Biol*. Epub ahead of print 2020. DOI: 10.1101/2020.06.17.156554.

26. **Krissinel E, Henrick K**. Inference of Macromolecular Assemblies from Crystalline State. *J Mol Biol* 2007;372:774–797.

27. **Guex N, Peitsch MC**. SWISS-MODEL and the Swiss-PdbViewer: An environment for comparative protein modeling. *Electrophoresis* 1997;18:2714–2723.

28. **Case DA, Belfon K, Ben-Shalom IY, Kollman PA, *et al.***(2020), AMBER 2020, University of California, San Francisco.

29. **Maier JA, Martinez C, Kasavajhala K, Wickstrom L, Hauser KE, *et al.*** ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *J Chem Theory Comput* 2015;11:3696–3713.

30. **Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML**. Comparison of simple potential functions for simulating liquid water. *J Chem Phys* 1983;79:926–935.

31. **Roe DR, Cheatham TE**. PTRAJ and CPPTRAJ: Software for processing and analysis of molecular dynamics trajectory data. *J Chem Theory Comput* 2013;9:3084–3095.

32. **Humphrey W, Dalke A, Schulten K**. VMD: Visual molecular dynamics. *J Mol Graph* 1996;14:33–38.

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.26.437194; this version posted March 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

33. **Miller BR, McGee TD, Swails JM, Homeyer N, Gohlke H, *et al.*** MMPBSA.py: An efficient program for end-state free energy calculations. *J Chem Theory Comput* 2012;8:3314–3321.

34. **Blondel VD, Guillaume JL, Lambiotte R, Lefebvre E**. Fast unfolding of communities in large networks. *J Stat Mech Theory Exp* 2008;2008:P10008.

35. **Yang J, Yan R, Roy A, Xu D, Poisson J, *et al.*** The I-TASSER suite: Protein structure and function prediction. *Nature Methods* 2014;12:7–8.

36. **Yang J, Zhang Y**. I-TASSER server: New development for protein structure and function predictions. *Nucleic Acids Res* 2015;43:W174–W181.

37. **Roy A, Kucukural A, Zhang Y**. I-TASSER: A unified platform for automated protein structure and function prediction. *Nat Protoc* 2010;5:725–738.

38. **Ramachandran GN, Ramakrishnan C, Sasisekharan V**. Stereochemistry of polypeptide chain configurations. *Journal of Molecular Biology* 1963;7:95–99.

39. **Colovos C, Yeates TO**. Verification of protein structures: Patterns of nonbonded atomic interactions. *Protein Sci* 1993;2:1511–1519.

40. **Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, *et al.*** UCSF Chimera - A visualization system for exploratory research and analysis. *J Comput Chem* 2004;25:1605–1612.

41. **Dominguez C, Boelens R, Bonvin AMJJ**. HADDOCK: A protein-protein docking approach based on biochemical or biophysical information. *J Am Chem Soc* 2003;125:1731–1737.

42. **Cai Q-C, Jiang Q-W, Zhao G-M, Guo Q, Cao G-W, *et al.*** Putative caveolin-binding sites in SARS-CoV proteins. *Acta Pharmacol Sin* 2016;24:1051059–1051059.

43. **Lu W, Zheng BJ, Xu K, Schwarz W, Du L, *et al.*** Severe acute respiratory syndrome-associated coronavirus 3a protein forms an ion channel and modulates virus release. *Proc Natl Acad Sci U S A* 2006;103:12540–12545.

44. **Ravid D, Leser GP, Lamb RA**. A Role for Caveolin 1 in Assembly and Budding of the Paramyxovirus Parainfluenza Virus 5. *J Virol* 2010;84:9749–9759.

45. **Padhan K, Tanwar C, Hussain A, Hui PY, Lee MY, *et al.*** Severe acute respiratory

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.26.437194; this version posted March 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

syndrome coronavirus Orf3a protein interacts with caveolin. *J Gen Virol* 2007;88:3067–3077.

46. **Pagadala NS, Syed K, Tuszynski J**. Software for molecular docking: a review. *Biophysical Reviews* 2017;9:91–102.

47. **Wu S, Tian C, Liu P, Guo D, Zheng W, *et al.*** Effects of SARS-CoV-2 mutations on protein structures and intraviral protein–protein interactions. *J Med Virol* 2020;jmv.26597.

48. **Zeng HL, Dichio V, Horta ER, Thorell K, Aurell E**. Global analysis of more than 50,000 SARS-CoV-2 genomes reveals epistasis between eight viral genes. *Proc Natl Acad Sci U S A* 2020;117:31519–31526.

49. **Zeberg H, Pääbo S**. The major genetic risk factor for severe COVID-19 is inherited from Neanderthals. *Nature* 2020;587:610–612.

50. **Inoue Y, Tanaka N, Tanaka Y, Inoue S, Morita K, *et al.*** Clathrin-Dependent Entry of Severe Acute Respiratory Syndrome Coronavirus into Target Cells Expressing ACE2 with the Cytoplasmic Tail Deleted. *J Virol* 2007;81:8722–8729.

51. **Wölfel R, Corman VM, Guggemos W, Seilmaier M, Zange S, *et al.*** Virological assessment of hospitalized patients with COVID-2019. *Nature* 2020;581:465–469.

52. **Fecchi K, Anticoli S, Peruzzu D, Iessi E, Gagliardi MC, *et al.*** Coronavirus Interplay With Lipid Rafts and Autophagy Unveils Promising Therapeutic Targets. *Frontiers in Microbiology* 2020;11:1821.

53. **Xing Y, Wen Z, Gao W, Lin Z, Zhong J, *et al.*** Multifaceted Functions of Host Cell Caveolae/Caveolin-1 in Virus Infections. *Viruses* 2020;12:487.

54. **Marquez-Miranda V, Rojas M, Duarte Y, Diaz-Franulic I, Holmgren M, *et al.*** Analysis of SARS-CoV-2 ORF3a structure reveals chloride binding sites. *bioRxiv Prepr Serv Biol*. Epub ahead of print 22 October 2020. DOI: 10.1101/2020.10.22.349522.

55. **Padhan K, Tanwar C, Hussain A, Hui PY, Lee MY, *et al.*** Severe acute respiratory syndrome coronavirus Orf3a protein interacts with caveolin. *J Gen Virol* 2007;88:3067–3077.

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.26.437194; this version posted March 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

56. **Gordon DE, Jang GM, Bouhaddou M, Xu J, Obernier K, *et al.*** A SARS-CoV-2 protein interaction map reveals targets for drug repurposing. *Nature* 2020;583:459–468.

57. **Mateus J, Grifoni A, Tarke A, Sidney J, Ramirez SI, *et al.*** Selective and cross-reactive SARS-CoV-2 T cell epitopes in unexposed humans. *Science (80-)*;370. Epub ahead of print 2 October 2020. DOI: 10.1126/science.abd3871.

58. **Grifoni A, Weiskopf D, Ramirez SI, Mateus J, Dan JM, *et al.*** Targets of T Cell Responses to SARS-CoV-2 Coronavirus in Humans with COVID-19 Disease and Unexposed Individuals. *Cell* 2020;181:1489-1501.e15.

59. **Le Bert N, Tan AT, Kunasegaran K, Tham CYL, Hafezi M, *et al.*** SARS-CoV-2-specific T cell immunity in cases of COVID-19 and SARS, and uninfected controls. *Nature* 2020;584:457–462.

60. **Bastian M, Bastian M, Heymann S, Jacomy M**. Gephi: An Open Source Software for Exploring and Manipulating Networks. *Int AAAI Conf WEBLOGS Soc MEDIA; THIRD Int AAAI Conf WEBLOGS Soc MEDIA*. http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.172.7704 (2009, accessed 6 February 2021).

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.26.437194; this version posted March 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

LIST OF TABLES

Table 1: The list of binding free energies for the three systems are given. The values in parentheses indicate their standard deviations.

System	Free Energy (kcal/mol)
WT	-5360.30 (12.95)
D155Y	-5263.66 (12.56)
S171L	-5375.66 (19.34)

Table 2: List of residues contributing to the overall stability of the systems

WT		D155Y		S171L	
Residue	Energy (kcal/mol)	Residue	Energy (kcal/mol)	Residue	Energy (kcal/mol)
ARG68	-173.181	ARG126’	-172.593	ARG126’	-173.754
ARG134	-172.764	ARG68	-171.184	ARG126	-172.424
ARG126	-170.217	ARG126	-170.909	ARG134	-172.153'
ARG134’	-169.607	ARG68’	-170.573	ARG68’	-171.78
ARG126’	-169.45	ARG134’	-168.276	ARG68	-170.225
ARG122	-168.216	ARG122	-166.244	ARG122	-168.148
ARG68’	-167.236	ARG122’	-165.746	ARG134’	-167.918
ARG122’	-164.614	ARG134	-164.769	ARG122’	-165.75
ASP238’	-99.689	ASP238	-100.33	ASP238’	-100.602
ASP238	-99.342	ASP238’	-99.694	ASP238	-100.401

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.26.437194; this version posted March 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Table 3: The list of binding free energies for the different mutant ORF3a proteins are given. The values in parentheses indicate their standard deviations.

System	Free Energy (kcal/mol)
Q57H	-5256.6964 (69.71)
Q57H-D155Y	-5113.0428 (43.07)
Q57H-S171L	-5251.6282 (68.42)
W131C	-5364.5810 (46.83)
W131R	-5740.5108 (53.43)
G172C	-5305.7174 (58.26)
G172V	-5379.1921 (61.07)

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.26.437194; this version posted March 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

LIST OF FIGURES

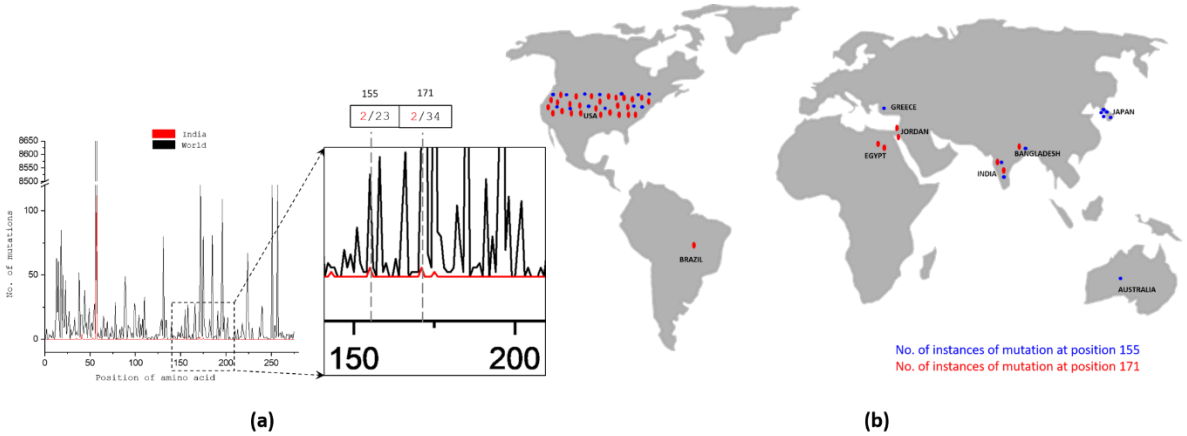

Fig. 1

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.26.437194; this version posted March 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

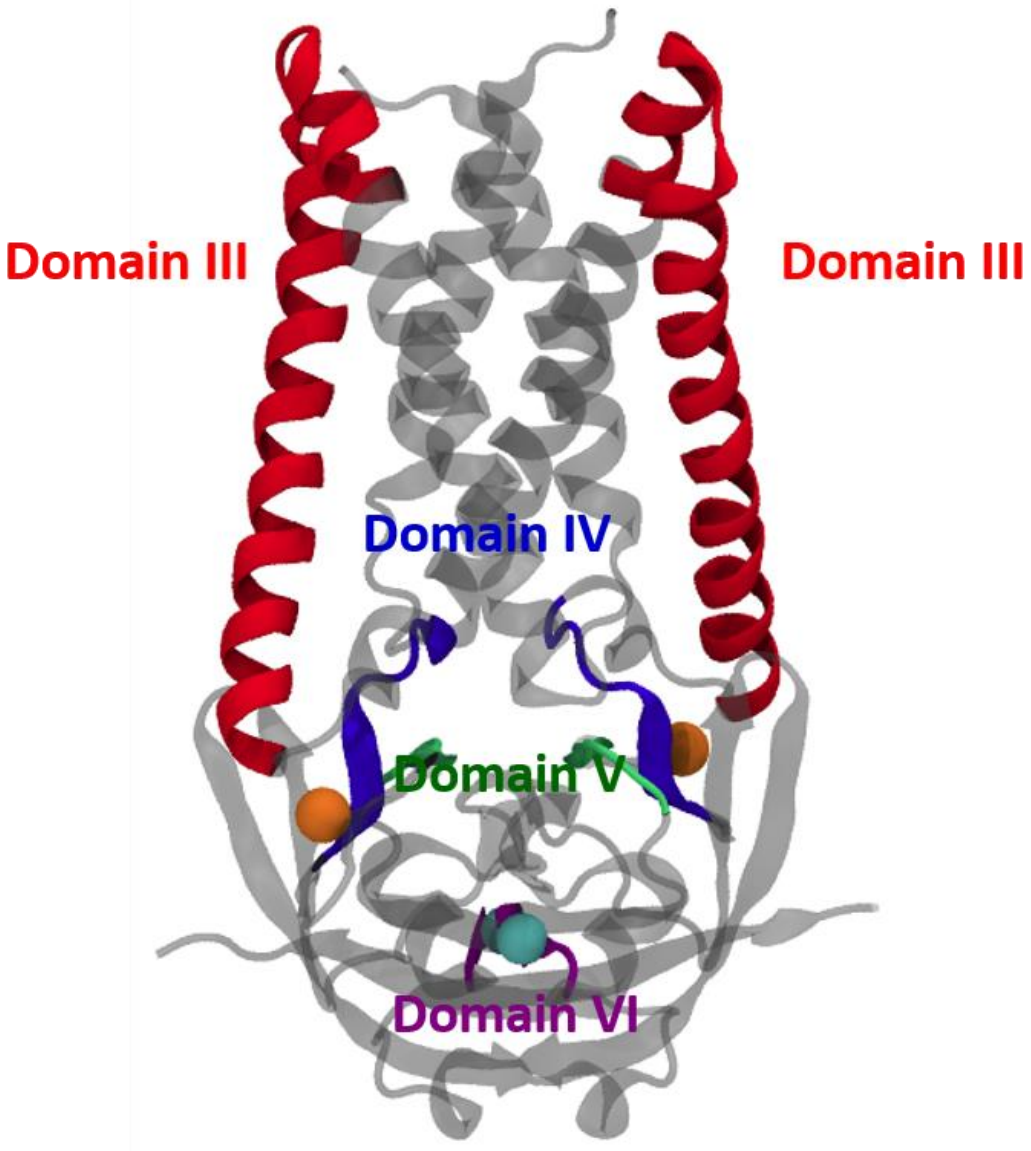

Fig. 2

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.26.437194; this version posted March 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

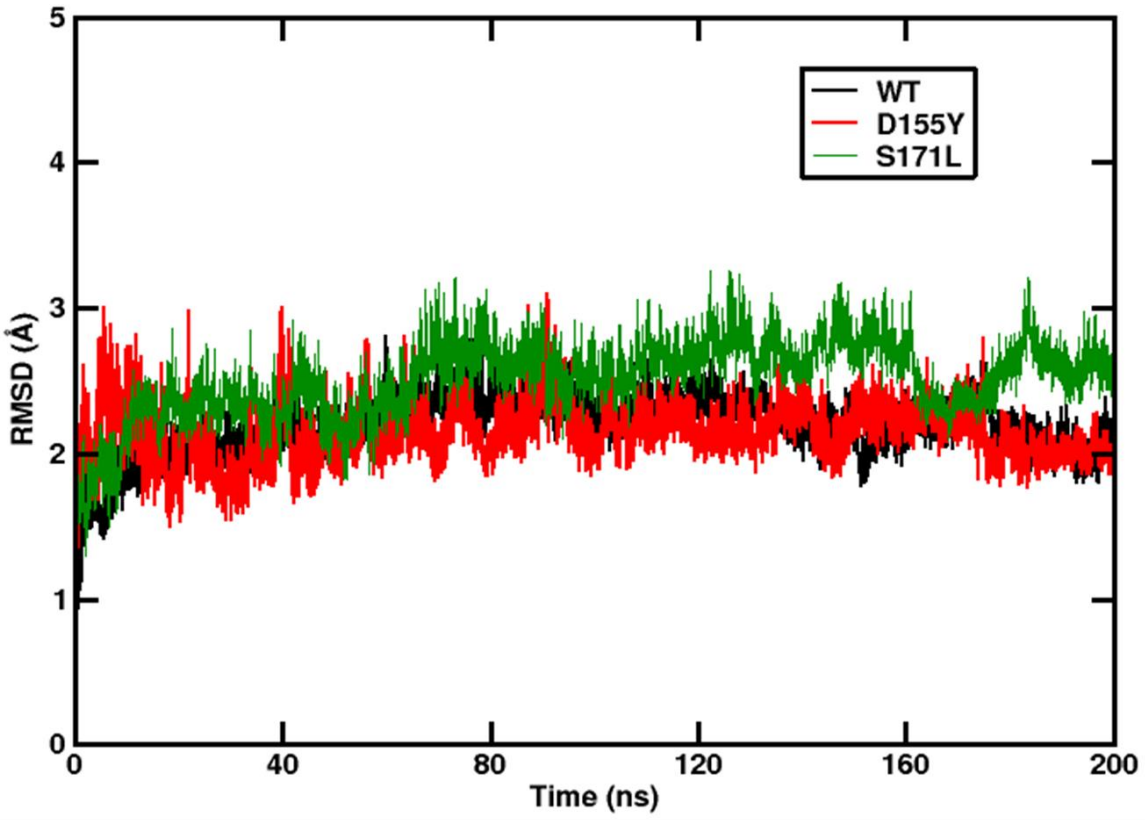

Fig. 3

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.26.437194; this version posted March 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

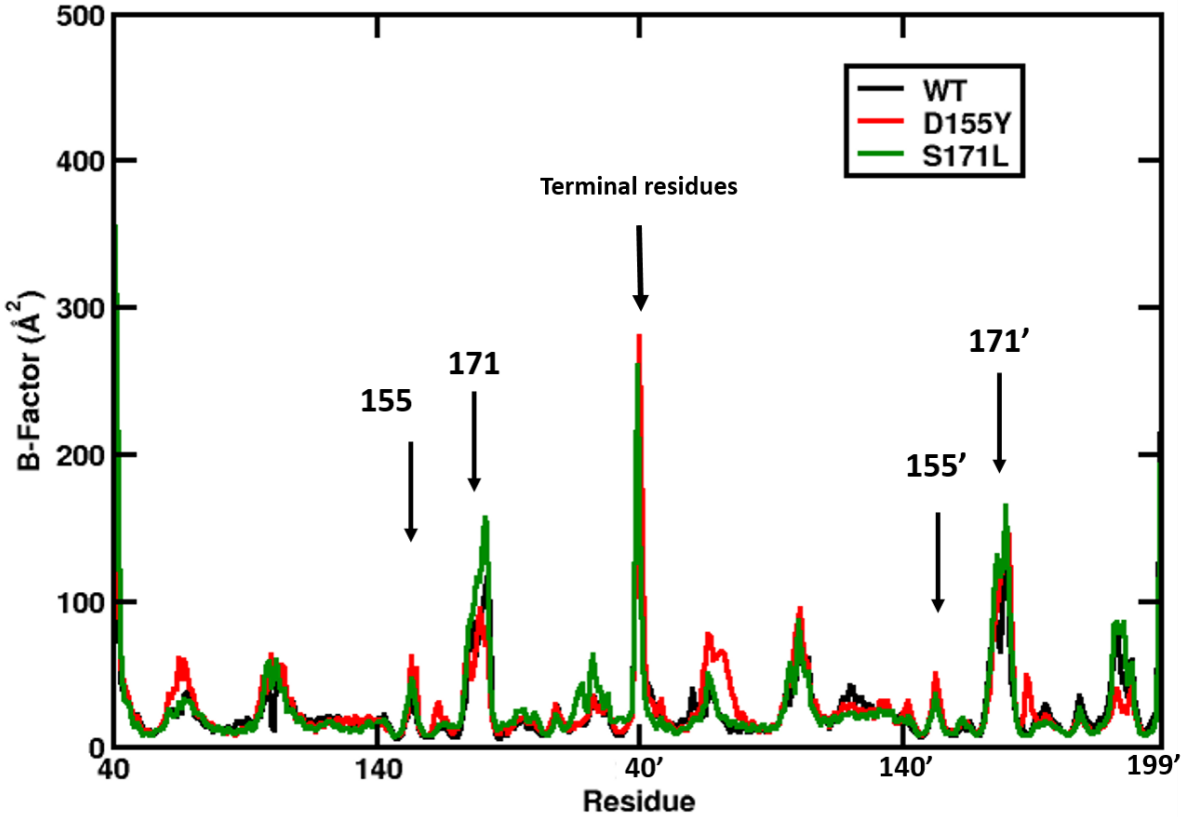

Fig. 4

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.26.437194; this version posted March 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

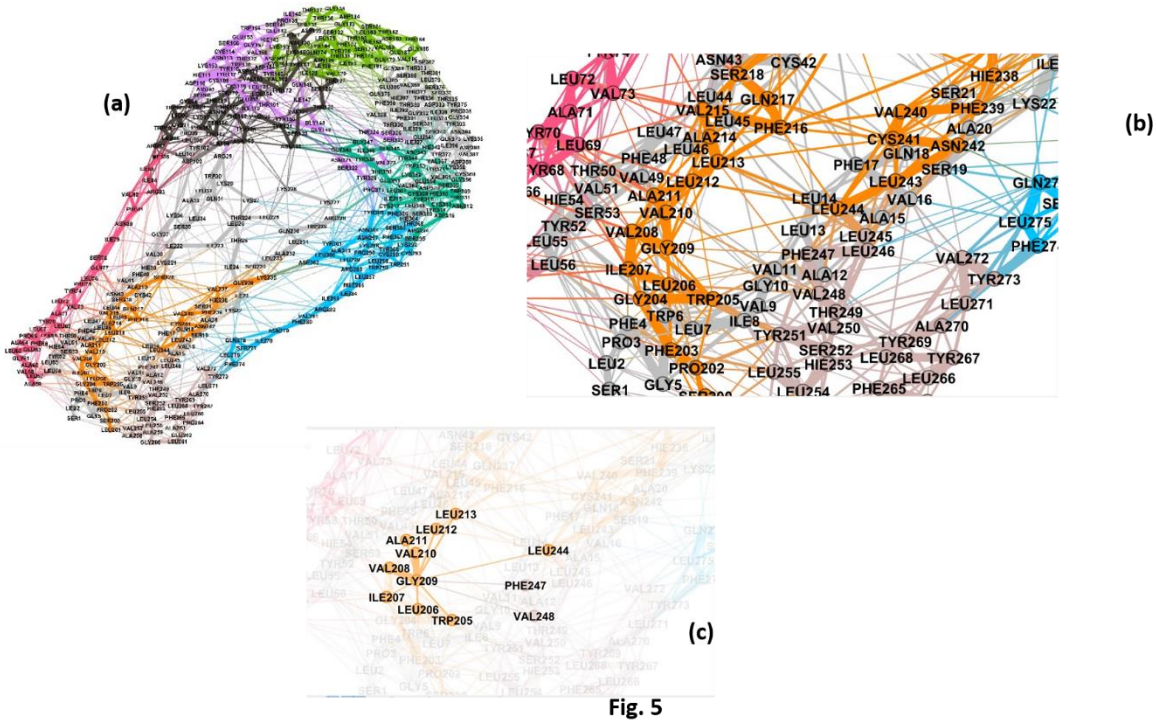

Fig. 5

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.26.437194; this version posted March 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

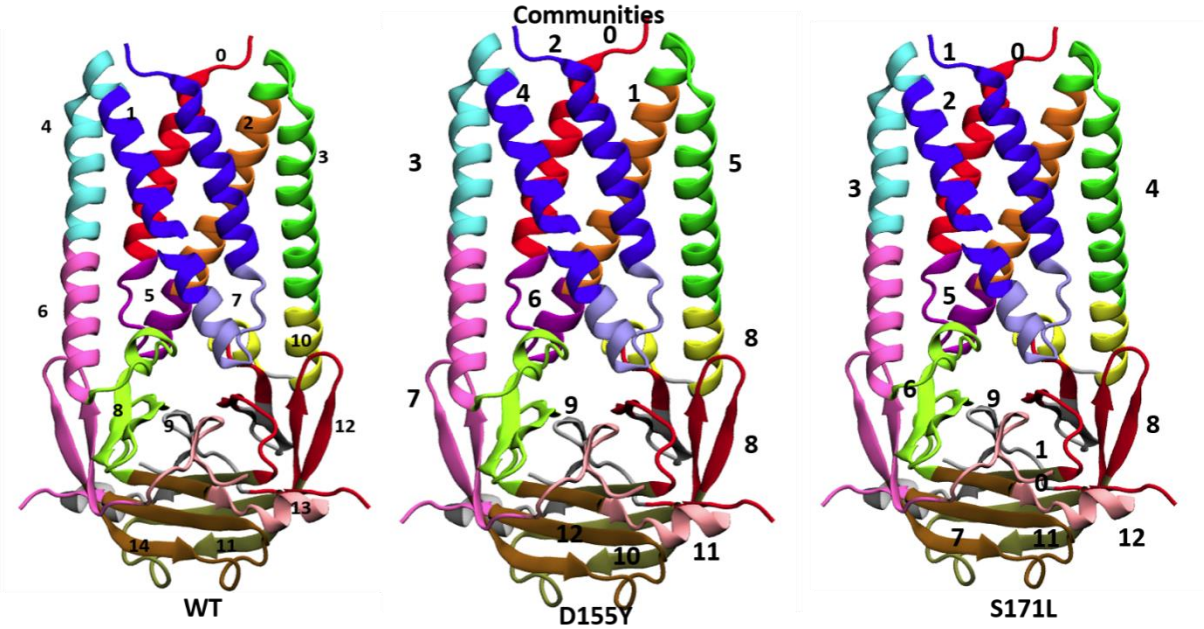

Fig. 6

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.26.437194; this version posted March 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

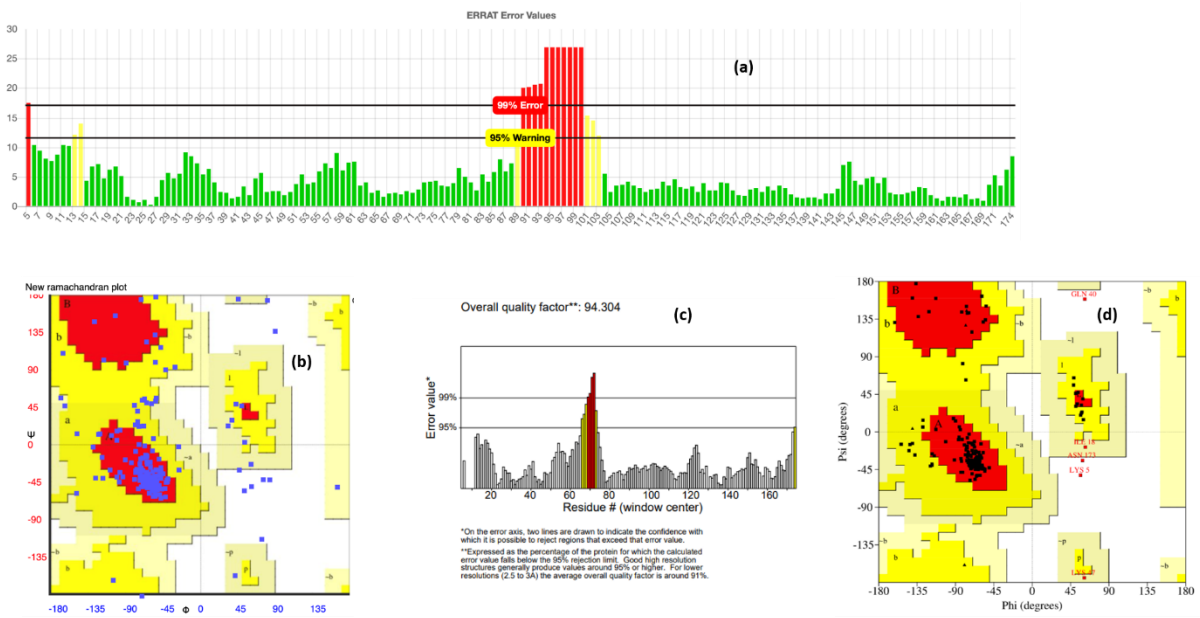

Fig. 7

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.26.437194; this version posted March 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

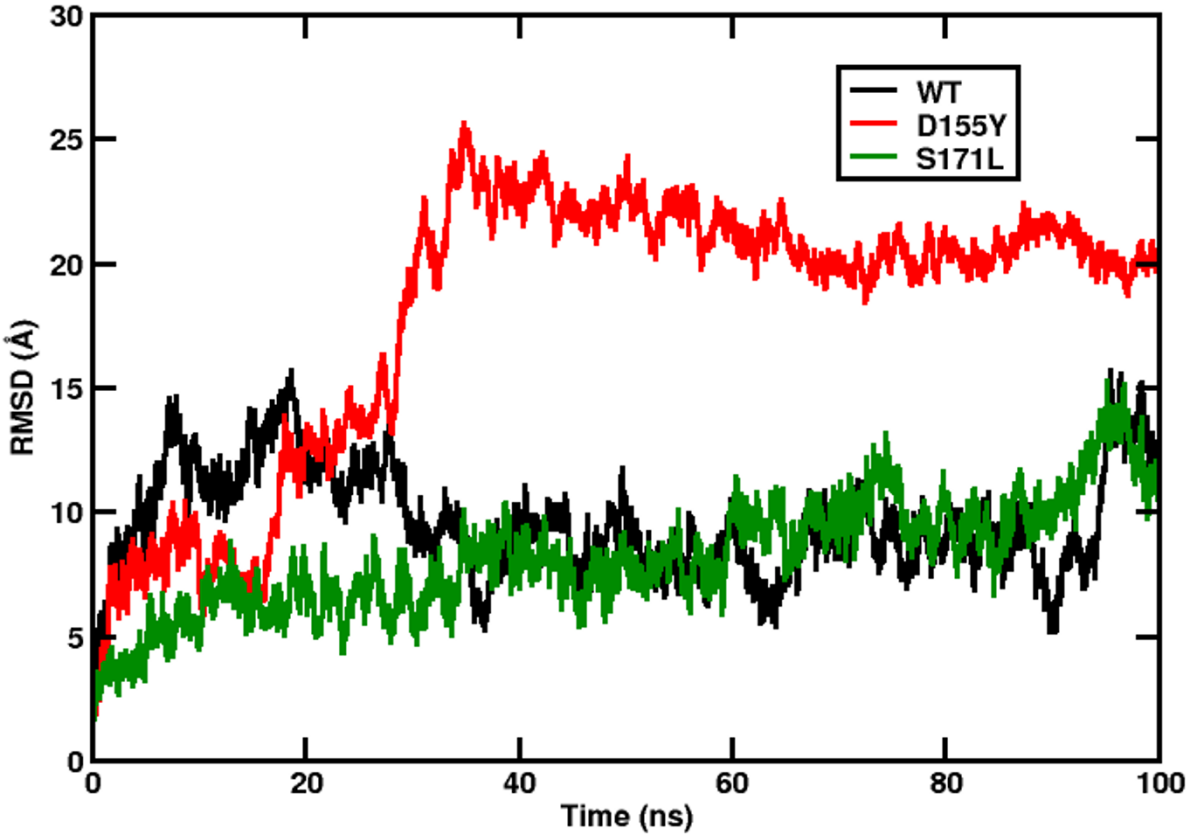

Fig. 8