Original article

Analyzing the effect of mutations in SARS-CoV2 papain-like protease from Saudi isolates on protein structure and drug-protein binding: Molecular modelling and dynamics studies

Alhumaidi B. Alabbas, Mubarak A. Alamri *

Department of Pharmaceutical Chemistry, College of Pharmacy, Prince Sattam Bin Abdulaziz University, Al-Kharj, Saudi Arabia

A B S T R A C T

The continuous and rapid development of the severe acute respiratory syndrome coronavirus-2 (SARS-CoV-2) virus remains a health concern especially with the emergence of numerous variants and mutations worldwide. As with other RNA viruses, SARS-CoV-2 has a genetically high mutation rate. These mutations have an impact on the virus characteristics, including transmissibility, antigenicity and development of drug and vaccine resistance. This work was pursued to identify the differences that exist in the papain-like protease (PL^{Pro}) from 58 Saudi isolates in comparison to the first reported sequence from Wuhan, China and determine their implications on protein structure and the inhibitor binding. PL^{pro} is a key protease enzyme for the host cells invasion and viral proteolytic cleavage, hence, it emerges as a valuable antiviral therapeutic target. Two mutations were identified including D108G and A249V and shown to increase the molecular flexibility of PL^{Pro} protein and alter the protein stability, particularly with D108G mutation. The effect of these mutations on the stability and dynamic behavior of PL^{Pro} structures as well as their effect on the binding of a known inhibitor; GRL0617 were further investigated by molecular docking and dynamic simulation.

© 2021 The Author(s). Published by Elsevier B.V. on behalf of King Saud University. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

1. Introduction

The pandemic outbreak of the 21st century caused by the severe acute respiratory syndrome coronavirus-2 (SARS-CoV-2) is still going on for over a year, causing millions of deaths worldwide. The viral transmission remains uncontrolled in many parts of the globe leading to high mortality. In fact, the number of deaths has nearly doubled in six months from approximately 1.8 to 4 million since the beginning of 2021 (https://covid19.who.int). Although substantial research progress and countermeasures have been made to manage the crisis, limiting the ongoing spread of the disease is still a major bottleneck as multiple waves of the outbreak hit many countries. Moreover, the emergence of thousands of mutations of the virus that may cause alterations in the virus characteristics has aroused significant concerns and their full impacts on virus characteristics remain unknown (Chand et al., 2020; Gu et al., 2020; Liu et al., 2021; Pachetti et al., 2020; Reshamwala et al., 2021; Salpini et al., 2021).

Monitoring the biological and pathological properties of the SARS-CoV-2 is crucial to better recognize the virus and offers means for appropriate counter-responses to the pandemic. The genomic sequences of SARS-CoV-2 were successfully obtained at an early stage of the outbreak and deposited into the National Center for Biotechnology Information (NCBI) GenBank sequence database (isolate Wuhan-Hu-1, accession number NC_045512). The SARS-CoV-2 belongs to the beta genera of coronaviruses consisting of a positive-sense single-stranded RNA whose genome is about 29.8 Kb in length and possesses almost 14 open reading frames (ORFs) encoding 4 structural and 16 nonstructural proteins (nsp1–16) and at least six accessory proteins. The structural proteins, including a spike protein (S), an envelope protein (E), a membrane protein (M), and a nucleocapsid protein (N), are encoded by the sub-genomic RNA (sgRNA) and play important roles in attachment to host cell, replication, and assembly of the mature virus (Wu, et al., 2020a; Wu, et al., 2020b). The nsps are generated

through a proteolytic cleavage of two large polyproteins (pp1a and pp1a/b) that are translated from the open reading frames (ORF) 1a/ab of the RNA. Nsps mediate multiple crucial roles in the life cycle of the virus such as replication cycle and host pathogenesis. Among these 16 nsps, the papain-like protease (PL^{Pro}; known as nsp3) and RNA dependent RNA polymerase (RdRp; known as nsp12) are considered as pivotal targets for antiviral drugs as they mediate vital overlapping roles inside the host cell starting from the polyprotein cleavage to viral replication and transcription (Henderson et al., 2020). On the other hand, the accessory proteins are thought to regulate interferon signaling pathways and the production of pro-inflammatory cytokines but are not required for replication (Liu et al., 2014).

SARS-CoV-2 adapts to a new host environment by genetic evolution and natural selection as all RNA viruses. This subsequently prompts the emergence of thousands of mutations throughout the epidemic that might cause problems in the development of successful vaccines, treatments, and accurate diagnostic assays. Some of these mutations have relatively high incidences, such as a non-synonymous mutation (D614G) in the virus's S protein (Korber et al, 2020; Zhang et al., 2020). In fact, the genome sequencing of SARS-CoV-2 has revealed an evolution rate of about 1.12×10^{-3} nucleotide substitution per year per site (Duchene et al., 2020; Koyama et al., 2020). The key question is whether the SARS-CoV-2 evolution may affect viral traits such as transmissibility and pathogenicity. Yet, because of the constant mutation, determining these characteristics with precision is fraught with many difficulties. Furthermore, according to a study, some SARS-CoV-2 subtypes have become more contagious and significantly more infectious strains may emerge (Chen et al., 2020). Thus, rigorous monitoring and tracking of SARS-CoV-2 mutations are necessary to fight the virus.

The papain-like protease (PL^{Pro}) is a key enzyme contained in the viral nonstructural protein 3 (nsp3). This protease is of vital importance for the viral replication through the ORF1ab polyprotein cleavage into functional components (nsp1-nsp3). It also provides the virus with an evasion mechanism against host immune response through the deubiquitination and deISGylation, removal of interferon-stimulating gene-15 protein (ISG15) from host proteins, disrupting the innate antiviral response of the host. Essential activities like deISGylation and deubiquitination of PL^{Pro} are also important for viral replication. (Freitas et al., 2020; Lindner et al., 2005; Shin et al., 2020). PL^{Pro} blocks downstream Interferon induction by reducing Interferon regulatory factor IRF3 phosphorylation and thereby suppressing type I interferon (IFN) (Matthews et al., 2014). Thus, PL^{Pro} serves as a target for anti-viral development.

Several mutations have been discovered in SARS CoV 1 and 2 PL^{Pro}. These mutations appear to often alter the enzyme specificity, activity and antiviral activity. Previously, R167E mutation in SARS CoV-1 PL^{Pro} significantly increases deubiquitinase activity and was nearly 20 times less efficient in hydrolyzing ISG15. Furthermore, Q233E PL^{Pro} favors a more robust deISGylase activity (Daczkowski et al., 2017). In SARS-CoV-2 PL^{Pro}, non-synonymous mutations (A889V and V843F) were recently found to alter the protein's flexibility and reduce the affinity for ISG-15 (Hossain et al., 2021). Additionally, mutations, such as Y269T and Y268G, in SARS-CoV-2 PL^{Pro} strongly reduced the inhibitory effect of the putative antiviral agent, GRL0617 (Shin et al., 2020). Mutations in other SARS-CoV-2 proteins such as the spike protein could also have substantial effects on the biochemical interactions (Amin et al., 2020). This shows that mutations must be closely monitored to ensure a better understanding of the virus behavior.

Here, we determined and characterized novel mutations in the PL^{Pro} protein sequences isolated in Saudi Arabia in comparison to the 'Wuhan wet seafood market' SARS-CoV-2 PL^{Pro} (Wu, F. et al., 2020). The effect of these mutations on the secondary structure

as well as on the molecular flexibility and dynamic properties of the 3D structure of PL^{Pro} were investigated. The observed effects on the structures of protein were further assisted using all atom molecular dynamics simulation. Since protein mutations often alter protein–ligand binding affinity that could lead to a drug resistance, we monitor the effect of these mutations on PL^{Pro} on the binding and interaction of known non-covalent inhibitor (GRL0617) using docking and molecular dynamics (MD) studies. Overall, our results provide a piece of preliminary evidence that mutations in the PL^{Pro} alter the protein's properties and highlight the need for further functional studies targeting this virus.

2. Materials and methods

We aim to identify and study the mutation of SARS-CoV-2 PL^{Pro} protein emerging in Saudi Arabia. Our method relies on detecting amino acid mutations in the protein sequence using a multiple sequence alignment platform. Then, to determine the exact implications of these mutations on the protein's secondary, tertiary structure, stability and flexibility, we implemented a secondary structure and free energy ($\Delta\Delta G$) and vibrational entropy ($\Delta\Delta Svib$) prediction tools for each mutated protein in comparison to the wild type PL^{Pro}. Next, molecular dynamic simulation for the wild-type (WT) sequence and the two identified mutations was conducted in two separate systems, unligated PL^{Pro} protein system and inhibitor-bound PL^{Pro} protein system as describes in full details below.

2.1. Sequence's retrieval

The sequences of the largest SARS-CoV-2 gene; ORF1ab (length: 7096) were download from the NCBI virus database (https://www.ncbi.nlm.nih.gov/labs/virus/vssi/#/). As of 21 May 2021, there were 58 SARS-CoV-2 sequences of ORF1ab polyprotein isolated in Saudi Arabia were available at the NCBI. These sequences were downloaded and their accession numbers for protein identification are included in supplementary **Table S1** and used for analyzing the PL^{Pro} protein sequences (1564–1878) in this study. The PL^{Pro} protein sequence of the first sequenced virus from the "Wuhan wet seafood market" region (accession number: YP_009724389) was utilized as a reference for the WT sequence (Wu F. et al., 2020).

2.2. Multiple sequence alignment

All sequences were aligned using a multiple sequence alignment platform of Clustal Omega online tool (Madeira et al., 2019). The SARS-CoV-2 PL^{Pro} sequences reported from Saudi Arabia were aligned and compared with the first PL^{Pro} protein sequence reported in Wuhan, China, to detect mutations present in the protein structure (Wu F. et al., 2020).

2.3. Secondary structure predictions

The prediction of the secondary structures of mutated PL^{Pro} proteins was performed using the Chou-Fasman secondary structure prediction (CFSSP) online server (Kumar, 2013). This tool applies Chou-Fasman algorithm using the primary amino acid sequences as an input. The secondary structure aspects (e.g., α helix, β sheet, and turns) are predicted based on the analyses of the amino acid relative frequency in secondary structures of proteins that have been solved by X-ray crystallography.

2.4. Protein structure stability and flexibility

The impact of identified mutations on the molecular stability and flexibility of PL^Pro protein structures was measure using the DynaMut web server (http://biosig.unimelb.edu.au/dynamut/) (Rodrigues et al., 2018). In brief, this server is used to analyze and visualize the protein conformational dynamics by sampling the protein's conformations using Normal Mode Analysis (NMA). It manifests the effect of mutations on dynamics and stability of proteins as a result of vibrational entropy changes (i.e., free energy ($\Delta\Delta G$) and vibrational entropy ($\Delta\Delta$Svib) ENCoM between the WT and the mutants are calculated). For 3D DynaMut protein modeling, the RCSB protein ID: 6W9C of WT PL^Pro was used (Osipiuk et al., 2020). DynaMut server also provides the three-dimensional visual representation of fluctuation in protein structure.

2.5. Molecular docking with Autodock Vina

The molecular docking was performed using Autodock Vina software (Trott and Olson, 2010). The X-ray structure of SARS-CoV-2 PL^Pro WT (RCSB ID: 6W9C), as well as the modeled 3D structures with indicated mutations by the 3D DynaMut protein modeling tool, were utilized in this study. The protein structure was cleaned by removing co-crystallized water molecules and saved in PDB format using BIOVIA Discovery Studio Visualizer 2019. The three-dimensional structure of GRL0617 was retrieved from PubChem (PubChem CID: 24941262) in PDB format. Before docking, the polar hydrogen atoms of protein and ligand were added by Autodock Tools (Huey et al., 2012) and structures were then saved in PDBQT format. The auto-grid tool was used for the determination of the docking grid map with 1.00 Å spacing and box dimensions of 20 X × 20 Y × 20 Z Å and centers of 43.186 X × - 8.55 Y × 7.144 Z Å. The interactions between the SARS-CoV-2 PL^Pro and the inhibitor GRL0617 were visualized and analyzed by PyMOL (DeLano, 2002) and UCSF Chimera 1.14 (Pettersen et al., 2004).

2.6. Molecular dynamic (MD) simulation

The GROMACS 2018.1 package (Van Der Spoel et al., 2005) was used to conduct the MD simulation in this study using the all-atom optimized potentials for liquid simulations (OPLS-AA) force field and TIP3P water model as described previously (Alamri et al., 2020). Briefly, the high-resolution structure of apo SARS-CoV-2 PL^Pro (RCSB ID: 6W9C) as well as with mutations were used as initial structures for MD summation. The PDB files of proteins were further optimized by the DockPerp tool in UCSF Chimera (Pettersen et al., 2004). The systems were solvated and neutralized with 150 mM of Na$^+$/Cl$^-$ as counter ions in a cubic box with at least 1 nm spacing from the system. The systems were subjected then to global energy minimization using the "steepest descent algorithm". The system was then equilibrated for 100 ps using "canonical ensembles NVT" followed by NPT (isothermal–isobaric ensemble). For the docking study, the top-predicted docked-pose of GRL0617 against each protein with the lowest binding score was used as an initial structure for the MD simulation. The ligand was parametrized using SwissParam webserver (Zoete et al., 2011). Finally, a 20 ns production run was performed for each system. Various MD parameters were analyzed such as root-mean-square deviations (RMSD), root-mean-square fluctuations (RMSF), number of hydrogen-bond (HB), and radius of gyration (Rg) and Solvent Accessible Surface Area (SASA) using GROMACS 2018.1 toolkits (Van Der Spoel et al., 2005).

3. Results

3.1. Identification of mutations in SARS-CoV-2 PL^Pro proteins in Saudi isolates

We retrieved all PL^Pro protein sequences submitted in the NCBI viral database from Saudi isolates up until May 21, 2021. Supplementary **Table S1** lists the accession numbers for the PL^Pro protein utilized in this investigation. The Clustal Omega program was used to align various sequences, and the Wuhan virus PL^Pro protein sequences were utilized as a reference (accession number: YP_009724389). Out of the 58 isolates, two mutations were found. One sequence was found to harbor a mutation Asp^{108} → Gly of PL^Pro (D108G), while the other mutation was found at position 249 Ala^{249} → Val (A249V).

3.2. Mutations cause alteration in the secondary structure of proteins

To evaluate the influence of these mutations on the secondary structure of PL^Pro, we carried out a secondary structure prediction using the CFSSP server. The secondary structure of mutant residues was then analyzed and compared to the WT sequence.

D108G mutation shows that the substitution of aspartate (D), a bulky amino acid to glycine (G), the simplest amino acid, at position 108 resulted in a distortion of a helical structure at position 107 of the PL^Pro (Fig. 1). Secondary structure prediction of both mutants showed changes in or nearby the site of mutation. A249V mutation, where alanine (A) is replaced with valine (V) residue at the position 249 distorts a helix backbone, hence, beta-sheet is formed at the same position as compared to the Wuhan type sequence (Fig. 1). Valine is a bulkier beta-branched amino acid making α-helical conformations difficult to achieve (Cornish et al., 1994). These findings imply that the mutations discovered in PL^Pro may result in alteration of the secondary structure. However, further investigation is needed to determine the significance of the observed mutations' effect on the secondary structure, hence, the protease function of PL^Pro.

3.3. Protein structure stability and flexibility

Proteins are extremely dynamic molecules with molecular motions that are inextricably linked to their activity. Afterward, the free energy differences ($\Delta\Delta G$) between Wuhan PL^Pro and mutants were calculated to probe the observed changes in secondary structure to the stability and tertiary structure dynamics of the PL^Pro, the differences in free energy ($\Delta\Delta G$) between Wuhan

Fig. 1. Prediction of the secondary structure of PL^Pro protein. The differences in the PL^Pro secondary structure between the WT and Saudi isolates are indicated in the box of each panel.

Table 1
The values of change in $\Delta\Delta G$ DynaMut (kcal mol^{-1}) and $\Delta\Delta S$ ENCoM (kcal mol^{-1} K^{-1}) due to the mutations in PL^{Pro}.

S. No	Wuhan Isolate	Saudi Isolate	AA Position	$\Delta\Delta G$ DynaMut	Effect	$\Delta\Delta S_{Vib}$ ENCoM	Effect
1	D	G	108	−0.418	Destabilizing	0.495	Increase of molecule flexibility
2	A	V	249	0.594	Stabilizing	0.053	Increase of molecule flexibility

WT PL^{Pro} and mutants were calculated. Analyzing and visualizing protein dynamics were accomplished using the DynaMut program, a web server (Rodrigues et al., 2018). The effect of mutations on the protein dynamicity and stability was established in terms of the changes in the vibrational entropy. The server produces the anticipated change in the stability (kcal mol^{-1}) as well as the change in entropy energy between the WT and the mutant structures (kcal mol^{-1} K^{-1}). The $\Delta\Delta G \geq 0$ indicates a stabilizing effect, while mutants that exhibit $\Delta\Delta G < 0$ cause a destabilization of the protein. The variations in vibrational entropy energy ($\Delta\Delta S_{Vib}$ENCoM) between the WT and the mutants were also assessed by DynaMut using the ENCoM tool (Frappier et al., 2014). The $\Delta\Delta S_{Vib}$ENCoM ≥ 0 of the PL^{Pro} indicates an increase in flexibility, while $\Delta\Delta S_{Vib}$-ENCoM < 0 corresponds to rigidification in the protein structure. These free energies for the WT and the mutants are tabulated in Table 1.

PL^{Pro} stability analysis showed that the D108G substitution resulted in a slightly more dynamic structure with $\Delta\Delta G$ of −0.418 kcal mol^{-1} (a destabilizing mutation). On the other hand, the change in the structural stability of PL^{Pro} induced by the A249V mutation indicates that the alanine to valine substitution leads to stabilization of protein structure, $\Delta\Delta G$ of 0.594 kcal mol^{-1}.

The vibrational entropy energy between the WT and the mutants (D108G and A249V) yielded $\Delta\Delta S_{Vib}$ ENCoM values of 0.495 and 0.053 kcal mol^{-1} K^{-1} respectively, which suggests an increase in overall molecular flexibility in the mutated PL^{Pro} protein (Table 1). However, A249V increased rigidity, while D108G confers structural flexibility at the site of the mutations in the PL^{Pro} structure (Fig. 2).

3.4. Effect of mutations on the PL^{Pro} intramolecular interactions

The mutations with a single amino acid substitution may alter the complex network of protein intramolecular interactions, affecting protein folding, stability, dynamics, and thus protein function. Therefore, to assess the alterations on intramolecular interactions triggered by mutant residues in the PL^{Pro} protein, we visualize all covalent and non-covalent interactions of the WT and the mutants in their respective 3D using DynaMut webserver. The substitution of the native D residue with mutant shorter G residue changes the chain length, causing a loss of various intramolecular interactions, which involves hydrogen bonds and hydrophobic interactions. It also disrupts the interactions between residues that present in the proximity to the mutation site, such as W93, V159 and G160 interaction (Fig. 3**A&B**). Thus, a change from bulkier D to G residue is likely inducing some changes in the protein folding capability (destabilizing). The exchange of the hydrophobic residue alanine with hydrophobic valine residue (A249V) also results in changes in intramolecular interactions. The valine side chain is bulkier than alanine, hence, an additional salt-bridge with P247 is introduced as a result of a shorter special distance, leading to a stabilization effect (Fig. 3**C&D**).

Therefore, it can be concluded that these mutations cause alterations in the stability of PL^{Pro} and impact several intramolecular interactions in the protein mutation site.

3.5. Effect of mutations on PL^{Pro} structural conformation and dynamic stability

The MD simulation approach was used to further analyze the structural stability and dynamic characteristics of the WT PL^{Pro} and the two mutations identified here at atomic resolution to gain a better appreciation of protein dynamics. We used the GROMACS program to run 20 ns MD simulations to estimate the root-mean-square deviation (RMSD), root-mean-square fluctuation (RMSF), the radius of gyration (Rg) and the solvent-accessible surface area (SASA). These parameters were employed to deduce the dynamic behavior of PL^{Pro} in each simulated system.

The variations observed during the simulation can be used to assess the protein's stability in relation to its conformation. The fewer the variations exist, the more stable the protein structure is. The MD analysis for the WT, D108G, and A249V showing the RMSD values of the PL^{Pro} backbone for the simulated structures is presented in Fig. 4**A**. The RMSD of the WT shows a deviation
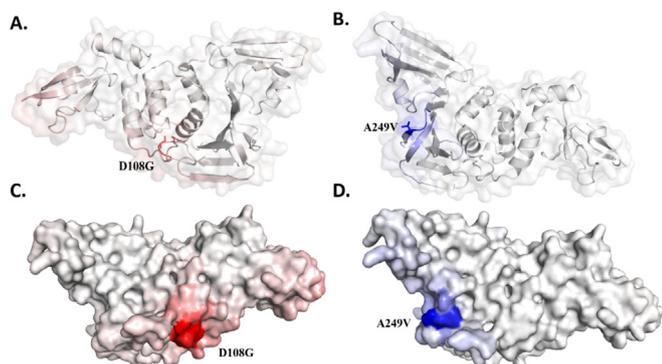
Fig. 2. Changes in Vibrational Entropy Energy between the wild-type and the mutants PL^{Pro}. Ribbon representation showing the change in Δ Vibrational Entropy Energy between the wild-type and the mutants PL^{Pro} due to (**A**) D108G and (**B**) A249V mutation. Surface representation showing the change in Δ Vibrational Entropy Energy between the wild-type and the mutant PL^{Pro} due to (**C**) D108G and (**D**) A249V mutation. The residues are colored blue and red, indicating an increased rigidification and flexibility, respectively.
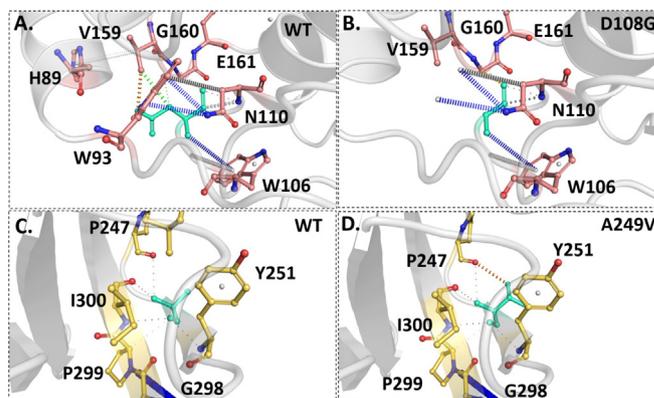
Fig. 3. Effect of mutation on interactions between atoms within PL^{Pro} at the mutation site (**A** and **B**) show the contribution of the interatomic interaction of the wild-type and D108G of PL^{Pro}. (**C** and **D**) show the contribution of the interatomic interaction of wild-type and A249V of PL^{Pro}. The wild-type and mutant amino acids are colored cyan and the surrounding amino acids, are shown as sticks.
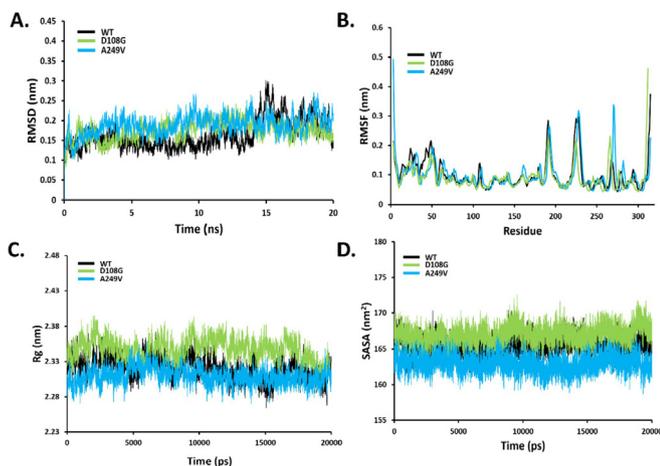
Fig. 4. MD simulations of the wild-type PL^{Pro}, D108G, and A249V. (**A**) (RMSD) the root-mean-square-deviation of the Cα atoms backbone of the native and mutants. (**B**) |(RMSF) the root-mean-square fluctuation. (**C**) (Rg) a plot of the radius of gyration. (**D**) (SASA) solvent Accessible Surface Area.

for the PL^{Pro} backbone from 0.15 to 0.30 nm in the initial 15 ns of the simulation then decreased to 0.15 at the end of the simulation with an average RMSD of 0.16 nm. This deviation might be attributed to the conformational flexibility nature of the protein. PL^{Pro}, being a cysteine protease, is rich in cysteine and histidine residues that, in addition to their catalytic contribution, mediate conformational plasticity due to their capability to induce protonation and deprotonation states of the protein (Henderson et al., 2020). For both mutants, the RMSD value in the first 10 ns fluctuated between ∼ 0.15 and 0.25 nm. Thereafter, the value remained almost the same within the next 10 ns with some fluctuations. The RMSD for D108G and A249V mutants were slightly higher than WT with average values of 0.17 ± 0.04, 0.18 ± 0.02, 0.19 ± 0.02 nm for WT, D108G and A249V, respectively (Table 2). These results point to the stability of these simulated systems.

The residues' RMSF was computed to pinpoint the residues' effect on the dynamic behavior of the mutant and WT protein. The flexibility of each residue is assessed by its root mean square fluctuation (RMSF). It was noticeable from the RMSF analysis that simulated structures display higher conformational fluctuations in certain PL^{Pro} domains (Fig. 4**B**). The PL^{Pro} has four main domains, the N-terminal ubiquitin-binding (UBL) domain (the first 62 aa), the α-helical thumb domain (residues 63–182), the β-stranded finger domain (zinc-binding site; residues 183–240) is important for the structural stability and the enzymatic activity, and the palm domain (residues 241–314), which coins the catalytic active site triad (Cys111, His272, and Asp286) and the essential blocking loop 2 (BL2; Gly266–Gly271) near the active site. BL2 is flexible and crucial in PL^{Pro} binding to host and viral protein substrate as it results in an open (unliganded), or closed conformation upon a substrate or an inhibitor binding to the active site (Osipiuk et al., 2021). The mean RMSF values calculated from 0 to 20 ns for the WT, D108G, and A249Vsystems are 0.10 ± 0.05, 0.10 ± 0.05, and 0.09 ± 0.04 nm, respectively (Table 2). Although the overall trend of the flexibility did not change significantly in all systems, both mutations mainly lead to a higher fluctuation peak of PL^{Pro} in residues

between ∼ 260–275, indicating more flexible residues (Fig. 4**B**). The higher fluctuation seen in this region indicates a more dynamic conformation that could result in a more flexible loop region making the active site more accessible. Moreover, this region encompasses key residues, such as Tyr269 and Gln270, for small molecule binding (Ghosh et al., 2010; Ratia et al., 2008) and the potential alteration induced by mutation may influence the ligand binding.

The compactness of PL^{Pro} WT and the discovered mutants was evaluated by assessing and comparing their radius of gyration (Rg). A compact packing of amino acid residues is known for the protein's stability and folding (Lobanov et al., 2008). The variation in Rg values among the WT and the mutants is depicted in Fig. 4**C**. The average Rg values of WT, D108G, and A249V mutations were found to be 2.32 ± 0.02, 2.35 ± 0.02, and 2.31 ± 0.01 nm, respectively (Table 2). Consistent with the above free energy (ΔΔG) calculation, D108G displayed a greater Rg, indicating lower compactness of protein, while A249V did not show a significant change in protein compactness. The compactness, stability, and ability of proteins to interact with other molecules as well as the extent of hydrophobicity inside the folded protein can also be assessed with SASA analysis. The tight folding process usually translates to a significant decrease in SASA value, which indicates the smaller accessibility of a protein to the solvent. The protein's accessibility to the solvent in the WT and mutant systems is shown in Fig. 4**D**. The average molecular SASA for the WT PL^{Pro} was 165 nm². While the D108G mutant exhibits a higher SASA of 167 nm², A249V demonstrated a lower SASA value of 162 nm² compared to the WT protein (Table 2). This indicates that D108G is less compact and more flexible, whereas A249V maintains a proper folding. It is possible that the transition from D to G in D108, which is housed in a solvent-exposed cleft close to the PL^{Pro} catalytic triad (Cys111, His272, and Asp286), may reduce steric hindrance, thus, a higher SASA is observed. Notably, the amino acid mutated from A to V increases the volume of the amino acid at this position in A249V. Yet, due to the larger side chain of V, the steric effect would be attributed to the less SASA. An alternative explanation is that the higher hydrophobicity of V249, which is situated near to a small druggable hydrophobic cavity containing P247 and P248 and Tyr264, increases hydrophobic interaction with these residues, hence, restricting the interaction with the external solvent and decreasing the SASA (Báez-Santos et al., 2014; Bosken et. al., 2020; Shen et. al., 2021). These data suggest that the identified mutations can influence PL^{Pro} spatial packing and dynamicity.

3.6. MD simulation on ligand binding at 20 ns.

To assess mutation-induced effects of ligand binding, we carried out an MD simulation on the inhibitor binding to the WT and mutant PL^{Pro}. One of the most promising inhibitors is GRL0617 (N-[1-(naphthalen-1-yl)ethyl]benzamide), a non-covalent inhibitor of SARS-CoV-1 and SARS-CoV-2 PL^{Pro} with an IC$_{50}$ value of 0.6 and 2.3 uM, respectively (Ratia et al., 2008; Osipiuk et al., 2021; Fu at al., 2020).

The GRL0617 binding site is in the palm domain of PL^{Pro}, occupying part of the catalytic active site but does not interact with the catalytic triad (Fig. 5). GRL0617 binds in the substrate cleft formed

Table 2
Effect of mutations on PL^{Pro} stability predicted by the molecular dynamic simulations.

System	RMSD (nm)	RMSF (nm)	Rg (nm)	SASA (nm²)
WT	0.17 ± 0.04	0.10 ± 0.05	2.32 ± 0.02	165.53 ± 1.21
D108G	0.18 ± 0.02	0.10 ± 0.05	2.35 ± 0.02	167.00 ± 1.27
A249V	0.19 ± 0.02	0.09 ± 0.04	2.31 ± 0.01	162.92 ± 1.18
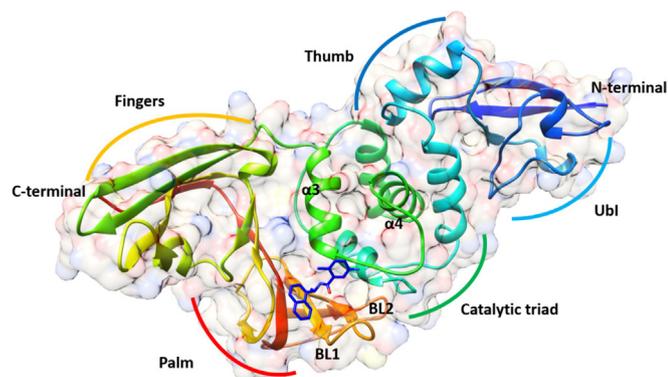
Fig. 5. 3D structure of SARS-CoV-2 PL[Pro] WT in complex with GRL0617 (blue) showing the secondary structure, domains and subdomains.

between the BL2 loop and the loop connecting α3 and α4 (α3-to-α4 loop) of PL[Pro]. Two important hydrogen bonds with D164 and Q269 are essential to PL[Pro]–GRL0617 interaction (Freitas et. al., 2020; Fu et al., 2020; Osipiuk et al., 2021; Ratia et al., 2008).

The stability of the protein–ligand complex was estimated by calculating the RMSD of backbone α carbon atoms of the WT and mutant PL[Pro] upon GRL0617 binding. The RMSD of PL[Pro] in complex with GRL0617 gradually increased from ∼ 0.15–0.3 nm from 2 ns to 5 ns. Afterward, the RMSD remained constant for the rest of the 20 ns MD run (Fig. 6A). The RMSD values for both D108G PL[Pro]-GRL0617 and A249V PL[Pro]-GRL0617 complexes exhibited roughly steady values (average ∼ 0.2 nm) through the 20 ns run with some negligible fluctuations. Therefore, it can be concluded that all systems remained in a stable equilibrium conformation. In essence, the inhibitor bound systems did not introduce any large conformational change as they remained stable through the time course of the simulation. Next, we examined the flexibility of different regions of PL[Pro] WT and mutants in complex with GRL0617 by calculating the RMSF of alpha carbon atoms for all systems (Fig. 6B). The RMSF profiles of all PL[Pro]-GRL0617 systems exhibited slightly

higher conformational fluctuations in comparison to their inhibitor-free counterparts (Table 2&Table 3). The BL2 loop region (residues 266–271) shows significantly higher RMSF upon inhibitor binding to WT (+0.23 nm) and D108G (+0.18 nm). This suggests a higher fluctuation of this loop. A249V exhibited a similar RMSF value for this loop of ∼ 0.3 nm in free and complex systems. Similarly, the Rg values of complexes were consistent with that of PL[Pro] in apo-form (no ligand-bound) (Fig. 6C). The D108G PL[Pro]-GRL0617 complex was less stable than WT PL[Pro]-GRL0617 and A249V PL[Pro]-GRL0617 with average Rg values of 1.66 ± 0.02, 1.69 ± 0.02 and 1.62 ± 0.01 nm for WT, A249V and D108G, respectively. Additionally, the formation of hydrogen bonds between GRL0617 and PL[Pro] was assessed over the simulation time (Fig. 6D). The hydrogen bond is an essential factor for stabilizing the ligand at the active site of the target protein. The GRL0617 interacts by up to 2 hydrogen bonds with WT and by up to 3 hydrogen bonds with A249V and D108G mutants (Table 3). However, hydrogen bonds between GRL0617 and with A249V and D108G mutants were not stable and reduced to 2 bonds during the last 15 ns (Fig. 6D). These results indicating that the binding of GRL0617 might be altered by D108G more than A249V mutation.

4. Discussion

Tracing the mutations occurring in SARS-CoV-2 vital component is necessary to understand the possible activity changes and lay the foundation for the development of potential inhibitors. Our screening has led us to identify mutational changes in a critical

Table 3
Effect of PL[Pro] mutations on the binding stability of ligand- PL[Pro] predicted by the molecular dynamic simulations.

System	RMSD (nm)	RMSF (nm)	Rg (nm)	HB
WT	0.42 ± 0.05	0.14 ± 0.06	1.66 ± 0.02	2
D108G	0.37 ± 0.04	0.13 ± 0.06	1.62 ± 0.01	3
A249V	0.28 ± 0.06	0.16 ± 0.06	1.69 ± 0.02	3
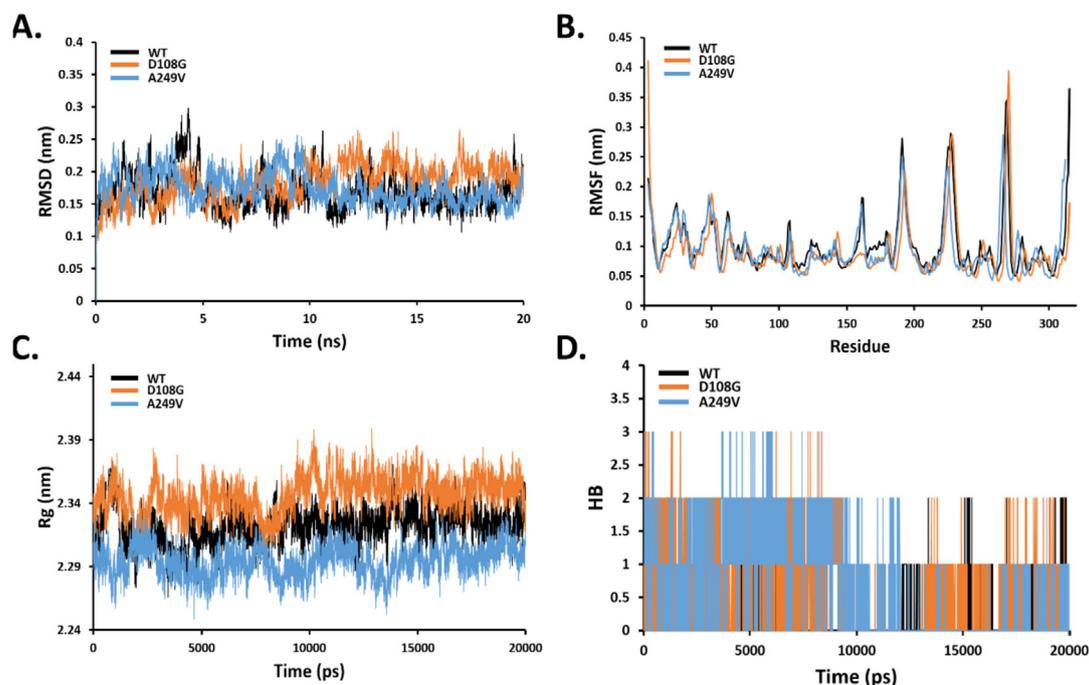
Fig. 6. MD simulations of GRL0617 in complex with the wild-type PL[Pro], D108G, and A249V. (**A**) The root-mean-square-deviation of the Cα atoms backbone of native and mutants. (**B**) The root-mean-square-fluctuation plot. (**C**) A plot of the radius of gyration. (**D**) The number of hydrogen bonds (HB) formed over the simulation time.

enzyme, PL^{Pro}, for SARS-CoV-2 with numerous roles in the viral pathogenesis host immune suppression (Freitas et al., 2020; Shin et al., 2020). Only properly folded proteins can supply a protein's functional characteristics. Protein scaffolds consist of secondary structural details, including α -helices, β strands, and turns. Minor changes in the size or characteristics of an amino acid side chain can modify or prohibit a protein's function and affect the ability to interact with other proteins (Studer et al., 2013; Zhou et al., 2020). Because the mutated amino acids found in this study differ in nature, such substitutions can induce a local conformation change in the PL^{Pro}, which can lead to changes in the protein–protein interactions, ligand interactions or even functional alterations (Studer et al., 2013). The observed secondary and 3D structural changes induced by D108G and A249V mutations may have significant implications on the protease activity, which requires further investigations.

Interestingly, the two mutation sequences display higher flexibility than the WT PL^{Pro}. This is of substantial importance because earlier reports show that spike protein substitution to glycine in D614G mutation increases the flexibility due to the shorter side chain allowing a more efficient cleavage of S protein subunits, which might explain the improved ACE binding affinity (Ozono et al., 2021). Furthermore, D614G, due to the higher flexibility, gains an increase in its thermostability as compared to the WT virus (Plante et al., 2021).

It is important to note that the activity of PL^{Pro}, is dependent on the availability of optimal substrate binding sites. Thus, the alterations observed in the intramolecular interactions in the mutation site, particularly with the D108G, are likely to play an important role in modifying the substrate binding or the protease enzyme preference for deubiquitination and deISGylation activity. A previous study has shown that D109 (in SARS-CoV-1; corresponding to D108 in SARS-CoV-2) makes an H-bond with W94 (W93 in SARS-CoV-2) to prevent the collapse of the blocking loop (BL1) into the active site (Ratia et. Al., 2006). The hydrogen bond between D108 and W93 (2.8 Å) also strengthens the conformation of the oxygen anion hole, which is critical for the stabilization of the tetrahedral cleavage intermediate. Detailed studies would be needed to address the substrate binding to these mutations.

5. Conclusion

The high mutation rate of the SARS-CoV-2 virus may hinder the progress for effective drugs and vaccines. Therefore, studying the impact of emerging mutations on the key viral proteins' structure and function is needed. In this study, two mutations within SARS-CoV-2 PL^{Pro} were identified including D108G and A249V by analyzing 58 Saudi isolates in comparison to the first sequence reported in Wuhan, China. These mutations showed to alter the protein's structural stability and increase its molecular flexibility by disrupting the local interaction networks. In addition, the binding of known inhibitor; GRL0617 to SARS-CoV-2 PL^{Pro} might be affected by these mutations to some extent due to the large flexibility around the active site, particularly with the D108G mutation. Overall, the results in this work suggest that these two mutations may have an observable impact on the protein structure and function and more importantly on the ligand-binding affinity. This would provide knowledge of future functional studies and more understanding for the development of antiviral therapeutics.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.sjbs.2021.09.028.

References

Alamri, M.A., Altharawi, A., Alabbas, A.B., Alossaimi, M.A., Alqahtani, S.M., 2020. Structure-based virtual screening and molecular dynamics of phytochemicals derived from Saudi medicinal plants to identify potential COVID-19 therapeutics. Arab. J. Chem. 13 (9), 7224–7234.

Amin, M., Sorour, M.K., Kasry, A., 2020. Comparing the Binding Interactions in the Receptor Binding Domains of SARS-CoV-2 and SARS-CoV. J. Phys. Chem. Lett. 11 (12), 4897–4900.

Báez-Santos, Y.M., Barraza, S.J., Wilson, M.W., Agius, M.P., Mielech, A.M., Davis, N.M., Baker, S.C., Larsen, S.D., Mesecar, A.D., 2014. X-ray structural and biological evaluation of a series of potent and highly selective inhibitors of human coronavirus papain-like proteases. J. Med. Chem. 57 (6), 2393–2412.

Bosken, Y.K., Cholko, T., Lou, Y.C., Wu, K.P., Chang, C.A., 2020. Insights Into Dynamics of Inhibitor and Ubiquitin-Like Protein Binding in SARS-CoV-2 Papain-Like Protease. Front. Mol. Biosci. 7, 174.

Chand, G. B., Banerjee, A., Azad, G. K., 2020. Identification of novel mutations in RNA- dependent RNA polymerases of SARS-CoV-2 and their implications on its protein structure. PeerJ, 8, e9492.

Chen, J., Wang, R., Wang, M., Wei, G.-W., 2020. Mutations Strengthened SARS-CoV-2 Infectivity. J. Mol. Biol 432 (19), 5212–5226.

Cornish, V.W., Kaplan, M.I., Veenstra, D.L., Kollman, P.A., Schultz, P.G., 1994. Stabilizing and destabilizing effects of placing beta-branched amino acids in protein alpha-helices. Biochem. 33, 12022–12031.

Daczkowski, C.M., Dzimianski, J.V., Clasman, J.R., Goodwin, O., Mesecar, A.D., Pegan, S.D., 2017. Structural Insights into the Interaction of Coronavirus Papain-Like Proteases and Interferon-Stimulated Gene Product 15 from Different Species. J. Mol. Biol. 429 (11), 1661–1683.

DeLano, W.L. (2002) The PyMOL molecular graphics system on world wide web. http://www.pymol.org.

Duchene, S., Featherstone, L., Haritopoulou-Sinanidou, M., Rambaut, A., Lemey, P., Baele, G., 2020. Temporal signal and the phylodynamic threshold of SARS-CoV-2. Virus Evol. 6 (2), veaa061.

Frappier, V., Najmanovich, R.J., MacKerell, A.D., 2014. A coarse-grained elastic network atom contact model and its use in the simulation of protein dynamics and the prediction of the effect of mutations. PLoS Comput. Biol. 10 (4), e1003569. https://doi.org/10.1371/journal.pcbi.100356910.1371/journal. pcbi.1003569.g00110.1371/journal.pcbi.1003569.g00210.1371/journal. pcbi.1003569.g00310.1371/journal.pcbi.1003569.g00410.1371/journal. pcbi.1003569.g00510.1371/journal.pcbi.1003569.g00610.1371/journal. pcbi.1003569.g00710.1371/journal.pcbi.1003569.g00810.1371/journal. pcbi.1003569.g00910.1371/journal.pcbi.1003569.g01010.1371/journal. pcbi.1003569.g01110.1371/journal.pcbi.1003569.g01210.1371/journal. pcbi.1003569.g01310.1371/journal.pcbi.1003569.s00110.1371/journal. pcbi.1003569.s00210.1371/journal.pcbi.1003569.s00310.1371/journal. pcbi.1003569.s00410.1371/journal.pcbi.1003569.s005.

Freitas, B.T., Durie, I.A., Murray, J., Longo, J.E., Miller, H.C., Crich, D., Hogan, R.J., Tripp, R.A., Pegan, S.D., 2020. Characterization and Noncovalent Inhibition of the Deubiquitinase and deISGylase Activity of SARS-CoV-2 Papain-Like Protease. ACS Infect. Dis. 6 (8), 2099–2109.

Fu, Z., Huang, B., Tang, J., Liu, S., Liu, M., Ye, Y., Liu, Z., Xiong, Y., Zhu, W., Cao, D., Li, J., Niu, X., Zhou, H., Zhao, Y.J., Zhang, G., Huang, H., 2021. The complex structure of GRL0617 and SARS-CoV-2 PLpro reveals a hot spot for antiviral drug discovery. Nat. Commun. 12 (1), 488.

Ghosh, A.K., Takayama, J., Rao, K.V., Ratia, K., Chaudhuri, R., Mulhearn, D.C., Lee, H., Nichols, D.B., Baliji, S., Baker, S.C., Johnson, M.E., Mesecar, A.D., 2010. Severe acute respiratory syndrome coronavirus papain-like novel protease inhibitors: design, synthesis, protein-ligand X-ray structure and biological evaluation. J. Med. Chem. 53 (13), 4968–4979.

Gu, H., Chen, Q., Yang, G., He, L., Fan, H., Deng, Y.Q., Wang, Y., Teng, Y., Zhao, Z., Cui, Y., Li, Y., Li, X.F., Li, J., Zhang, N.N., Yang, X., Chen, S., Guo, Y., Zhao, G., Wang, X., Luo, D.Y., Wang, H., Yang, X., Li, Y., Han, G., He, Y., Zhou, X., Geng, S., Sheng, X., Jiang, S., Sun, S., Qin, C.F., Zhou, Y., 2020. Adaptation of SARS-CoV-2 in BALB/c mice for testing vaccine efficacy. Science 369 (6511), 1603–1607.

Henderson, J. A., Verma, N., & Shen, J., 2020. Assessment of Proton-Coupled Conformational Dynamics of SARS and MERS Coronavirus Papain-like Proteases: Implication for Designing Broad-Spectrum Antiviral Inhibitors. bioRxiv : the preprint server for biology, 2020.06.30.181305.

Hossain, M.U., Bhattacharjee, A., Emon, M., Chowdhury, Z.M., Ahammad, I., Mosaib, M.G., Moniruzzaman, M., Rahman, M.H., Islam, M.N., Ahmed, I., Amin, M.R., Rashed, A., Das, K.C., Keya, C.A., Salimullah, M., 2021. Novel mutations in NSP-1 and PLPro of SARS-CoV-2 NIB-1 genome mount for effective therapeutics. J. Genet. Eng. Biotechnol. 19 (1), 52.

Https://covid19.who.int; (Accessed July 19, 2021).

Huey, R., Morris, G.M., Forli, S., 2012. Using AutoDock 4 and AutoDock vina with AutoDockTools: a tutorial. The Scripps Research Institute Molecular Graphics Laboratory 10550, 92037.

Korber, B., Fischer, W. M., Gnanakaran, S., Yoon, H., Theiler, J., Abfalterer, W., Hengartner, N., Giorgi, E. E., Bhattacharya, T., Foley, B., Hastie, K. M., Parker, M. D., Partridge, D. G., Evans, C. M., Freeman, T. M., de Silva, T. I., Sheffield COVID-19 Genomics Group, McDanal, C., Perez, L. G., Tang, H., Moon-Walker, A., Montefiori, D. C., 2020. Tracking Changes in SARS-CoV-2 Spike: Evidence that D614G Increases Infectivity of the COVID-19 Virus. Cell, 182(4), 812–827.

Koyama, T., Platt, D., Parida, L., 2020. Variant analysis of SARS-CoV-2 genomes. Bull. World Health Organ. 98 (7), 495–504.

Kumar, T.A., 2013. CFSSP: Chou and Fasman secondary structure prediction server. Wide Spectrum 1 (9), 15–19.

Lindner, H.A., Fotouhi-Ardakani, N., Lytvyn, V., Lachance, P., Sulea, T., Ménard, R., 2005. The papain-like protease from the severe acute respiratory syndrome coronavirus is a deubiquitinating enzyme. J. Virol. 79 (24), 15199–15208.

Liu, D.X., Fung, T.S., Chong, K.-L., Shukla, A., Hilgenfeld, R., 2014. Accessory proteins of SARS-CoV and other coronaviruses. Antivir. Res. 109, 97–109.

Liu, H., Zhang, Q., Wei, P., Chen, Z., Aviszus, K., Yang, J., Downing, W., Peterson, S., Jiang, C., Liang, B., Reynoso, L., Downey, G. P., Frankel, S. K., Kappler, J., Marrack, P., Zhang, G., 2021. The basis of a more contagious 501Y.V1 variant of SARS-COV-2. bioRxiv: the preprint server for biology, 2021.02.02.428884

Lobanov, M., Bogatyreva, N.S., Galzitskaia, O.V., 2008. Radius of gyration is indicator of compactness of protein structure. Mol. Biol. 42 (4), 701–706.

Madeira, F., Park, Y.M., Lee, J., Buso, N., Gur, T., Madhusoodanan, N., Basutkar, P., Tivey, A., Potter, S.C., Finn, R.D., Lopez, R., 2019. The EMBL-EBI search and sequence analysis tools APIs in 2019. Nucleic acids Res. 47 (W1), 636–641.

Matthews, K., Schäfer, A., Pham, A., Frieman, M., 2014. The SARS coronavirus papain like protease can inhibit IRF3 at a post activation step that requires deubiquitination activity. Virol. J. 11, 209.

Osipiuk, J., Jedrzejczak, R., Tesar, C., Endres, M., Stols, L., Babnigg, G., Kim, Y., Michalska, K., Joachimiak, A., 2020. The crystal structure of papain-like protease of SARS CoV-2. RSCB PDB.

Osipiuk, J., Azizi, S.A., Dvorkin, S., Endres, M., Jedrzejczak, R., Jones, K.A., Kang, S., Kathayat, R.S., Kim, Y., Lisnyak, V.G., Maki, S.L., Nicolaescu, V., Taylor, C.A., Tesar, C., Zhang, Y.A., Zhou, Z., Randall, G., Michalska, K., Snyder, S.A., Dickinson, B.C., Joachimiak, A., 2021. Structure of papain-like protease from SARS-CoV-2 and its complexes with non-covalent inhibitors. Nat. Commun. 12 (1), 743.

Ozono, S., Zhang, Y., Ode, H., Sano, K., Tan, T.S., Imai, K., Miyoshi, K., Kishigami, S., Ueno, T., Iwatani, Y., Suzuki, T., Tokunaga, K., 2021. SARS-CoV-2 D614G spike mutation increases entry efficiency with enhanced ACE2-binding affinity. Nat. Commun. 12 (1), 848.

Pachetti, M., Marini, B., Benedetti, F., Giudici, F., Mauro, E., Storici, P., Masciovecchio, C., Angeletti, S., Ciccozzi, M., Gallo, R.C., Zella, D., Ippodrino, R., 2020. Emerging SARS-CoV-2 mutation hot spots include a novel RNA-dependent-RNA polymerase variant. J. Transl. Med. 2020 (18), 1–9.

Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C., Ferrin, T.E., 2004. UCSF Chimera—a visualization system for exploratory research and analysis. J. Comput. Chem. 25 (13), 1605–1612.

Plante, J.A., Liu, Y., Liu, J., Xia, H., Johnson, B.A., Lokugamage, K.G., Zhang, X., Muruato, A.E., Zou, J., Fontes-Garfias, C.R., Mirchandani, D., Scharton, D., Bilello, J.P., Ku, Z., An, Z., Kalveram, B., Freiberg, A.N., Menachery, V.D., Xie, X., Plante, K.

S., Weaver, S.C., Shi, P.-Y., . Spike mutation D614G alters SARS-CoV-2 fitness. Nature 592 (7852), 116–121.

Ratia, K., Pegan, S., Takayama, J., Sleeman, K., Coughlin, M., Baliji, S., Chaudhuri, R., Fu, W., Prabhakar, B.S., Johnson, M.E., Baker, S.C., Ghosh, A.K., Mesecar, A.D., 2008. A noncovalent class of papain-like protease/deubiquitinase inhibitors blocks SARS virus replication. Proc. Natl. Acad. Sci. U. S. A. 105 (42), 16119–16124.

Ratia, K., Saikatendu, K. S., Santarsiero, B. D., Barretto, N., Baker, S. C., Stevens, R. C., & Mesecar, A. D., 2006. Severe acute respiratory syndrome coronavirus papain-like protease: Structure of a viral deubiquitinating enzyme. Proc. Natl. Acad. Sci., 103(15), 5717 LP – 5722.

Reshamwala, S.M.S., Likhite, V., Degani, M.S., Deb, S.S., Noronha, S.B., 2021. Mutations in SARS-CoV-2 nsp7 and nsp8 proteins and their predicted impact on replication/transcription complex structure. J. Med. Virol. 93 (7), 4616–4619.

Rodrigues, C.H., Pires, D.E., Ascher, D.B., 2018. DynaMut: predicting the impact of mutations on protein conformation, flexibility and stability. Nucleic Acids Res. 46 (W1), 350–355.

Salpini, R., Alkhatib, M., Costa, G., Piermatteo, L., Ambrosio, F.A., Di Maio, V.C., Scutari, R., Duca, L., Berno, G., Fabeni, L., Alcaro, S., Ceccherini-Silberstein, F., Artese, A., Svicher, V., 2021. Key genetic elements, single and in clusters, underlying geographically dependent SARS-CoV-2 genetic adaptation and their impact on binding affinity for drugs and immune control. J. Antimicrob. Chemother. 76 (2), 396–412.

Shen, Z., Ratia, K., Cooper, L., Kong, D., Lee, H., Kwon, Y., Li, Y., Alqarni, S., Huang, F., Dubrovskyi, O., Rong, L., Thatcher, G. R., Xiong, R., 2021. Potent, Novel SARS-CoV-2 PLpro Inhibitors Block Viral Replication in Monkey and Human Cell Cultures. bioRxiv: the preprint server for biology,

Shin, D., Mukherjee, R., Grewe, D., Bojkova, D., Baek, K., Bhattacharya, A., Schulz, L., Widera, M., Mehdipour, A.R., Tascher, G., Geurink, P.P., Wilhelm, A., van der Heden van Noort, G.J., Ovaa, H., Müller, S., Knobeloch, K.-P., Rajalingam, K., Schulman, B.A., Cinatl, J., Hummer, G., Ciesek, S., Dikic, I., 2020. Papain-like protease regulates SARS-CoV-2 viral spread and innate immunity. Nature 587 (7835), 657–662.

Trott, O., Olson, A.J., 2010. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. J. Comput. Chem. 31 (2), 455–461.

Van Der Spoel, D., Lindahl, E., Hess, B., Groenhof, G., Mark, A.E., Berendsen, H.J.C., 2005. GROMACS: fast, flexible, and free. J. Comput. Chem. 26 (16), 1701–1718.

Wu, A., Peng, Y., Huang, B., Ding, X., Wang, X., Niu, P., Meng, J., Zhu, Z., Zhang, Z., Wang, J., Sheng, J., Quan, L., Xia, Z., Tan, W., Cheng, G., Jiang, T., 2020a. Genome Composition and Divergence of the Novel Coronavirus (2019-nCoV) Originating in China. Cell Host Microbe 27 (3), 325–328.

Wu, F., Zhao, S.u., Yu, B., Chen, Y.-M., Wang, W., Song, Z.-G., Hu, Y.i., Tao, Z.-W., Tian, J.-H., Pei, Y.-Y., Yuan, M.-L., Zhang, Y.-L., Dai, F.-H., Liu, Y.i., Wang, Q.-M., Zheng, J.-J., Xu, L., Holmes, E.C., Zhang, Y.-Z., 2020b. A new coronavirus associated with human respiratory disease in China. Nature 579 (7798), 265–269.

Zhang, L., Jackson, C. B., Mou, H., Ojha, A., Rangarajan, E. S., Izard, T., Farzan, M., Choe, H., 2020. The D614G mutation in the SARS-CoV-2 spike protein reduces S1 shedding and increases infectivity. bioRxiv: the preprint server for biology, 2020.06.12.148726.

Studer, R.A., Dessailly, B.H., Orengo, C.A., 2013. Residue mutations and their impact on protein structure and function: detecting beneficial and pathogenic changes. Biochem. J. 449 (3), 581–594.

Zhou, G., Chen, M., Ju, C.J., Wang, Z., Jiang, J.Y., Wang, W., 2020. Mutation effect estimation on protein–protein interactions using deep contextualized representation learning. NAR Genom. Bioinform. 2 (2), lqaa015.

Zoete, V., Cuendet, M.A., Grosdidier, A., Michielin, O., 2011. SwissParam: a fast force field generation tool for small organic molecules. J. Comput. Chem. 32 (11), 2359–2368.