bioRxiv preprint doi: https://doi.org/10.1101/2021.07.17.452787; this version posted July 19, 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 The copyright holder for this preprint (which was not certified by peer review) is the author/funder.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.

Unveiling Mutation Effects on the Structural Dynamics of the Main Protease from SARS-CoV-2 with Hybrid Simulation Methods

Philot EA[1], Gasparini P[1], Magro AJ[2,3], Mattos JC[1], Torres-Bonfim NESM[1], Kliousoff A[1], Quiroz RCN[1], Perahia D[4], Scott AL[1,*]

[1] Computational Biology and Biophysics Laboratory, Federal University of ABC - UFABC, Santo André, São Paulo, Brazil.

[2] Department of Bioprocesses and Biotechnology, School of Agriculture (FCA), Unesp, Botucatu, São Paulo, Brazil.

[3] Institute of Biotechnology (IBTEC), Unesp, Botucatu, São Paulo, Brazil.

[4] ÉcoleNormale Supérieure Paris-Saclay, LBPA, Scaly, France.

* Corresponding author: Ana Ligia Scott

Email: analigiascott@gmail.com

Telephone: +551149960041

Address: UFABC - Universidade Federal do ABC, Centro de Matemática, Computação e Cognição, Laboratório de Biofísica e Biologia Computacional, Campus Santo André, Avenida dos Estados 5001, Bairro Bangu, Santo André – SP, Brazil, CEP 09606-070.

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.17.452787; this version posted July 19, 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 The copyright holder for this preprint (which was not certified by peer review) is the author/funder.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.

Highlights

- Single surface mutations lead to changes in M^{pro} structural dynamics.
- Mutants can be more stable than WT according to the structural dynamics properties.
- M^{pro}mutants can present a distinct functionality in relation to the wild-type.
- Potential viral markers for more pathogenic or transmissible SARS-CoV-2 variants.

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.17.452787; this version posted July 19, 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 The copyright holder for this preprint (which was not certified by peer review) is the author/funder.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 main protease of SARS-CoV-2 (called M^{pro} or $3CL^{pro}$) is essential for processing polyproteins encoded by viral RNA. Macromolecules adopt several favored conformations in solution depending on their structure and shape, determining their dynamics and function. Integrated methods combining the lowest-frequency movements obtained by Normal Mode Analysis (NMA), and the faster movements from Molecular Dynamics (MD), and data from biophysical techniques, are necessary to establish the correlation between complex structural dynamics of macromolecules and their function. In this article, we used a hybrid simulation method to sample the conformational space to characterize the structural dynamics and global motions of WT SARS-CoV-2 M^{pro} and 48 mutants, including several mutations that appear in P.1, B.1.1.7, B.1.351, B.1.525 and B.1.429+B.1.427 variants. Integrated Hybrid methods combining NMA and MD have been useful to study the correlation between the complex structural dynamics of macromolecules and their functioning mechanisms. Here, we applied this hybrid approach to elucidate the effects of mutation in the structural dynamics of SARS-CoV-2 M^{pro}, considering their flexibility, solvent accessible surface area analyses, global movements, and catalytic dyad distance. Furthermore, some mutants showed significant changes in their structural dynamics and conformation, which could lead to distinct functional properties.

Keywords: Covid-19, Protease M^{pro}, Variants, Structural Dynamics, Normal Modes, Molecular Dynamics.

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.17.452787; this version posted July 19, 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 2019 coronavirus disease (COVID-19) pandemic is the most severe health crisis in the past 100 years.As of June 07, 2021, approximately 173 million cases have been reported worldwide, with more than 3.72 million deaths (WHO Coronavirus (COVID-19) Dashboard). In addition to the zoonotic agents responsible for Severe Acute Respiratory Syndrome (SARS) and Middle East Respiratory Syndrome (MERS), the novel coronavirus, SARS-CoV-2, is a spherical-enveloped virus with a positive-sense single-stranded RNA (+ssRNA) which belongs to the broad *Coronaviridae*family and the *Nidovirales* order [1,2]. The SARS-CoV-2 genome comprises about 30,000 nucleotides, including open reading frames (ORFs) responsible for encoding structural and non-structural proteins. ORF1a and ORF1b are translated into two overlapping polyproteins, pp1a and pp1ab, respectively [1]. These polyproteins are processed by the viral proteases M^{pro} (main protease) and PL^{pro} (papain-like protease) into 16 non-structural proteins (NSPs). Other ORFs undergo discontinuous transcription, and subgenomic mRNAs are translated into other non-structural, accessory, and structural molecules, such as envelope (E), spike (S), membrane (M) and nucleocapsid (N) proteins [3,4].

SARS-CoV-2 M^{pro} is a dimer composed of three domains (the catalytic domains I and II and the C-terminal domain III), with the catalytic dyad formed by the residues His41 and Cys145 located in a cleft between domains I (residues 10-99) and II (residues 100-182) [5]. Dimerization is essential for SARS-CoV-2 M^{pro} enzymatic activity since the NH_2-terminal residues (N-finger) of each monomer interact with the Glu166 residue of the other monomer, which effectively contributes to the correct arrangement of the S1 pocket of the substrate binding site. Thus, the C- and N-terminus of the monomers form the dimeric interface of the active protein and are less flexible in comparison to their higher mobility in the free monomers [5]. This general protein architecture is highly conserved in M^{pro}s of diverse coronaviruses [6]; therefore, despite the extensive mutagenesis of these viruses in general, these key proteins are well conserved [7] and, therefore, are good targets for preventing virus replication and proliferation and reducing the risk of mutation-mediated drug resistance in new viral strains. Thus, the inhibition of M^{pro} enzymatic activity could be an interesting strategy for exploring new therapeutic approaches to treat COVID-19. Indeed, non-structural proteins have been reported as potential targets for the design and development of antiviral agents against SARS and MERS [8].

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.17.452787; this version posted July 19, 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.

Recently, six worrisome SARS-CoV-2 variants, commonly associated with a potential escape of the immune response and a more transmissible and infectious capacity, were first detected in the UK (Alpha, B.1.1.7) [9], in South Africa (Beta, B.1.351) [10,11], in Brazil/Japan (Gama, P.1) [12], in India (Delta, B.1.617+) [13], in USA/California (Epsilon, B.1.429+B.1.427) [14] and in UK/Nigeria (Eta, B.1.525) [15]. Interestingly, the genomic analysis of these variants listed in the GISAID database (https:gisaid.org) [16] highlights a set of frequent mutations such as K90R, P108S, K236R, L220F and R279C, in the SARS-CoV-2 M^{pro}. Also, Amamuddy et al. (2020) [17] identified other non-synonymous mutations in all M^{pro} domains, including substitutions in solvent accessible residues and some in the N-finger region. Moreover, these last authors demonstrated the collective effects of various M^{pro} mutations in several SARS-CoV-2 isolates using different approaches and techniques, as geometry calculations, cavity compaction analyses, molecular dynamics simulations, anisotropic network model (ANM) calculations, and coarse-grained Monte Carlo simulations, among others.

From a protein dynamics perspective, it would be worth investigating the structural effects of these amino acid substitutions on the stability, activation, reactivity, and catalytic activity related to SARS-CoV-2 M^{pro}. Hence, in this work, we analyzed the molecular characteristics as solvent accessible surface areas, the conformation of the catalytic dyad, and the flexibility of the regions involved in dimerization and substrate binding in wild-type and different SARS-CoV-2 M^{pro} mutants. A set of conformations of these molecules was generated using *in silico* methods based on all-atom molecular normal modes calculation. The results indicated that specific single mutations in SARS-CoV-2 M^{pro} could cause important changes in structural and dynamical characteristics regarding collective movements, the solvent accessible surface area (SASA) of the dimer interface, energetically accessible conformations, and arrangement of the catalytic dyad. These changes may influence the function and stability of SARS-CoV-2 M^{pro} present in important viral variants and could contribute to the emergence of harmful strains of SARS-CoV-2. This work analyzed the wild-type (WT) SARS-CoV-2 M^{pro} and several mutants using a hybrid approach *in silico* protocol to understand the possible influence of single mutations on this essential viral protein. The results shown here indicated that some of these mutations affect the SARS-CoV-2 M^{pro} structural dynamics and may potentially alter the functional properties of this macromolecule.

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.17.452787; this version posted July 19, 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. Material and methods

A Hybrid Method (VMOD, a module of the CHARMM program) that combines Normal Mode Analysis and low-temperature Molecular Dynamics simulations (described in the next section) was used to analyze the global motions and conformational changes for both WT and mutants of SARS-CoV-2 M^{pro}. The characterization of the protein structural dynamics was done with the following analyses: a) flexibility with the Root Mean Square Fluctuation (RMSF); b) analysis of collective modes of wild-type and mutants (correspondence/correlation calculation); c) conformational sampling using hybrid methods; d) distance analysis between the two residues of the catalytic dyad; e) SASA analysis for different regions; f) analysis of the frequency of occurrence of mutants in the patient data deposited in the GISAID database.

2.1. SARS-CoV-2 M^{pro} mutants modeling

A total of 98 SARS-CoV-2 M^{pro} structures deposited in PDB until November 2020 were selected for initial analyses. Wild-type and apo crystal SARS-CoV-2 M^{pro} structures with 1.91 Å resolution (PDB ID: 7C2Y [DOI: 10.2210/pdb7C2Y/pdb]) were used to generate 48 mutants with single point mutations [17]using PyMOL software [The PyMOL Molecular Graphics System, Version 2.0 Schrödinger, LLC]. The most likely mutant rotamers were selected using the PyMOL Mutagenesis tool. The residue protonation states WT and mutants of SARS-CoV-2 M^{pro} at physiological pH (pH= 7.0) were assigned in the PDB2PQR server [18]. The protonated structures were submitted to the Solution Builder module of the CHARMM-GUI web server [19–21] to generate the input files for the molecular dynamics simulation used to analyze the Normal Modes [22–24]. We followed the protocol as outlined below:

- Selection of a dimeric wild-type and apo SARS-CoV-2 M^{pro} structure in PDB.
- Selection of SARS-CoV-2 M^{pro} mutants described by Amamuddy et al, 2020 [17].
- Building 48 mutants using Pymol and protonation with PDB2PQR server (see Figure 1). It is important to highlight that the start point of this study is the mutants reported by Amamuddy et al. (2020) [17].
- Energy minimization of the WT and mutants of SARS-CoV-2 M^{pro}.

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.17.452787; this version posted July 19, 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.

- Calculation of the first 18 lowest frequency modes with an all-atom approach. The lowest frequency normal modes represent the collective motions of the protein.
- Analysis and selection of the mutants with a significant difference of RMSF compared to wild-type. We considered four important regions [5,17] presented in Table 1 to accomplish this analysis.
- Generation of energetically relaxed conformations along the lowest frequency normal modes by carrying out a low temperature MD and energy minimizations (the first 6 modes).

Table 1: Important regions of the SARS-CoV-2 M^{pro} used as reference to compare the RMSFs and select the most significant mutants [5,17].

Region	Residues	Functionality
Catalytic Dyad (CatDy)	His41, Cys145	Responsible for substrate cleavage
Substrate binding (Bsite)	His41, Met49, Gly143, Ser144, His163, His164, Met165, Glu166, Leu167, Asp187, Arg188, Gln189, Thr190, Ala191, Gln192	Region that undergoes conformational modifications to binding
Dimerization (Dinterf)	Arg4, Ser10, Gly11, Glu14, Asn28, Ser139, Phe140, Ser147, Glu290, Arg298	Dimer Interface
N-Finger (N-Fi)	Ser1, Gly2, Phe3, Arg4, Lys5, Met6, Ala7, Phe8, Pro9	Important for the Dimer Stability and Catalysis

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.17.452787; this version posted July 19, 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.

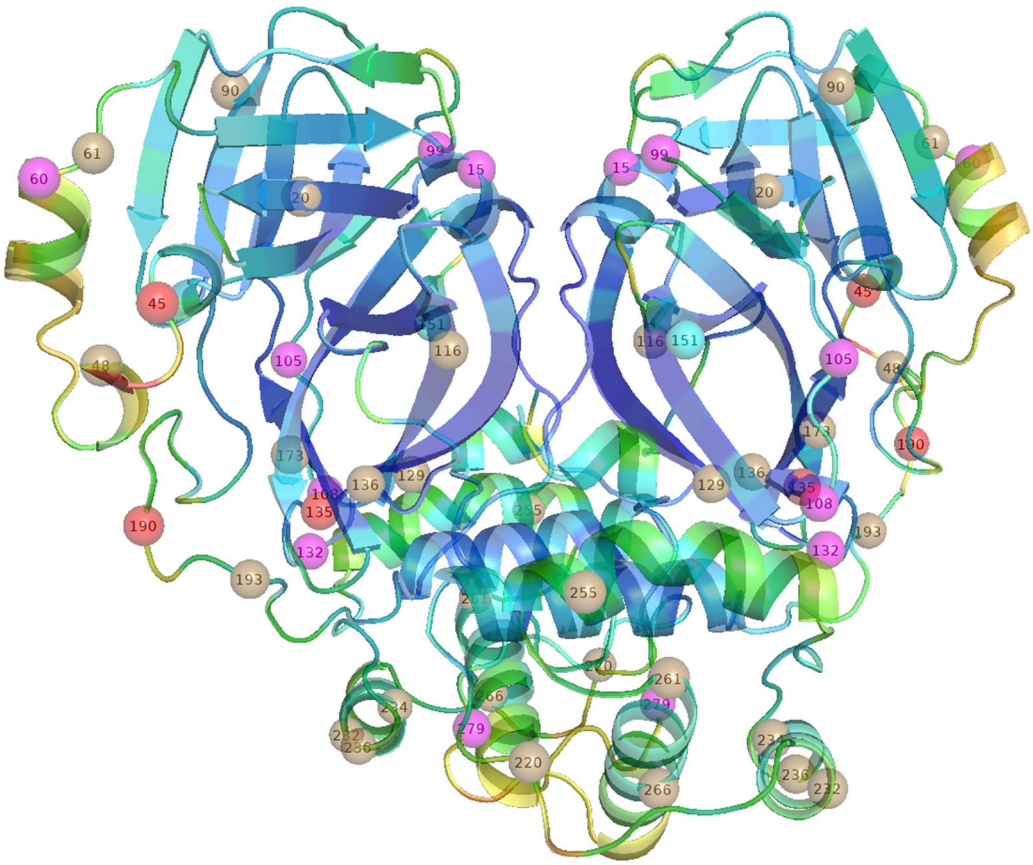


Figure 1: SARS-CoV-2 M^{pro}cartoon representation with the mutated amino acid residues along the molecular tertiary structure. The backbone color corresponds to the RMSFs values for the WT where blue and red represent more rigid and flexible respectively. The positions of mutation for the mutants selected by RMSF (described in Table 2) are highlighted by spheres colored according to the characteristic of mutation: a) conservative (no charge change) in brown, b) charge inversion in cyan, c) polar to apolar without charge in red, d) special residue in magenta. Figure generated in PyMOL.

2.2. Normal Modes Analysis (NMA) and conformational selection of the most significant mutations

Initially, the protein structure was minimized in two stages. In the first one harmonic constraints were applied and progressively decreased (250 to 0 kcal $mol^{-1}Å^{-2}$) and for each constraint value 500 steps of gradient conjugate (CG) were applied. After the CG minimization the constraints were removed and $2x10^5$ steps of adopted basis NeWTon Raphson (ABNR) were used with the convergence criterion of 10^{-6} kcal $mol^{-1}Å^{-1}$ RMS to energy gradient. The normal modes were calculated using the DIMB (Iterative Mixed-Basis Diagonalization) module [24] implemented in CHARMM software [25,26] considering all the atoms of the protein and a force field corresponding to the CHARMM36m [27].

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.17.452787; this version posted July 19, 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.

A set of the first six normal modes describing the internal movements were computed to ensure that all relevant large amplitude motions were included in our analysis. This is a sufficient set to ensure that all relevant large amplitude movements are included in our analysis. The modes corresponding to global rotation and translation were discarded. In the treatment of nonbonded interactions the Van der Waals (VDW) ones were computed up to 5 Å and a switching function was used to approximate these interactions until 9 Å. A distance-dependent dielectric constant ($e=2r_{i,j}$) was employed to treat the shielding of electrostatic interactions by the solvent. In the work described here such a function has a good performance to predict realistic intrinsic global movements for all structures (WT and mutants of SARS-CoV-2 M^{pro}).

2.3. Conformational space sampling along the collective modes

To obtain energetically allowed displaced structures along selected normal modes, we used the VMOD module from CHARMM [24] that makes use of harmonic restraints applied only to the Cα atoms. The restraints are set to target a given position along a specific normal mode vector. The MD and energy minimization were carried out for all atoms, enabling complete freedom of motion of the protein, including side chains. The structures were displaced from -1.0 to +1.0 Å (Mass Weighted Root Mean Square - MRMS values) along modes with steps of 0.2 Å, resulting in 11 intermediate structures for each mode. The displacements along the modes are achieved by using a series of low temperature MD simulations followed by energy minimizations. The procedure is similar to the one used by Batista et al. (2011) [28] and it is briefly described here. At each stage, the force constant of the harmonic restraining potential (K_d) was increased to ensure that the desired normal mode displacement was obtained. This procedure ensures that the structure slowly converges to the desired displacement. The restrain force K_d value was increased from 1,000 to 10,000 kcal·mol^{-1} Å^{-2} during successive 10 ps MD simulations. Velocities were assigned at random corresponding to a temperature of 30 K in the MD. The Berendsen thermostat with a coupling constant of 0.1 ps was used in all low temperature simulations.

Using a low-temperature dynamic allows for a more efficient conformational search than using only energy minimizations. A final MD simulation was achieved with a K_d value of 20,000 kcal·mol^{-1} Å^{-2} followed by 1,000 steepest descent and 1000 conjugate gradient steps of minimizations to reach the final displacement target along the

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.17.452787; this version posted July 19, 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.

mode.Using this procedure 66 structures were generated (11 structures per mode x 6 modes) for each molecular system considered.

2.4. Distance between the catalytic dyad atoms

According to Ramos-Guzmán et al. (2020) [29], the reaction mechanism proposed to proteolysis catalyzed by SARS-CoV-2 M^{pro} involves two steps, the acylation and deacylation steps. Briefly, in the first one the Sγ atom of Cys145 achieves nucleophilic attack to substrate peptide bond, before that occurs the formation of catalytic dyad ion pair by proton transfer from Sγ atom to the Nε atom of His41. Lastly, in the deacylation process, the bond between the M^{Pro} and the substrate is broken by hydrolysis that enables the enzyme to participate in a new catalytic cycle [29]. Regarding the first step of the reaction, the distance between the histidine nitrogen atom (Nε) and the cysteine sulfur (Sγ) was calculated to check its variability in the set of structures generated along the normal modes considered.

2.5. Solvent accessible surface area (SASA)

The CHARMM program was used to calculate the Solvent Accessible Surface Area (SASA) for the set of structures generated along the normal modes. We have reported the SASA values to the entire protein (SASAt - total), hydrophobic residues (SASAhp), dimeric interface residues (SASAdim), substrate binding site residues (SASAsb), His41 and Cys145 residue.

2.6. Mantel test

The lowest frequency normal modes calculated for WT and mutants of SARS-CoV-2 M^{pro} may not be completely redundant or possess matching. Several works in the literature show and report that a simple mutation can alter the motions present in the mutant, causing the emergence of new modes or the loss of some motion present in the WT. This same phenomenon can be noticed in the case of the SARS-CoV-2 M^{pro}.

Comparisons between the movements of the WT and the chosen mutants of SARS-CoV-2 M^{pro} were performed using the protocol described in the Mantel test. In this protocol described by Louet et al. [29], motions in 3D space were analyzed by calculating 2D matrices, exchanging the relative displacements of all pairs of Cα atoms between two structures [30,31]. Briefly, the method consisted of calculating 2D matrices reflecting the

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.17.452787; this version posted July 19, 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.

relative displacements of all pairs of Cα atoms between two structures. For each mode, these matrices were calculated from the two structures obtained at a displacement amplitude of 1.0 Å in each direction. The correlation coefficients between two matrices were then calculated with Mantel's test [30]. With this method, it was assumed that two 2D maps sharing a correlation coefficient greater than 0.6 described highly related motions in cartesian space.

2.7 Statistical analysis

The RMSF arithmetic average was calculated for each analyzed region (N-finger, catalytic dyad, substrate binding and dimerization) of the WT and mutants of SARS-CoV-2 M^{pro}. The selected mutants in which the average RMSF of at least one chain (A or B) exceeded the threshold of 10% compared to the RMSF values of the control group (WT).

The data obtained related to the potential energy, distance between the catalytic dyad atoms (Cys145, His41), and SASA from (i) binding site, (ii) hydrophobic residues, (iii) dimer interface residues, and (iv) total molecular surface were compared between the 48 mutants and WT of SARS-CoV-2 M^{pro}using ANOVA (Analysis of Variance) followed by Dunnett's test (with significance at $p \leq 0.05$). All statistical analyzes were performed using the multcomp package of ANOVA [32] implemented by R-Studio package [33]. The Heatmap, Violin and Boxplot representations were generated in Python using the Pandas [34,35], NumPy [36], Matplotlib [37] and Seaborn [38] libraries.

2.8 Analysis of dimer interface

We used the PDBePISA (Proteins, Interfaces, Structures and Assemblies) to investigate the interactions in the dimer interface for the WT and those mutants of SARS-CoV-2 M^{pro}. Using the PISA web server, we calculate the solvation free energy gain upon formation of the interface, interface area and number of potential hydrogen bonds and salt bridge [39]. Complementary to the PISA results, we applied the TKSA-MC [40] to estimate the contribution of electrostatic interaction to free energy for the polar residues, especially those that are localized in the interface of the dimer (Arg4, Glu14, Glu240 and Arg298). The TKSA-MC calculates protein charge–charge interactions via the Tanford–Kirkwood Surface Accessibility model with the Monte Carlo method for sampling different protein protonation states. We used these two softwares to improve the

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.17.452787; this version posted July 19, 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.

discussion about the stability of the dimer for 4 mutants showing significant SASA and Potential Energy reduction (see Table 2).

3. Results and discussion

3.1. General characteristics of the SARS-CoV-2 M^{pro} mutants identified in GISAID database

Several works have reported variants of SARS-CoV-2 with mutations in different regions of M^{pro} and their effect on its functionality [5,17]. Thereby, in this paper we intend to study the collective motions for the wild-type and different mutant SARS-CoV-2 M^{pro} molecules shed some light on the contribution of a group of known amino acid substitutions on the dynamics characteristics of this essential viral protein. For this, initially we calculated the normal modes for the WT SARS-CoV-2 $M^{pro}$and 48 mutants of SARS-CoV-2 M^{pro} (set 1) using an all-atom approach. Following, we analyzed the structural dynamics of these molecules based on their Cα flexibility of key regions, such as the N-finger, dimeric interface, substrate binding site and His41-Cys145 catalytic dyad. As shown in Figure 2, both chains A and B from mutants of SARS-CoV-2 M^{pro} present significant fluctuations/differences in comparison to the wild-type molecule considering the average Cα RMSF (root mean square deviation) of the mentioned key regions. Additionally, there is a clear variation of the flexibility even between the chains A and B from the WT and mutants of SARS-CoV-2 M^{pro} (Figures S1, S2 and S3). In chain B, it is evident that the substrate binding region and the catalytic dyad of the mutants presented, on average, a higher flexibility in comparison to the wild-type molecule. This asymmetric/distinct behavior between the chains in homodimeric enzymes were also identified by molecular dynamics simulations involving the SARS-CoV-2 M^{pro}[17] and SARS-CoVM^{pro}[41] as well as NMA calculation for HIV-1 protease [42] and DPP-IV diabetes related protein [43]. Thus, this analysis indicated that some mutants may present characteristics that can influence their catalytic activity, substrate binding affinity and/or dimer stability, as also shown previously by Amamuddy et al. (2020) [17].

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.17.452787; this version posted July 19, 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.

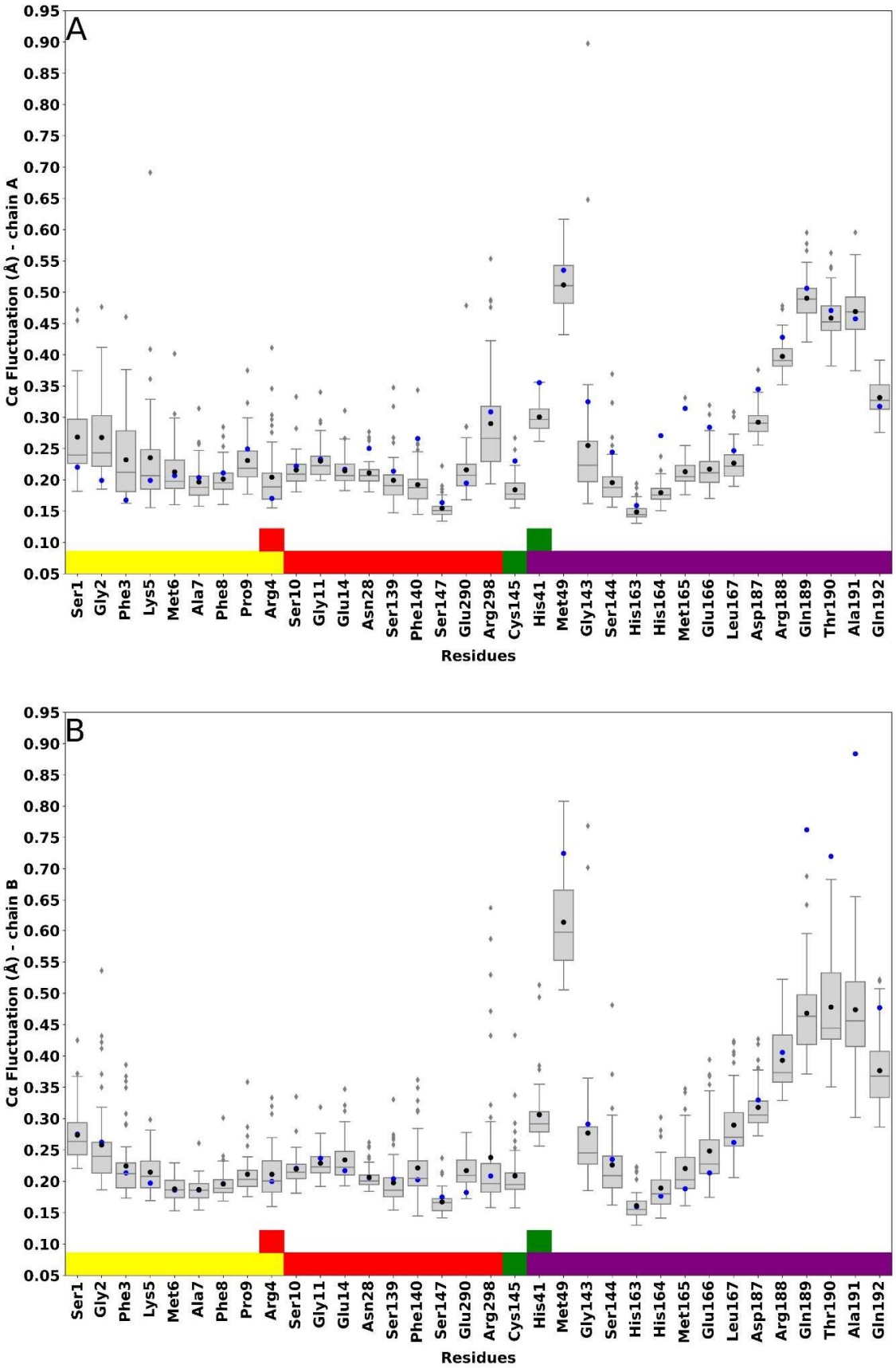


Figure 2: Boxplot RMSF of important regions of the SARS-CoV-2 M^{pro} reported in Table 1. (**A**) Chain A and (**B**) chain B. N-finger in yellow, dimeric interface residues in red, catalytic dyad in green and substrate binding site in purple. Black circle represents the mean of mutantsand the

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.17.452787; this version posted July 19, 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.

blue circle represents the measure to WT. The residues Arg4 and Cys145 are not in ascending order in N-finger and catalytic dyad regions, respectively, to clarify that they are present in more than one region.

The initial analysis of the 48 SARS-CoV-2 M^{pro} mutants allowed the selection of 35 mutants with a high degree of flexibility (set 2), which were later evaluated considering some additional criteria as potential energy (Figure S4), SASAt (Figure S5), SASAhp (Figure S6A), SASAdim (Figure S6B), SASA of His41 and Cys 145 (Figure S7A and S7B), SASAsb (Figure S8), and catalytic dyad distance (Figure S9). Thus, this new screening of the calculated structures allowed to underscore 26 mutants of SARS-CoV-2 M^{pro} (set 3) with significant statistical variation in at least three of the mentioned criteria in comparison to WT SARS-CoV-2 M^{pro}. Additionally, these 26 mutants from set 3 were divided into four groups according to the type of amino acid substitution, i.e., involving conservative substitution (when a given amino acid residue is replaced by a different one with similar biochemical properties), electric charge or polarity modification, and special cases in which occurs the inclusion/exclusion of amino acid residues with unique properties (cysteine, glycine, histidine or proline) (Table 2).

The most number of mutants from set 3 (15 structures) was classified in the first case, involving conservative replacement. In this group, besides significant Cα RMSF alterations, nine mutants showed relevant modifications of potential energy. More specifically, seven of these 15 mutants (D48E, K90R, I136V, A173V, L220F, and L232F) presented a negative variation of the potential energy and only one (L220F) showed a positive variation of the potential energy in relation to the WT SARS-CoV-2 M^{pro}. From these structures, the amino acid substitution D48E was the unique which did not present a significant negative variation in at least one of the considered additional criteria (SASAdim, SASAsb, SASAhp, and SASAt). Considering the occurrence of potential energy and SASAdim variation and at least one of the parameters related to surface exposition (SASAhp and SASAt), it is possible to point out five mutants of SARS-CoV-2 $M^{pro}$with statistically significant differences in comparison to the WT SARS-CoV-2 M^{pro} (K90R, I136V, A173V, L220F, and L232F). Particularly, the mutation K90R is present in the variants B.1.1.7 (first detected in the UK), B.1.351 (first detected in South Africa), P.1 (first detected in Brazil/Japan), B1.617 (first detected in India), B.1.429+B.1.427 (first detected in USA/California), and B.1.525 (first detected in UK/Nigeria) and also is markedly prevalent in the SARS-CoV-2 M^{pro}sequences deposited

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.17.452787; this version posted July 19, 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 the GISAID database (around 70% of frequency). Another relatively prevalent mutation identified in the GISAID database was the K236R amino acid substitution, which presents 1,08 % of frequency and could be found in the variants B.1.1.7, B.1.351, P.1, B1.617, and B.1.429+B.1.427. In this case, there is no significant difference of the SASAdim in comparison to the wild-type molecule, however it was verified SASAt and SASAhp alterations. Curiously, the mutation I136V is the only one which presents significant variation in the SASAdim, SASAt and SASAhp, but was found in none of the variants from the GISAID database.

The other SARS-CoV-2 M^{pro} amino acid substitutions from set 3 (11) were grouped according to the electric charge/polarity alteration or associated to the inclusion/exclusion of cysteine, glycine, histidine, or proline amino acid residues. Only one substitution lead to the introduction of a negative-charged amino acid residue (N151D); it is important to highlight that this mut SARS-CoV-2 M^{pro} molecule was just found in the variant B.1.1.7 and showed a significant variation of the potential energy, SASAdim, SASAt and SASAhp in relation to the WT SARS-CoV-2 M^{pro}. Another three mutants of SARS-CoV-2 M^{pro} presented polar (not electrically charged) to non-polar amino acid substitutions involving the change of threonines by isoleucine residues (T45I, T135I, and T190I). In this case, the amino acid substitution T190I was found in four variants (B.1.1.7, B.1.351, P.1, and B.1429+B.1427) and could be considered that with the more important alterations compared to WT SARS-CoV-2 M^{pro}, since it introduces important modifications related to the potential energy, SASAdim, SASAsb, SASAhp, and SASAt. Finally, it were identified in the GISAID database seven amino acid substitutions related to the inclusion/exclusion of cysteine, glycine, histidine, or proline amino acid residues (G15D, R60C, P99L, R105H, P108S, P132L, and R279C). Here, it is possible to underline the amino acid substitution P108S, which can be identified in six variants (B.1.1.7, B.1.351, P.1, B1.617, B.1.429+B.1.427, and B.1525) and is also the most prevalent among the seven mutants of this subgroup, with 5,71 % of frequency considering the SARS-CoV-2 M^{pro} sequences deposited in the GISAID database. In addition, the substitution P108S promotes significant alterations of the SASAdim, SASAsb, and SASAt, as well as the substitutions R60C and R279C. However, the substitutions R60C and R279C are associated with an increase of the potential energy in comparison to the WT SARS-CoV-2 M^{pro}, whereas the substitution P108S decreases the potential energy.

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.17.452787; this version posted July 19, 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 2 - The table lists the mutants presenting at least three structural characteristics with significant variations compared with WT where the data are grouped by: 1) no charge change: apolar to apolar (AP -> AP: A->V, V->A, I->V, L->F, V->L), positive polar to positive polar (P(+) -> P(+): K->R, R->H), negative apolar to negative apolar (P(-) -> P(-): D->E); **2) charge inversion:** neutral to negative polar (NT -> P(-): N->D); **3)polar to apolar without charge:** polar to apolar (P -> AP: T->I); **4) special residue:** special to neutral (spe -> NT: P->S), special to apolar (spe -> AP: P->L), negative polar to special (P(-) ->spe: R->C), special to negative polar (spe -> P(-): G->D). **Columns means: Mutant:** mutation and residue; **Res.:** type of mutation (residues); **RMSF:** Root mean square fluctuations of chain A and B; **Energy:** energy calculated of each mutant; **Distance Nε:** distance between **Sγ** atom of Cys145 and **Nε** atom of His41; **Cys145 and His45:** catalytic dyad;**Bind. site:** SASA substrate binding site region; **Hidrop. SASA:** SASA ofhydrophobic residues; **Total SASA:** accessible total area; **Freq.:** frequency of each mutant on variants (%), by sequences deposited on GISAID (last access: 31/05/2021). **Variant:** variants which the mutant are present in GISAID. The used symbols to denote significant statistics of parameter versus WT: "*****" (presence of significant statistics); "**-**" **(**decreased parameter), "**+**" (increased parameter), "**A-**" (decreased parameter on chain A), "**B-**" **(**decreased parameter on chain B), "**B+**" (increased parameter on chain B). In yellow are highlighted the probable stable mutant dimers.

Mutant	RMSF	Energy	Distance Nε2	CYS 145	HIS 41	Bind. site	SAS Hydrop.	Dimer interface	Total SAS	Freq.	Variants
Conservative substitutions											
A116V	*	* -	* A-	* B-		* A-	* -			0,67%	B117; P1; B1429+B1427; B1525
A129V	*		* A+	* B-	* A+ B-		* -	* +	* -	0,62%	B117; B1351; P1; B1617; B1429+B1427; B1525
A173V	*	* -	* A-	* A+	* A-	* A-	* +		* -	0,04%	B117
A193V	*			* A-					* -	0,78%	B117; B1351; P1; B1617; B1429+B1427
A234V	*		* A-				* +			0,29%	B117; P1; B1617; B1429+B1427
A255V	*		* A-				* +			0,15%	B117; B1351; P1; B1617; B1429+B1427
A266V		* +								0,24%	B117; B1351; P1; B1617; B1429+B1427
V261A	*	* -		* B-					* -	0,02%	B117; B1429+B1427
I136V	*	* -	* A- B+	* A+	* A-	* A-	* -	* -	* -	0%	none
L220F	*	* +	* A+	* A+ B-	* A+		* -	* -	* -	0,71%	B117; B1351; P1; B1617; B1429+B1427; B1525
L232F	*		* B-		* B-			* -	* -	0,32%	B117; B1351; P1; B1617; B1429+B1427
V20L	*		* B-	* A+	* B-		* +	* +		0,04%	B117
K236R	*	* -	* A-				* -		* -	1,08%	B117; B1351; P1; B1617; B1429+B1427
K61R	*	* -				* A-	* -			0,01%	B117

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.17.452787; this version posted July 19, 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.

K90R	*	* -	* A-					* -	* -	70,54%	B117; B1351; P1; B1617; B1429+B1427; B1525
D48E	*	* -			* B+	* B+				0,71%	B117; B1617
Electric charge inversion substitutions											
N151D	*	* -	* A-		* A-		* -	* -	* -	0,01%	B117
Polarity change substitutions											
T135I	*		* B+					* -		0,03%	B117; P1; B1617
T190I	*	* -			* B+	* A- B+	* +	* +	* -	0,32%	B117; B1351; P1; B1429+B1427
T45I	*	* +	* A- B+		* B+	* A- B+	* +			0,10%	B117; B1351; B1617; B1429+B1427
Special substitutions											
P108S	*	* -					* -	* -	* -	5,71%	B117; B1351; P1; B1617; B1429+B1427; B1525
P132L	*	* -				* A- B+			* -	0,79%	B117; B1351; P1; B1429+B1427; B1525
P99L	*	* -	* B-	* B-	* B-			* -	* -	0,10%	B117; B1429+B1427
R279C	*	* +	* A+	* A+ B-	* A+		* -	* -	* -	0,93%	B117; P1; B1617; B1429+B1427; B1525
R60C	*	* +		* B-	* B-	* A- B-	* -	* +	* -	0,27%	B117; B1351; B1429+B1427
G15D	*	* -		* A-		* B+		* -		0,01%	B117; B1429+B1427
R105H	*	* +	* B-	* B-			* -		* -	0,04%	B117; B1617; B1429+B1427

3.2. Modifications of the SARS-CoV-2 M^{pro} dimeric interface induced by amino acid substitutions at protein surface

Up to now, the solved crystallographic structures of M^{pro} molecules from coronavirus indicate they are structurally very similar and basically dimeric. The M^{pro} protomers from coronavirus consists of three well-characterized domains, being two with a chymotrypsin fold and a third extra helix domain, which is essential for dimerization of this viral protease [44-48]. Also, since the mid 2000s, a series of scientific works has confirmed that the dimerization process is fundamental to the full catalytic activity of the main proteases from coronavirus [5]. Hence, it is reasonable to suppose that the combination of low levels of potential energy, lesser SASA values and larger dimeric interfaces could favor the protein stabilization and, in the specific case of the mutants of SARS-CoV-2 M^{pro}, to affect the activity of these enzymes. About this issue, Hu et al. (2009) [49] showed that two specific and neighbouring mutations at the SARS-CoV-1 M^{pro} dimeric interface lead this molecule to assume distinct oligomeric conformations. In fact, this work concluded that certain key amino acid residues control the SARS-CoV-1 M^{pro} dimerization and also suggested the dimeric stability of this protease is heavily dependent on the extent and integrity of the contacts between the monomers. In general,

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.17.452787; this version posted July 19, 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.

the calculated mutant structures from set 3 showed an expected tendency of SASAt decreasing associated with a negative variation of the potential energy. Although the mutants of SARS-CoV-2 M^{pro} do not present amino acid substitutions at their dimeric interfaces, some of the calculated mutant structures showed wider contact areas in relation to the wild-type molecules. More specifically, four surface amino acid substitutions from set 3 (K90R, P99L, P108S, and N151D) (Table 2) allowed to reach the mentioned parameter combination (lower energy, smaller SASAt, larger dimeric interfaces) considering their respective calculated structures. Thus, these mutants of SARS-CoV-2 M^{pro} were carefully examined to assess their differences to the WT SARS-CoV-2 M^{pro}, particularly those related to the dimeric interfaces and the influence of the K90R, P99L, P108S, and N151D amino acid substitutions in the conformational space accessed by the mutant molecules.

The PDBePISA web server indicated that some calculated K90R and N151D mutant structures presented a higher dimeric interface contribution to the total solvation free energy (Δ^iG) not including the effect of satisfied hydrogen bonds and salt bridges across the interface) in relation to the wild-type molecule and the other mutants (Figure 3A). Therefore, at least these two mutants are able to assume oligomeric conformations which are kept by a dimeric interface more hydrophobic than the remaining wild-type and mutant structures generated by the computed normal mode displacements. On the other hand, as shown in Figure 3A, there is a visible concentration of wild-type molecules which reach Δ^iG values around -11.0 kcal/mol whereas the calculated mut SARS-CoV-2 M^{pro} structures showed mostly higher Δ^iG values. Hence, the K90R, P99L, P108S, and N151D mutants can also form, with a higher probability, dimeric interfaces which offer a lesser hydrophobic contribution to the Δ^iG in comparison to the WT SARS-CoV-2 M^{pro}. Indeed, the Figure 3C indicates the K90R, P99L, P108S, and N151D mutants can potentially establish at their dimeric interfaces a higher number of hydrogen bonds than the wild-type protease, indicating thus the formation of more stable and specific contacts between the protomers. An example regarding this difference between the wild-type and mutant calculated structures is the influence of the mutation P99L. In this case, it was possible to generate a dimer presenting approximately ten hydrogen bonds at the interface more than the wild-type calculated structure with the highest number of hydrogen bonds connecting its monomers.Moreover, according to the TKSA-MC server [40], the electrostatic contribution to the total free energy (ΔGqq) of four amino acid residues

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.17.452787; this version posted July 19, 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.

present at the mut SARS-CoV-2 M^{pro} dimeric interfaces (residues) considering both protease chains was clearly distinct. As shown in Figure S10 (Supplementary Material), the ΔGqq values related to the mutant chains A were lower in comparison to the corresponding chains of the calculated wild-type structures. Otherwise, this difference is not observed for the chains B from the calculated wild-type and mutant K90R, P99L, P108S, and N151D structures. Although the ΔGqq dimeric interface electrostatic contribution is based on the amino acid residues Arg4, Glu14, Glu290 and Arg298, this finding corroborates the results calculated by the PDBePISA web server that also indicated the prevalence of polar dimeric interface interactions in several mutant structures. Indeed, Ding et al. (2005) [50] also highlighted that hydrophobic contacts and electrostatic interactions play major roles in the binding of a SARS-CoVM^{pro} dimerization inhibitor based on affinity capillary electrophoresis experiments. Then, based on the preceding findings, it is feasible to suggest that the surface amino acid substitutions K90R, P99L, P108S, and N151D were able to induce a higher density of structures with different dimeric interfaces (and consequently distinct oligomers), which also showed larger extensions and a considerable number of polar contacts. Thus, the comparison with the calculated wild-type structures indicates that the mutations in question may lead to more stable SARS-CoV-2 M^{pro}s.

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.17.452787; this version posted July 19, 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.

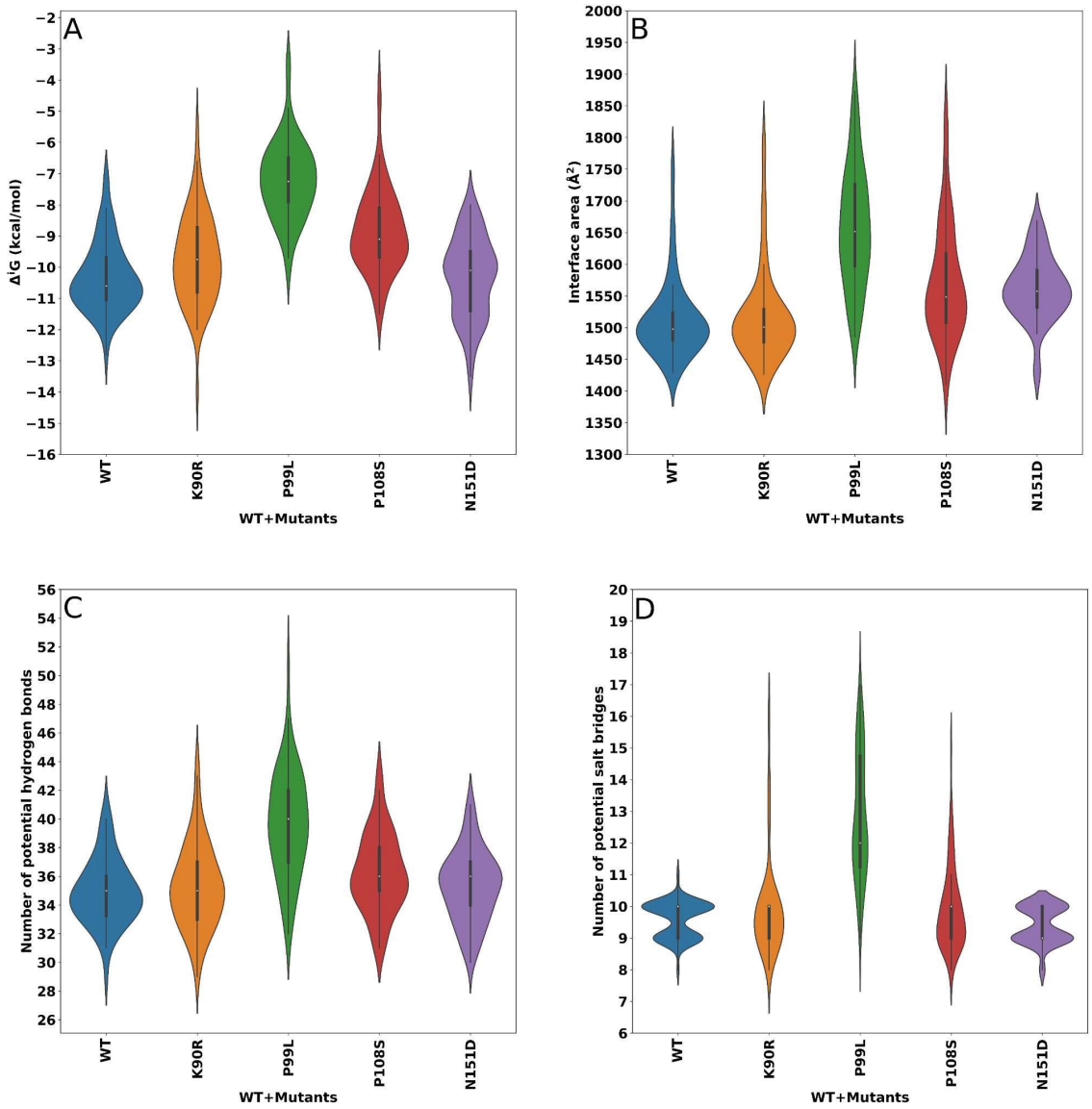


Figure 3: Dimeric interface analysis using PISA web server. (**A**) $\Delta^{i}G$ indicates the solvation free energy gain upon formation of the interface (kcal/mol). (**B**) Interface area (Å^2). (**C**) Number of potential hydrogen bonds. (**D**) Number of potential salt bridges.

3.3. Amino acid substitutions X catalytic dyad distance

It is well known that some amino acid residues are essential to the dimerization and activity of the main proteases from SARS-CoV-1. Bacha et al. (2004) [51] were the first authors to demonstrate this fact by identifying a vital cluster of conserved serine residues at the neighborhoods of the SARS-CoV-1 M^{pro} active site. The substitution of these serine residues by alaninesimpaired significantly the enzymatic activity of this protease. Simultaneously to the previous work, Chou et al. (2004) [52] reported the importance of a salt bridge interaction between Arg4 and Glu290 for SARS-CoV-1

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.17.452787; this version posted July 19, 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.

M^{pro}subunit association; under high salt concentration and low pH conditions this ionic interaction is missed leading to loss of dimerization and reduced activity of this CoV-1 protease. Hsu et al. (2005) [45] reported that the truncation of the residues 1 to 4 located at the N-terminal region promoted SARS-CoV-1 M^{pro} monomerization and substantially diminished its catalytic activity. Chen et al. (2008) [53] identified critical amino acid residues related to SARS-CoV-1 M^{pro} dimerization based on mutagenesis experiments. In this work, the authors found the Ser10 and Glu14 residues located at SARS-CoV-1 M^{pro} domain I are crucial to the formation of interface interactions, being highly conserved in several coronavirus main proteases.

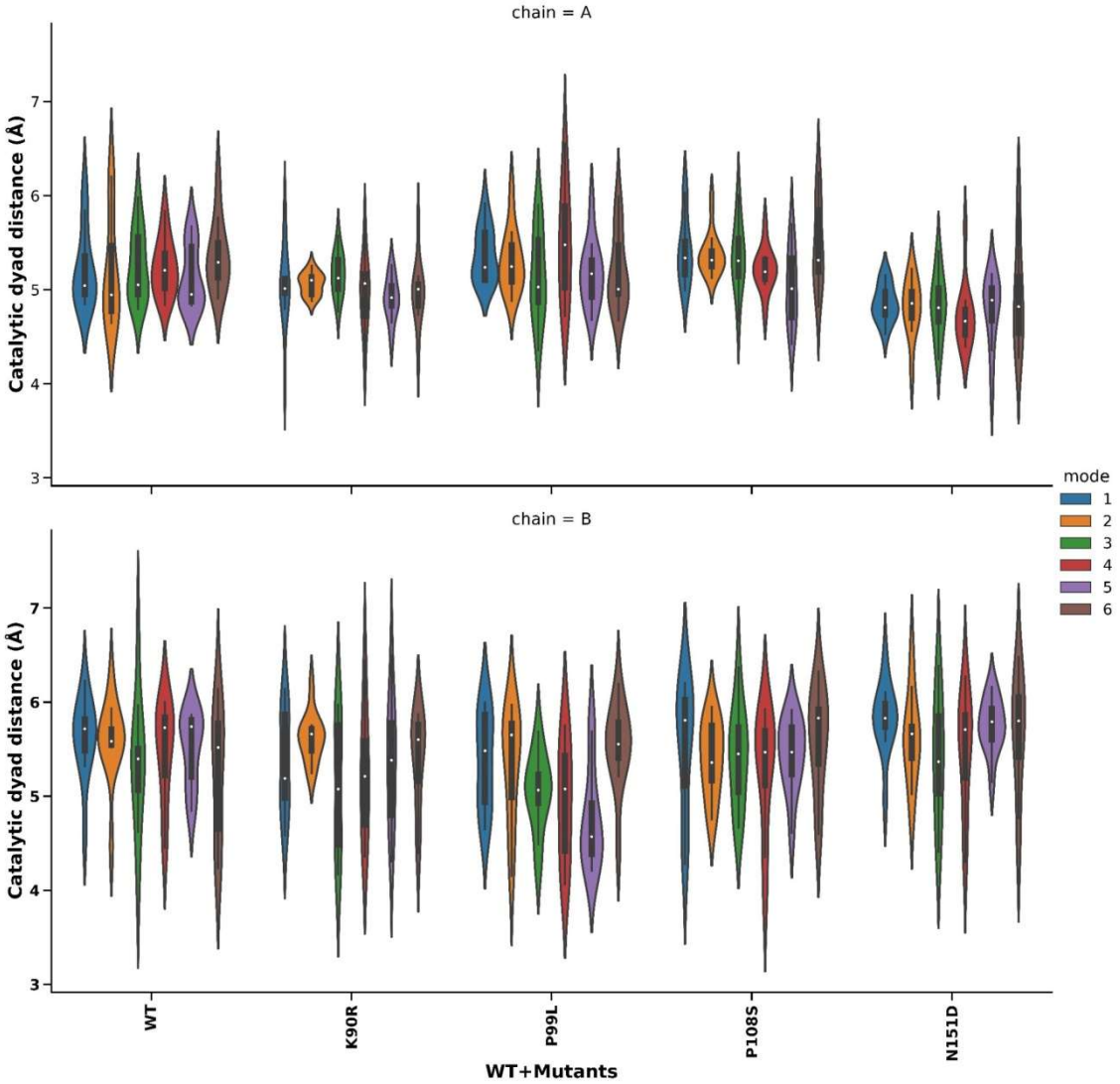
Additionally, Chen et al. (2008) [54] reported that the amino acid substitution G11A led to a total monomer separation after crystallization of this mutant SARS-CoV-1. Next, Shi et al. (2008) [55] showed that the SARS-CoV-1 M^{pro} R298A mutation produced crystals with monomeric molecules. In this study, the authors highlighted that Arg298 is a key amino acid residue involved in the dimerization of this protease. Remarkably, Zhong et al. (2008) [56] demonstrated that the truncated-N-finger M^{pro} C-terminal domain from SARS-CoV-1 can remain in a monomeric form and also dimerize forming a new interface. However, in this case, the new dimer is inactive, thus confirming the N-finger relevance for the formation of catalytically active SARS-CoV-1 M^{pro} molecules. Also, Hu et al. (2009) [49] identified two close mutations (S139A and F140A) at the SARS-CoV-1 M^{pro} dimer interface which resulted in different conformations of the crystal structure of this enzyme.

Most of the previous works based their conclusions on amino acid residues close to the dimeric interface or directly involved in the formation of dimeric contacts of the SARS-CoV-1 M^{pro}. However, our results indicated that the distinct conformational behaviour of the calculated K90R, P99L, P108S, and N151D mutants and wild-type SARS-CoV-2 M^{pro} structures are exclusively due to surface mutations since there was no alteration of the amino acid residues which compose the dimeric interface of these molecules. This is an interesting point, mainly considering the notable sequence and structural similarity between the SARS-CoV-1 and SARS-CoV-2 M^{pro} molecules. Indeed, it is possible to suppose that the surface K90R, P99L, P108S, and N151D amino acid substitutions may not only alter the dimeric interface but also the functionality of the SARS-CoV-2 M^{pro}. From this perspective, it would be noteworthy to investigate how the

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.17.452787; this version posted July 19, 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.

structural alterations related to these mutations could affect the nucleophilic attack of the Sγ atom of the Cys145 to the substrate scissile peptide bond.

According to Ramos-Guzmán (2020) [29], multiscale simulation methods showed the most probable between the catalytic dyad amino acid residues (Sγ atom from and Nε atom from H41) is around 3.3 Å, considering a conformer which keeps the shortest possible distance of the C45 side chain to a bound substrate molecule. The analysis of different apo-form dimeric SARS-CoV-1 and SARS-CoV-2 M^{pro} molecules available at the Protein Data Bank (PDB) also showed similar catalytic dyad distances (3.5 to 4.3 Å) (Table S1). As shown in Figure 4, some of the calculated K90R, P99L, P108S, and N151D mutant structures were able to reach catalytic dyad distances below 4.0 Å. Therefore, these results indicated the proposed protocol involving displacement of the initial wild-type and mutant structures generated molecules with a catalytic dyad geometry prone to trigger the substrate catalysis.



bioRxiv preprint doi: https://doi.org/10.1101/2021.07.17.452787; this version posted July 19, 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 4: Catalytic Distance: Distance Violin plot by chain. Distance between the Cys 145 Sγ atom and His 41 Nε atom for both chains of the structures generated by VMOD. Chain A in blue and chain B in orange. The WT SARS-CoV-2 M^{pro} measures are highlighted by a red rectangle.

Remarkably, new layers of information emerge based on the analysis of the collective motions of the calculated wild-type and K90R, P99L, P108S, and N151D mutant structures. Only the modes 3, 4, and 5 from mutant and wild-type molecules present some conservation degree (Table S2). This finding is very significant since it demonstrates the four surface mutations in question modify the SARS-CoV-2 M^{pro} dynamics, a fact that is a strong indication of functional diversity. Besides, there is also a higher catalytic dyad distance variation in the protomers B from the mentioned four mutants. Except for the normal mode 6, all of the other normal modes (modes 1 to 5) allowed the generation of mutant protomers B with shorter catalytic dyad distances in comparison to the corresponding wild-type protomers (Figure 4). It is also noteworthy to highlight that the K90R, P99L, P108S, and N151D mutants presented movements which brought their protomer B catalytic dyads to distances close to theoretical and experimental values.

These results may be analyzed under the light of the work executed by Chen et al. (2006) [41], which employed MD simulations and mutational studies to demonstrate that only one protomer of the dimeric SARS-CoV-1 M^{pro} is catalytically active. Hence, the number of calculated mutant structures (K90R, P99L, P108S, and N151D) with catalytic dyads presenting closer distances to that adequate for nucleophile activation (the C145 Sγ atom) also indicated that one of the SARS-CoV-2 M^{pro} protomers (in this case, the protomer B) has a higher probability to reach conformations more appropriate for catalysis. Further, other important information comes from a more detailed comparison between the wild-type protease and the K90R, P99L, P108S, and N151D mutants. As verified in Figure 4, the modes which generate protomers B with catalytic dyad distances around or below 4.0 Å are the following: (i) 1, 2, 3, 4, and 6 for WT SARS-CoV-2 M^{pro}; (ii) 1, 3, 4, 5 and 6 for K90R mut SARS-CoV-2 M^{pro}, (ii) 2, 3, 4, 5, and 6 for P99L SARS-CoV-2 M^{pro}, (iii) 1, 4 and 6 for P108S SARS-CoV-2 M^{pro}; and (iv) 3, 4 and 6 for N151D SARS-CoV-2 $M^{pro.}$ However, as plainly indicated by the distribution of the calculated structures (Figure 4), the modes do not present the same probability of generating protomer B molecules with catalytic dyad distances adequate for catalysis reaction. As

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.17.452787; this version posted July 19, 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.

example, it is possible to compare the modes 5 from the protomers B from wild-type and P99L mutant molecules; in this case, the wild-type movement do not generate structures with catalytic distances close to 4.0 Å, whereas the movements correspondent to the mode 5 lead to a higher number of mutant conformers with catalytic dyad distances in the appropriate range for nucleophile formation. In truth, most of the collective motions related to the conformations with shorter catalytic dyad distances correspond to a set of different mixed bending/twisting movements, including bending motions between monomers and twist torsions around the axes of the monomers (Figure S11). Thus, the results shown here point to some hints that could help to identify eventual differences between the action mechanism of wild-type and mutant SARS-CoV-2 M^{pro} molecules. As highlighted by Goyal and Goyal (2020) several works have been based on the design of substrate binding pocket-ligands to inhibit the SARS-CoV-1 M^{pro}, but none of them has reached clinical trials to date. Thus, a good therapeutic alternative to fight against SARS-CoV-2 could be to target the dimerization of its main protease. In this work, the importance of this process for SARS-CoV-2 M^{pro} is strengthened by the evident influence of surface mutations on the oligomeric conformation of this essential viral protease and, probably, to its molecular evolution and adaptation. Finally, the results shown here may also help to identify factors related to the emergence of more pathogenic or transmissible viral types.

5. Conclusions

This work compared the WT and mutants of SARS-CoV-2 M^{pro} using a normal modes-based protocol involving the analysis of simple parameters as Cα flexibility, potential energy, hydrophobic and electrostatic contributions to free energy, solvent accessible surface areas, and catalytic dyad distance. These parameters allowed the selection of potential structurally stable dimers, which were posteriorly analyzed and remarkably showed that single surface amino acid substitutions are able to induce significant dimeric interface changes. Also, the mutants showed a low conservation of collective motions in relation to the wild-type molecule according to the Mantel test. Additionally, the K90R, P99L, P108S, and N151D mutants presented modes with different probabilities of generating conformers with catalytic dyad prone to trigger the catalysis reaction in comparison to the wild-type protease. Therefore, this finding allowed us to suppose the mutants could present a distinct functionality in relation to the original SARS-CoV-2 M^{pro}.

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.17.452787; this version posted July 19, 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.

Finally, it is important to emphasize again the importance of the dimerization process for SARS-CoV-2 M^{pro} and the evident influence of surface mutations on the oligomeric conformation of this protein. Thus, this work may shed some light on the molecular evolution of this viral protease and the adaptation process of the SARS-CoV-2 to their human hosts and help to identify markers related to more dangerous variants and strains of this virus.

Funding

Thisworkwassupportedby Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) - Brazil; and Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) - Brazil, Fundação de Amparo à Pesquisa (Fapesp) [FinanceCode 001] - São Paulo, Brazil.

Declaration of interests

The authors declare no conflict of interest.

Appendix A. Supplementary data

Supplementary data to this article can be found online.

References

[1] S. Xiu, A. Dick, H. Ju, S. Mirzaie, F. Abdi, S. Cocklin, P. Zhan, X. Liu, Inhibitors of SARS-CoV-2 Entry: Current and Future Opportunities, J. Med. Chem. 63 (2020) 12256–12274. https://doi.org/10.1021/acs.jmedchem.0c00502.
[2] A. Ali, R. Vijayan, Dynamics of the ACE2–SARS-CoV-2/SARS-CoV spike protein interface reveal unique mechanisms, Sci. Rep. 10 (2020) 14214. https://doi.org/10.1038/s41598-020-71188-3.
[3] P.K. Panda, M.N. Arul, P. Patel, S.K. Verma, W. Luo, H.-G. Rubahn, Y.K. Mishra, M. Suar, R. Ahuja, Structure-based drug designing and immunoinformatics approach for SARS-CoV-2, Sci. Adv. 6 (2020) eabb8097. https://doi.org/10.1126/sciadv.abb8097.
[4] P. V'kovski, A. Kratzel, S. Steiner, H. Stalder, V. Thiel, Coronavirus biology and replication: implications for SARS-CoV-2, Nat. Rev. Microbiol. 19 (2021) 155–170. https://doi.org/10.1038/s41579-020-00468-6.
[5] B. Goyal, D. Goyal, Targeting the Dimerization of the Main Protease of Coronaviruses: A Potential Broad-Spectrum Therapeutic Strategy, ACS Comb. Sci. 22 (2020) 297–305. https://doi.org/10.1021/acscombsci.0c00058.
[6] J.S. Morse, T. Lalonde, S. Xu, W.R. Liu, Learning from the Past: Possible Urgent Prevention and Treatment Options for Severe Acute Respiratory Infections Caused by 2019-nCoV, Chembiochem Eur. J. Chem. Biol. 21 (2020) 730–738. https://doi.org/10.1002/cbic.202000047.

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.17.452787; this version posted July 19, 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.

[7] H.M. Mengist, T. Dilnessa, T. Jin, Structural Basis of Potential Inhibitors Targeting SARS-CoV-2 Main Protease, Front. Chem. 9 (2021). https://doi.org/10.3389/fchem.2021.622898.
[8] A. Zumla, J.F.W. Chan, E.I. Azhar, D.S.C. Hui, K.-Y. Yuen, Coronaviruses - drug discovery and therapeutic options, Nat. Rev. Drug Discov. 15 (2016) 327–347. https://doi.org/10.1038/nrd.2015.37.
[9] Rambaut, et al., Preliminary genomic characterisation of an emergent SARS-CoV-2 lineage in the UK defined by a novel set of spike mutations - SARS-CoV-2 coronavirus / nCoV-2019 Genomic Epidemiology, Virological. (2020). https://virological.org/t/preliminary-genomic-characterisation-of-an-emergent-sars-cov-2-lineage-in-the-uk-defined-by-a-novel-set-of-spike-mutations/563 (accessed June 9, 2021).
[10] H. Tegally, E. Wilkinson, M. Giovanetti, A. Iranzadeh, V. Fonseca, J. Giandhari, D. Doolabh, S. Pillay, E.J. San, N. Msomi, K. Mlisana, A. von Gottberg, S. Walaza, M. Allam, A. Ismail, T. Mohale, A.J. Glass, S. Engelbrecht, G. Van Zyl, W. Preiser, F. Petruccione, A. Sigal, D. Hardie, G. Marais, M. Hsiao, S. Korsman, M.-A. Davies, L. Tyers, I. Mudau, D. York, C. Maslo, D. Goedhals, S. Abrahams, O. Laguda-Akingba, A. Alisoltani-Dehkordi, A. Godzik, C.K. Wibmer, B.T. Sewell, J. Lourenço, L.C.J. Alcantara, S.L.K. Pond, S. Weaver, D. Martin, R.J. Lessells, J.N. Bhiman, C. Williamson, T. de Oliveira, Emergence and rapid spread of a new severe acute respiratory syndrome-related coronavirus 2 (SARS-CoV-2) lineage with multiple spike mutations in South Africa, Epidemiology, 2020. https://doi.org/10.1101/2020.12.21.20248640.
[11] H. Tegally, E. Wilkinson, R.J. Lessells, J. Giandhari, S. Pillay, N. Msomi, K. Mlisana, J.N. Bhiman, A. von Gottberg, S. Walaza, V. Fonseca, M. Allam, A. Ismail, A.J. Glass, S. Engelbrecht, G. Van Zyl, W. Preiser, C. Williamson, F. Petruccione, A. Sigal, I. Gazy, D. Hardie, N. Hsiao, D. Martin, D. York, D. Goedhals, E.J. San, M. Giovanetti, J. Lourenço, L.C.J. Alcantara, T. de Oliveira, Sixteen novel lineages of SARS-CoV-2 in South Africa, Nat. Med. 27 (2021) 440–446. https://doi.org/10.1038/s41591-021-01255-3.
[12] Naveca, et al., SARS-CoV-2 reinfection by the new Variant of Concern (VOC) P.1 in Amazonas, Brazil - SARS-CoV-2 coronavirus / nCoV-2019 Genomic Epidemiology, Virological. (2021). https://virological.org/t/sars-cov-2-reinfection-by-the-new-variant-of-concern-voc-p-1-in-amazonas-brazil/596 (accessed June 9, 2021).
[13] E.C. Wall, M. Wu, R. Harvey, G. Kelly, S. Warchal, C. Sawyer, R. Daniels, P. Hobson, E. Hatipoglu, Y. Ngai, S. Hussain, J. Nicod, R. Goldstone, K. Ambrose, S. Hindmarsh, R. Beale, A. Riddell, S. Gamblin, M. Howell, G. Kassiotis, V. Libri, B. Williams, C. Swanton, S. Gandhi, D.L. Bauer, Neutralising antibody activity against SARS-CoV-2 VOCs B.1.617.2 and B.1.351 by BNT162b2 vaccination, The Lancet. (2021). https://doi.org/10.1016/S0140-6736(21)01290-3.
[14] M. McCallum, J. Bassi, A.D. Marco, A. Chen, A.C. Walls, J.D. Iulio, M.A. Tortorici, M.-J. Navarro, C. Silacci-Fregni, C. Saliba, M. Agostini, D. Pinto, K. Culap, S. Bianchi, S. Jaconi, E. Cameroni, J.E. Bowen, S.W. Tilles, M.S. Pizzuto, S.B. Guastalla, G. Bona, A.F. Pellanda, C. Garzoni, W.C. Van Voorhis, L.E. Rosen, G. Snell, A. Telenti, H.W. Virgin, L. Piccoli, D. Corti, D. Veesler, SARS-CoV-2 immune evasion by variant B.1.427/B.1.429, Immunology, 2021. https://doi.org/10.1101/2021.03.31.437925.
[15] E.A. Ozer, L.M. Simons, O.M. Adewumi, A.A. Fowotade, E.C. Omoruyi, J.A. Adeniji, T.J. Dean, J. Zayas, P.P. Bhimalli, M.K. Ash, A. Godzik, J.R. Schneider, J.I. Mamede, B.O. Taiwo, J.F. Hultquist, R. Lorenzo-Redondo, Coincident rapid expansion of two SARS-CoV-2 lineages with enhanced infectivity in Nigeria, MedRxiv. (2021) 2021.04.09.21255206. https://doi.org/10.1101/2021.04.09.21255206.
[16] S. Elbe, G. Buckland-Merrett, Data, disease and diplomacy: GISAID's innovative contribution to global health, Glob. Chall. Hoboken NJ. 1 (2017) 33–46. https://doi.org/10.1002/gch2.1018.
[17] O. Sheik Amamuddy, G.M. Verkhivker, Ö. Tastan Bishop, Impact of Early Pandemic Stage Mutations on Molecular Dynamics of SARS-CoV-2 M^{pro}, J. Chem. Inf. Model. 60 (2020) 5080–5102. https://doi.org/10.1021/acs.jcim.0c00634.
[18] T.J. Dolinsky, J.E. Nielsen, J.A. McCammon, N.A. Baker, PDB2PQR: an automated pipeline for the setup of Poisson–Boltzmann electrostatics calculations, Nucleic Acids Res. 32 (2004) W665–W667. https://doi.org/10.1093/nar/gkh381.

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.17.452787; this version posted July 19, 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.

[19] S. Jo, T. Kim, V.G. Iyer, W. Im, CHARMM-GUI: A web-based graphical user interface for CHARMM, J. Comput. Chem. 29 (2008) 1859–1865. https://doi.org/10.1002/jcc.20945.
[20] J. Lee, X. Cheng, J.M. Swails, M.S. Yeom, P.K. Eastman, J.A. Lemkul, S. Wei, J. Buckner, J.C. Jeong, Y. Qi, S. Jo, V.S. Pande, D.A. Case, C.L. Brooks, A.D. MacKerell, J.B. Klauda, W. Im, CHARMM-GUI Input Generator for NAMD, GROMACS, AMBER, OpenMM, and CHARMM/OpenMM Simulations Using the CHARMM36 Additive Force Field, J. Chem. Theory Comput. 12 (2016) 405–413. https://doi.org/10.1021/acs.jctc.5b00935.
[21] J. Lee, M. Hitzenberger, M. Rieger, N.R. Kern, M. Zacharias, W. Im, CHARMM-GUI supports the Amber force fields, J. Chem. Phys. 153 (2020) 035103. https://doi.org/10.1063/5.0012280.
[22] I. Bahar, T.R. Lezon, L.-W. Yang, E. Eyal, Global Dynamics of Proteins: Bridging Between Structure and Function, Annu. Rev. Biophys. 39 (2010) 23–42. https://doi.org/10.1146/annurev.biophys.093008.131258.
[23] J.M. Krieger, P. Doruker, A.L. Scott, D. Perahia, I. Bahar, Towards gaining sight of multiscale events: utilizing network models and normal modes in hybrid methods, Curr. Opin. Struct. Biol. 64 (2020) 34–41. https://doi.org/10.1016/j.sbi.2020.05.013.
[24] D. Perahia, L. Mouawad, Computation of low-frequency normal modes in macromolecules: Improvements to the method of diagonalization in a mixed basis and application to hemoglobin, Comput. Chem. 19 (1995) 241–246. https://doi.org/10.1016/0097-8485(95)00011-G.
[25] B.R. Brooks, R.E. Bruccoleri, B.D. Olafson, D.J. States, S. Swaminathan, M. Karplus, CHARMM: A program for macromolecular energy, minimization, and dynamics calculations, J. Comput. Chem. 4 (1983) 187–217. https://doi.org/10.1002/jcc.540040211.
[26] Brooks et al., CHARMM: The biomolecular simulation program - Brooks - 2009 - Journal of Computational Chemistry - Wiley Online Library, (n.d.). https://onlinelibrary.wiley.com/doi/full/10.1002/jcc.21287 (accessed June 9, 2021).
[27] J. Huang, S. Rauscher, G. Nawrocki, T. Ran, M. Feig, B.L. de Groot, H. Grubmüller, A.D. MacKerell, CHARMM36m: an improved force field for folded and intrinsically disordered proteins, Nat. Methods. 14 (2017) 71–73. https://doi.org/10.1038/nmeth.4067.
[28] P.R. Batista, G. Pandey, P.G. Pascutti, P.M. Bisch, D. Perahia, C.H. Robert, Free Energy Profiles along Consensus Normal Modes Provide Insight into HIV-1 Protease Flap Opening, J. Chem. Theory Comput. 7 (2011) 2348–2352. https://doi.org/10.1021/ct200237u.
[29] C.A. Ramos-Guzmán, J.J. Ruiz-Pernía, I. Tuñón, Unraveling the SARS-CoV-2 Main Protease Mechanism Using Multiscale Methods, ACS Catal. 10 (2020) 12544–12554. https://doi.org/10.1021/acscatal.0c03420.
[30] M. Louet, D. Perahia, J. Martinez, N. Floquet, A concerted mechanism for opening the GDP binding pocket and release of the nucleotide in hetero-trimeric G-proteins, J. Mol. Biol. 411 (2011) 298–312. https://doi.org/10.1016/j.jmb.2011.05.034.
[31] P.T. Resende-Lara, D. Perahia, A.L. Scott, A.S.K. Braz, Unveiling functional motions based on point mutations in biased signaling systems: A normal mode study on nerve growth factor bound to TrkA, PLOS ONE. 15 (2020) e0231542. https://doi.org/10.1371/journal.pone.0231542.
[32] T. Hothorn, F. Bretz, P. Westfall, Simultaneous inference in general parametric models, Biom. J. Biom. Z. 50 (2008) 346–363. https://doi.org/10.1002/bimj.200810425.
[33] R Core Team, R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria., (2020). https://www.R-project.org/ (accessed July 13, 2021).
[34] W. McKinney, Data Structures for Statistical Computing in Python, in: Austin, Texas, 2010: pp. 56–61. https://doi.org/10.25080/Majora-92bf1922-00a.
[35] J. Reback, jbrockmendel, W. McKinney, J.V. den Bossche, T. Augspurger, P. Cloud, S. Hawkins, gfyoung, Sinhrks, M. Roeschke, A. Klein, T. Petersen, J. Tratner, C. She, W. Ayd, P. Hoefler, S. Naveh, M. Garcia, J. Schendel, A. Hayden, D. Saxton, M.E. Gorelli, R. Shadrach, V. Jancauskas, A. McMaster, F. Li, P. Battiston, S. Seabold, attack68, K. Dong, pandas-dev/pandas: Pandas 1.3.0, Zenodo, 2021. https://doi.org/10.5281/zenodo.5060318.
[36] C.R. Harris, K.J. Millman, S.J. van der Walt, R. Gommers, P. Virtanen, D. Cournapeau, E. Wieser, J. Taylor, S. Berg, N.J. Smith, R. Kern, M. Picus, S. Hoyer, M.H. van Kerkwijk, M. Brett, A. Haldane, J.F. del Río, M. Wiebe, P. Peterson, P. Gérard-Marchant, K. Sheppard, T. Reddy, W.

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.17.452787; this version posted July 19, 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.

Weckesser, H. Abbasi, C. Gohlke, T.E. Oliphant, Array programming with NumPy, Nature. 585 (2020) 357–362. https://doi.org/10.1038/s41586-020-2649-2.
[37] J.D. Hunter, Matplotlib: A 2D Graphics Environment, Comput. Sci. Eng. 9 (2007) 90–95. https://doi.org/10.1109/MCSE.2007.55.
[38] M. Waskom, seaborn: statistical data visualization, J. Open Source Softw. 6 (2021) 3021. https://doi.org/10.21105/joss.03021.
[39] E. Krissinel, K. Henrick, Inference of Macromolecular Assemblies from Crystalline State, J. Mol. Biol. 372 (2007) 774–797. https://doi.org/10.1016/j.jmb.2007.05.022.
[40] V.G. Contessoto, V.M. de Oliveira, B.R. Fernandes, G.G. Slade, V.B.P. Leite, TKSA-MC: A web server for rational mutation through the optimization of protein charge interactions, Proteins. 86 (2018) 1184–1188. https://doi.org/10.1002/prot.25599.
[41] H. Chen, P. Wei, C. Huang, L. Tan, Y. Liu, L. Lai, Only One Protomer Is Active in the Dimer of SARS 3C-like Proteinase*, J. Biol. Chem. 281 (2006) 13894–13898. https://doi.org/10.1074/jbc.M510745200.
[42] A.S.K. Braz, P. Tufanetto, D. Perahia, L.P.B. Scott, Relation between flexibility and positively selected HIV-1 protease mutants against inhibitors, Proteins Struct. Funct. Bioinforma. 80 (2012) 2680–2691. https://doi.org/10.1002/prot.24151.
[43] S.Q. Pantaleão, E.A. Philot, P.T. de Resende-Lara, A.N. Lima, D. Perahia, M.A. Miteva, A.L. Scott, K.M. Honorio, Structural Dynamics of DPP-4 and Its Influence on the Projection of Bioactive Ligands, Mol. J. Synth. Chem. Nat. Prod. Chem. 23 (2018) 490. https://doi.org/10.3390/molecules23020490.
[44] H. Yang, M. Yang, Y. Ding, Y. Liu, Z. Lou, Z. Zhou, L. Sun, L. Mo, S. Ye, H. Pang, G.F. Gao, K. Anand, M. Bartlam, R. Hilgenfeld, Z. Rao, The crystal structures of severe acute respiratory syndrome virus main protease and its complex with an inhibitor, Proc. Natl. Acad. Sci. 100 (2003) 13190–13195. https://doi.org/10.1073/pnas.1835675100.
[45] W.-C. Hsu, H.-C. Chang, C.-Y. Chou, P.-J. Tsai, P.-I. Lin, G.-G. Chang, Critical Assessment of Important Regions in the Subunit Association and Catalytic Action of the Severe Acute Respiratory Syndrome Coronavirus Main Protease, J. Biol. Chem. 280 (2005) 22741–22748. https://doi.org/10.1074/jbc.M502556200.
[46] H. Yang, W. Xie, X. Xue, K. Yang, J. Ma, W. Liang, Q. Zhao, Z. Zhou, D. Pei, J. Ziebuhr, R. Hilgenfeld, K.Y. Yuen, L. Wong, G. Gao, S. Chen, Z. Chen, D. Ma, M. Bartlam, Z. Rao, Design of Wide-Spectrum Inhibitors Targeting Coronavirus Main Proteases, PLOS Biol. 3 (2005) e324. https://doi.org/10.1371/journal.pbio.0030324.
[47] T.-W. Lee, M.M. Cherney, C. Huitema, J. Liu, K.E. James, J.C. Powers, L.D. Eltis, M.N.G. James, Crystal Structures of the Main Peptidase from the SARS Coronavirus Inhibited by a Substrate-like Aza-peptide Epoxide, J. Mol. Biol. 353 (2005) 1137–1151. https://doi.org/10.1016/j.jmb.2005.09.004.
[48] T. Xu, A. Ooi, H.C. Lee, R. Wilmouth, D.X. Liu, J. Lescar, Structure of the SARS coronavirus main proteinase as an active C2 crystallographic dimer, Acta Crystallograph. Sect. F Struct. Biol. Cryst. Commun. 61 (2005) 964–966. https://doi.org/10.1107/S1744309105033257.
[49] T. Hu, Y. Zhang, L. Li, K. Wang, S. Chen, J. Chen, J. Ding, H. Jiang, X. Shen, Two adjacent mutations on the dimer interface of SARS coronavirus 3C-like protease cause different conformational changes in crystal structure, Virology. 388 (2009) 324–334. https://doi.org/10.1016/j.virol.2009.03.034.
[50] L. Ding, X.-X. Zhang, P. Wei, K. Fan, L. Lai, The interaction between severe acute respiratory syndrome coronavirus 3C-like proteinase and a dimeric inhibitor by capillary electrophoresis, Anal. Biochem. 343 (2005) 159–165. https://doi.org/10.1016/j.ab.2005.04.027.
[51] U. Bacha, J. Barrila, A. Velazquez-Campoy, S.A. Leavitt, E. Freire, Identification of novel inhibitors of the SARS coronavirus main protease 3CLpro, Biochemistry. 43 (2004) 4906–4912. https://doi.org/10.1021/bi0361766.
[52] C.-Y. Chou, H.-C. Chang, W.-C. Hsu, T.-Z. Lin, C.-H. Lin, G.-G. Chang, Quaternary structure of the severe acute respiratory syndrome (SARS) coronavirus main protease, Biochemistry. 43 (2004) 14958–14970. https://doi.org/10.1021/bi0490237.
[53] S. Chen, J. Zhang, T. Hu, K. Chen, H. Jiang, X. Shen, Residues on the Dimer Interface of SARS Coronavirus 3C-like Protease: Dimer Stability Characterization and Enzyme Catalytic

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.17.452787; this version posted July 19, 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 Analysis, J. Biochem. (Tokyo). 143 (2008) 525–536. https://doi.org/10.1093/jb/mvm246.
[54] S. Chen, T. Hu, J. Zhang, J. Chen, K. Chen, J. Ding, H. Jiang, X. Shen, Mutation of Gly-11 on the Dimer Interface Results in the Complete Crystallographic Dimer Dissociation of Severe Acute Respiratory Syndrome Coronavirus 3C-like Protease, J. Biol. Chem. 283 (2008) 554–564. https://doi.org/10.1074/jbc.M705240200.
[55] J. Shi, J. Sivaraman, J. Song, Mechanism for Controlling the Dimer-Monomer Switch and Coupling Dimerization to Catalysis of the Severe Acute Respiratory Syndrome Coronavirus 3C-Like Protease, J. Virol. 82 (2008) 4620–4629. https://doi.org/10.1128/JVI.02680-07.
[56] N. Zhong, S. Zhang, P. Zou, J. Chen, X. Kang, Z. Li, C. Liang, C. Jin, B. Xia, Without Its N-Finger, the Main Protease of Severe Acute Respiratory Syndrome Coronavirus Can Form a Novel Dimer through Its C-Terminal Domain, J. Virol. 82 (2008) 4227–4234. https://doi.org/10.1128/JVI.02612-07.

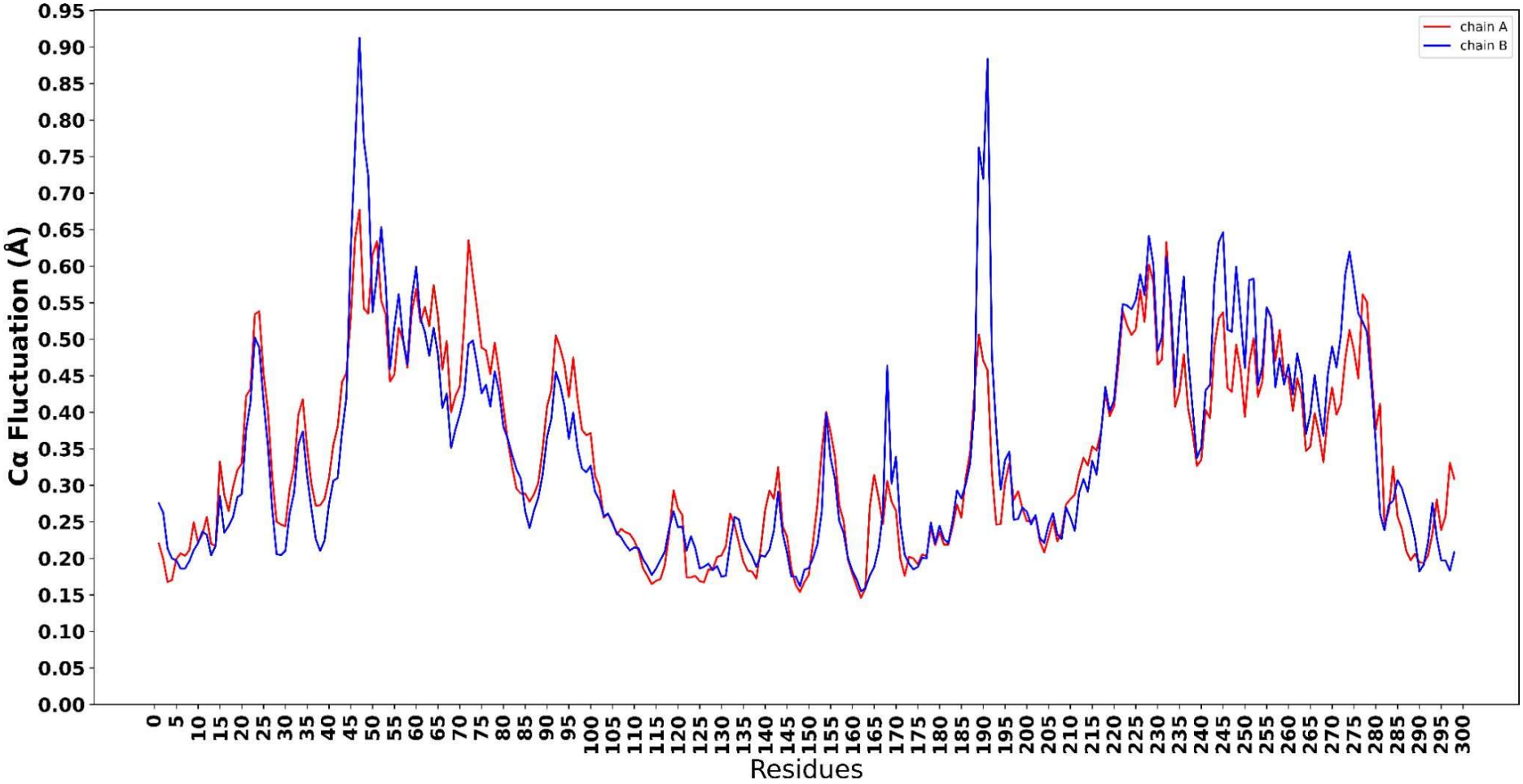


Figure S1: WT SARS-CoV-2 M^{pro} Cα fluctuation obtained by NMA. Chain A in red and chain B in blue.

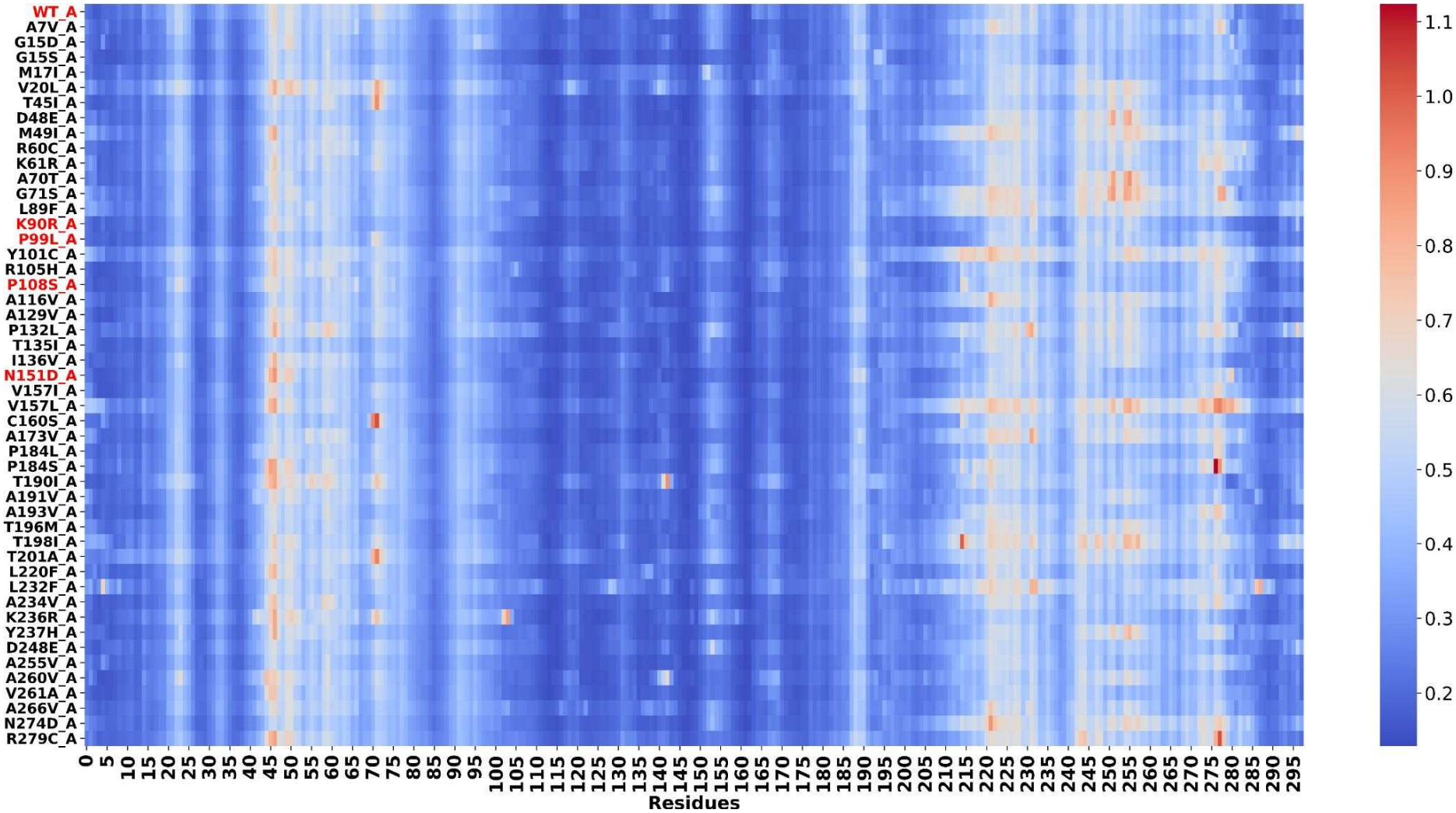


Figure S2: Heat map for Cα fluctuation of chain A of the WT and 48 mutants of SARS-CoV-2 M^{pro}.

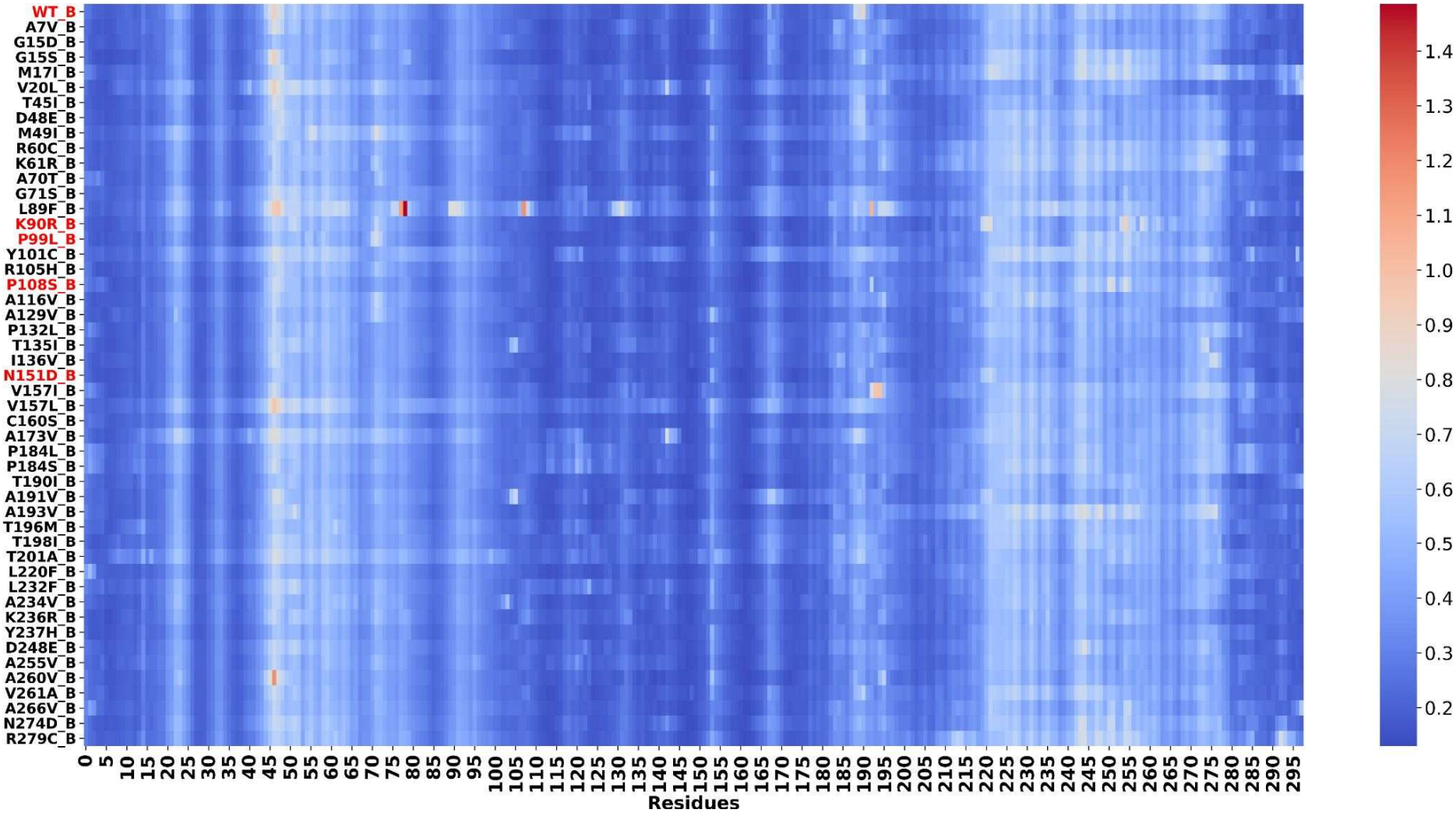


Figure S3: Heat map for Cα fluctuation of chain B of the WT and 48 mutants of SARS-CoV-2 M^{pro}.

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.17.452787; this version posted July 19, 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.

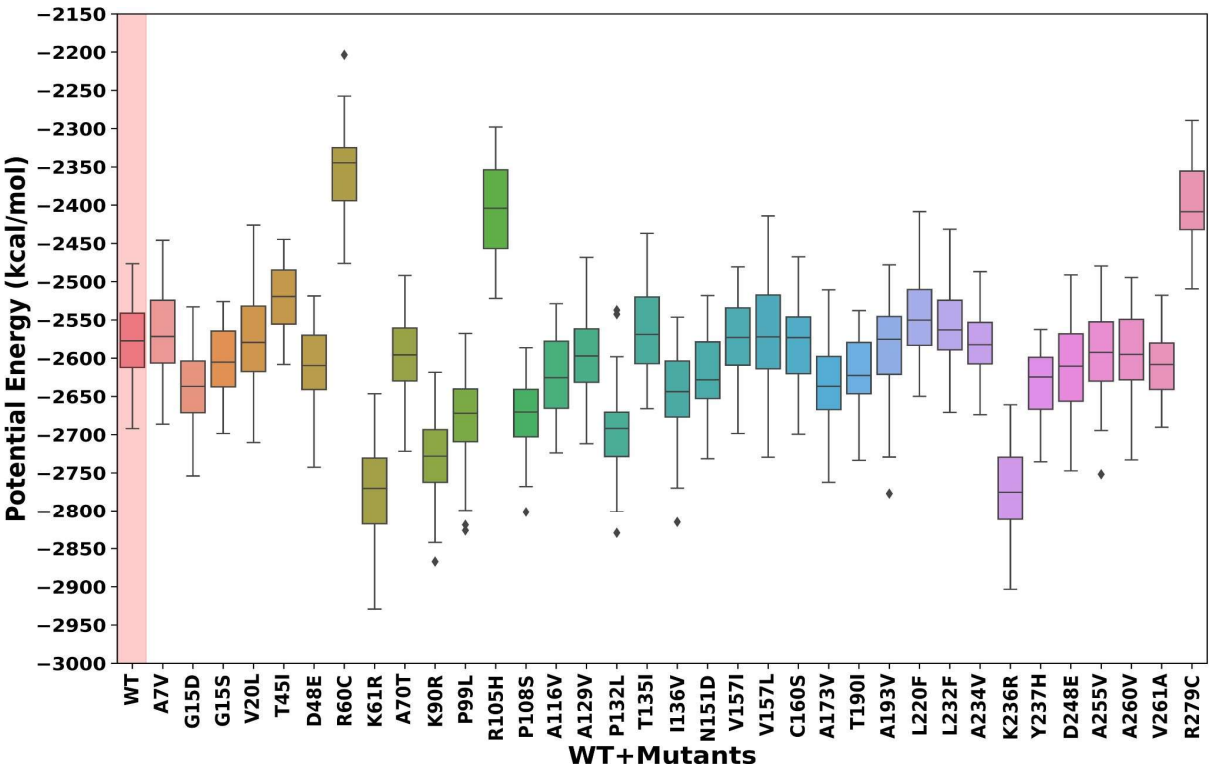


Figure S4: PotentialEnergy Boxplot for the structures generated by VMOD.The WT SARS-CoV-2 M^{pro} measure is highlighted by a red rectangle.

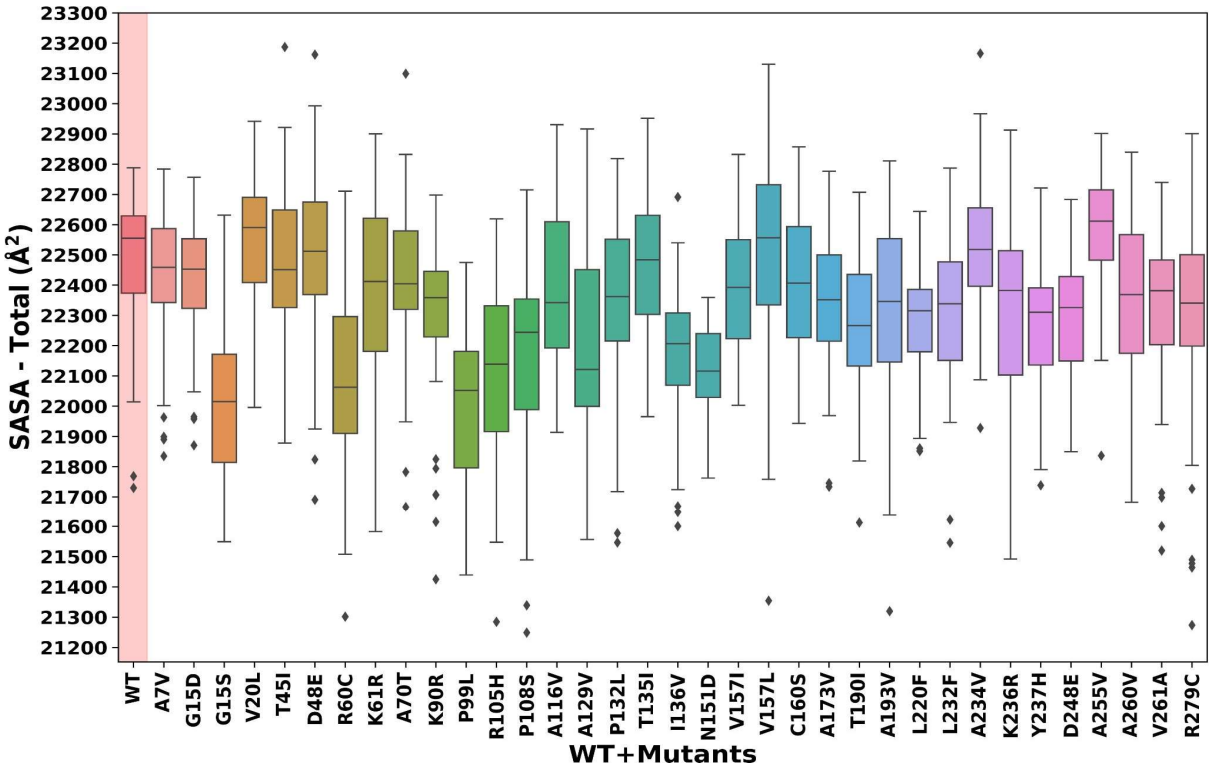


Figure S5: SASA Total Boxplot for the structures generated by VMOD. The WT SARS-CoV-2 M^{pro} measure is highlighted by a red rectangle.

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.17.452787; this version posted July 19, 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.

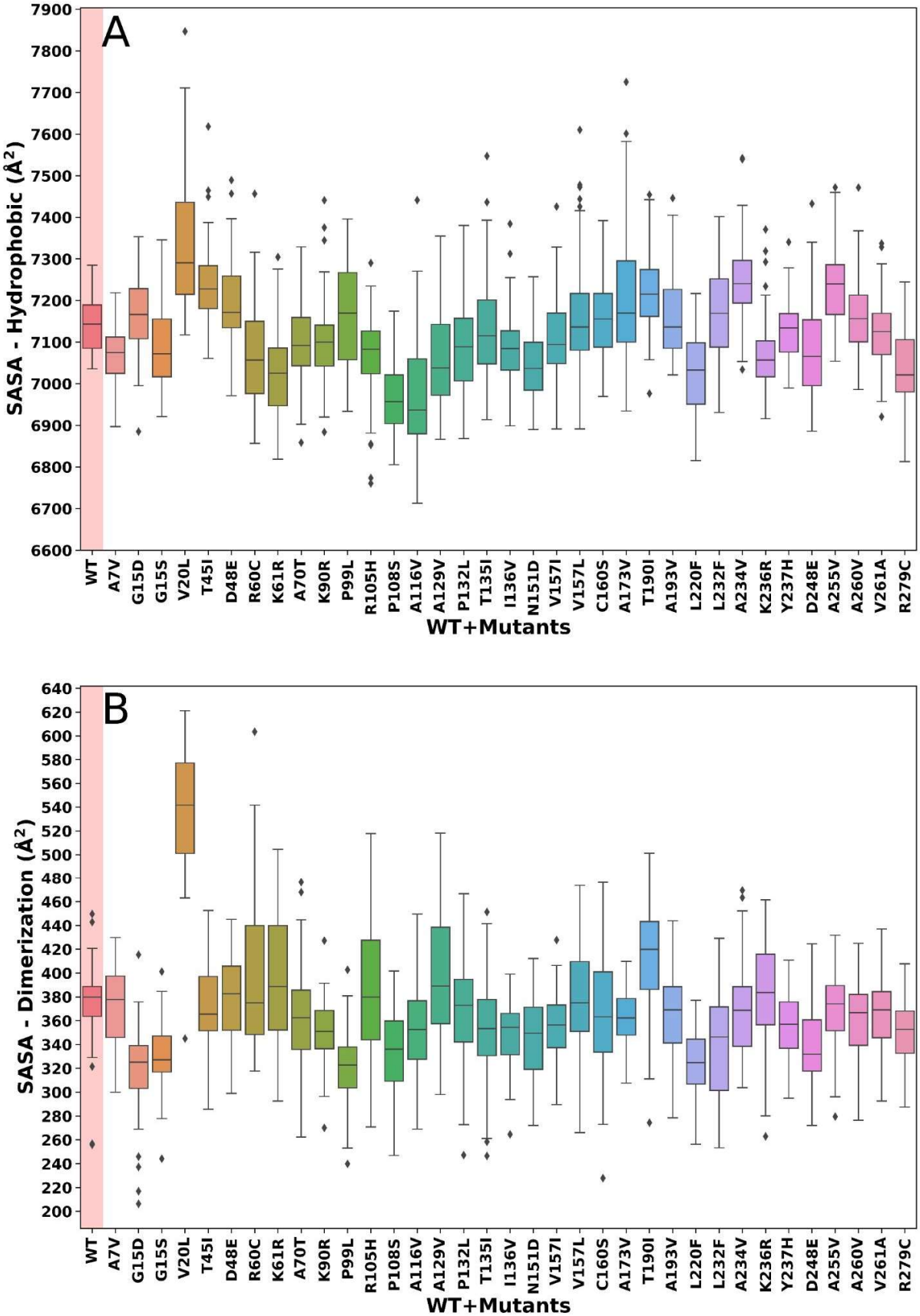


Figure S6: (A) Hydrophobic residues SASAand **(B)** Dimeric interface residues SASA Boxplot for the structures generated by VMOD. The residues involved in the dimer interface are listed in Table 1. The WT SARS-CoV-2 M^{pro} measure is highlighted by a red rectangle.

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.17.452787; this version posted July 19, 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.

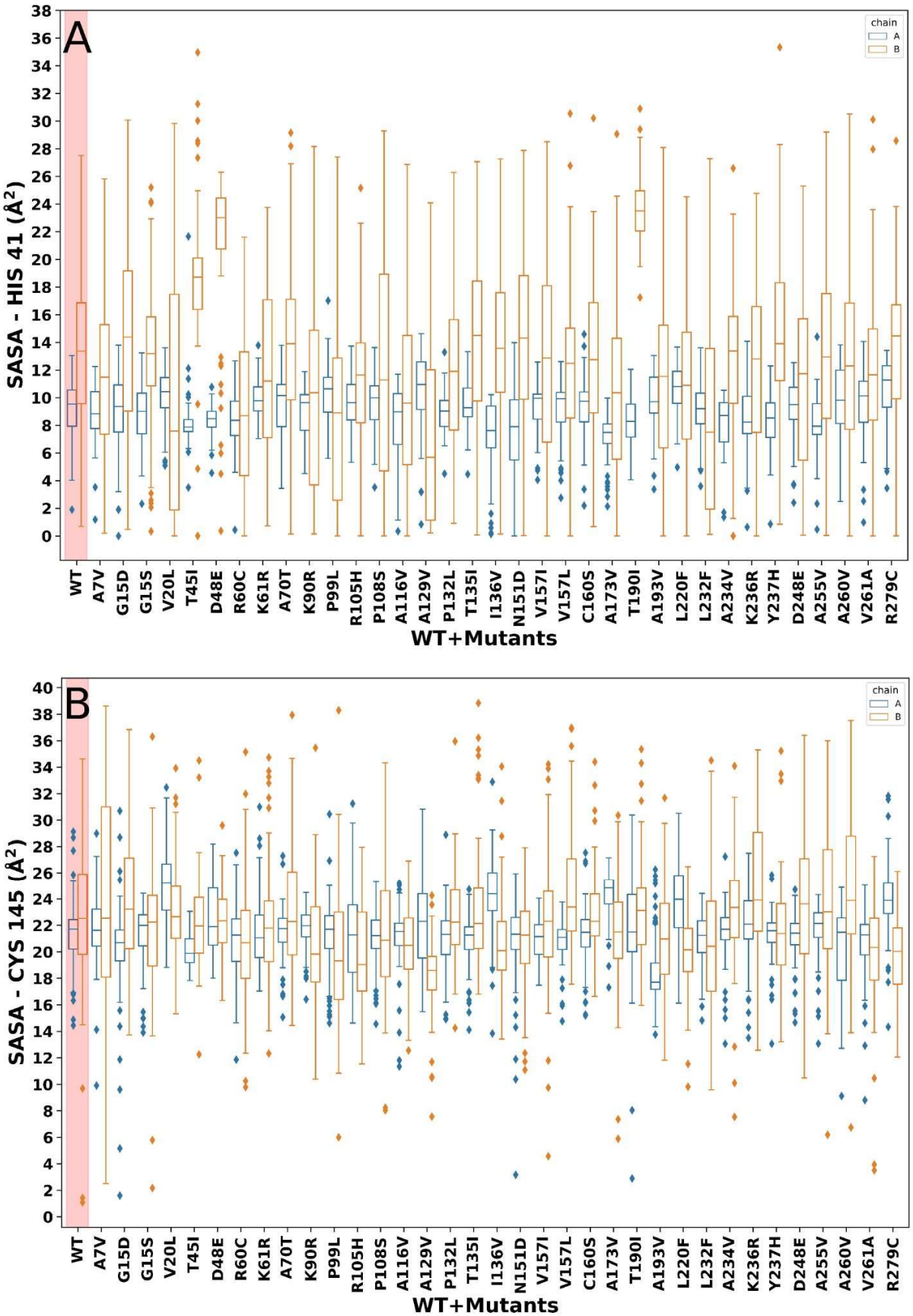


Figure S7: **(A)** SASA of His 41 and **(B)** SASA of Cys 145 catalytic dyad residues by chain. Chain A in blue and chain B in orange. The WT SARS-CoV-2 M^{pro} measures are highlighted by a red rectangle.

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.17.452787; this version posted July 19, 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.

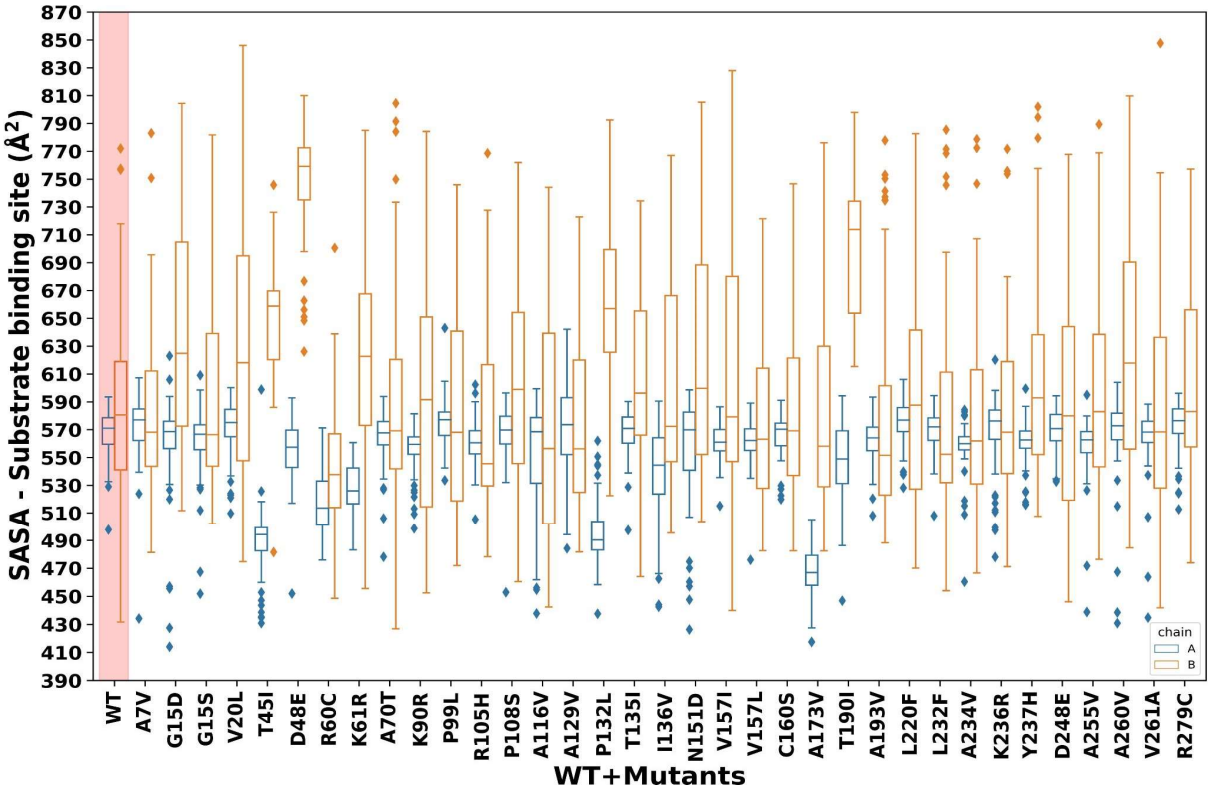


Figure S8: SASA of substrate binding sites presented by chain. Chain A in blue and chain B in orange. The WT SARS-CoV-2 M^{pro} measure is highlighted by a red rectangle.

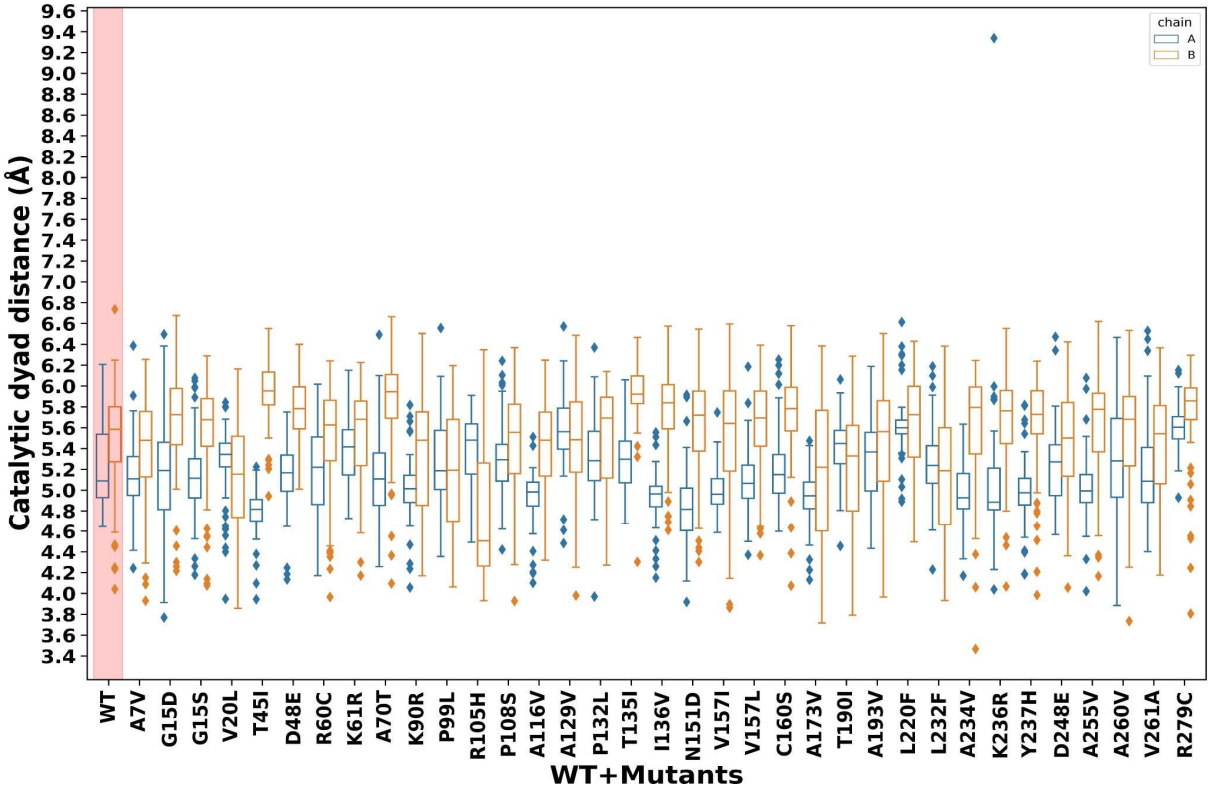


Figure S9: Catalytic Distance Distance Boxplot by chain. Distance between the Cys 145 Sγ atom and His 41 Nε atom for both chains of the structures generated by VMOD. Chain A in blue and chain B in orange. The WT SARS-CoV-2 M^{pro} measures are highlighted by a red rectangle.

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.17.452787; this version posted July 19, 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.

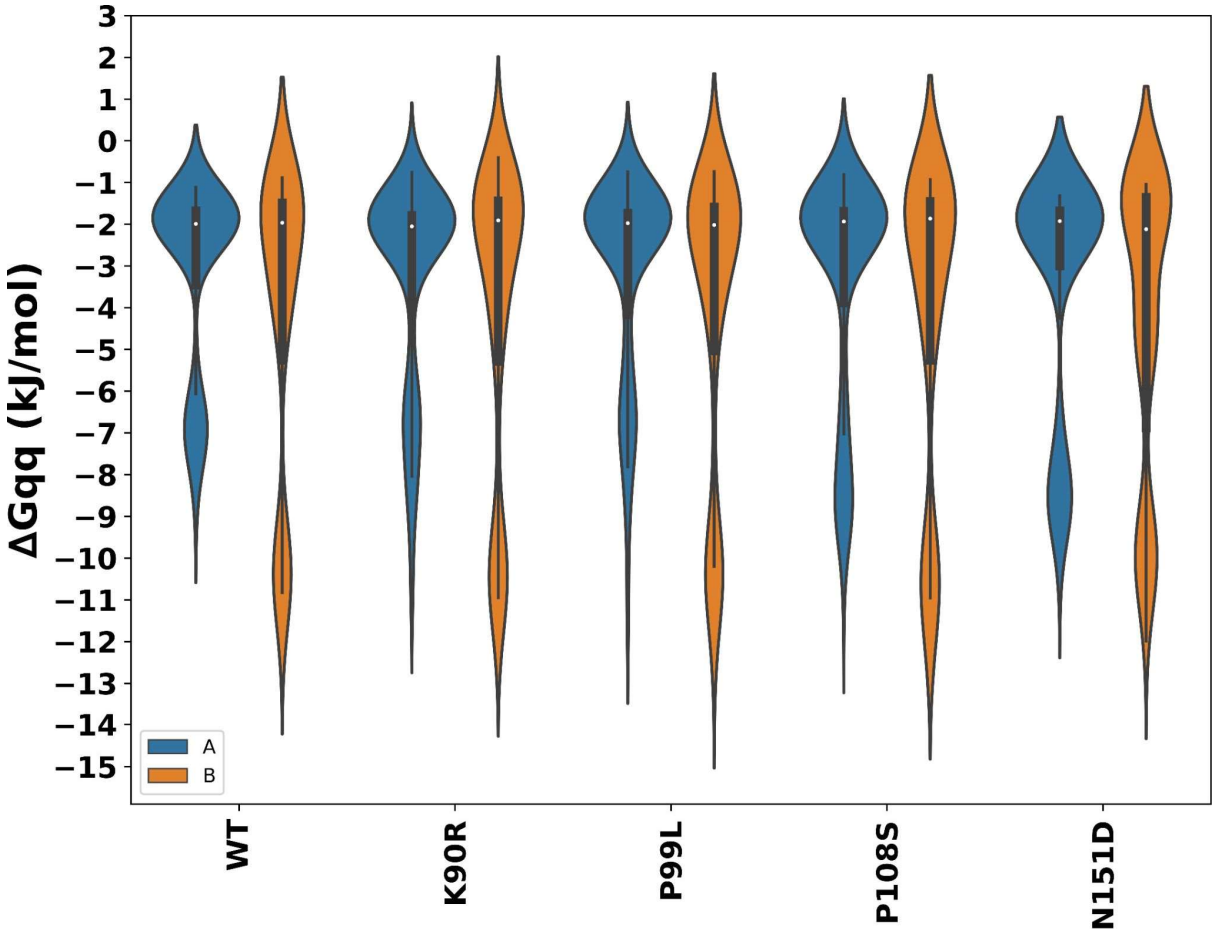


Figure S10: TKSA-MC ΔGqq (kJ/mol) of the interface residues Arg4, Glu14, Glu 290 and Arg 298 for WT and four mutants SARS-CoV-2 M^{pro}.

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.17.452787; this version posted July 19, 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.

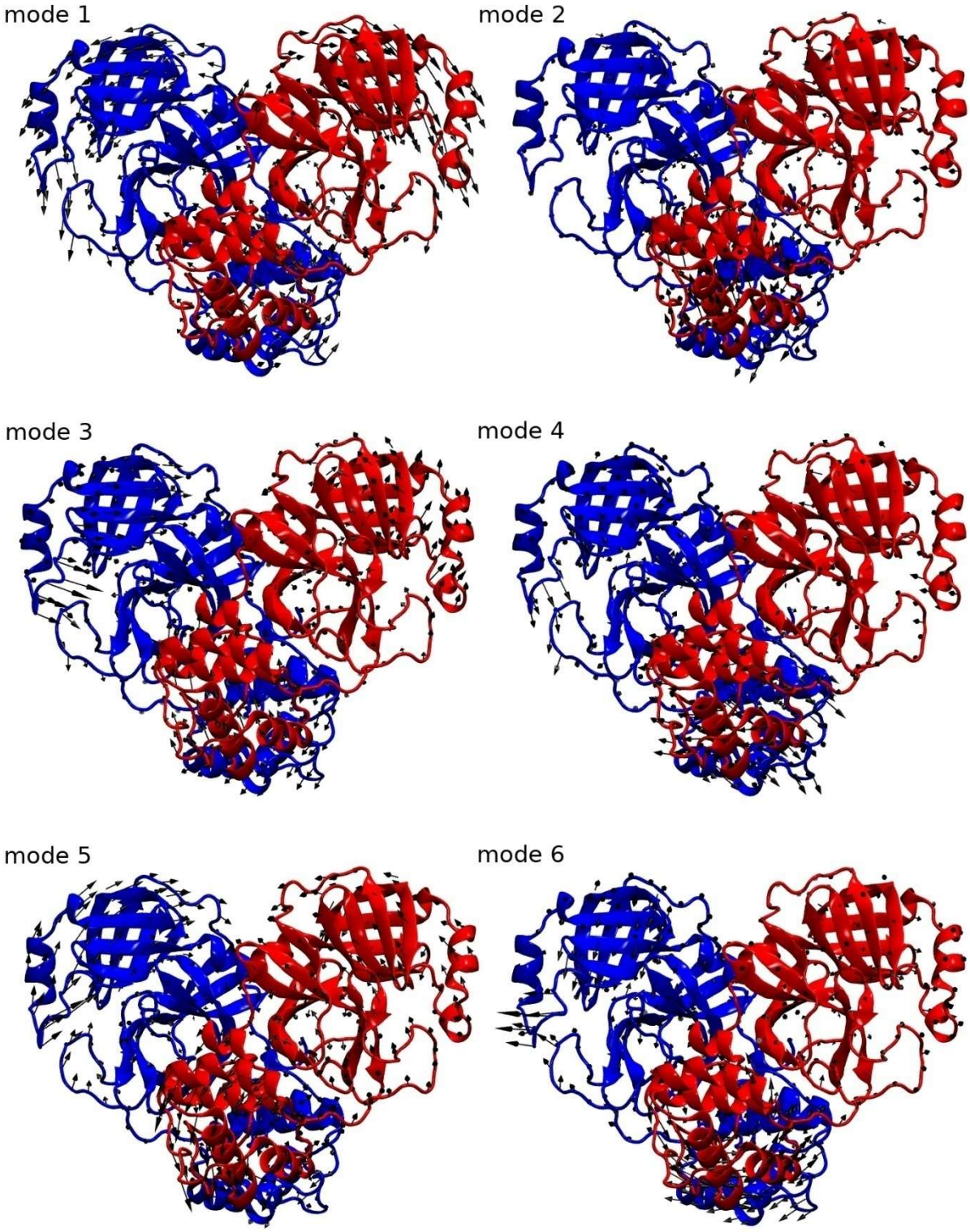


Figure S11: Representation of WT SARS-CoV-2 M^{pro} normal modes. The directions and the amplitude of the movements are represented by arrows. Figure generated in VMD [1].

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.17.452787; this version posted July 19, 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 S1: M^{pro} catalytic dyad distance (Å) for some dimeric apo structures of SARS-CoV-1 and SARS-CoV-2 available in PDB. Distance measured in PyMOL.

PDB ID	**Virus**	**chain A**	**chain B**
6WTM [2]	SARS-CoV-2	3.6	3.7
7ALI [3]	SARS-CoV-2	3.7	3.7
7C2Y [4]	SARS-CoV-2	3.7	3.9
7C2Q [5]	SARS-CoV-2	3.9	4.0
7LKD [6]	SARS-CoV-2	3.5	3.7
7CAM [7]	SARS-CoV-2	3.6	3.8
7MBG [8]	SARS-CoV-2	3.7	3.8
7CWC [9]	SARS-CoV-2	4.1	4.3
7KFI [8]	SARS-CoV-2	3.8	3.8
7BB2 [10]	SARS-CoV-2	3.8	3.8
6XHU [11]	SARS-CoV-2	3.8	3.8
3VB3 [12]	SARS-CoV-1	3.9	3.8
2DUC [13]	SARS-CoV-1	3.6	4.1
1UJ1 [14]	SARS-CoV-1	3.9	3.8
1UK3 [14]	SARS-CoV-1	3.9	3.8
1UK2 [14]	SARS-CoV-1	3.8	3.4
1Q2W [15]	SARS-CoV-1	3.8	3.9
2GT7 [16]	SARS-CoV-1	3.5	3.7

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.17.452787; this version posted July 19, 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 S2: Results from Mantel teste showing the modes of WT SARS-CoV-2 M^{pro} and their matching with the modes of the four mutants.

	Mutants			
Wild-Type	**K90R**	**P99L**	**P108S**	**N151D**
mode 1	no matching	no matching	no matching	no matching
mode 2	no matching	no matching	no matching	no matching
mode 3	no matching	mode 4	no matching	mode 2
mode 4	modes 2 and 4	mode 4	mode 4	mode 5
mode 5	mode 3	mode 5	no matching	no matching
mode 6	no matching	no matching	no matching	no matching

References:

[1] W. Humphrey, A. Dalke, K. Schulten, VMD: Visual molecular dynamics, J. Mol. Graph. 14 (1996) 33–38. https://doi.org/10.1016/0263-7855(96)00018-5.
[2] W. Vuong, M.B. Khan, C. Fischer, E. Arutyunova, T. Lamer, J. Shields, H.A. Saffran, R.T. McKay, M.J. van Belkum, M.A. Joyce, H.S. Young, D.L. Tyrrell, J.C. Vederas, M.J. Lemieux, Feline coronavirus drug inhibits the main protease of SARS-CoV-2 and blocks virus replication, Nat. Commun. 11 (2020) 4282. https://doi.org/10.1038/s41467-020-18096-2.
[3] R.P.D. Bank, RCSB PDB - 7ALI: Crystal structure of the main protease (3CLpro/Mpro) of SARS-CoV-2 at 1.65A resolution (spacegroup P2(1))., (n.d.). https://www.rcsb.org/structure/7ali (accessed July 16, 2021).
[4] R.P.D. Bank, RCSB PDB - 7C2Y: The crystal structure of COVID-2019 main protease in the apo state, (n.d.). https://www.rcsb.org/structure/7c2y (accessed July 16, 2021).
[5] X. Zhou, F. Zhong, C. Lin, X. Hu, Y. Zhang, B. Xiong, X. Yin, J. Fu, W. He, J. Duan, Y. Fu, H. Zhou, P.J. McCormick, Q. Wang, J. Li, J. Zhang, Structure of SARS-CoV-2 main protease in the apo state, Sci. China Life Sci. 64 (2021) 656–659. https://doi.org/10.1007/s11427-020-1791-3.
[6] R.P.D. Bank, RCSB PDB - 7LKD: X-ray crystal structure of the SARS-CoV-2 main protease in space group P21., (n.d.). https://www.rcsb.org/structure/7lkd (accessed July 16, 2021).
[7] Y.-C. Wang, W.-H. Yang, C.-S. Yang, M.-H. Hou, C.-L. Tsai, Y.-Z. Chou, M.-C. Hung, Y. Chen, Structural basis of SARS-CoV-2 main protease inhibition by a broad-spectrum anti-

bioRxiv preprint doi: https://doi.org/10.1101/2021.07.17.452787; this version posted July 19, 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.

coronaviral drug, Am. J. Cancer Res. 10 (2020) 2535–2545.
[8] G.D. Noske, A.M. Nakamura, V.O. Gawriljuk, R.S. Fernandes, G.M.A. Lima, H.V.D. Rosa, H.D. Pereira, A.C.M. Zeri, A.F.Z. Nascimento, M.C.L.C. Freire, D. Fearon, A. Douangamath, F. von Delft, G. Oliva, A.S. Godoy, A Crystallographic Snapshot of SARS-CoV-2 Main Protease Maturation Process, J. Mol. Biol. 433 (2021) 167118. https://doi.org/10.1016/j.jmb.2021.167118.
[9] R.P.D. Bank, RCSB PDB - 7CWC: Ambient-Temperature Serial Femtosecond X-ray Crystal structure of SARS-CoV-2 Main Protease at 2.1 A Resolution (P212121), (n.d.). https://www.rcsb.org/structure/7cwc (accessed July 16, 2021).
[10] RCSB PDB - 7BB2: Crystal structure of the main protease (3CLpro/Mpro) of SARS-CoV-2 at 1.6A resolution (spacegroup P2(1)2(1)2(1)), (n.d.). https://www.rcsb.org/structure/7bb2 (accessed July 16, 2021).
[11] D.W. Kneller, G. Phillips, H.M. O'Neill, K. Tan, A. Joachimiak, L. Coates, A. Kovalevsky, Room-temperature X-ray crystallography reveals the oxidation and reactivity of cysteine residues in SARS-CoV-2 3CL Mpro: insights into enzyme mechanism and drug design, IUCrJ. 7 (2020) 1028–1035. https://doi.org/10.1107/S2052252520012634.
[12] C.-P. Chuck, C. Chen, Z. Ke, D. Chi-Cheong Wan, H.-F. Chow, K.-B. Wong, Design, synthesis and crystallographic analysis of nitrile-based broad-spectrum peptidomimetic inhibitors for coronavirus 3C-like proteases, Eur. J. Med. Chem. 59 (2013) 1–6. https://doi.org/10.1016/j.ejmech.2012.10.053.
[13] T. Muramatsu, C. Takemoto, Y.-T. Kim, H. Wang, W. Nishii, T. Terada, M. Shirouzu, S. Yokoyama, SARS-CoV 3CL protease cleaves its C-terminal autoprocessing site by novel subsite cooperativity, Proc. Natl. Acad. Sci. 113 (2016) 12997–13002. https://doi.org/10.1073/pnas.1601327113.
[14] H. Yang, M. Yang, Y. Ding, Y. Liu, Z. Lou, Z. Zhou, L. Sun, L. Mo, S. Ye, H. Pang, G.F. Gao, K. Anand, M. Bartlam, R. Hilgenfeld, Z. Rao, The crystal structures of severe acute respiratory syndrome virus main protease and its complex with an inhibitor, Proc. Natl. Acad. Sci. 100 (2003) 13190–13195. https://doi.org/10.1073/pnas.1835675100.
[15] A. Pollack, Company Says It Mapped Part of SARS Virus, N. Y. Times. (2003). https://www.nytimes.com/2003/07/30/business/company-says-it-mapped-part-of-sars-virus.html (accessed July 16, 2021).
[16] T.-W. Lee, M.M. Cherney, J. Liu, K.E. James, J.C. Powers, L.D. Eltis, M.N.G. James, Crystal Structures Reveal an Induced-fit Binding of a Substrate-like Aza-peptide Epoxide to SARS Coronavirus Main Peptidase, J. Mol. Biol. 366 (2007) 916–932. https://doi.org/10.1016/j.jmb.2006.11.078.