

Article

In Silico Substrate-Binding Profiling for SARS-CoV-2 Main Protease (M^{Pro}) Using Hexapeptide Substrates

Sophakama Zabo  and Kevin Alan Lobb * 

Department of Chemistry, Rhodes University, Makhanda 6139, South Africa; sophakamazabo@gmail.com
* Correspondence: k.lobb@ru.ac.za; Tel.: +27-46-6039693

Abstract: The SARS-CoV-2 main protease (M^{Pro}) is essential for the life cycle of the COVID-19 virus. It cleaves the two polyproteins at 11 positions to generate mature proteins for virion formation. The cleavage site on these polyproteins is known to be Leu-Gln↓(Ser/Ala/Gly). A range of hexapeptides that follow the known sequence for recognition and cleavage was constructed using RDKit libraries and complexed with the crystal structure of M^{Pro} (PDB ID 6XHM) through extensive molecular docking calculations. A subset of 131 of these complexes underwent 20 ns molecular dynamics simulations. The analyses of the trajectories from molecular dynamics included principal component analysis (PCA), and a method to compare PCA plots from separate trajectories was developed in terms of encoding PCA progression during the simulations. The hexapeptides formed stable complexes as expected, with reproducible molecular docking of the substrates given the extensiveness of the procedure. Only Lys-Leu-Gln*** (KLQ***) sequence complexes were studied for molecular dynamics. In this subset of complexes, the PCA analysis identified four classifications of protein motions across these sequences. KLQ*** complexes illustrated the effect of changes in substrate on the active site, with implications for understanding the substrate recognition of M^{Pro} and informing the development of small molecule inhibitors.

Keywords: SARS-CoV-2 main protease; protein substrate; multi-conformer substrate library; molecular docking; molecular dynamics; PCA



Citation: Zabo, S.; Lobb, K.A. In Silico Substrate-Binding Profiling for SARS-CoV-2 Main Protease (M^{Pro}) Using Hexapeptide Substrates. *Viruses* **2023**, *15*, 1480. <https://doi.org/10.3390/v15071480>

Academic Editor: Nicholas C. Wu

Received: 30 May 2023
Revised: 26 June 2023
Accepted: 27 June 2023
Published: 29 June 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Since its emergence in late 2019, the death toll due to Coronavirus disease (COVID-19) has reached over 6.9 million people, with more than 760 million cases recorded globally according to the WHO [1]. COVID-19 generally manifests as a pulmonary disease caused by Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) infection [2], but it also has fatal extrapulmonary manifestations [3,4]. SARS-CoV-2 belongs to a diverse family of enveloped viruses consisting of large, single-stranded, positive-sense RNA genomes of around 27–32 kb in length—typically composed of a 5'-methylguanosine cap at the beginning, a 3'-poly-A tail at the end, and a total of 6–10 genes in between [5]. The viral genome has high frequencies of genomic recombinations and mutations, as demonstrated by the existence of its variants [6,7]. The SARS-CoV-2 genome encodes at least 27 proteins, including 16 non-structural proteins (nsp1–10, nsp12–16), 4 structural proteins, and 8 accessory proteins (ORF3a, ORF3b, ORF6, ORF7a, ORF7b, ORF8, ORF9b, and ORF14) [3,8].

Among the 16 encoded nsps is the main protease (M^{Pro}), also referred to as chymotrypsin-like protease (3CL^{pro}), which is a homodimeric, 33.8 kDa cysteine protease encoded as nsp5 by the open reading frame ORF1a/b [9,10]. Each protomer of the main protease consists of catalytic dyads (HIS41 and CYS145), which cleave the polyproteins into mature proteins translated by ORF1a and ORF1ab. The main protease is first autocleaved from the polyprotein (pp1ab) and then downstream nsps (nsp4–nsp16) are cleaved at 11 different sites of the pp1ab—recognizing the sequence Leu-Gln↓(Ser/Ala/Gly) (↓ shows the cleavage site) [6,11]. This

protein is vital to the life cycle of coronaviruses, as it directly mediates the maturation of the nsps essential for viral replication [6].

With the emergence of variants with increased virulence and advanced evasive mechanisms against host immunological defences [12–14], existing pharmacological measures have lost their efficacy, necessitating the development of broad-spectrum chemotherapies. There are no vaccines that confer long-term immunity and consistently elicit immunological protection against a broad range of SARS-CoV-2 variants. Optimal pharmacological measures to control the spread and SARS-CoV-2-related morbidity have yet to be discovered and developed. Simulations on the cleavage points of pp1ab with proteases show interesting changes in stability depending on the protease complex [15]. Hence, in this study, we explored SARS-CoV-2 M^{Pro} proteolytic substrate recognition as a viable technique to gain an insight into the molecular cleavage of pp1ab substrates at the active site. It is hoped that this in turn will rationally inform the development of anti-coronavirus compounds.

Viral proteases have proven to be excellent drug targets that have led to the development of effective drugs against chronic infections. The functional importance of the main protease makes it a prudent target for antiviral drug discovery against SARS-CoV-2, as its inhibition could potentially hinder the viral replication cycle, and overall, stall the production of infectious SARS-CoV-2 virions [6,10]. Importantly, the structural architecture of the main protease is highly conserved across various coronaviruses, despite the extensive mutagenesis undergone by coronaviruses [11]. Therefore, the development of main protease inhibitors could lead to the creation of broad-spectrum antiviral therapeutic agents against SARS-CoV-2 and other coronaviruses, while reducing the risk of mutation-mediated drug resistance in future deadly viral strains [6,11]. Moreover, there are no human proteases with analogous substrate specificity to the protein and therefore, main protease inhibitors are more likely to be harmless to patients [11]. The present work hypothesized profiling the substrate binding at the catalytic site to understand the function and mechanisms of SARS-CoV-2 M^{Pro}. As per the hypothesis, a multi-conformer hexapeptide library—following the recognition and cleavage sequences for M^{Pro}—were generated and used to evaluate binding efficiency and substrate recognition and specificity against the crystallographic structure of M^{Pro}. Accordingly, the behavior of M^{Pro} in the presence of the hexapeptides was assessed using molecular dynamics simulations.

2. Materials and Methods

2.1. Creation of Multi-Conformer Hexapeptide Library

Due to the unavailability of oligopeptide libraries consisting of peptide chains with SARS-CoV-2 M^{Pro} recognition sequences, a hexapeptide (P3-P3') library was thus generated on a local server using RDKit (v. 2019.09.1) within the Python scripting environment. The constituent amino acid residues were selected from the findings of Ullrich and Nitsche [10], prioritizing high-occurrence amino acids in the consensus sequence over all cleavage sites. The peptide structures were constructed using the MolFromSequence method and subsequently capped with acetyl (ACE) and methylamine (NME) constructs at the C- and N- termini, respectively. Hydrogen atoms were removed and added to ensure each atom valency was satisfied. The three-dimensional (3D) coordinates of the atoms were generated using the EmbedMolecule method and the molecules were optimized using the implemented UFFOptimizeMolecule method. A structural conformational search was performed to generate 100 conformers per system. The resulting conformers were stored in an SDF file.

2.2. Molecular Docking

To construct the 3D M^{Pro}–hexapeptide complexed structures, molecular docking simulations were performed using AutoDock Vina [16]. The crystallographic 3D structure of SARS-CoV-2 M^{Pro} was retrieved from the RCSB Protein Data Bank (PDB) (PDB id: 6XHM). All of the co-crystallized ligands and waters were removed. The hexapeptide structures were subjected to geometry optimization employing a semi-empirical quantum

mechanical method (XTB semi-empirical method) using xTB software [17] and converted to PDB format using openBabel [18]. Gasteiger charges and atom types were assigned to the receptor, and the peptide substrates were prepared using Python scripts provided by AutoDockTools [19]. Preliminary molecular docking simulations resulted in favorable binding to protomer B over protomer A (Supplementary Figure S14). Therefore, a grid box size was set at $20 \times 20 \times 20$ Å and centered at -18.444 , -16.361 , and 7.944 in the x , y , and z directions, respectively, targeting the protomer B-binding pocket of the receptor. The energy range and exhaustiveness were set at 4 and 480, respectively. The docking simulations were performed in parallel using 24 central processing unit (CPU) cores per job at the Centre for High Performance Computing (CHPC). After the hexapeptides were docked into the binding pocket, the ligand PDBQT files were split into separate files and the poses with the lowest docking scores were prioritized for further study. The validation and reproducibility of the docking results were confirmed for the best poses per hexapeptide by means of structural superimposition, subsite mapping, and intermolecular interactions, which were explored using PyMOL [20] and BIOVIA Discovery Studio 2020 Client [21]. Respective M^{Pro}–hexapeptide complexed structures were constructed using conformers with the best pose.

2.3. Molecular Dynamics Simulations

For the molecular dynamics simulations, a focused subset of the 810 hexapeptides was used. The 20 ns molecular dynamics (MD) simulations were performed using GROMACS (version 2018.1) for SARS-CoV-2 M^{Pro} in the absence and presence of hexapeptides containing the sequence KLQ^{***}, for a total of 131 complexes. Together with the *apo* system, this accounted for 132 MD simulations. Substrates were combined as a third protein chain in the receptor pdb file. The topology and coordinate files for the *apo*–M^{Pro} and M^{Pro}–hexapeptide systems were created using the AMBER03 protein, nucleic AMBER94 force field [22]. The 132 systems were solvated in a cubic box of 10 nm dimension using the TIP3P water model and subsequently neutralized with 0.15 M NaCl. The solvated systems were minimized for 50,000 steps using the steepest descent minimization algorithm until maximum forces of <10.0 kJ/mol were achieved. After minimization, the systems were equilibrated with an NVT ensemble (constant number of atoms, volume, and temperature) at 300 K for 100 ps, followed by an NPT ensemble (constant number of atoms, pressure, and temperature) equilibration at 1.0 bar for 100 ps. A modified Berendsen thermostat algorithm was used for both equilibration steps. For production runs, the time steps were set at 2 fs, whereas the trajectory and coordinate information were saved every 10 ps, resulting in 2000 frames saved for every system simulated. The LINCS holonomic constraints algorithm was deployed to constrain all bonds, whereas the Particle Mesh Ewald algorithm was set for all long-range electrostatics. The entire procedure was automated on the CHPC cluster, using 8 nodes per job with 24 cores per node and 24 processes per node. Following completion of production runs, the periodic boundary conditions were removed and the trajectories were analyzed according to root mean square deviation (RMSD), root mean square fluctuation (RMSF), radius of gyration (Rg) and principal component analysis (PCA).

2.4. Principal Component Analysis and Protein Motion Classification

Conformational changes and structural protein motions over the course of the trajectory for the protein backbone were monitored using PCA. Due to the high volume of PCA data, a custom pairwise comparison of the molecular dynamics simulations was conducted in an attempt to classify the similarities and differences in M^{Pro} during the 132 simulations. The Cartesian coordinates of PC1 and PC2 for each system were fitted in a 5×5 grid to demarcate the protein motions as defined by PCA over the course of the simulation. In intervals of 2 ns (200 frames), the PC1 and PC2 coordinates of the protein were averaged and subsequently used to position the PCA within the grid. These averaged figures were assigned a letter code corresponding to the position within the grid of the PCA in the particular time interval. The resulting string of letter codes (for the full simulation time)

thus tracked the progression of the PCA for the respective simulation. The differences in these codes between separate molecular dynamics simulations were used to create a pairwise comparison that included all 132 PCA analyses, under the assumption that the protein motions described by PC1 and PC2 s were the same in the compared systems. The differences between these PCA components were clustered using correlation as a measure of distance. Hierarchical clustering was performed to indicate the hierarchical relationships in the PCA progression with regard to dynamic protein motions, thereby grouping all systems in terms of similar molecular dynamics. The molecular dynamics of two systems from each class/group were visually inspected using Visual Molecular Dynamics (VMD) [23] to verify protein motion similarity and dissimilarity in appropriate pairs of the 132 PCA plots.

3. Results and Discussion

3.1. The Multi-Conformer Hexapeptide Library

A total of 810 hexapeptides were generated from the high-occurrence amino acids in the consensus sequence over all cleavage sites derived from Ullrich and Nitsche [10]. Using the information from Ullrich and Nitsche [10], octapeptides were initially considered, which would have resulted in 61,000 possible substrates. That many substrates would have required computational power and resources beyond the feasibility of our study. Hexapeptides, on the other hand, provided the appropriate computational manageability and intensity to study the substrate-binding profiles of SARS-CoV-2 M^{Pro}. Since only the P3-P1' residues are required for recognition, specificity, and cleavage, working with hexapeptides (the P3-P3') provided the largest set of substrates that was still computationally manageable. The resulting library consisted of peptide chains demonstrating physicochemical diversity that allowed for broad-spectrum characterization of M^{Pro}-substrate interaction (Table 1). The combinations of high-occurrence amino acids used to create the multi-conformer library allowed for the assessment and study of potential/promising substrates that are not necessarily naturally occurring within the SARS-CoV-2 genome yet still adhere to the substrate recognition and specificity of M^{Pro}; most importantly, they possessed desirable properties that could serve as a rational basis for drug design and discovery. By doing so, this study incorporated wide-range characterization of M^{Pro} functionality (keeping in mind its genomic and structural conservation across variants and species), and in turn, provides insight into the protein's behavior in the presence of mutant and homologous forms of the SARS-CoV genome and polyprotein substrates. Such information could serve as a basis for the development of broad-spectrum chemotherapies that retain efficacy in spite of emerging mutations.

Table 1. The amino acid residues in the polyprotein cleavage sites recognized by SARS-CoV-2 M^{Pro} used in the construction of hexapeptides. The amino acids are colored following the coloring scheme of WebLogo*.

Residue Position	Amino Acids
P3	Arg; Lys; Met; Thr; Val
P2	Leu
P1	Gln
P1'	Ala; Ser
P2'	Ala; Asn; Glu; Gly; Leu; Ser; Thr; Val
P3'	Ala; Asn; Asp; Gln; Glu; Gly; Met; Phe; Val

* Green: polar; Blue: basic; Red: acidic; Black: hydrophobic; Purple: neutral.

Following the construction of each hexapeptide, terminal capping was performed to enhance the structural stability of the peptide substrates, particularly in dynamic environments. The subsequent conformational search provided an opportunity to identify the most favorable geometries prior to binding. This could have future use in the design of peptidomimetics displaying potent antiviral activity. Ultimately, docking studies were performed with 81,000 systems, comprising 100 conformers for each of the 810 hexapeptides.

3.2. Virtual High-Throughput Screening of the Hexapeptides

A quality crystal structure of SARS-CoV-2 M^{Pro} was retrieved from the RCSB PDB (PDB ID: 6XHM) showing good resolution of 1.41 Å and no mutations or missing residues. The protein was reported as encoded using the SARS-CoV-2 MN908947.3 genome and synthesized via *Escherichia coli* expression [24]. Moreover, the crystal structure provided three rotamers of M^{Pro} and this study utilized the rotamer designated “A”.

The high-throughput screening of the hexapeptides was performed on protomer B targeting the active site pocket, after preliminary docking studies assessing both protomers (Supplementary Figure S14). Overall, the best docked poses (or docked pose with minimum energy) registered binding free energies ranging between -8.7 and -7.0 kcal.mol⁻¹ across all 810 substrates. The mean binding free energies of the hexapeptides—grouped by P3-P1 residues—alongside their respective ligand efficiencies are reported in Table 2. These docking results were arguably good and alluded to high-affinity binding of the substrates onto the active site of SARS-CoV-2 M^{Pro}, as to be expected. The ligand efficiencies indicated that all of the P3-P1 residue combinations displayed similar goodness of interaction with the receptor and thus showed similar potential improvement in binding affinity for structural- and efficiency-driven drug design [25–27]. Therefore, all of these P3-P1 residue combinations suggested a viable basis for characterization of the active site for rational antiviral drug design, given their ligand efficiency (Table 2).

Table 2. The docking scores and ligand efficiencies of the hexapeptide substrates docked onto SARS-CoV-2 M^{Pro} on the basis of P3-P1 variable amino acid residues.

Sequence (P3-P1)	Docking Score (kcal.mol ⁻¹)	HA *	LE ** (kcal.mol ⁻¹ /HA)
Lys-Leu-Gln	-8.0 ± 0.26	28	-0.3 ± 0.01
Met-Leu-Gln	-7.9 ± 0.30	27	-0.3 ± 0.01
Arg-Leu-Gln	-7.9 ± 0.33	30	-0.3 ± 0.01
Thr-Leu-Gln	-7.9 ± 0.26	26	-0.3 ± 0.01
Val-Leu-Gln	-8.0 ± 0.23	26	-0.3 ± 0.01

* HA: heavy atoms (non-hydrogen atoms); ** LE: ligand efficiency.

Further evaluation and validation of the docking results revealed the reproducibility of the best docking pose together with conducive binding interactions with the active site residues for these particular poses. The visualization of the best poses of conformers per hexapeptide showed overlap in the backbone (α -carbons) of the peptide substrates when superimposed, with only a few substrates displaying alternative poses with minimum energy (Supplementary Figure S1). This was an impressive feat considering the flexible nature of the peptide macromolecules. Figure 1A shows the superimposed poses with minimum binding free energy occupying the same docking/binding space. The use of a high degree of search exhaustiveness (of 480) substantially increased the probability of finding plausible poses that were reproducible.

Mapping of the substrates in the binding pocket of M^{Pro} was used to visually assess the substrate recognition in accordance with the nomenclature of Schechter and Berger [28] and Ullrich and Nitsche [10]. Ullrich and Nitsche [10] stipulated that SARS-CoV-2 M^{Pro} mainly recognizes substrate residues P4-P1'; however, substrate specificity is determined by residues P2-P1', as demonstrated by the highest degree of conservation across the pp1a/ab cleavage sites. The substrate residues were recognized and anchored within the binding pocket by specific active site residues that comprised subsites. Table 3 lists these subsites as well as their constituent active site residues. These subsites surrounded the catalytic dyad consisting of His41 and Cys145, which mediates digestion of the polyproteins [10]. The key active site residues that essentially mediate substrate binding and processing include His41, Met49, Gly143, Ser144, His163, His164, Met165, Glu166, Leu167, Asp187, Arg188, Gln189, Thr190, Ala191, and Gln192 [11].

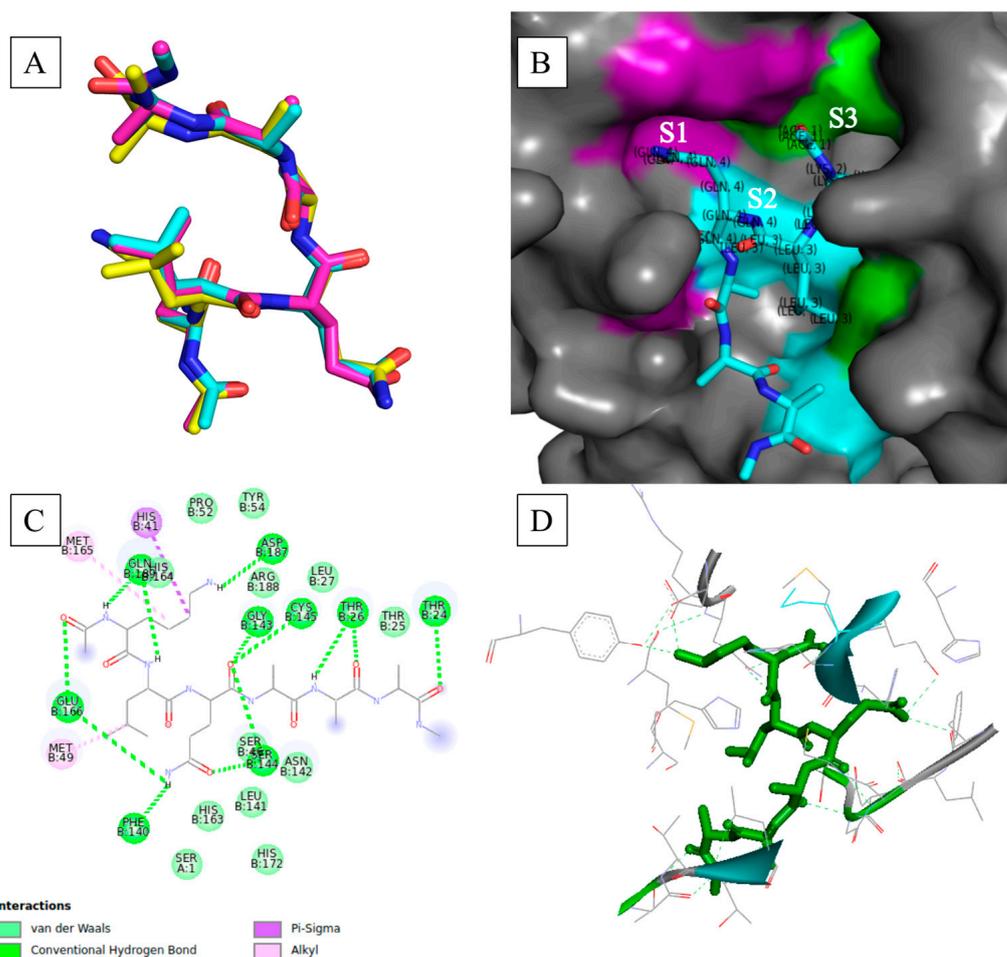


Figure 1. The assessment and validation of docking results for the SARS-CoV-2 main protease. (A) Validation of the reproducibility of the docking results. The visualization of the best poses of substrate Lys-Leu-Gln-Ala-Ala-Ala. Image was generated using PyMOL. (B) The surface of SARS-CoV-2 M^{Pro} (PDB ID:6XHM) showing docked substrate and the substrate-binding subsites are color-coded as follows: purple: S1, cyan: S2; green: S3. The image was generated using PyMOL. (C) Resolution of intermolecular interactions between M^{Pro} and substrate at the active site. Two-dimensional representation of the protein–ligand interactions at the active site for M^{Pro} complexed with Lys-Leu-Gln-Ala-Ala-Ala substrate. The image was generated using BIOVIA Discovery Studio 2020 Client. (D) Hydrogen anchoring of the best docked pose. Three-dimensional representation of the Lys-Leu-Gln-Ala-Ala-Ala substrate residues forming hydrogen interaction with active site residues of M^{Pro}. The Lys-Leu-Gln-Ala-Ala-Ala substrate attained a binding free energy of $-8.3 \text{ kcal.mol}^{-1}$. The image was generated using BIOVIA Discovery Studio 2020 Client.

Table 3. The active site residues constituting the substrate-binding subsites.

Subsites	Active Site Residues	Function	References
S3	His41; Met49; Met165	Shallow subsite that tolerates different functionalities.	[6,29]
S2	Thr25; His41; Met49; Cys145	Deeply buried subsite involved in hydrophobic and electrostatic interactions.	[6,30]
S1	Phe140; Gly143; Cys145; His163; Glu166; His172	Deeply buried subsite involved in hydrophobic and electrostatic interactions.	[6,30]
S1'	Thr25; Thr26; Leu27; Cys145	Generally forms polar contact interactions with substrates.	[30]

Figure 1B shows the binding mode of the hexapeptide within the binding pocket with respect to the active site subsites [11,29,30]. The P3-P1 residues were mounted desirably onto the active site, with each residue being anchored onto its appropriate subsite. The desired binding modes and adherence of the hexapeptide binding modes to the nomenclature of Schechter and Berger [28] were prevalent across the best docked poses. Violation of the nomenclature of Schechter and Berger [28] was only apparent in some of the top docked poses. Hexapeptides such as RLQATF and RLQSTF—the top poses ($-8.7 \text{ kcal}\cdot\text{mol}^{-1}$)—showed the side chain of P3 anchored in S1, whereas the side chains of P2 and P1 were anchored in S3 and S2, respectively. Others, such as RLQAAN, showed S1 anchoring the side chains of P1 and P3, whilst S2 rightfully anchored P2 (Supplementary Figures S2 and S3). Nonetheless, most binding modes showed that the M^{Pro} crystal structure followed the substrate recognition commonly associated with SARS-CoV-2 M^{Pro} . The side chains of the hexapeptides were appropriately accommodated in the subsites of the active site, regardless of significant differences in docking scores. Furthermore, these modes confirmed that these hexapeptides, as informed by the literature, were recognized by M^{Pro} and appropriately bound to the receptor protein. Thus, substrate recognition by M^{Pro} towards the generated hexapeptides was confirmed. Ultimately, the establishment of this recognition for the hexapeptides will assist in efforts to design potent and therapeutic peptidomimetics.

Resolution of the intermolecular interactions at the binding interface showed the high prevalence of hydrogen bonds and van der Waals interactions forming across all hexapeptides and active site residues, as shown in Figure 1C. The formed hydrogen bonds included conventional hydrogen bonds, carbon–hydrogen bonds, and Pi-donor hydrogen bonds. Hydrogen bonds, especially conventional hydrogen bonds, are the fundamental stabilizing force in biomolecular structures that underpin structure, function, and conformational dynamics [31,32], and they also regulate complementarity and stability in protein–ligand complexes at the binding interface [33,34]. Therefore, the prevalence of hydrogen bond formation at the active site indicated shape and electrostatic complementarities between M^{Pro} and the hexapeptides, and it also pointed to high affinity for the peptides, as evidently shown by the binding free energies and desirable binding modes (Figure 1B–D; Supplementary Figures S2–S7). Furthermore, the stabilizing effect of hydrogen bonds would prove to be integral in maintaining the complex structures in dynamic processes.

Notably, the key active site residues (including the catalytic dyad) frequently participated in hydrogen bonds and van der Waals interactions with the hexapeptides. Cys145 typically formed conventional hydrogen bonds with the oxygen atoms of the carboxyl group of P1, which in turn placed the catalytic residue in close proximity to the scissile peptide bond. Key residues 138 to 146, which constitute the oxyanion loop that confers substrate stability during the proteolytic process, participated in key stabilizing forces (hydrogen bonds and van der Waals forces of attraction) that promoted the formation of stable complexes. Other key residues, such as Met49, His163, His164, Met165, Glu166, Leu167, Asp187, Arg188, Gln189, Thr190, Ala191, and Gln192, which underpinned the subsites and accommodated the appropriate binding of the substrate residues via side-chain rearrangement [11,35–37], also formed the aforementioned stabilizing interactions with the hexapeptides. Figure 1D shows that the entirety of the hexapeptide was anchored by various hydrogen-bonding interactions formed with active site residues. Conclusively, stable M^{Pro} –hexapeptide complexed structures were constructed as shown by many stabilizing interactions at the binding interface. Moreover, the prevalent formation of hydrogen bonds at the binding interface pointed to the substrate specificity that a coronavirus main protease would typically exhibit towards peptide structures with the Leu-Gln↓(Ala/Ser) cleavage sequence.

3.3. Global Stability of the M^{Pro} Systems

The 20 ns MD simulations were performed on 131 Lys-Leu-Gln*** (KLQ***)- M^{Pro} complexed structures and an unbound M^{Pro} (*apo*-) structure. The exclusive use of KLQ*** complexed structures for molecular dynamics was to preserve comparability between systems—maintaining that the ligands shared identical P3-P1 residues and that they held

similar ligand efficiencies with other P3-P1 ligand groups in this study. The global stability of the M^{PTO} systems was assessed through the calculation of root mean square deviation (RMSD) and radius of gyration (Rg) values for the protein backbone (α -carbons) of M^{PTO} based on the 20 ns MD simulation trajectories, which were plotted as violin plots to highlight significant changes in conformation and stability (Figures 2 and 3).

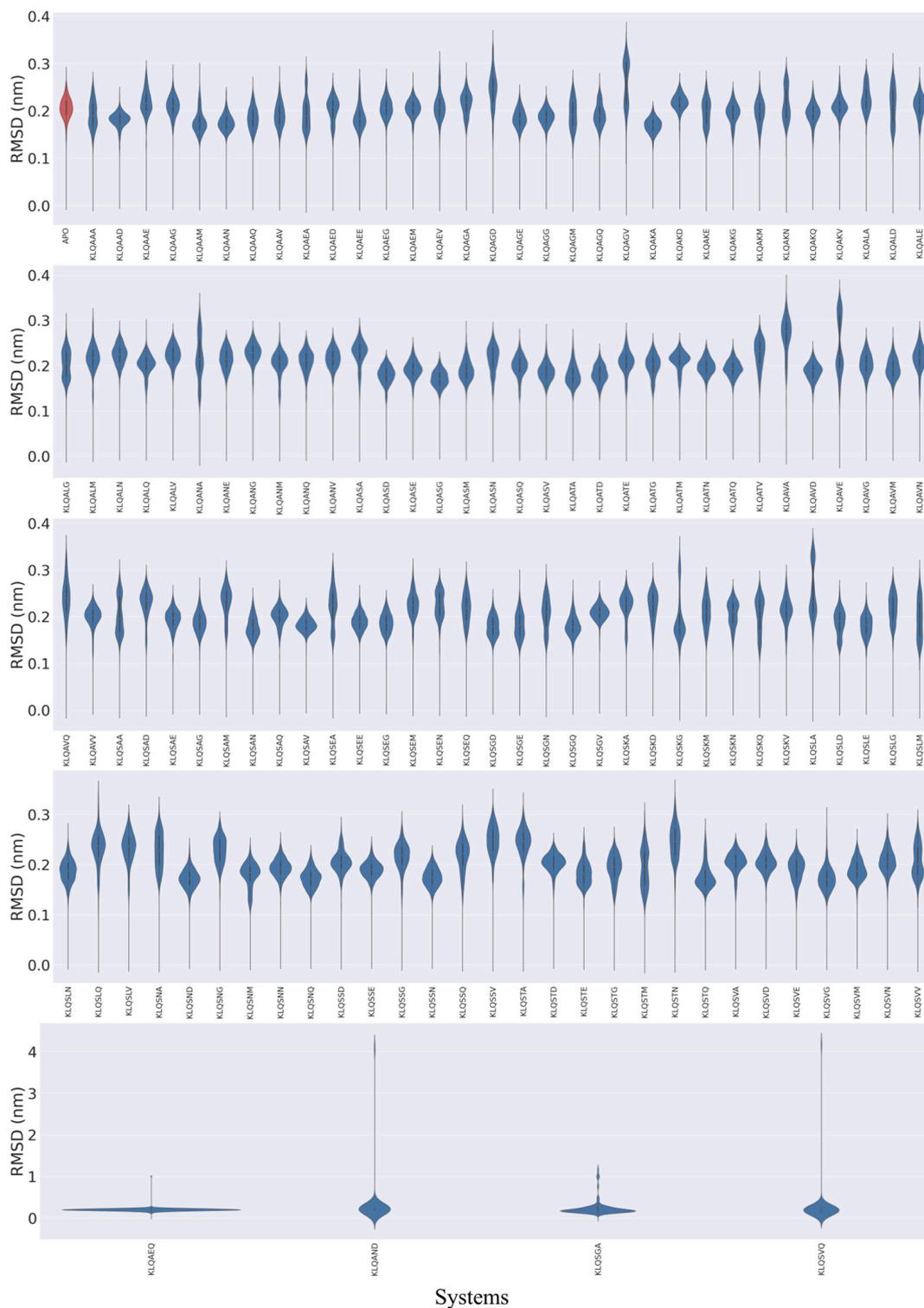


Figure 2. The violin plots of M^{PTO} RMSD values in unbound (red) and hexapeptide-bound (blue) systems during the 20 ns MD simulation. Images were generated using Seaborn on Python.

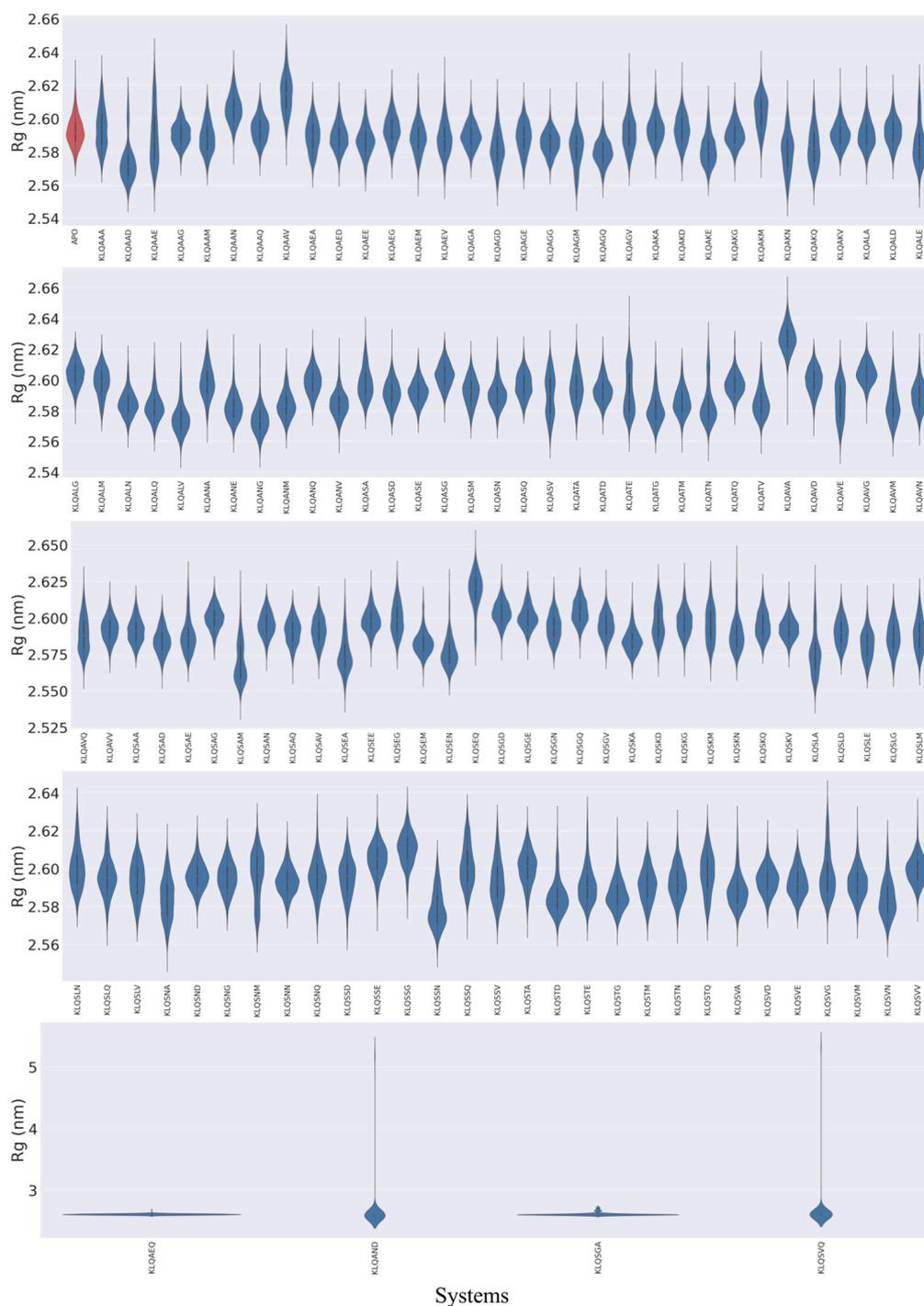


Figure 3. The violin plots of M^{PT0} Rg values in unbound (red) and hexapeptide-bound (blue) systems during the 20 ns MD simulation. Images were generated using Seaborn on Python.

The RMSD plots (Figure 2) showed the various behaviors of the protein across the systems and mainly displayed three distinct trends in structural stability listed in the following order of prevalence: substantial fluctuation, stable conformation, and significant changes in location (dissociation). Varying degrees of backbone fluctuation and conformational shifts were prevalent among the KLQ*** systems, as illustrated by bimodal and multimodal clustering and clustering at high RMSD values (>0.2 nm). Bimodal clustering, as clearly depicted in the KLQAGV, KLQAVE, KLQSAA, and KLQSLA systems, suggested that the protein converged on two alternative stable conformations during the simulation—an indication of reduced overall stability. Multimodal clustering (KLQAAA, KLQAEA, KLQAGM,

revealed the positions of these regions within the 3D protein structure (Figures 4–6). Across all M^{Pro} systems, the RMSF values of both protomers of the protein approximated each other, with protomer A registering slightly higher RMSF values despite having no bound hexapeptide (Figures 4A,B, 5A,B and 6A,B). Protomer B is where the bound hexapeptide resided. Overall, the highest RMSF values were obtained in the KLQAND and KLQSVQ systems (Figure 6) due to the dissociation of the protomers. Since these values were disproportionately higher than those of the rest of the systems, the data were separated to optimize visualization as follows: (i) systems with KLQ*** substrates with Ala at P1' (Figure 4); (ii) systems KLQ*** substrates with Ser at P1' (Figure 5); and (iii) KLQAND and KLQSVQ systems (Figure 6).

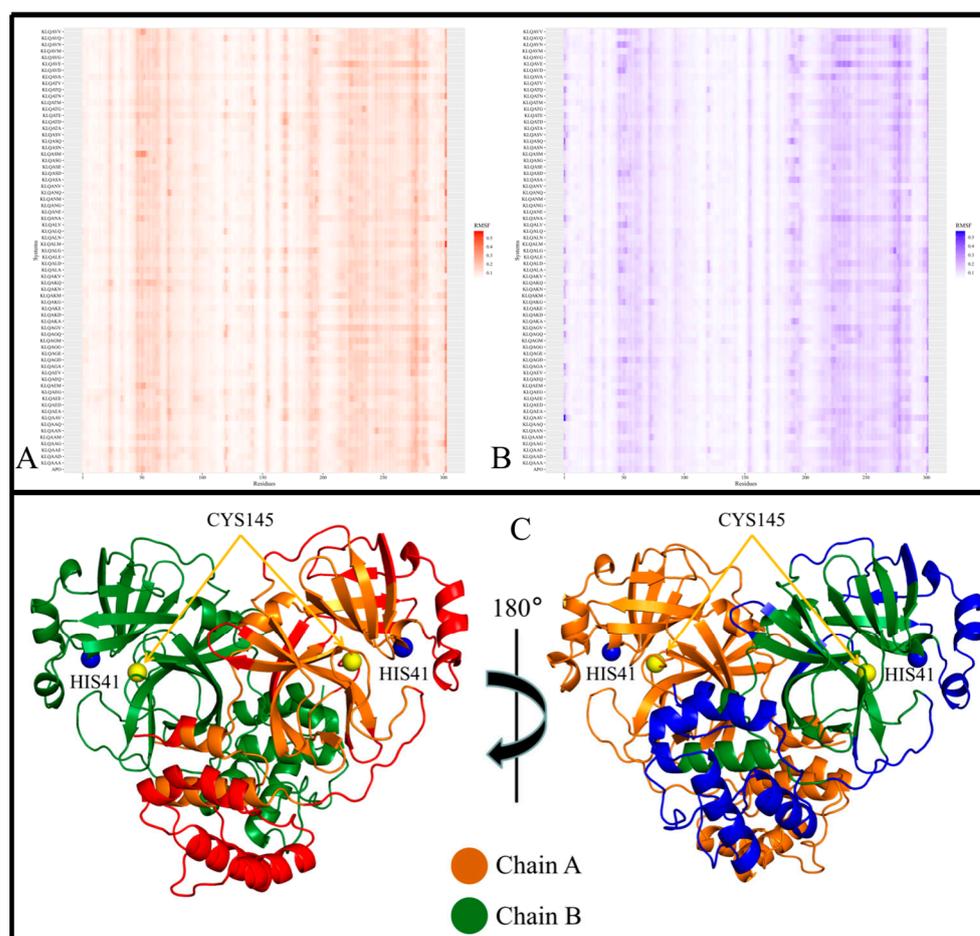


Figure 4. The local stability of M^{Pro} and M^{Pro} -Lys-Leu-Gln-Ala** (KLQA**) hexapeptide complexes. RMSF values of the backbone α -carbon atoms for the *apo*-protein and KLQA** hexapeptide-bound M^{Pro} systems during the 20 ns MD simulation. (A) The heatmap distribution of RMSF values for protomer A. (B) The heatmap distribution of RMSF values for protomer B. (C) The crystal structure of M^{Pro} with regions of consistent higher RMSF values colored in red for protomer A and in blue for protomer B. Images were created using RStudio and PyMOL.

Figure 4 shows the heatmaps for both protomers of the Lys-Leu-Gln-Ala** (KLQA**) systems alongside *apo*- M^{Pro} . While the RMSF values slightly differed from one system to the next, both protomers of M^{Pro} displayed high flexibility in the same regions in all systems (Figure 4A,B). Overall, the RMSF values for the KLQA** systems ranged between 0.0384 and 0.5805 nm for protomer A and 0.0394–0.5698 nm for protomer B. High flexibility was observed in residues 21–26, 44–80, 92–97, 118–127, 141–144, 152–156, 167–171, 188–198, 215–288, and 298–302 in protomer A (Figure 4A) and residues 1–4, 22–24, 44–80, 92–96, 118–125, 153–156, 168–171, 188–197, 212–288, and 297–301 in protomer B (Figure 4B), respectively. The Lys-

Leu-Gln-Ser** (KLQS**) systems showed similar trends as the KLQA** systems in terms of RMSF values and the localization of flexible residues (Figure 5A,B). The overall range of RMSF values was between 0.0399 and 0.4223 nm for protomer A and 0.0372–0.3726 nm for protomer B. Similarly, flexibility was demonstrated in residues 21–26, 32–35, 44–80, 92–97, 119–123, 139–143, 153–156, 167–171, 187–197, 212–288, and 297–302 in protomer A and residues 1–5, 21–26, 33–35, 44–66, 70–80, 92–98, 152–156, 167–170, 186–197, 212–238, 241–286, and 297–301 for protomer B.

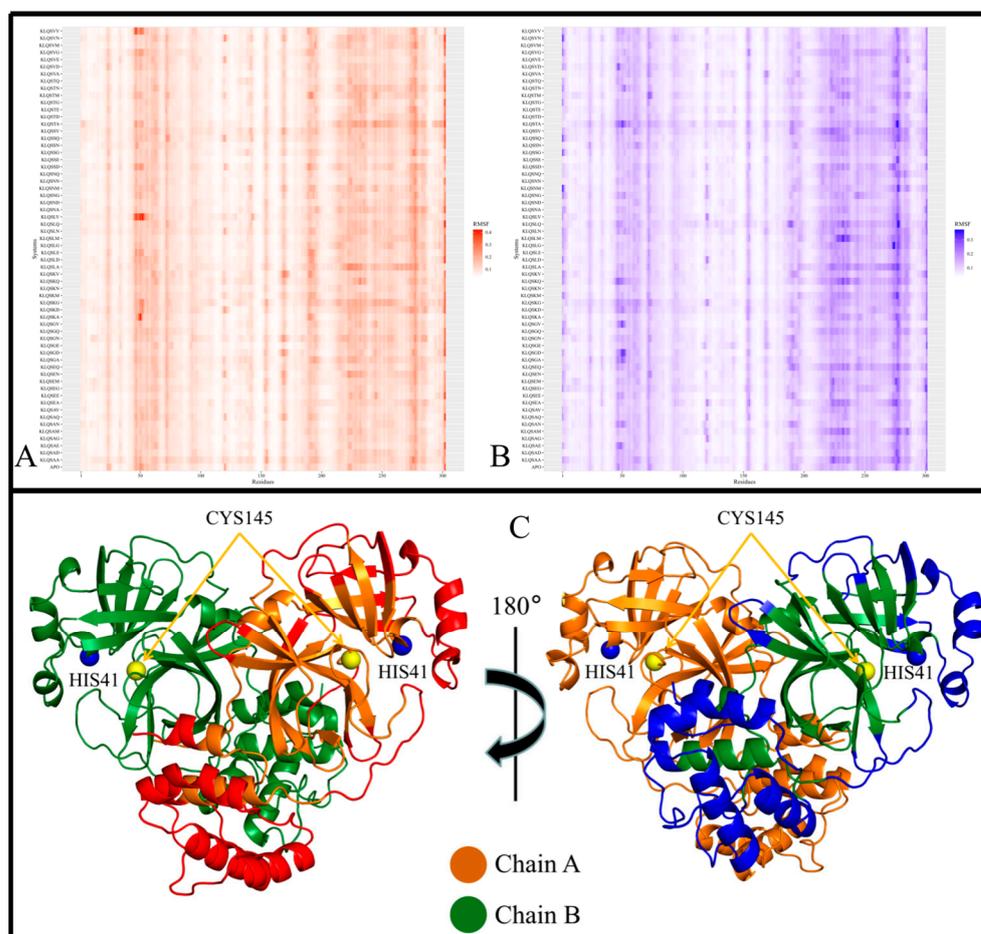


Figure 5. The local stability of the MP^{Pro} and MP^{Pro}-Lys-Leu-Gln-Ser** (KLQS**) hexapeptide complexes. The RMSF values of the backbone α -carbon atoms for the *apo*-protein and KLQS** hexapeptide-bound MP^{Pro} systems during the 20 ns MD simulation. (A) The heatmap distribution of RMSF values for protomer A. (B) The heatmap distribution of RMSF values for protomer B. (C) The crystal structure of MP^{Pro} with regions of consistent higher RMSF values colored in red for protomer A and in blue for protomer B. Images were created using RStudio and PyMOL.

The majority of the highly fluctuating residues constituted loop regions in both protomers of MP^{Pro}, as shown (as red, and blue) in Figures 4C and 5C. This high fluctuation could be attributed the flexible nature of loop structures. However, of all the flexible loop regions, residues 1–4 displayed flexibility exclusively in protomer B. Residues 1–9 comprise the N-finger terminal region, which plays a crucial role in dimer formation through interactions with Domain II of protomer A [38]. Semi-flexibility in β -sheets (Domains I and II) was displayed on either ends of the structure in both protomers—connecting to or from loop regions. In addition, α -helices displayed semi-flexibility in residues 44–80 (domain I) and 212–288 (domain III). Of note, residues 44–80 are part of the catalytic domain responsible for catalysis and MP^{Pro} autocleavage [30]. Sequentially, this α -helix comes after catalytic His41 and consists of key binding residues such as Met49, which contributes to substrate

stabilization. While the RMSD and Rg values did not indicate overall destabilization of the M^{PRO} structure, this apparent semi-flexibility could be attributed to functional flexibility and the intrinsic side-chain rearrangement mechanisms that accommodate the hexapeptide in the binding pocket. This point was further supported by the absence of hexapeptide ejection from the binding pocket. The α -helices of Domain III (Figures 4C and 5C; the lower region of protein) consistently displayed high flexibility in both protomers. The high fluctuation could be attributed to their connection to long loop regions, as α -helices typically demonstrate restricted protein motion but can confer great flexibility, which is essential to protein function [39].

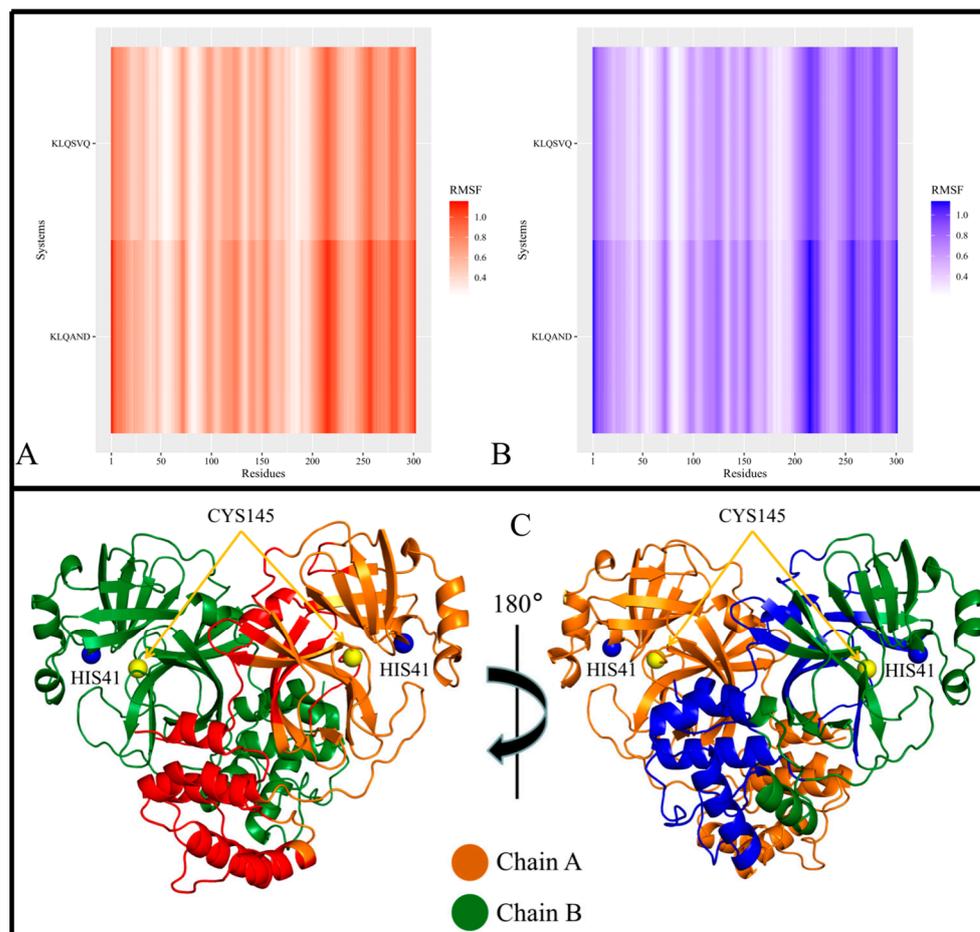


Figure 6. The local stability of the M^{PRO}–KLQ*** hexapeptide complexes. RMSF values of the backbone α -carbon atoms for the KLQAND and KLQSVQ hexapeptide-bound M^{PRO} systems during the 20 ns MD simulation. (A,B) Refer to the protomers of the M^{PRO} homodimer. (C) Refers to the crystal structure of M^{PRO} with regions of consistent higher RMSF values colored in red for protomer A and in blue for protomer B. Images were created using RStudio and PyMOL.

In the context of the final two systems, KLQAND and KLQSVQ—for which the residue fluctuations were the highest among all systems due to dimer dissociation—RMSF values ranging between 0.2261 and 1.1565 nm for protomer A and 0.2174–1.1333 nm for protomer B were registered. However, the flexible residues were similar to those in the KLQA** and KLQS** systems, including residues 1–17, 69–73, 96–100, 111–127, 138–144, 152–157, 202–209, 210–223, 224–234, 236–237, 242–254, 255–259, 260–276, 277–285, 286–298, and 299–302 in protomer A and residues 1–19, 24–29, 69–74, 95–100, 111–128, 138–143, 151–157, 170–173, 199–206, 207–223, 224–227, 247–288, 291–299, and 300–301 in protomer B. Notably, these were the only instances where the N-finger terminal residues displayed flexibility in both protomers (Figure 6), as expected, since dimer dissociation occurred in both of these systems.

Furthermore, the localization of the flexible residues showed greater residue fluctuation in Domain II involving entire β -sheet structures (Figure 6C). This was not evident in the other systems, as fluctuations in β -sheets were only registered for their terminal ends connecting to loop regions. The semi-flexibility in α -helices (residues 44–63) surrounding the catalytic dyad was absent in these two systems, pointing to reduced activity and/or inactivity of the active site residues. Moreover, the α -helix residues constituting Domain III demonstrated much greater flexibility than any of the other systems, further indicating destabilization of the M^{Pro} structure.

In conclusion, the RMSF results supported that the M^{Pro} residues typically displayed similar fluctuation patterns in the presence of the KLQ*** substrates, with the exception of the KLQAND and KLQSVQ systems. The RMSF values of both protomers approximated one another in magnitude, and the locations of the high-fluctuation regions on the M^{Pro} protomers were similar. Interestingly, the KLQAEQ and KLQSGA systems did not register high RMSF values like the KLQAND and KLQSVQ systems, but instead behaved similar to the majority of complexed systems. Therefore the unbinding of the hexapeptides did not cause significant changes to the motions and conformations of M^{Pro} in these two systems (KLQAEQ and KLQSGA). Consequently, the steep hikes in RMSD and Rg values were only attributed to the fact the MD program interpreted the hexapeptide chain as a third chain in the overall protein system.

3.5. Principal Component Analysis and Protein Motion Classification

The prominent structural motions and conformational changes of the M^{Pro} backbone during the 20 ns MD simulations were assessed using PCA calculations. PCA divided the overall protein motions of the trajectories into principal components that described the essential functional protein motions during the simulation. Since the first two principal components, PC1 and PC2, retained the majority of the variance of the original data, they could be used to provide a meaningful description of the protein motions throughout the course of the simulations. Thus, 2D projections of these principal components were plotted using the Cartesian coordinates of all backbone atoms to visualize and examine these conformational changes (Supplementary Figure S8). The direction of change for the movement of M^{Pro} against time was unique to each system. For example, some systems displayed clockwise changes in PCA, while others showed anti-clockwise changes in PCA. Seemingly, almost all systems retained steady conformational changes throughout the simulations as the distribution of the PCA coordinates was generally compact/clustered. Without regard to their pattern of progression, the typical range of the coordinates was between -5 and 5 for both PC1 and PC2. The KLQAEQ, KLQAND, KLQSGA, and KLQSVQ systems displayed the most drastic changes in protein motions due to the aforementioned changes in stability, structure, and conformation.

In an attempt to classify these PCA data, a custom pairwise comparison of the M^{Pro} systems was performed as detailed in Section 2.4. Figure 7 shows the cluster map illustrating the correlations of the protein motions for all systems with respect to one another (bearing in mind the underlying assumption). In the cluster map, systems sharing similarities in protein motions were registered in hues between black and violet, whereas systems with dissimilarities were registered in hues between white and orange. The accompanying dendrogram illustrates the similarities or correlations among the prominent protein motions of the clustered M^{Pro} systems. Hierarchical clustering yielded four main classifications of correlation among all the systems that depicted similarity in proteins motions that describe the trajectories of the M^{Pro} systems. The four classifications of similarity in the M^{Pro} systems are listed in Table 4.

Although three classifications of the systems shared varying similarities within and between them, the Group 1 PCA data (comprising the KLQSVQ, KLQAEQ, and KLQAND systems) demonstrated the highest dissimilarity to the rest of the PCA groups, as expected, considering their patterns in RMSD, Rg, and even RMSF values (for the KLQAND and KLQSVQ systems). The PCA of the KLQAND and KLQAEQ systems showed high similar-

ity, as opposed to that of the KLQSVQ system. The KLQAND and KLQSVQ systems both showed dimer dissociation, although not similarly, during their simulations. Nevertheless, the clustering was successful in identifying highly dissimilar protein motions in the M^{Pro} systems, as mutually confirmed by the RMSD, R_g , and RMSF data. Interestingly, the KLQSGA system, which also showed structural instability in its RMSD and R_g values, was classified in Group 3 and not alongside the KLQSVQ, KLQAEQ, and KLQAND systems. Of note, the dendrogram showed the KLQSGA system displaying the highest dissimilarity with the rest of the systems within Group 3, showing a weak correlation in terms of protein motions and conformational changes with its “similar” systems.

Systems in Groups 2 to 4 all showed varying degrees of similarity within and across themselves. Systems in Group 3 displayed the least amount of divergence in the hierarchical dendrogram when compared to Groups 2 and 4 (Figure 7). Notably, Group 3 was occupied by systems that attained high medians and/or multimodal distribution in their RMSD and R_g values. Systems that demonstrated stability were split across the three groups.

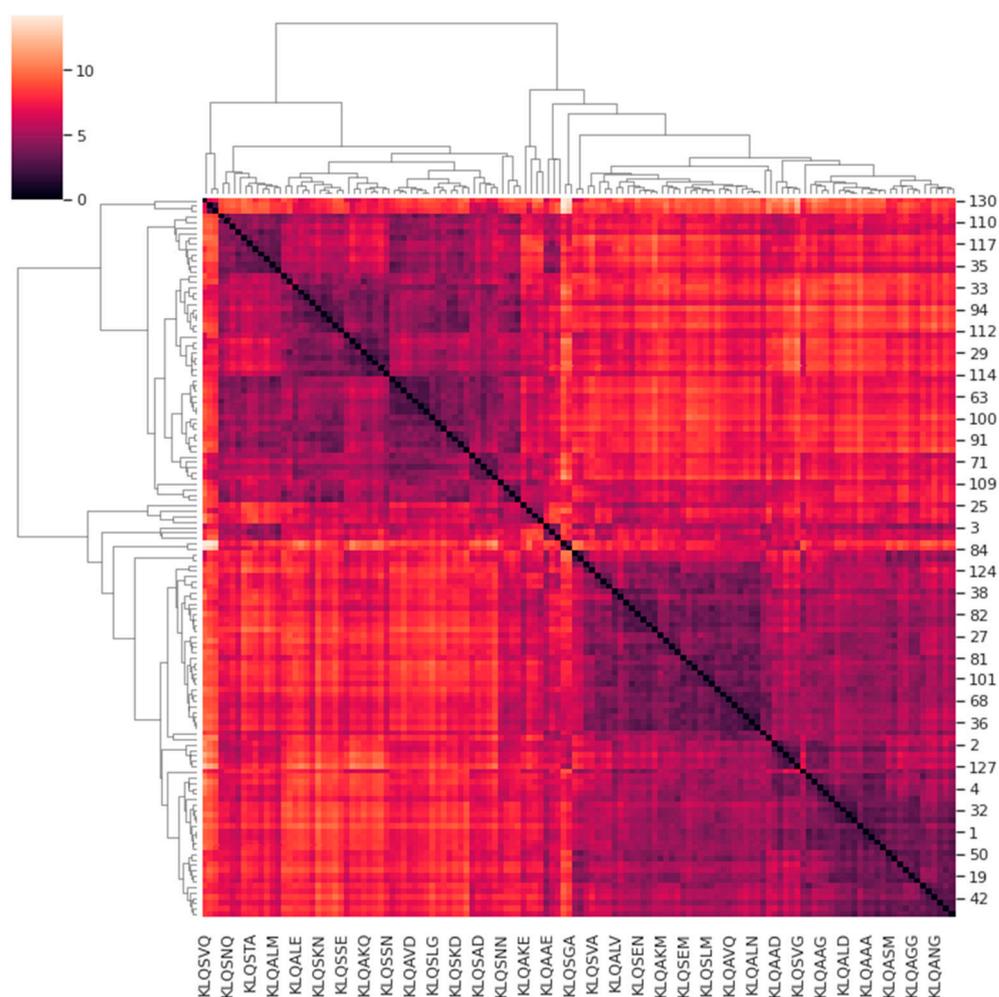


Figure 7. The correlations and similarities in protein motions of the M^{Pro} dynamic systems. The clustering of differences in protein motions of the M^{Pro} systems using correlation as a metric. The dendrogram illustrates hierarchical clustering of dynamic protein motions in the *apo*- M^{Pro} and M^{Pro} -hexapeptide systems as a measure of similarity. The figure was generated using Seaborn in Python.

The PCA results highly aligned with the RMSD, R_g , and RMSF values in terms of overall trends. Most of the M^{Pro} systems registered PCA coordinates within the range of 5 and -5 for PC1 and PC2, just as they attained RMSD and R_g equilibration around 0.2 nm and 2.60 nm, respectively. Importantly, there were only two unbinding events registered to

the 131 KLQ***-bound M^{Pro} systems—an indication of strong intermolecular interaction in the binding pocket. The retention of hexapeptides by the protein could be attributed to the substrate recognition and specificity postulated in the docking studies. Despite binding substrates with minor chemical differences, the protein achieved and retained similar flexibility, compaction, and prominent protein motions, which altogether caused similar behaviors and patterns of stability in the M^{Pro} systems. Consequently, these results suggested that SARS-CoV-2 M^{Pro} has an intrinsic mechanism that enables the binding of different peptide substrates without rendering drastic instability to the entire structure. Even in the PCA, the majority of the systems seemingly occupied the same positions on the PCA plot throughout the course of the simulation. The exceptions to these trends consistently included the KLQAEQ, KLQAND, KLQSGA, and KLQSVQ systems, as clearly demonstrated by their high dissimilarity to the rest of the systems. While the dissimilarities in protein motions depicted in the dendrogram (Figure 7) could be easily explained, the similarities also required verification. Henceforth, at least two relevant systems from each hierarchical classification (or group) were selected for trajectory visualization using VMD.

Table 4. The similarities among the dynamic M^{Pro} systems on the basis of prominent protein motions and conformational changes during the 20 ns MD simulation.

	Group 1	Group 2	Group 3	Group 4
				KLQAAG; KLQSGV; KLQANA; KLQAKN; KLQAAA; KLQSEA; KLQAVA; KLQAAD; KLQALQ; KLQSVG; KLQSAE; KLQSAQ; KLQAGM; KLQSN; KLQSVV; KLQANG; KLQSGQ; KLQSLN; KLQALD; KLQAGV; KLQAKD; KLQATA; KLQAGG; KLQASV; KLQSKV; KLQANM; KLQAVN; KLQSKG; KLQSTQ; KLQSLA; KLQSNM; KLQSN; KLQASM; KLQSTM; KLQANQ; KLQSLD; KLQSEE; KLQSEM; KLQSLM; KLQSVN; KLQAEE; KLQAGA; KLQAVM; KLQSN; KLQATQ; KLQSSG; KLQALN; KLQAEG; KLQATG; KLQATN; KLQAVQ; KLQASG; KLQAAV; KLQASD; KLQSKM; KLQAVE; KLQSEN; KLQASE; KLQALV; KLQANE; KLQAEV; KLQSVA; KLQAAN; KLQAKM; KLQSKQ; KLQSSQ; KLQSVM
		KLQANV; KLQSN; KLQAKE; APO; KLQSEG; KLQSA; KLQSAV; KLQSSV; KLQALG; KLQATD; KLQSGE; KLQSNQ; KLQAAM; KLQAEA; KLQSKN; KLQSLG; KLQSLV; KLQSS; KLQALA; KLQSA; KLQSGD; KLQSKD; KLQSTD; KLQAGE; KLQAKV; KLQATM; KLQAVG; KLQAKG; KLQAVD; KLQATV; KLQSTE; KLQSTG; KLQAKA; KLQSLQ; KLQSTN; KLQSD; KLQASN; KLQSSD; KLQALE; KLQAED; KLQSGN; KLQAKQ; KLQASA; KLQATE; KLQAGD; KLQSSN; KLQSV; KLQSAM; KLQSKA	KLQSGA; KLQSL; KLQAAE; KLQAGQ; KLQSAN; KLQALM; KLQSV; KLQASQ; KLQAAQ; KLQAEM; KLQAVV; KLQSTA	
M ^{Pro} Systems	KLQSVQ; KLQAEQ; KLQAND			

The visualization of the trajectories for the KLQAEQ and KLQSVQ systems in Group 1 confirmed the weak correlation in protein motions reported in the dendrogram. The protein backbone for both systems showed no overlap at any point in the simulations (Supplementary Figure S9). Dissimilarity in these systems was expected as the KLQAEQ

system showed substrate unbinding while the KLQSVQ system showed dimer dissociation. Therefore, these systems only fell within the same classification due to their drastic changes in conformation and location, as shown by their RMSD and Rg values. While the ejection of non-covalently bound ligands from the binding pocket normally does not register such extreme changes in RMSD and Rg values for the protein backbone, the steep hikes displayed in the KLQAEQ system were attributed to the inclusion of substrates as third chains of M^{PRO} by the GROMACS software prior to the MD simulations. The protein motions of the hexapeptide when it was ejected and later bound again affected the values for RMSD, Rg, and PCA. The KLQAND system seemingly displayed similarity to the KLQAEQ system, even though the former demonstrated dimer dissociation. The visualization of these systems refuted this similarity as the protein backbones did not exhibit overlap during the simulations (Supplementary Figure S9). However, the relationship between their PCA results could be attributed to the substrate unbinding (KLQAEQ) and dissociation (KLQAND) that were later restored for both systems. Therefore, the similarity between the PCA results was owing to both systems having registered steep hikes in RMSD and Rg values rather than steep gradients in RMSD and Rg values.

The visualization of the systems (*apo*-M^{PRO} and M^{PRO}-KLQSEG) in Group 2 showed a strong similarity in protein motions as the structures overlapped throughout all frames of the trajectories (Supplementary Figure S10). Particularly, Domains I and II (the chymotrypsin-like structure) consistently overlapped throughout the simulation, whereas the helices of Domain III displayed the most structural deviation in both structures. The mapping of high-RMSF-value residues indicated that residues in this domain were highly flexible. The KLQAVV and KLQSGA systems in Group 3 displayed little similarity in protein motions at any point of the simulations (Supplementary Figure S11). This was supported by the placement of these systems in the dendrogram; wherein the KLQSGA system exhibited high dissimilarity with the rest of the systems in Group 3 owing to the unbinding of the hexapeptide towards the end of the simulation. However, the similarity between the KLQSGA system and other Group 3 systems could have been the result of overlapping M^{PRO} motions prior to unbinding of the hexapeptide. The KLQAVV system demonstrated structural stability as well as ligand stability throughout the simulation and also showed conformational similarity with the closely related KLQSAN system (Supplementary Figure S11).

Similarly, the systems (M^{PRO}-KLQAAA and M^{PRO}-KLQSKG) in Group 4 also showed little similarity in protein motions (Supplementary Figure S12). While overlapping of the structures was visible in the β -sheets of Domain I, a strong similarity was highly unlikely considering their placement in the dendrogram (Figure 7). It could be, in general, that the protein motions were within a broad range, but it was visually impossible to assess this detail. Lastly, the visualization of the systems from all four groups revealed little to no similarity in protein motions between groups (Supplementary Figure S13). Structural overlaps rarely occurred during the simulations. The dissimilarity in protein motions was highly expected given the fact that these classifications diverged from the highest point of dissimilarity, as illustrated in Figure 7. Thus, this observation validated the hierarchical clustering since these systems exhibited no similarity in their prominent protein motions throughout the 20 ns MD simulation.

The visualization of the four clades of PCA progression allowed the inspection and confirmation of the similarities in protein motions of the systems. The outcomes of the visualization indicated that this type of PCA progression analysis is able, certainly in some cases, to identify both similar and dissimilar protein motions in comparable systems. With refinement, it could provide information for general use. In this case, it certainly enabled us to focus on particular systems given the number of simulations to assess.

4. Conclusions

COVID-19, as a disease and pandemic, continues to pose a tremendous threat to global stability as it has the potential to strain public health services to extraordinary proportions, cause economic disruptions and/or inactivity leading to financial crises, and

predominantly sabotage the livelihood of millions of people around the world. Moreover, the existing measures in place to control the spread of the virus are losing efficiency due to the emergence of novel variants with increased virulence and immunological evasive mechanisms. SARS-CoV-2 M^{PRO} has proven to be a promising drug target as it exhibits high degrees of conservation in sequence, structure, and specificity. However, there are details of the mechanism of proteolysis employed by SARS-CoV-2 M^{PRO} that still require study. Thus, this study was carried out to profile the binding of hexapeptide substrates onto SARS-CoV-2 M^{PRO} in preparation for the elucidation of its proteolytic mechanism and, ultimately, rational substrate-based and efficiency-driven drug design to combat SARS-CoV-2 and its variants. In summary, the virtual high-throughput screening of the hexapeptides showed favorable binding in terms of binding free energy score, ligand efficiency, as well as substrate recognition and specificity. Based on the MD simulations performed, M^{PRO} displayed various conformational behaviors but overall persistent stability in all complexed systems, with the exception of four outliers. The PCA showed compact distribution of PC1 and PC2 in all but the same four systems. The custom pairwise comparison for the quantification of PCA progression revealed four main classifications of similarity in protein motions, which were confirmed by visualization of the trajectories.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/v15071480/s1>, Supplementary Figure S1. Validation of reproducibility of the best poses in the docking results; Supplementary Table S1: The ligand efficiencies of the hexapeptide substrates docked onto SARS-CoV-2 M^{pro} on basis of recognition sequence; Supplementary Figure S2. Confirmation of SARS-CoV-2 M^{pro} substrate recognition in binding poses for substrates RLQATF, RLQSGA and RLQSTF; Supplementary Figure S3. Confirmation of SARS-CoV-2 M^{pro} substrate recognition in binding poses for substrates RLQAAN, RLQAAF, RLQAGA, RLQALG, RLQAVN, TLQAGF, TLQAVA, VLQAAF and VLQAVF; Supplementary Figure S4. Confirmation of SARS-CoV-2 M^{pro} substrate recognition in binding poses for substrates KLQSKM, KLQAEM and TLQSLM; Supplementary Figure S5. Resolution of intermolecular interactions between M^{pro} and substrates at the active site; Supplementary Figure S6. Resolution of intermolecular interactions between M^{pro} and substrates at the active site; Supplementary Figure S7. Resolution of intermolecular interactions between M^{pro} and substrates at the active site; Supplementary Figure S8. The 2D projections of the principal components for M^{pro} (apo and KLQ***-substrate-bound) systems over the duration of the 20 ns MD simulations; Supplementary Figure S9. Visualisation of the MD trajectories for Group 1 systems; Supplementary Figure S10. Visualisation of the MD trajectories for Group 2 systems; Supplementary Figure S11. Visualisation of the MD trajectories for Group 3 systems; Supplementary Figure S12. Visualisation of the MD trajectories for Group 4 systems; Supplementary Figure S13. Visualisation of the MD trajectories for systems in all hierarchical groups; Supplementary Table S2: RMSD from substrate dynamics ordered by Mean and ordered by Standard Deviation; Supplementary Table S3: R_g from substrate dynamics ordered by Mean and ordered by Standard Deviation; Supplementary Table S4: RMSD and R_g from substrate dynamics ordered by substrate; Supplementary Figure S14. Preliminary docking studies to determine the protein chain to prioritise for docking studies.

Author Contributions: Conceptualization, K.A.L.; methodology, K.A.L. and S.Z.; software, S.Z.; validation, K.A.L. and S.Z.; resources, K.A.L.; data curation, S.Z.; writing—original draft preparation, S.Z.; writing—review and editing, K.A.L.; visualization, S.Z.; supervision, K.A.L.; project administration, K.A.L.; funding acquisition, S.Z. All authors have read and agreed to the published version of the manuscript.

Funding: Special thanks to the National Research Fund grant MND190725459450 and LINK-SA Trust for providing funds for this work and related studies.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not Applicable.

Data Availability Statement: Data is contained within the article or Supplementary Material.

Acknowledgments: This work acknowledges the Centre for High Performance Computing (CHPC) in Cape Town, South Africa, for providing the computational resources to perform this study (project CHEM0802).

Conflicts of Interest: The authors declare no conflict of interest.

References

1. WHO. Coronavirus Disease (COVID-19) Pandemic. World Health Organization. 2023. Available online: <https://www.who.int/emergencies/diseases/novel-coronavirus-2019> (accessed on 6 May 2023).
2. Liu, C.; Zhou, Q.; Li, Y.; Garner, L.V.; Watkins, S.P.; Carter, L.J.; Smoot, J.; Gregg, A.C.; Daniels, A.D.; Jervey, S.; et al. Research and development on therapeutic agents and vaccines for COVID-19 and related human coronavirus diseases. *ACS Cent. Sci.* **2020**, *6*, 315–331. [[CrossRef](#)] [[PubMed](#)]
3. Guo, G.; Ye, L.; Pan, K.; Chen, Y.; Xing, D.; Yan, K.; Chen, Z.; Ding, N.; Li, W.; Huang, H.; et al. New insights of emerging SARS-CoV-2: Epidemiology, etiology, clinical features, clinical treatment, and prevention. *Front. Cell Dev. Biol.* **2020**, *8*, 410. [[CrossRef](#)] [[PubMed](#)]
4. Cascella, M.; Rajnik, M.; Cuomo, A.; Dulebohn, S.C.; Di Napoli, R. Features, Evaluation, and Treatment of Coronavirus (COVID-19). In *StatPearls*; StatPearls Publishing: Treasure Island, FL, USA, 2020. Available online: <https://www.ncbi.nlm.nih.gov/books/NBK554776/> (accessed on 18 July 2020).
5. Tu, Y.F.; Chien, C.S.; Yarmishyn, A.A.; Lin, Y.Y.; Luo, Y.H.; Lin, Y.T.; Lai, W.Y.; Yang, D.M.; Chou, S.J.; Yang, Y.P.; et al. A review of SARS-CoV-2 and the ongoing clinical trials. *Int. J. Mol. Sci.* **2020**, *21*, 2657. [[CrossRef](#)] [[PubMed](#)]
6. Khan, S.A.; Zia, K.; Ashraf, S.; Uddin, R.; Ul-Haq, Z. Identification of chymotrypsin-like protease inhibitors of SARS-CoV-2 via integrated computational approach. *J. Biomol. Struct. Dyn.* **2020**, *39*, 2607–2616. [[CrossRef](#)]
7. CDC. SARS-CoV-2 Variant Classifications and Definitions. Centers for Disease Control and Prevention. 2023. Available online: <https://www.cdc.gov/coronavirus/2019-ncov/variants/variant-classifications.html> (accessed on 6 May 2023).
8. Mousavizadeh, L.; Ghasemi, S. Genotype and phenotype of COVID-19: Their roles in pathogenesis. *J. Microbiol. Immunol. Infect.* **2020**, *54*, 159–163. [[CrossRef](#)]
9. Ampornanai, K.; Meng, X.; Shang, W.; Jin, Z.; Rogers, M.; Zhao, Y.; Rao, Z.; Liu, Z.J.; Yang, H.; Zhang, L.; et al. Inhibition mechanism of SARS-CoV-2 main protease by ebselen and its derivatives. *Nat. Commun.* **2021**, *12*, 3061. [[CrossRef](#)]
10. Ullrich, S.; Nitsche, C. The SARS-CoV-2 Main Protease as Drug Target. *Bioorg. Med. Chem. Lett.* **2020**, *30*, 127377. [[CrossRef](#)]
11. Goyal, B.; Goyal, D. Targeting the dimerization of main protease of coronaviruses: A potential broad-spectrum therapeutic strategy. *ACS Comb. Sci.* **2020**, *22*, 297–305. [[CrossRef](#)]
12. Aleem, A.; Akbar Samad, A.B.; Vaqar, S. Emerging Variants of SARS-CoV-2 And Novel Therapeutics Against Coronavirus (COVID-19). In *StatPearls*; StatPearls Publishing: Treasure Island, FL, USA, 2021.
13. Azgari, C.; Kilinc, Z.; Turhan, B.; Circi, D.; Adebali, O. The mutation profile of SARS-CoV-2 is primarily shaped by the host antiviral defense. *Viruses* **2021**, *13*, 394. [[CrossRef](#)]
14. Chen, J.; Wang, R.; Gilby, N.B.; Wei, G.W. Omicron variant (B. 1.1. 529): Infectivity, vaccine breakthrough, and antibody resistance. *J. Chem. Inf. Model.* **2021**, *62*, 412–422. [[CrossRef](#)]
15. Sanusi, Z.K.; Lobb, K.A. Insights into the Dynamics and Binding of Two Polyprotein Substrate Cleavage Points in the Context of the SARS-CoV-2 Main and Papain-like Proteases. *Molecules* **2022**, *27*, 8251. [[CrossRef](#)]
16. Trott, O.; Olson, A.J. AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J. Comput. Chem.* **2010**, *31*, 455–461. [[CrossRef](#)]
17. Bannwarth, C.; Ehlert, S.; Grimme, S. GFN2-xTB—An accurate and broadly parametrized self-consistent tight-binding quantum chemical method with multipole electrostatics and density-dependent dispersion contributions. *J. Chem. Theory Comput.* **2019**, *15*, 1652–1671. [[CrossRef](#)]
18. O’Boyle, N.M.; Banck, M.; James, C.A.; Morley, C.; Vandermeersch, T.; Hutchison, G.R. Open Babel: An open chemical toolbox. *J. Cheminform.* **2011**, *3*, 33. [[CrossRef](#)]
19. Morris, G.M.; Huey, R.; Lindstrom, W.; Sanner, M.F.; Belew, R.K.; Goodsell, D.S.; Olson, A.J. AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. *J. Comput. Chem.* **2009**, *30*, 2785–2791. [[CrossRef](#)]
20. DeLano, W.L. PyMOL. 2002. Available online: https://www.ccp4.ac.uk/newsletters/newsletter40/11_pymol.pdf (accessed on 10 February 2022).
21. BIOVIA, Dassault Systèmes. *BIOVIA Discovery Studio 2020 Client, Version 20.1*; Dassault Systèmes: San Diego, CA, USA, 2020.
22. Duan, Y.; Wu, C.; Chowdhury, S.; Lee, M.C.; Xiong, G.; Zhang, W.; Yang, R.; Cieplak, P.; Luo, R.; Lee, T.; et al. A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations. *J. Comput. Chem.* **2003**, *24*, 1999–2012. [[CrossRef](#)]
23. Humphrey, W.; Dalke, A.; Schulten, K. VMD: Visual molecular dynamics. *J. Mol. Graph.* **1996**, *14*, 33–38. [[CrossRef](#)]
24. Hoffman, R.L.; Kania, R.S.; Brothers, M.A.; Davies, J.F.; Ferre, R.A.; Gajiwala, K.S.; He, M.; Hogan, R.J.; Kozminski, K.; Li, L.Y.; et al. Discovery of ketone-based covalent inhibitors of coronavirus 3CL proteases for the potential therapeutic treatment of COVID-19. *J. Med. Chem.* **2020**, *63*, 12725–12747. [[CrossRef](#)]

25. Hopkins, A.L.; Groom, C.R.; Alex, A. Ligand efficiency: A useful metric for lead selection. *Drug Discov. Today* **2004**, *9*, 430–431. [[CrossRef](#)]
26. Chen, H.; Zhou, X.; Gao, Y.; Zhou, J. 2.08—Fragment-based drug design: Strategic advances and lessons learned. In *Comprehensive Medicinal Chemistry III*; Chackalamannil, S., Rotella, D., Ward, S.E., Eds.; Elsevier Ltd.: Amsterdam, The Netherlands, 2017; Volume 8, pp. 212–232. ISBN 978-0-12-803201-5.
27. Orita, M.; Ohno, K.; Warizaya, M.; Amano, Y.; Niimi, T. Lead generation and examples: Opinion regarding how to follow up hits. *Methods Enzymol.* **2011**, *493*, 383–419.
28. Abramowitz, N.; Schechter, I.; Berger, A. On the size of the active site in proteases II. Carboxypeptidase-A. *Biochem. Biophys. Res. Commun.* **1967**, *29*, 862–867. [[CrossRef](#)] [[PubMed](#)]
29. Lu, I.L.; Mahindroo, N.; Liang, P.H.; Peng, Y.H.; Kuo, C.J.; Tsai, K.C.; Hsieh, H.P.; Chao, Y.S.; Wu, S.Y. Structure-based drug design and structural biology study of novel nonpeptide inhibitors of severe acute respiratory syndrome coronavirus main protease. *J. Med. Chem.* **2006**, *49*, 5154–5161. [[CrossRef](#)] [[PubMed](#)]
30. Mengist, H.M.; Dilnessa, T.; Jin, T. Structural basis of potential inhibitors targeting SARS-CoV-2 main protease. *Front. Chem.* **2021**, *9*, 622898. [[CrossRef](#)]
31. Horowitz, S.; Trievel, R.C. Carbon-oxygen hydrogen bonding in biological structure and function. *J. Biol. Chem.* **2012**, *287*, 41576–41582. [[CrossRef](#)] [[PubMed](#)]
32. Herschlag, D.; Pinney, M.M. Hydrogen bonds: Simple after all? *Biochemistry* **2018**, *57*, 3338–3352. [[CrossRef](#)]
33. Lippert, T.; Rarey, M. Fast automated placement of polar hydrogen atoms in protein-ligand complexes. *J. Cheminform.* **2009**, *1*, 13. [[CrossRef](#)]
34. Norel, R.; Petrey, D.; Wolfson, H.J.; Nussinov, R. Examination of shape complementarity in docking of unbound proteins. *Proteins* **1999**, *36*, 307–317. [[CrossRef](#)]
35. Lee, J.; Worrall, L.J.; Vuckovic, M.; Rosell, F.I.; Gentile, F.; Ton, A.T.; Caveney, N.A.; Ban, F.; Cherkasov, A.; Paetzel, M.; et al. Crystallographic structure of wild-type SARS-CoV-2 main protease acyl-enzyme intermediate with physiological C-terminal autoprocessing site. *Nat. Commun.* **2020**, *11*, 5877. [[CrossRef](#)]
36. Muramatsu, T.; Takemoto, C.; Kim, Y.T.; Wang, H.; Nishii, W.; Terada, T.; Shirouzu, M.; Yokoyama, S. SARS-CoV 3CL protease cleaves its C-terminal autoprocessing site by novel subsite cooperativity. *Proc. Natl. Acad. Sci. USA* **2016**, *113*, 12997–13002. [[CrossRef](#)]
37. Hsu, M.F.; Kuo, C.J.; Chang, K.T.; Chang, H.C.; Chou, C.C.; Ko, T.P.; Shr, H.L.; Chang, G.G.; Wang, A.H.J.; Liang, P.H. Mechanism of the Maturation Process of SARS-CoV 3CL Protease. *J. Biol. Chem.* **2005**, *280*, 31257–31266. [[CrossRef](#)]
38. Sheik Amamuddy, O.; Verkhivker, G.M.; Tastan Bishop, O. Impact of early pandemic stage mutations on molecular dynamics of SARS-CoV-2 Mpro. *J. Chem. Inf. Model.* **2020**, *60*, 5080–5102. [[CrossRef](#)]
39. Skipper, L. PROTEINS|Overview☆. In *Encyclopedia of Analytical Science*, 3rd ed.; Worsfold, P., Poole, C., Townshend, A., Miró, M., Eds.; Elsevier Ltd.: Amsterdam, The Netherlands, 2005; pp. 412–419, ISBN 978-0-08-101984-9.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.