Article

Structure-Based Virtual Screening to Discover Potential Lead Molecules for the SARS-CoV-2 Main Protease

Anuj Gahlawat,[⊥] Navneet Kumar,[⊥] Rajender Kumar,[⊥] Hardeep Sandhu, Inder Pal Singh, Saranjit Singh, Anders Sjöstedt, and Prabha Garg*

Cite This: https://dx.doi.org/10.1021/acs.jcim.0c00546

Read Online

ACCESS | 📊 Metrics & More | 📖 Article Recommendations | 🔵 Supporting Information

ABSTRACT: The COVID-19 disease is caused by a new strain of the coronavirus family (SARS-CoV-2), and it has affected at present millions of people all over the world. The indispensable role of the main protease (M^{pro}) in viral replication and gene expression makes this enzyme an attractive drug target. Therefore, inhibition of SARS-CoV-2 M^{pro} as a proposition to halt virus ingression is being pursued by scientists globally. Here we carried out a study with two objectives: the first being to perform comparative protein sequence and 3D structural analysis to understand the effect of 12 point mutations on the active site. Among these, two mutations, viz., Ser46 and Phe134, were found to cause a significant change at the active sites of SARS-CoV-2. The Ser46 mutation present at the entrance of the S5 subpocket of SARS-CoV-2 increases the contribution of other two hydrophilic residues, while the Phe134 mutation, present in the catalytic cysteine loop, can cause an increase in catalytic efficiency of M^{pro} by facilitating fast proton transfer from the Cys145 to His41 residue. It was observed that active site remained conserved among M^{pro} of both SARS-CoVs, except at the entrance of the S5 subpocket, suggesting sustenance of substrate specificity. The second objective was to screen the inhibitory effects of three different data sets (natural products, coronaviruses main protease inhibitors, and FDA-approved drugs) using a structure-based virtual screening approach. A total of 73 hits had a combo score >2.0. Eight different structural scaffold classes were identified, such as one/two tetrahydropyran ring(s), dipeptide/tripeptide/oligopeptide, large (approximately 20 atoms) cyclic peptide, and miscellaneous. The screened hits showed key interactions with subpockets of the active site. Further, molecular dynamics studies of selected screened compounds confirmed their perfect fitting into the subpockets of the active site. This study suggests promising structures that can fit into the SARS-CoV-2 M^{pro} active site and also offers direction for further lead optimization and rational drug design.
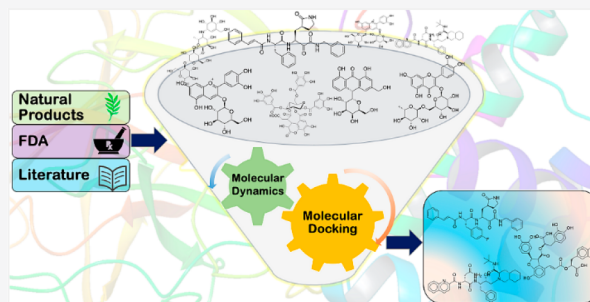
INTRODUCTION

Coronaviruses possess a positive-sense single-stranded RNA genome and cause disease in birds and mammals. Coronaviruses belongs to the family coronaviridae, and its genome size of approximately 30 kb is larger than any other reported RNA virus. Initially two strains of human coronavirus, HCoV-229E and HCoV-OC43, were identified in 1960 and 1967, respectively.[1,2] In the last two decades, some other members of the genus have been identified, such as SARS-CoV in 2003, HCoV NL63 in 2004, HKU1 in 2005, MERS-CoV in 2012, and more recently SARS-CoV-2 in 2019.[3,4] In humans, coronaviruses generally infect the lower respiratory tract and cause the common cold, but some coronavirus strains like SARS-CoV and MERS-CoV are more virulent with considerable mortality of 10−30%, and lethality has been observed not only in elderly and immunocompromised individuals but even in healthy persons.

The corona virus comprises an external envelope lipoprotein bilayer, where the three structural proteins, i.e., membrane (M), envelope (E), and spike (S) are anchored. Inside the envelope, there is another type of structural protein, i.e., nucleocapsid (N). Multiple copies of N are bound to the positive-sense single-stranded RNA genome in a continuous manner, similar to a string of beads. The lipid bilayer envelope, membrane proteins, and nucleocapsid are responsible for protection of the virus, when it is outside the host cell. S glycoproteins comprise two subunits (S1 and S2), which are involved in the entry of virus inside the host cell. The RNA of virus encodes for these structural proteins for further replication.[5]

COVID-19 has assumed pandemic status as of now. After its initialization in December 2019 in Wuhan, Hubei, China, it has spread to the whole world with >4 million reported cases and

nearly 300 000 deaths as of mid-May, 2020. Therefore, finding an effective intervention to the disease has assumed global priority. Once it was recognized that spreading sickness is due to SARS-CoV-2, one finds continuous efforts from the scientific community to develop treatment for COVID-19, either in the form of a drug therapy or a vaccine, which yet are elusive. Currently, only symptomatic therapy is being used for managing fever, respiratory distress, and associated symptoms like diarrhea.

The main protease (M^{pro}) of SARS-CoV-2 has very high sequence similarity with the M^{pro} of SARS-CoV, albeit 12 point mutations. As reported, the genomes of β-coronaviruses transcribe a ~800 kDa polypeptide, and two proteolytic enzymes are required for the cleavage into various functional proteins. One of them is 3-chymotrypsin-like-protease (3CLpro) or M^{pro}, which cleaves the polypeptide chain at 11 distinct sites. The transformation gives rise to several nonstructural proteins, considered important for virus replication. The human proteases perhaps have minimal effect, owing to their substrate specificities being distinct. Hence the inhibition of M^{pro} can be considered to be one of the targets to block viral replication.

The experimental data available on specific inhibitors of M^{pro} of SARS-CoV-2 is very limited. If the active site of M^{pro} of SARS-CoV and SARS-CoV-2 share high structural similarity, then reported inhibitors of M^{pro} of SARS-CoV can be targeted to SARS-CoV-2. To understand this, the comparative structural analysis of active sites of M^{pro} of both proteins was performed. Then, structure-based virtual screening of three data sets of different compounds (natural products isolated from diverse families of plants, coronaviruses main protease inhibitors from the literature, and FDA approved drugs) on M^{pro} target of SARS-CoV-2 was performed with the purpose to identify potential lead molecules. The approach involved docking based virtual screening, molecular dynamics (MD) simulations and binding free energy calculations. There were 73 hits that had a combo score >2.0, and also 8 different structural subclasses were identified using heatmap and manual inspection. Details are provided herein.

METHODOLOGY

**Comparative Structural Analysis of the Active Site.** The SARS-CoV-2 and its homologous protein sequences and 3D structures were retrieved from the Brookhaven Protein Data Bank.[6] After removing duplicates, multiple sequence alignment (MSA) was performed on the remaining sequences using a built-in module in MEGA X software.[7] The evolutionary phylogenetic was inferred by using the Maximum Likelihood method using MEGA X software.[7] To understand the effect of point mutations on the active site of M^{pro} of SARS-COV-2, two PDBs, 2AMQ and 6LU7, were selected for SARS-CoV and SARS-CoV-2, respectively, with both PDBs having the same crystallized peptidomimetic ligand, i.e., N3. These two PDBs were imported to the Maestro package (Schrödinger, LLC, New York, NY) to remove the extra chains, all the water molecules and also the ligand. The sequence alignment was performed for both the PDBs using Clustal Omega tool,[8] whereas details about active site residues and atoms were obtained using publicly available CASTp tool[9] with default parameters. Furthermore, in-depth analysis of active sites of both M^{pro} was performed through manual inspection in Maestro visualizer. The effect of point mutations on protein stability were analyzed using DUET,[10] mCSM,[11]

and SDM (site directed mutator).[12] The mCSM predicts the effect of mutation in the protein using the graph based signature; SDM is based on statistical potential energy method, and DUET employs a combined/consensus results by combining the predictions from two methods (mCSM and SDM) in a nonlinear way.[10−12]

Molecular Docking-Based Screening. The 3D crystal structure of SARS-CoV-2 having PDB ID 6W63 was used for structure-based virtual screening. Protein preparation wizard of the Schrödinger software package (Schrödinger, LLC, New York, NY) was used to prepare the protein, by adding missing hydrogens, assigning right bond orders, removing water molecules, and optimizing the correct orientation of hydroxyl and amino groups. Further, PROPKA and Epik modules were utilized for determination of the ionization state of the amino acids at pH 7.0 ± 2. The resulting structure was further subjected to restrained minimization with a cutoff root mean square deviation (RMSD) of 0.3 Å.

Three different libraries of chemical compounds—(i) natural products isolated from diverse families of plants, (ii) coronavirus main protease inhibitors from the literature,[13−26] and (iii) the FDA approved drugs obtained from the DrugBank database[27] (https://www.drugbank.ca/)—were used for structure based virtual screening. All these ligands were subjected to ligand preparation using the LigPrep module of Schrödinger package (Schrödinger, LLC, New York, NY). The different possible ionization states were generated using the Epik module at pH 7.0 ± 2. Almost 32 conformers of all the combinations of each ligand were generated while keeping other parameters as default. In the next step, the receptor grid generation module of the Schrödinger software package was used to define the site for molecular docking. The grid center was defined using the centroid of cocrystallized ligand, i.e., X77. The grid was standardized by redocking the cocrystallized ligand X77. Further, the docking protocol was validated by superimposing the redocked ligand with cocrystallized ligand and the RMSD was calculated between them. Molecular docking was performed into the 6W63 protein grid with defaults setup. First, approved drugs were subjected for HTVS (high-throughput virtual screening) docking to reduce computational time; later a cutoff was used to select top molecules of HTVS and then subjected to extra precision docking, while the other two classes were directly subjected to extra precision docking due to smaller size of data sets.[28] The MM-GBSA free-energies were calculated using the Prime MM-GBSA module that utilized an optimized implicit solvation model and OPLS3 force field for ligand-protein complex.[29,30] The absolute values of each energy score, i.e., docking score, XP GScore, glide gscore, MMGBSA dG bind and glide emodel were normalized/scaled between 0 to 1 using min−max normalization method using R statistical software. A combo score for each molecule was obtained by summation of normalized values of each energy scores. The combo score ranging between 0 and 5 was used to prioritize the compounds, where higher score indicates higher priority.

Scaffold Identification. Based on the combo score, molecules with cutoff value >2.0 were considered as potential hits. The 2D structures of these 73 molecules were manually inspected to identify their common classes. Further, the Tanimoto similarity score[31] was calculated for clustering the screened compounds. The atom-pair based fingerprints of compounds were obtained using the ChemmineR package[32] of
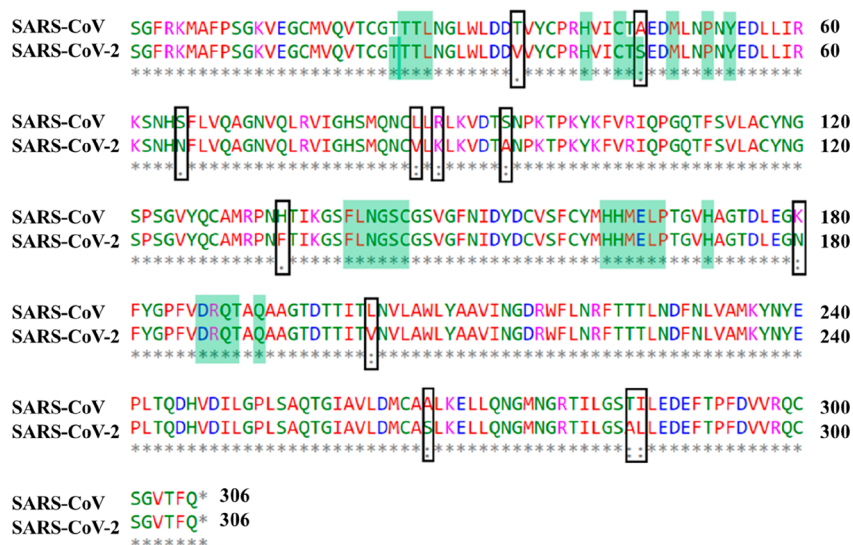
Figure 1. Sequence alignment of main protease (M^{pro}) of SARS-CoV-2 (PDB 6lu7) and SARS-CoV (PDB 2amq) using Clustal Omega. Black boxes represent 12 point mutations (Thr35Val, Ala46Ser, Ser65Asn, Leu86Val, Arg88Lys, Ser94Ala, His134Phe, Lys180Asn, Leu202Val, Ala267Ser, Thr285Ala, and Ile286Leu), and green shaded residues are active site residue of respective sequences.

R language. A heatmap was generated for visualization of clusters.

Molecular Dynamics Simulations. All-atom molecular dynamics simulations were performed using the Desmond-v6.1 module[33] of Schrödinger software package (Schrödinger, LLC, New York, NY). The system builder panel was used to prepare the initial system for MD simulations. SARS-CoV-2 M^{pro} and all docked complexes were placed in a cubic box of 1.0 nm size. Each box was solvated with TIP3P water models,[34] and negative charge on the system was neutralized using Na$^+$ ions. An ionic strength of 0.15 M was maintained by adding Na$^+$/Cl$^-$ ions to the system. Further, the solvated system was minimized and equilibrated under an NPT ensemble using the default protocol of Desmond, and it included a total of nine stages, among which there were two minimizations and four short simulations (equilibration phase) steps.[35] All well minimized and equilibrated systems were subjected to an MD run with periodic boundary conditions in the NPT ensemble using OPLS_2005 force field parameters[36] for 100 ns. During the simulation, the pressure (1 atm) and temperature (300 K) of the system were maintained by the Martyna−Tobias−Klein barostat and Nose−Hoover chain thermostat,[37,38] respectively. Binding energy between the SARS-CoV-2 M^{pro} and all docked ligands was calculated using the script thermal_mmgbsa.py.[29,30] The average binding energy between the protein and ligand was calculated from the last 20 ns of trajectory.

RESULTS AND DISCUSSION

Comparative Structural Analysis of the Active Site. The comparative sequence and structural analysis performed here will provide comprehensive structural information regarding the active binding site, which in turn can be used to develop/design target specific drugs. The SARS coronavirus M^{pro} enzyme exists in homodimer form where two protomers orient itself in a perpendicular position.[39−41] Each monomer has three domains. Domain I and II (residues 1−100 and 101−183, respectively) have a two-β-barrel fold, which is similar to chymotrypsin-like serine proteases. The third domain (residues 198−303), which is required for the

dimerization process, consists of five α-helices and a long loop connects it to domain II.[40] Zhang et al. reported that the enzyme catalytic activity depends on the dimerization which causes a structural change in the substrate binding subpocket S1.[39] The N-terminal finger (residues 1−7) also plays an important role in dimerization of M^{pro} and Ala mutation or deletion in this region causes loss in dimerization.[39] The substrate binding cleft contains a catalytic dyad of His41 and Cys145, which is present in between domain I and II at N-terminal and a hydrogen bonded water molecule to His 41 acts as a third component of the catalytic triad.[42]

Molecular phylogenetic analysis was performed on M^{pro} of all available coronavirus strains and some other homologues viruses using the Maximum Likelihood method. It was found that the SARS-CoV and SARS-CoV-2 M^{pro} sequences are the most similar (Figure S1). The sequence identity and the RMSD difference between SARS-CoV and SARS-CoV-2 were found to be 96% and 0.884 Å, respectively. During the sequence analysis, we identified 12 point mutations in the SARS-CoV-2 when compared with SARS-CoV sequence (Figure 1). To elaborate regarding the effect of these point mutations, we analyzed the crystallized structures of M^{pro} of SARS-CoV-2 (PDB ID: 6lu7) and SARS-CoV (PDB ID: 2amq) as both structures have the same N3 peptidomimetic inhibitor bound to their active site. In a biological environment, M^{pro} of both coronaviruses exhibits a glutamine site-specific cleavage of the polypeptide chain and this glutamine residue remains deeply embedded in the S1 subpocket.[39,43,44] The N3 is a peptide-like inhibitor containing a γ lactam ring as a surrogate for glutamine, which is more similar to the biological substrate of M^{pro} of coronaviruses, from this it can be expected that M^{pro} of coronaviruses exit in the active conformation with the N3 inhibitor.[39]

We employed the CASTp tool[9] to find out atoms of residues which constitute the active site surface for the N3 inhibitor in M^{pro} of both viruses. This lead to the identification of three residues, i.e. Thr24, Thr45, and Ser46 which constitute the active site of M^{pro} in SARS-CoV-2 (shaded in green in Figure 1), while these three residues do not make any contribution to the active site of M^{pro} in SARS-CoV as predicted by CASTp.
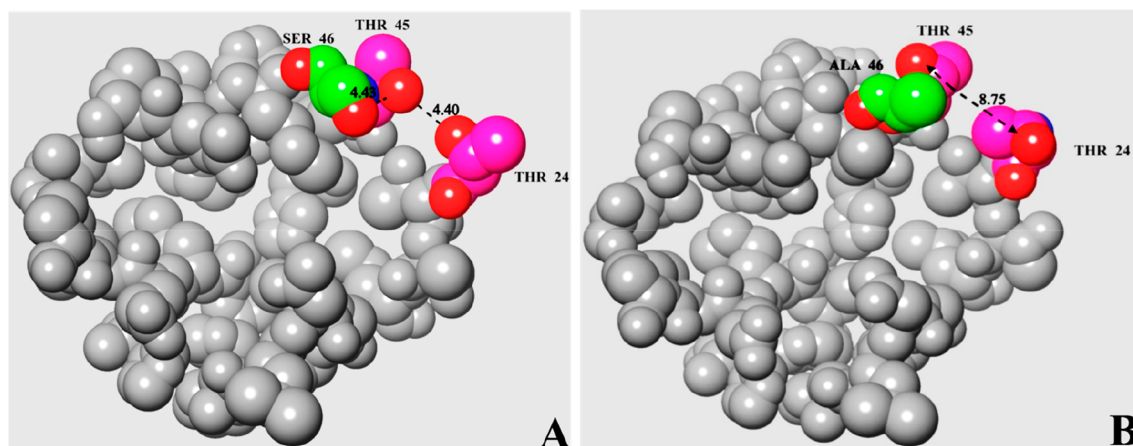
Figure 2. Active site residues of M^pro of SARS-CoV-2 (A) and SARS-CoV (B). Green represents mutated residue Ala46Ser, pink represents conserved residues (Thr24 and Thr45), which were affected by Ser46 mutation, and red shows oxygen atom of specific residues. Remaining positionally conserved residues are shown in gray.

All these three residues are hydrophilic in nature and present at the entrance of the S5 subpocket of SARS-CoV-2. The M^pro active site surface area and volume of SARS-CoV-2 were found to be 351 unit^2 and 319 unit^3, while SARS-CoV had 293 unit^2 and 224 unit^3, respectively. This considerable differences in surface area and volume of active site may be explained by the presence of the mutated residue Ser46 in SARS-CoV-2. The mutated residue Ser46 attracts two adjacent neighboring residues (Thr24 and Thr45) toward the active site, and it increases the active site surface area and volume in SARS-CoV-2, as compared to SARS-CoV. This can be evident from the fact that the distance between the side chain oxygens of Ser46 and Thr45 was found to be 4.43 Å in SARS-CoV-2 which was absent in case of SARS-CoV protein, and simultaneously, the distance between the side chain oxygens of Thr45 and Thr24 decreased to 4.40 Å, compared to 8.75 Å in the SARS-CoV protein (Figures 2A and B). These observations imply that the hydrogen bonding could be the reason for the participation of these residues (Thr24 and Thr45) in the active site of SARS-CoV-2. If these interactions exist in M^pro of SARS-CoV-2, then it can reduce the flexibility of the active site as compared to SARS-CoV. It also can be concluded that the active site entrance of SARS-CoV-2 is more hydrophilic than that of SARS-COV M^pro.

We also performed atomwise comparison of atoms that were present on active site surfaces, and essentially, the same atoms were present in the active sites of both M^pro (shown in gray color spheres in Figures 2A, 2B). While the mutation Ser65Asn present in the β-sheet near the His41 loop in SARS-CoV-2 has no impact on residues of the active site, small changes could be seen in atoms of these residues. Therefore, it can be concluded that active sites of M^pro of SARS-CoV-2 and SARS-CoV viruses have similar structural topology.

To explore the effect of point mutations within the catalytic dyad (His41 and Cys145 residues), we analyzed two loops, on each loop a catalytic residue was present (Figure 3). The two nearest mutation present on these loops (one at His loop Ala46Ser and another at Cys loop His134Phe) were analyzed to study the change at the catalytic dyad. Effect of the Ser46 mutation remained limited to the SARS-CoV-2 active site entrance, due to its similar size to the Ala residue, since it does not produce any significant change in the His loop. It is clear from Figure 3 that the position of catalytic residue His46
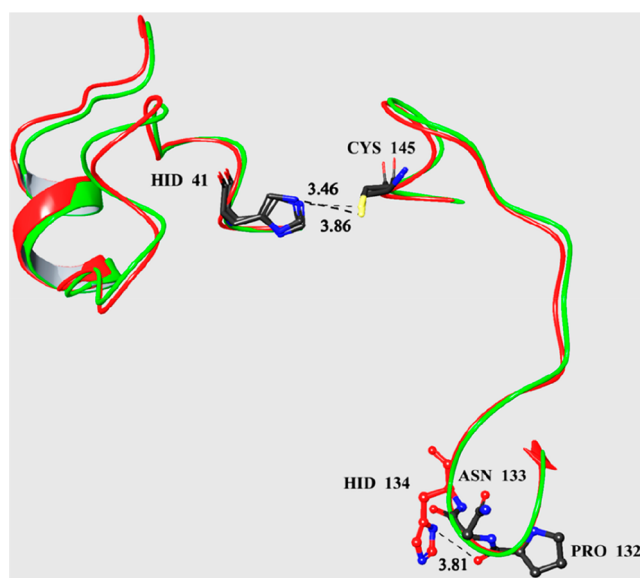
Figure 3. Superimposition of catalytic loops (i.e His loop (39−65 residues) and Cys loop (130−147 residues)) of SARS-CoV-2 (green) and SARS-CoV (red) protein. Mutated His134 residue of SARS-CoV shown in red, while other unmutated residues are shown in gray.

remains the same in the SARS-CoV proteins. When we analyzed the Cys loop mutation i.e. His134Phe, it was observed that the phenylalanine residue is very distinct from histidine residue and produced a significant change at Cys145 residue.

In M^pro of the SARS-CoV protein, the nitrogen of imidazole ring of His134 residue interacts with the backbone carboxyl group of Pro132 residue via the hydrogen bond as the distance between these atoms was found to be 3.81 Å. Jeffrey suggested that there are three categories of hydrogen bonds, the 3.2−4.0 Å donor−acceptor distance can be categorized as a weak hydrogen bond, and the distance can be increased in crystal structures as no hydrogen atom was present there.[45] This interaction in the SARS-CoV protein pulled the whole Cys loop slightly downward and that increased the distance between the His41 and Cys145 residues to 3.86 Å. Whereas in SARS-CoV-2, the His134 is mutated with Phe134 and phenylalanine does not have the capability to form a hydrogen

bond. It was also observed that it was not interacting with any other surrounding residues. Therefore, this mutation set Cys loop free in M^{pro} of SARS-CoV-2, and this may be the reason why the distance between the catalytic dyad residues get closer i.e. <3.46 Å. The reduced distance can result in faster proton transfer from sulfur of Cys145 to the nitrogen of the imidazole ring of the His41 residue, which is the first step of a catalytic mechanism and the change can accelerate the catalytic proteolysis of polypeptide chains.

While other mutations (Thr35Val, Ser65Asn, Leu86Val, Arg88Lys, Ser94Ala, and Lys180Asn) present in domains I and II can cause backbone strain and alter surrounding interactions, any direct effects of these mutations on the active site were not observed. Domain III mutations (Leu202Val, Ala267Ser, Thr285Ala, Ile286Leu) can produce a significant change in dimerization process. Lim et al. reported that mutations of Ser284, Thr285, Ile286, with the Ala residue in SARS CoV M^{pro} protein caused a 3.6-fold increase in activity.[46] In M^{pro} of the SARS-CoV-2 protein, there is an Ala285 mutation and simultaneously other nearby mutations that can allow tight packing of domain III of different monomers.

The impact of point mutations on the stability of protein can be predicted by estimating the free energy change ($\Delta\Delta G$). The two important point mutations, i.e. Ala46Ser and His134Phe, which have shown a significant changes in active site of M^{pro} were predicted for change in the stability. It was found that mutation His134Phe increases the stability, while mutation Ala46Ser imparts a very little destabilization effect in M^{pro} (Table S1). Generally, destabilizing mutation have free energy change $\Delta\Delta G < -0.5$ kcal/mol, stabilizing mutation have free energy changes $\Delta\Delta G > 0.5$ kcal/mol and values $0.5 \leq \Delta\Delta G \leq 0.5$ is considered to have neutral stability. The higher the magnitude of $\Delta\Delta G$ either on the positive or negative side of the scale, the higher the stability/instability of the protein.[47]

Here we suggest that one should perform a detailed study with the Apo form of M^{pro} using molecular simulation/ quantum methods on catalytic residues and altered residues present at active site entrance (Thr45, Ser46, and Thr24 residues) to precisely understand the differences between the SARS-CoV and SARS-CoV-2 main protease.

Molecular Docking Based Virtual Screening. Recently, the 3D crystal structure of SARS-CoV-2 M^{pro} was resolved with a noncovalent inhibitor (PDB ID 6W63) was used for molecular docking study. Like SARS-CoV M^{pro}, SARS-CoV-2 M^{pro} also has a wide active site with five subpockets, named as S1 to S5. The information about the residues which constitute each subpocket are given in Table 1, and a few common residues were shared by walls of subpockets. The S2 subpocket has a catalytic center i.e. sulfur from Cys145 and nitrogen from

Table 1. Information about Different Residues Belonging to Specific Subpockets of M^{pro} SARS-CoV-2

subpocket	residues
S1	Cys145, His164, Met165, Glu166, His172, His163, Phe140, Leu141, Ser144
S2 (catalytic center)	Cys145, Gly143, Leu27, Thr26, Thr25, His41,
S3	His41, His164, Met165, Asp187, Arg188, Gly189, Tyr54, Met49
S4	Asp187, Met165, Glu166, Leu167, Pro168, Gln192, Thr190, Arg188
S5	Thr25, Thr26, Thr24, Thr45, Ser46, Cys44, Met49, His41

imidazole ring of His41 residue. The mutation Ala46Ser and its effect related to the contribution of Thr24 and Thr45 to active site of M^{pro} of SARS-CoV-2 was found to be in S5 subpocket. While the remaining subpockets S1, S2, S3, and S4 remain unchanged.

First, the cocrystallized ligand was redocked into the active site of M^{pro} and it was found that the redocked conformation of ligand perfectly superimposed on the cocrystallized ligand (PDB ID: 6W63, ligand X77) with a RMSD value of 0.636 Å (Figure S2). The crystallized ligand mainly interacts with four subpockets (S1−S4). It forms two hydrogen bonds with the Gly143 and Glu166 residues of the S2 and S1 subpockets, respectively, and a $\pi-\pi$ interaction with His41 of the S2 subpocket. Our analysis of others crystal structures of SARS-CoV-2 M^{pro} identified other potential residues Gly143, His163, Glu166, Thr190, Gln192, Cys145, His41, Gln189, Thr26, and Asn142, which can also interact with inhibitors. The validated docking protocol was further used for molecular docking-based virtual screening of three different data sets. A total of 73 molecules were obtained with the combo score cutoff >2.0, and more detailed information about their score and chemical structures are provided in Table S2. We have listed the four top screened compounds with their combo score from each data set, their interaction, and binding details with subpockets in Table 2.

FDA Approved Drugs. A total of 2454 FDA approved drugs were screened using HTVS (high-throughput virtual screening). A docking score cutoff of −5.5 kcal/mol was used to select top 1190 potential molecules. The compounds like dyes, metal-containing, and diagnostic chemicals were manually removed. The remaining 1111 molecules were further subjected to molecular docking using glide in extra precise mode (XP). The MM-GBSA calculations were performed on the XP docked complexes that have a XP GScore score lower than −7 kcal/mol (201 compounds). Based on the combo score cutoff (>2.0), a total of 21 hits were obtained, including antiviral drugs, e.g., saquinavir, atazanavir, nelfinavir, lopinavir, and indinavir. Interestingly all these drugs are HIV protease inhibitors. Yamamoto et al. performed quantitative RT-PCR analysis of nelfinavir and ritonavir antiviral drugs on SARS-CoV virus and found that nelfinavir effectively inhibited SARS-CoV replication.[48] No other experimental activity data was found for these anti-HIV drugs for the coronaviruses, but the lopinavir drug has been used in many countries for COVID-19 patients. Acarbose is a linear oligosaccharide which binds to three subpockets (S2, S3, and S4) of SARS-CoV-2 M^{pro}. Octreotide, a peptide drug, binds to all subpockets due to its high flexibility. There were some other FDA-approved drugs which also have active site mimicking chemical scaffold. The screened compounds were observed to form key interactions with Cys145, Asn142, Glu66, Gln189, Arg188, and other residues.

Coronaviruses M^{pro} Inhibitors from the Literature. A total of 144 compounds were collected from the literature, among them 11 MERS-CoV, 26 SARS-CoV-2, and the remaining were SARS-CoV main protease inhibitors. The experimental activity has been reported for almost all SARS-CoV and MERS-CoV inhibitors, whereas only computational evidence was available for the SARS-CoV-2 inhibitors. SARS-CoV M^{pro} inhibitors were targeted on the main protease of SARS-CoV-2, by considering their active site similarity (discussed in the comparative analysis section). A total of 30 hits were identified from the literature. All five Chinese herbal medicines

Table 2. Molecular Docking Results of Top Four Hits from All Three Datasets with Their Interaction with Residues and Subpockets of M^{pro}

compound	combo score	data	hydrogen bond	$\pi-\pi$ interaction	subpockets
lithospermic acid B	4.759	literature	Cys145, Gly143, His163, Glu166, Gln189, Thr190, Gln192		S1, S3, S4, S5
chebulinic acid	4.756	NP	Gly143, His163, His164, Glu166, Gln189, Thr190, Gln192	His41	S1 to S5
rutin	4.088	literature	Thr24, Thr26, His41, Leu141, Gly143, His163, Glu166, Arg188		S1 to S5
delphinidin-3,5-diglucoside	4.017	NP	Gly143, His163, His164, Glu166, Gln189, Thr190, Gln192	His41	S3 to S5
cyanidin-3,5-diglucoside	3.979	NP	Cys44, Asn142, His164, Glu166, Thr190, Gln192	His41	S3 to S5
acteoside	3.964	NP	Thr26, Asn142, Cys44, Glu166, Thr190, Gln192		S2, S3, S4
neonuezhenide	3.708	literature	Asn142, Gly143, His163, Glu166, Thr190, Gln192		S1 to S4
acarbose	3.547	FDA	Thr24, Thr26, Gly143, Cys145, Glu166, Arg188, Thr190		S2, S3, S4
specnuezhenide	3.33	literature	Cys44, Phe140, Glu166, Thr190		S1, S3, S4, S5
saquinavir	3.146	FDA	Phe140, Asn142, Cys145, His163, Glu166, Gln189	His41	S1 to S4
octreotide	3.128	FDA	Thr26, His41, Cys44, Glu166	His41	S1 to S5
colistin	3.028	FDA	Thr26, His41, Leu141, Asn142, Glu166, Gln189		S1 to S5
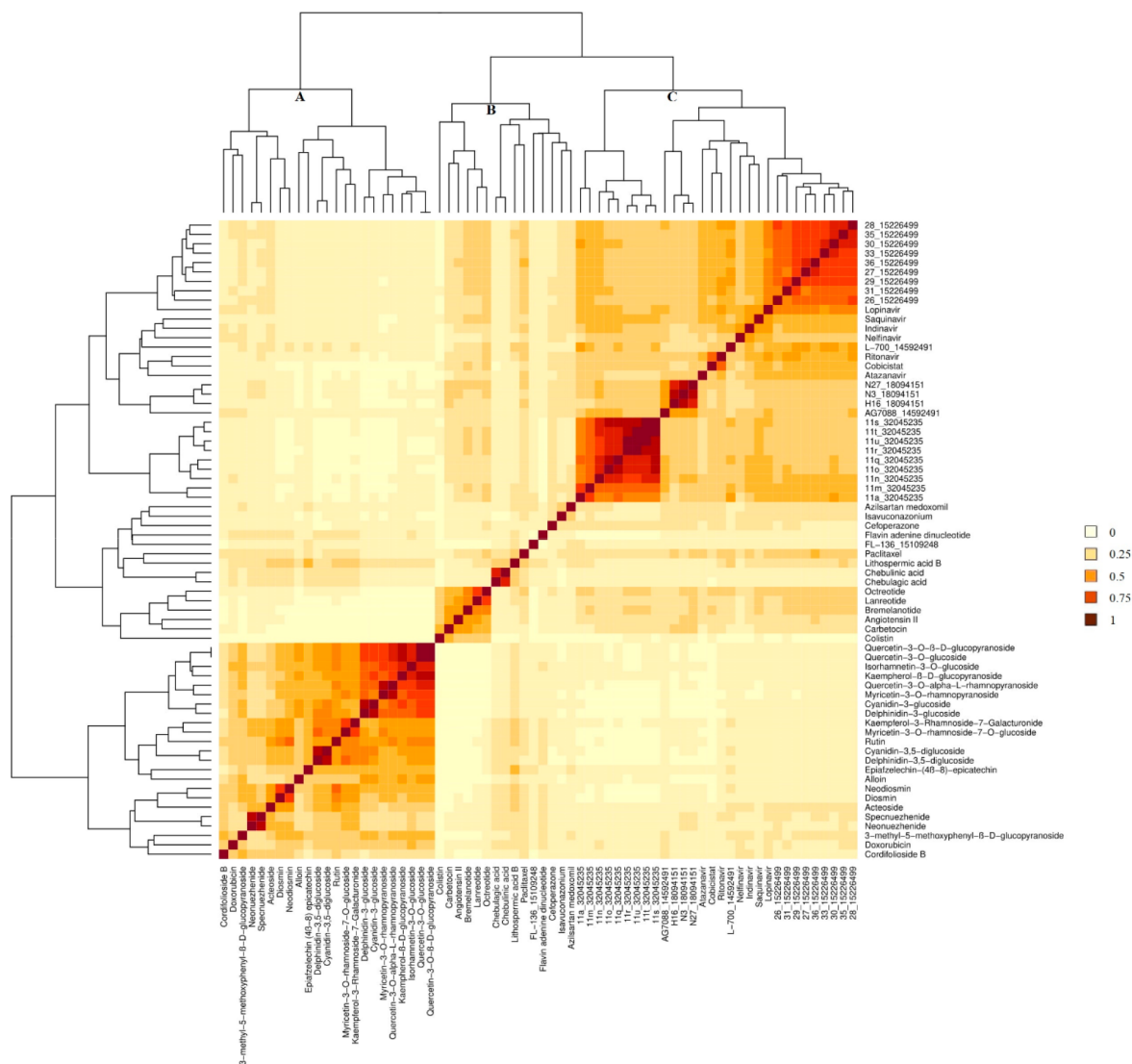
Figure 4. Heatmap generated based on Tanimoto similarity score.

(lithospermic acid B, rutin, neonuezhenide, specnuezhenide, and neodiosmin) reported for SARS-CoV-2 M^{pro} have performed well over SARS-CoV and MERS inhibitors, whereas SARS-CoV inhibitors were also among the best screened compounds. Among these drugs, rutin is an FDA-approved drug use to decrease capillary fragility. This drug was also present in FDA-approved database, and it interacts with all the subpockets. Compounds from this database form pivotal interactions with Cys145, Gly166, Gln189, His41, Thr190, Thr24, Gly143, and other residues.

Natural Product Data Set Isolated from Diverse Families of Plants. An in-house library of 138 compounds isolated from

Figure 5. Sunburst showing screened compounds grouped into scaffolds and substructure classes.

various plants of diverse families was also screened. The screened compounds belonged to various phytochemical classes and included anthocyanins and anthocyanidins, flavonoids, flavonoid glycosides, phenolic acids, phloroglucinols, and phloroglucinol-terpene adducts, triterpenoids, alkaloids, and phenyl propanoids. This natural product library having compounds belonging to diverse chemical classes was screened for the first time. A total of 18 compounds were screened from the natural database. Among them, chebulinic acid interacts with all subpockets, whereas other compounds (delphinidin-3,5-diglucoside, cyanidin-3,5-diglucoside, and acteoside) shown in Table 2, have interactions with limited subpockets. It was found that residues like Gly143, Asn142, Glu166, Gln192, His41, and other residues mainly form hydrogen bond interactions with the compounds of this data set, whereas in some compounds $\pi-\pi$ interaction with His41 residue was also seen.

From Table 2, it is clear that compounds from three different data sets have interactions with key residues, but variability was been seen toward their binding to various

subpockets. A hydrogen bonding interaction of rutin and acarbose with the Thr24 residue again asserts that in M^{pro} SARS-CoV-2, the contribution of Thr24, Thr45, and Ser46 has been increased with respect to SARS-CoV. In SARS-CoV-2 M^{pro}, various electronegative atoms are present that can act as donor/acceptor to form hydrogen bonds, but only His41 residue can act as a source for both $\pi-\pi$ interaction and hydrogen bonds. We had selected six molecules from different data sets and subjected to 100 ns molecular simulation for more detailed analysis.

Scaffold Identification. A total of 73 molecules obtained (Combo score >2.0) via structure based virtual screening were manually inspected to identify their eight common classes (Table S2). These common classes are comprised of similar scaffolds which were identified using manual visualization of small and large clusters present on a heatmap generated using Tanimoto score matrix (Figure 4). The eight classes are as oligopeptide, one tetrahydropyran/sugar rings, two tetrahydropyran ring, tetrahydropyran/sugar and gallate, cyclic peptides, dipeptides tripepetide and miscellaneous. The
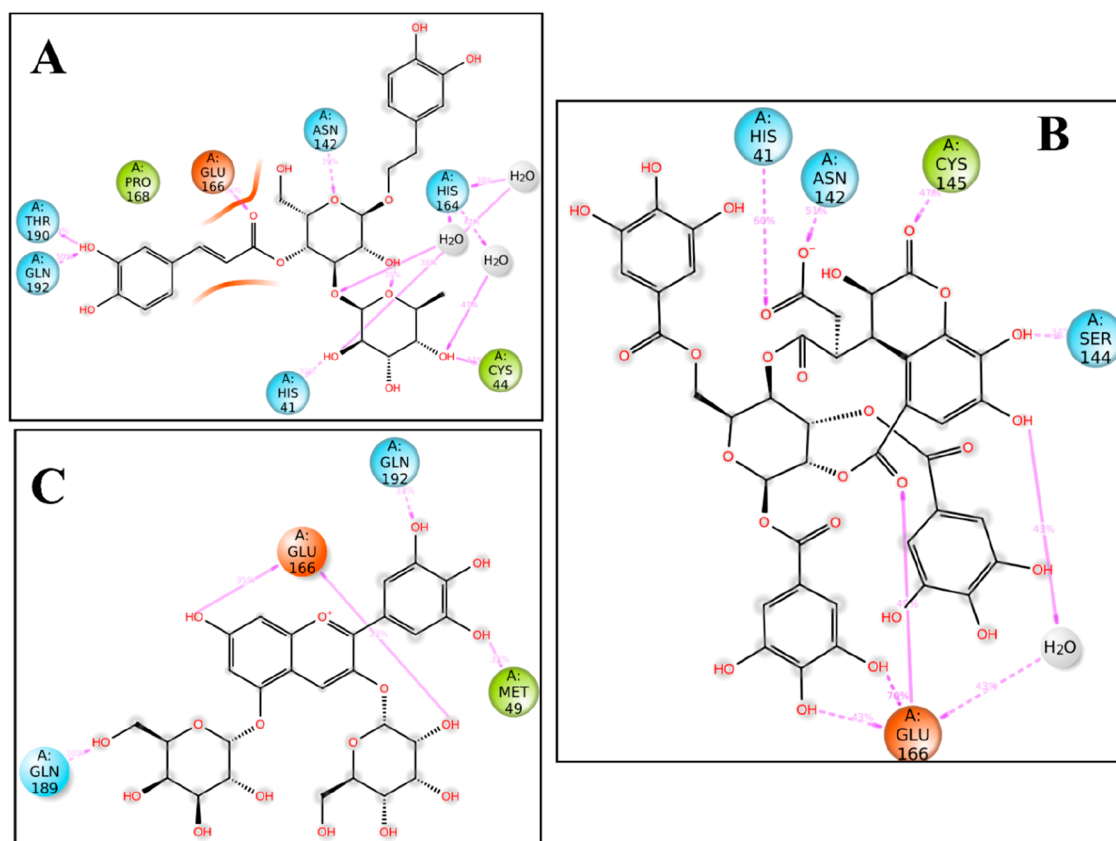
Figure 6. 2D interaction diagram of SARS-CoV-2 M^{pro} with (A) acteoside, (B) chebulinic acid, and (C) delphinidin-3,5-diglucoside.

sunburst chart in Figure 5 summarizes the information on all scaffolds belonging to their respective subclasses along with the name of compounds and their source. Mainly two classes, i.e. one and two tetrahydropyran ring/sugar which contribute a considerably larger portion among the top 73 hits, have a central ring and other three/four rings/side chains attached to its different position through linker. Peptide-like molecules has been classified into four categories viz oligopeptide, dipeptide, tripeptide, and cyclic peptides also constitiute one of the major group. All the scaffolds belonging to peptidomimetic class except the cyclic peptide, have aromatic rings joined by linker having amide bonds, and the flexibility of linkers let aromatic ring occupy important subpockets present in the active site of M^{pro}. The molecules belonging to cyclic peptide class are generally bigger in size and have a flexible cyclic ring with side chains and aromatics groups attached around to it. One remarkable observation shows that compounds from all three data sets are distributed among all major classes and subclasses. The potential noncovalent M^{pro} inhibitors can be rationally designed using these scaffolds, since all these scaffolds were obtained from the top hits from data sets having diverse range of compounds. To design new compounds from these scaffolds, one must consider subpocket information and the R groups present in various scaffolds represent major sites for change whereas other minor change can also be done at other sites. The substitutions can be made with numerous six/five member heterocyclic ring, and aliphatic side chains containing hydrophilic functional groups.

Molecular Dynamics Simulations. The MD simulation was performed for some selected hits for further structural insight about the active site binding mode analysis that guide to optimization of screened hits and rational design. The top

screened hits from natural products isolated from diverse families of plants, coronaviruses main protease inhibitors from the literature, and FDA approved drugs i.e. acteoside, chebulinic acid, and delphinidin-3,5-diglucoside, saquinavir, lithospermic acid B, and 11m_32045235, were subjected to MD simulations, and the residual interaction and structural fluctuations were monitored.

Binding Mode Analysis of Different Ligands with M^{pro}. *Acteoside, Chebulinic Acid, and Delphinidin-3,5-diglucoside.* Acteoside contains hydroxytyrosol, caffeic acid, and rhamnose attached to glucose unit. The 2D interaction diagram of acteoside with M^{pro} is summarized in Figure 6A. The 3-hydroxyl group and oxygen of the carbonyl group present in the caffeic acid moiety exhibit two hydrogen bonds with Thr190, Gln192, and Glu166 resulting in a stable interaction of the caffeic acid moiety with M^{pro}, which was consistent and maintained during the entire simulation run (Figures S3A and S5A). The rhamnose moiety of acteoside makes interactions with multiple residues. The C-2 hydroxy group in rhamnose exhibits one hydrogen bond with catalytic residue His41 and one water-mediated hydrogen bond with His164. The protein−ligand interaction histogram and time-line (Figures S3A and S5A) show that the above interactions were consistent and stable during the simulation. Further, the C-4 hydroxy group of rhamnose also showed interaction with two residues, one hydrogen bond directly with Cys44 and one water-mediated hydrogen bond with His164. Asn142 showed hydrogen bonding with the glucose moiety.

Chebulinic acid is a gallotannin consisting of a β-D-glucose unit attached with three galloyl groups (C-1, 3, and 6) and a chebuloyl group (C-2 and 4). The catalytic residues His41 and Cys145 showed hydrogen bonding interaction with chebuloyl
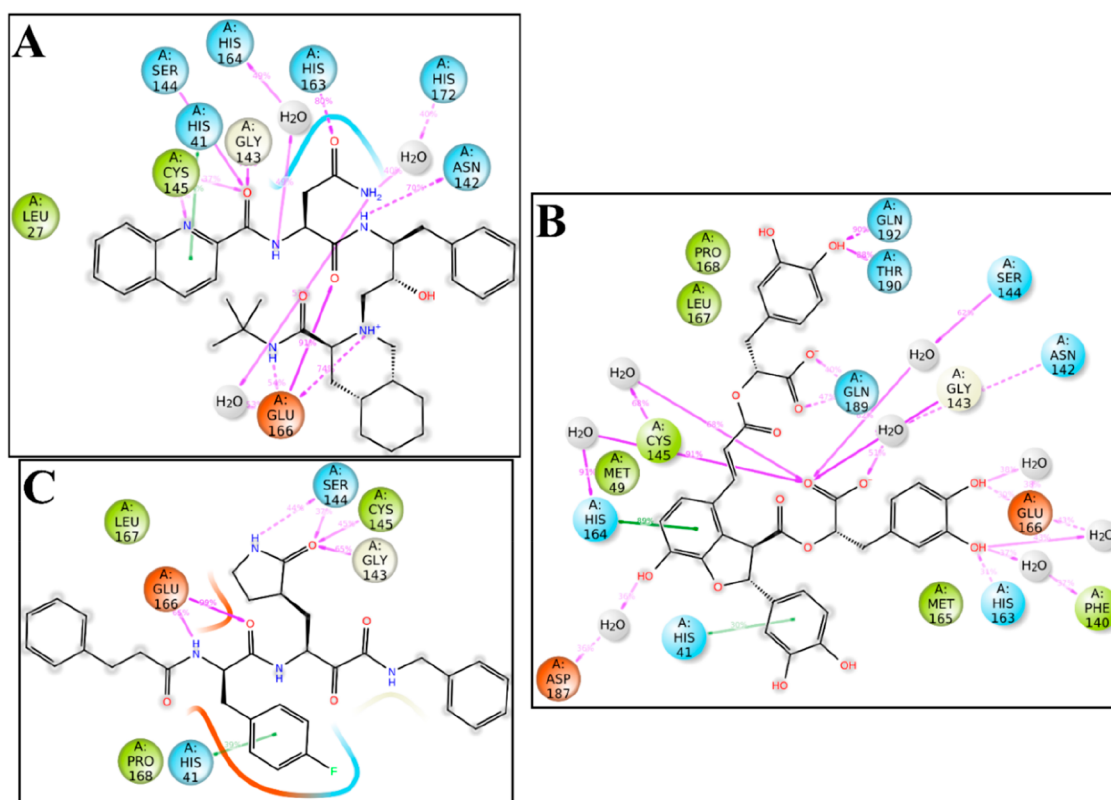
Figure 7. 2D interaction diagram of SARS-CoV-2 M^{pro} with (A) saquinavir, (B) lithospermic acid, and (C) 11m_32045235.

moiety that was maintained for 60% and 47% of total simulation time suggesting this interaction as stable and consistent (Figure 6B and S5B) The free hydroxyl groups present in galloyl moiety attached at C-1 of β-D-glucose showed a stable hydrogen bond interaction with Glu166 (Figure 6B). Apart from interacting with galloyl group, Glu166 also makes one hydrogen bond and one water-mediated hydrogen bond with the chebuloyl moiety.

Delphinidin-3,5-diglucoside is an anthocyanin found in many berries and contains two sugar molecules that are attached to the benzopyrylium nucleus of delphinidin. The 3,4,5-trihydroxyphenyl ring attached directly to benzopyrylium nucleus in delphinidin interacts with two residues of M^{pro}. The C-3′ and C-5′ hydroxyl group of anthocyanin interact with Gln192 and Met49, respectively (Figure 6C). The C-7 hydroxyl group in delphinidin moiety makes hydrogen bond with Glu166, and this interaction occurs more than 30% of the simulation time. Another interaction of residue Glu166 and the glucose molecule attached at C-3 of delphinidin moiety via hydrogen bond was also observed during the simulation. Another sugar molecule attached to the C-5 position of the delphinidin moiety forms a hydrogen bond with Gln189. The individual occurrence of the interaction between the sugar moiety and the M^{pro} residue was found to be more than 30% of the total simulation time. The protein ligand contact histogram and 2D interaction diagram (Figures S3B and 6C) shows that Glu166 is one of the prominent residue interacting with the both delphinidin and sugar moiety present in delphinidin-3,5-diglucoside.

Saquinavir, Lithospermic Acid B, and 11m_32045235. The 2D interaction diagram in Figure 7A shows that saquinavir's quinaldic acid moiety interacts with both catalytic residues of M^{pro}. This moiety of saquinavir forms a hydrogen

bond with Cys145 and π−π stacking with His41. The carbonyl group of the same moieties also forms a hydrogen bond with Gly143 and Ser144. One of the interactions between Glu166 and the different chemical moiety of saquinavir was observed to be stable and consistent during the whole simulation time. This residue, Glu166, interacts with the *tert*-butylcarbamoyl, octahydroisoquinoline, and asparginyl moieties of the saquinavir.

The lithospermic acid B or salvianolic acid B is one of the major components of *Salvia miltiorrhiza* and used as popular herbal traditional medicine in Asian countries. This molecule mainly has dihydrobenzofuran nucleus directly attached to one catechol and two 3-(3,4-dihydroxyphenyl) lactic acid moiety. The imidazole ring of catalytic residue His41 make π−π stacking interaction with the catechol moiety. Another catalytic residue Cys145 showed water mediated hydrogen bond with the 3-(3,4-dihydroxyphenyl) lactic acid moiety (Figure 7B). The hydroxyl groups in 3-(3,4-dihydroxyphenyl) lactic acid moiety attached to furan ring forms multiple hydrogen bonds and a water-mediated hydrogen bond with Glu166, His163, and Phe 140. The protein−ligand contact histogram and timeline (Figures S4B and S7A) obtained from the MD simulations show that the interactions of Glu166, His163, and Phe140 with lithospermic B were maintained consistently during the whole simulation time. Another lactic acid derivative moiety attached to the benzene ring of the central dihydrobenzofuran nucleus interacts with Gln189, Thr190, and Gln192 via a hydrogen bond.

The molecule 11m_32045235 belongs to the peptidomimetic class and is mainly a derivative of α-ketoamide. During MD simulation, it was observed that both catalytic residues, i.e. His41 and Cys145 of M^{pro} interact with 11m_32045235. The imidazole ring of the catalytic residue His41 form the π−π

stacking interaction with fluorobezyl moiety of 11m_32045235 and other catalytic residues Cys145, make hydrogen bond with oxygen of pyrrolidone moiety of 11m_32045235. The protein−ligand contact timeline (Figure S7B) shows that the interaction of both catalytic residues with the 11m_32045235 remain consistent and stable during the whole simulation. Apart from interacting with catalytic residue, the pyrrolidone moiety of 11m_32045235 also forms a hydrogen bond with Gly143 and Gly144. Another important interaction between the α-ketoamide and Glu166 was also observed during MD simulation. All the interactions of proteins with different ligands are summarized in the form of a 2D interaction diagram in Figures 6 and 7.

A perusal of chemical structures of the active natural products i.e. lithospermic acid B, chebulinic acid, rutin, delphinidin-3,5-diglucoside, acteoside, and myrecetin-3-O-rhamnopyranoside suggested that all these active compounds possessed an ortho-dihydroxy phenyl ring as a common structural feature. It may be inferred that two hydroxyl groups in ortho position on an aromatic ring seem to be essential for this activity. Four of these active compounds also contained a sugar moiety suggesting their possible role in demonstrating activity. The sugar ring adopts different pucker conformations, so it might be a specific conformation that binds into the subpocket. Similarly, saquinavir and 11m_32045235 also contain multiple amide linkages capable of showing interactions through hydrogen bonds with active site residues. One more thing can be inferred from the analysis that interaction of different ligands with aforementioned residues like His41, Cys145, His164, Glu166, Thr190, and Gln192 are already reported to be important in the various computational and experimental study.[49−51]

Analysis of Structural Stability and Binding Free Energy. The RMSD plot (Figure 8) of simulated complexes shows that all the M^{pro}−ligand complexes fluctuate within the 3 Å during the whole period of simulation. The comparison of average RMSD value of apo-M^{pro} with M^{pro} complexes (Figure 8) indicates that the binding of ligands overall stabilized the protein. The RMSF of M^{pro}-apo state provides a baseline for comparing the fluctuations with different ligand bound complex. It was observed that the lithospermic acid B increases the RMSF value for the residue 50−62 in domain-I and residues 272−279 belonging to domain-III of M^{pro} are responsible for the dimerization process. The RMSF value of the carboxy terminus loop residue in apo-M^{pro} was found to be up to 9 Å, but in the presence of ligands, a marked decrease in RMSF value was observed (Figure S8). An overall MD simulation indicates the stabilization of end terminus loop in the presence of the ligand.

The binding free energy of all simulated complexes with respect to time (last 20 ns) is plotted in Figure S9 and Table 3 which depicts the average binding free energy and its different contributing terms. The plot shows that the binding energies of all complexes with respect to time fluctuate around a stable value.

Our study shows the ligands like lithospermic acid B and 11m_32045235 which have comparatively higher binding free energy calculated from last 20 ns of simulations occupy four to five subpockets of the active site in M^{pro}. Both ligands show interactions with different residues including a catalytic one via a hydrogen bond and $\pi−\pi$ stacking. These interactions maintain the occupancy of particular moiety to a specific subsite of pocket. It was also observed that a ligand like
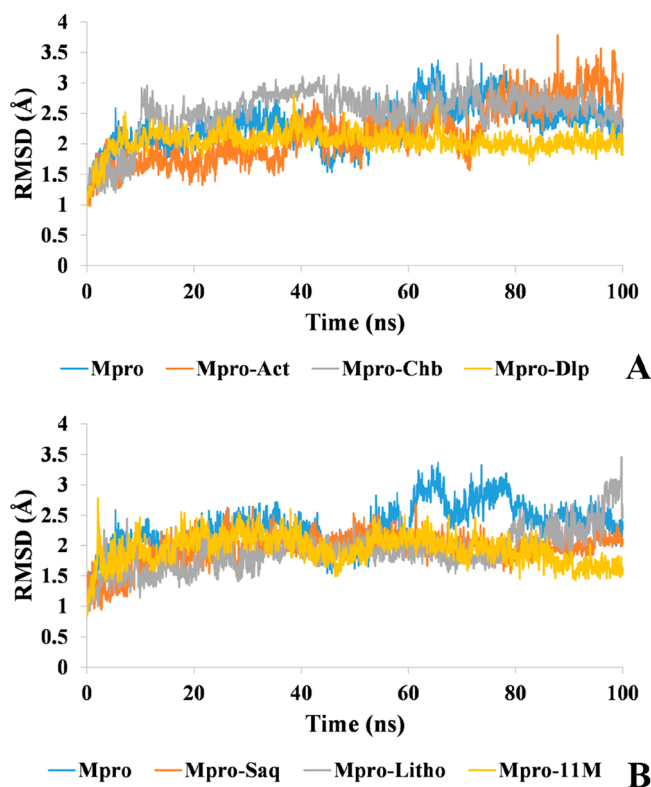
Figure 8. Showing the RMSD plot of (A) M^{pro}-apo, M^{pro}-acteoside, M^{pro}-chebulinic acid, and M^{pro}-delphinidin-3,5-diglucoside and (B) M^{pro}-apo, M^{pro}-saquinavir, M^{pro}-lithospermic acid, and M^{pro}-11m_32045235 plotted with respect to simulation time

acteoside, which occupies the three subpockets S4, S2, and S5 by caffeic acid and rhamnose moieties, has comparatively lower binding free energy. The reason behind the lower binding affinity of acteoside with M^{pro} is due to the higher flexibility of solvent exposed terminal pyrocatechol group. This group would have probably weakened some of the critical interactions during the simulation resulting in lesser binding free energy as compared to ligands which tops the list in term of binding free energy. So our MD simulation study on these ligands suggests that the higher the occupancies of subpockets by different moieties of the ligand, the better the M^{pro} inhibitor. But even a ligand that occupies lesser subpockets should have strong interactions via hydrogen bond and $\pi−\pi$ stacking, which stabilizes the ligand inside the subpockets. This study also suggest that terminal aromatic groups should be connected to the rest of ligand with a linker having an optimum number of carbons which will limit the flexibility, extra deviation, and maintain the critical interactions.

■ CONCLUSION

The 3D structural information of one of the crucial enzyme M^{pro} for protein synthesis in SARS-CoV-2 and SARV-CoV with peptide-like inhibitor was used for the comparative structural analysis of the active site. The study revealed that the M^{pro} active sites of SARS-CoV-2 and SARS-CoV are highly conserved. The comparative analysis study showed that out of 12 point mutations, 2 mutations present near the active site can cause considerable changes in SARS-CoV-2. The mutation of Ala46Ser in SARS-CoV-2 protein can attract two surrounding neighbor residues Thr24 and Thr45 and increase their contribution to the entrance of S5 subpocket. However,

Table 3. Average ΔG_{bind} (kcal/mol) and its Contributing Energy Terms for the Top Six Hits against SARS-CoV M^{pro} Calculated from MD Trajectories (last 20 ns)

complex	avg ΔG_{bind} coulomb[a]	avg ΔG_{bind} covalent[b]	avg ΔG_{bind} Hbond[c]	avg ΔG_{bind} lipo[d]	avg ΔG_{bind} packing[e]	avg ΔG_{bind} solv GB[f]	avg ΔG_{bind} vdW[g]	ΔG_{bind}total[h]
M^{pro}-Act	−27.18	−0.52	−1.59	−34.00	−1.04	33.53	−42.98	−73.75 ± 6.25
M^{pro}-Chb	−23.11	2.61	−3.82	−20.90	−0.45	26.63	−49.21	−68.24 ± 8.55
M^{pro}-Dlp	−50.60	0.16	−1.24	−21.89	−0.51	38.56	−34.42	−69.95 ± 5.16
M^{pro}-Saq	−72.76	3.07	−3.93	−32.96	−0.94	58.68	−55.21	−104.06 ± 6.30
M^{pro}-Litho	−14.69	0.31	−3.53	−52.20	−2.70	19.30	−65.18	−118.69 ± 6.87
M^{pro}-11M	−30.58	1.94	−1.15	−47.40	−1.72	33.10	−70.25	−116.06 ± 9.46

[a]Coulomb energy, [b]Covalent binding energy. [c]Hydrogen bonding correction. [d]Lipophilic energy. [e]$\pi-\pi$ packing correction. [f]Generalized Born electrostatic solvation energy. [g]van der Waals energy. [h]Total binding free energy.

these hydrophilic changes at entrance of the S5 subpocket cannot disturb the entry of biological substrate or inhibitors, because M^{pro} has a wide active site at its surface. Another mutation His134Phe, present on the catalytic Cys loop, can affect catalytic efficiency of M^{pro} of SARS-CoV-2 by facilitating fast proton transfer from the Cys145 to His41 residue. Since our analysis and comparison of binding pocket between SARS-CoV and SARS-CoV-2 suggest that SARS-CoV M^{pro} inhibitors can also be used to target the SARS-CoV-2. This is the reason that few potential SARS-CoV inhibitors have been identified as the top hits for SARS-CoV-2 in our docking based virtual screening and MD simulations based on binding free energy calculations. The docking-based virtual screening of three different data set (i) natural products data set isolated from diverse families of plants, (ii) coronaviruses M^{pro} inhibitors from the literature, and (iii) FDA approved drugs led to the identification of a number of potential hits. The analysis of the top-docked complex showed that Gly143, His163, Glu166, Thr190, Gln192, Cys145, His41, Gln189, Thr26, and Asn142 are potential residues that can form hydrogen bonds with the ligands, whereas the $\pi-\pi$ interaction can only be formed by the His41 residue. A total of 73 molecules having the Combo score >2.0 obtained from docking-based virtual screening in combination with MM-GBSA binding free energy were classified in eight classes having different kind of scaffolds in each subclass. These scaffolds which constitute eight major classes were identified using manual visualization of small and large clusters present on the heatmap generated using the Tanimoto score matrix. These different scaffolds represent the common molecular topographical properties which will be ideal for designing novel potential M^{pro} inhibitors. Further, MD simulations of M^{pro} with the six top hits explain the important and stable interactions of different chemical moieties with critical residues in terms of occupancy and histogram. This study also suggests that the compound occupying the higher subpockets in the active site have higher binding energy (observed for lithospermic acid B and 11m_32045235) unless until any extra flexible moiety does not disturb the critical interaction between the ligand and M^{pro} (observed for acteoside). This information about the scaffold and its behavior in different subpockets of the M^{pro} active site offers a new direction to the researcher to design target-specific inhibitors as drug repurposing has not been clinically effective in COVID-19 patients.

■ ASSOCIATED CONTENT

ⓈI Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jcim.0c00546.

Phylogenetic relationship of SARS-CoV-2 M^{pro}; superimposed cocrystallized and redocked ligand; protein thermodynamic stability information upon point mutations; list of top M^{pro} inhibitors with their structures, source and scaffold information; interaction fraction histogram obtained from MD simulations; protein-ligand contact timeline information; RMSF plot for protein in presence of different ligands; binding free energy plot; and analysis of impact of mutation on active site residues of M^{pro} (PDF)

■ AUTHOR INFORMATION

Corresponding Author

Prabha Garg − *Department of Pharmacoinformatics, National Institute of Pharmaceutical Education and Research (NIPER), S.A.S. Nagar 160062, Punjab, India;* ⓘ orcid.org/0000-0002-6922-4809; Phone: +91-2214683-87 extn 2016; Email: prabhagarg@niper.ac.in

Authors

Anuj Gahlawat − *Department of Pharmacoinformatics, National Institute of Pharmaceutical Education and Research (NIPER), S.A.S. Nagar 160062, Punjab, India;* ⓘ orcid.org/0000-0001-8227-4440

Navneet Kumar − *Department of Pharmacoinformatics, National Institute of Pharmaceutical Education and Research (NIPER), S.A.S. Nagar 160062, Punjab, India;* ⓘ orcid.org/0000-0003-4104-5187

Rajender Kumar − *Department of Clinical Microbiology and Laboratory for Molecular Infection Medicine Sweden (MIMS), Umeå University, SE-90185 Umeå, Sweden;* ⓘ orcid.org/0000-0002-3322-8621

Hardeep Sandhu − *Department of Pharmacoinformatics, National Institute of Pharmaceutical Education and Research (NIPER), S.A.S. Nagar 160062, Punjab, India*

Inder Pal Singh − *Department of Natural Products, National Institute of Pharmaceutical Education and Research (NIPER), S.A.S. Nagar 160062, Punjab, India*

Saranjit Singh − *Department of Pharmaceutical Analysis, National Institute of Pharmaceutical Education and Research (NIPER), S.A.S. Nagar 160062, Punjab, India;* ⓘ orcid.org/0000-0002-8433-4598

Anders Sjöstedt − *Department of Clinical Microbiology and Laboratory for Molecular Infection Medicine Sweden (MIMS), Umeå University, SE-90185 Umeå, Sweden*

Author Contributions

[⊥]Equal contribution.

Notes

The authors declare no competing financial interest.

◼ REFERENCES

(1) Hamre, D.; Beem, M. Virologic Studies of Acute Respiratory Disease in Young Adults: V. Coronavirus 229e Infections During Six Years of Surveillance. *Am. J. Epidemiol.* **1972**, *96*, 94−106.

(2) Myint, S. Human Coronaviruses: A Brief Review. *Rev. Med. Virol.* **1994**, *4*, 35−46.

(3) Pyrc, K.; Berkhout, B.; van der Hoek, L. The Novel Human Coronaviruses NL63 and HKU1. *J. Virol.* **2007**, *81*, 3051−3057.

(4) Yang, X.; Yu, Y.; Xu, J.; Shu, H.; Liu, H.; Wu, Y.; Zhang, L.; Yu, Z.; Fang, M.; Yu, T. Clinical Course and Outcomes of Critically Ill Patients with SARS-CoV-2 Pneumonia in Wuhan, China: A Single-Centered, Retrospective, Observational Study. *Lancet Respir. Med.* **2020**, *8*, 475.

(5) Chen, Y.; Liu, Q.; Guo, D. Emerging Coronaviruses: Genome Structure, Replication, and Pathogenesis. *J. Med. Virol.* **2020**, *92*, 418−423.

(6) Berman, H.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T.; Weissig, H.; Shindyalov, I.; Bourne, P.; Rose, P.; Prlic, A. RCSB Protein Data Bank: Structural Biology Views for Basic and Applied Research. *Nucleic Acids Res.* **2000**, *28*, 235−242.

(7) Kumar, S.; Stecher, G.; Li, M.; Knyaz, C.; Tamura, K. Mega X: Molecular Evolutionary Genetics Analysis across Computing Platforms. *Mol. Biol. Evol.* **2018**, *35*, 1547−1549.

(8) Sievers, F.; Wilm, A.; Dineen, D.; Gibson, T. J.; Karplus, K.; Li, W.; Lopez, R.; McWilliam, H.; Remmert, M.; Söding, J.; et al. Fast, Scalable Generation of High-Quality Protein Multiple Sequence Alignments Using Clustal Omega. *Mol. Syst. Biol.* **2011**, *7*, 539.

(9) Dundas, J.; Ouyang, Z.; Tseng, J.; Binkowski, A.; Turpaz, Y.; Liang, J. CASTp: Computed Atlas of Surface Topography of Proteins with Structural and Topographical Mapping of Functionally Annotated Residues. *Nucleic Acids Res.* **2006**, *34*, W116−W118.

(10) Pires, D. E.; Ascher, D. B.; Blundell, T. L. DUET: A Server for Predicting Effects of Mutations on Protein Stability Using an Integrated Computational Approach. *Nucleic Acids Res.* **2014**, *42*, W314−W319.

(11) Pires, D. E.; Ascher, D. B.; Blundell, T. L. mCSM: Predicting the Effects of Mutations in Proteins Using Graph-Based Signatures. *Bioinformatics* **2014**, *30*, 335−342.

(12) Worth, C. L.; Preissner, R.; Blundell, T. L. SDM—a Server for Predicting Effects of Mutations on Protein Stability and Malfunction. *Nucleic Acids Res.* **2011**, *39*, W215−W222.

(13) Bacha, U.; Barrila, J.; Velazquez-Campoy, A.; Leavitt, S. A.; Freire, E. Identification of Novel Inhibitors of the SARS Coronavirus Main Protease 3CL[pro]. *Biochemistry* **2004**, *43*, 4906−4912.

(14) Kao, R. Y.; To, A. P.; Ng, L. W.; Tsui, W. H.; Lee, T. S.; Tsoi, H.-W.; Yuen, K.-Y. Characterization of SARS-CoV Main Protease and Identification of Biologically Active Small Molecule Inhibitors Using a Continuous Fluorescence-Based Assay. *FEBS Lett.* **2004**, *576*, 325−330.

(15) Jenwitheesuk, E.; Samudrala, R. Identifying Inhibitors of the SARS Coronavirus Proteinase. *Bioorg. Med. Chem. Lett.* **2003**, *13*, 3989−3992.

(16) Zhang, L.; Lin, D.; Kusov, Y.; Nian, Y.; Ma, Q.; Wang, J.; Von Brunn, A.; Leyssen, P.; Lanko, K.; Neyts, J. α-Ketoamides as Broad-Spectrum Inhibitors of Coronavirus and Enterovirus Replication: Structure-Based Design, Synthesis, and Activity Assessment. *J. Med. Chem.* **2020**, *63*, 4562.

(17) Chen, L.; Gui, C.; Luo, X.; Yang, Q.; Günther, S.; Scandella, E.; Drosten, C.; Bai, D.; He, X.; Ludewig, B.; et al. Cinanserin Is an Inhibitor of the 3C-like Proteinase of Severe Acute Respiratory Syndrome Coronavirus and Strongly Reduces Virus Replication in Vitro. *J. Virol.* **2005**, *79*, 7095−7103.

(18) Wu, C.-Y.; Jan, J.-T.; Ma, S.-H.; Kuo, C.-J.; Juan, H.-F.; Cheng, Y.-S. E.; Hsu, H.-H.; Huang, H.-C.; Wu, D.; Brik, A.; et al. Small Molecules Targeting Severe Acute Respiratory Syndrome Human Coronavirus. *Proc. Natl. Acad. Sci. U. S. A.* **2004**, *101*, 10012−10017.

(19) Blanchard, J. E.; Elowe, N. H.; Huitema, C.; Fortin, P. D.; Cechetto, J. D.; Eltis, L. D.; Brown, E. D. High-Throughput Screening Identifies Inhibitors of the SARS Coronavirus Main Proteinase. *Chem. Biol.* **2004**, *11*, 1445−1453.

(20) Dooley, A. J.; Shindo, N.; Taggart, B.; Park, J.-G.; Pang, Y.-P. From Genome to Drug Lead: Identification of a Small-Molecule Inhibitor of the SARS Virus. *Bioorg. Med. Chem. Lett.* **2006**, *16*, 830−833.

(21) Liang, R.; Wang, L.; Zhang, N.; Deng, X.; Su, M.; Su, Y.; Hu, L.; He, C.; Ying, T.; Jiang, S.; Yu, F. Development of Small-Molecule MERS-CoV Inhibitors. *Viruses* **2018**, *10*, 721.

(22) Turlington, M.; Chun, A.; Tomar, S.; Eggler, A.; Grum-Tokars, V.; Jacobs, J.; Daniels, J. S.; Dawson, E.; Saldanha, A.; Chase, P.; et al. Discovery of N-(benzo [1, 2, 3] triazol-1-yl)-N-(benzyl) acetamido) phenyl) Carboxamides as Severe Acute Respiratory Syndrome Coronavirus (SARS-CoV) 3CLpro Inhibitors: Identification of ML300 and Noncovalent Nanomolar Inhibitors with an Induced-Fit Binding. *Bioorg. Med. Chem. Lett.* **2013**, *23*, 6172−6177.

(23) 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.

(24) De Clercq, E. Potential Antivirals and Antiviral Strategies against SARS Coronavirus Infections. *Expert Rev. Anti-Infect. Ther.* **2006**, *4*, 291−302.

(25) Xue, X.; Yu, H.; Yang, H.; Xue, F.; Wu, Z.; Shen, W.; Li, J.; Zhou, Z.; Ding, Y.; Zhao, Q.; et al. Structures of Two Coronavirus Main Proteases: Implications for Substrate Binding and Antiviral Drug Design. *J. Virol.* **2008**, *82*, 2515−2527.

(26) Thuy, B. T. P.; My, T. T. A.; Hai, N. T. T.; Hieu, L. T.; Hoa, T. T.; Thi Phuong Loan, H.; Triet, N. T.; Anh, T. T. V.; Quy, P. T.; Tat, P. V.; et al. Investigation into SARS-CoV-2 Resistance of Compounds in Garlic Essential Oil. *ACS omega* **2020**, *5*, 8312−8320.

(27) Wishart, D. S.; Knox, C.; Guo, A. C.; Cheng, D.; Shrivastava, S.; Tzur, D.; Gautam, B.; Hassanali, M. DrugBank: A Knowledgebase for Drugs, Drug Actions and Drug Targets. *Nucleic Acids Res.* **2008**, *36*, D901−D906.

(28) Friesner, R. A.; Murphy, R. B.; Repasky, M. P.; Frye, L. L.; Greenwood, J. R.; Halgren, T. A.; Sanschagrin, P. C.; Mainz, D. T. Extra Precision Glide: Docking and Scoring Incorporating a Model of Hydrophobic Enclosure for Protein- Ligand Complexes. *J. Med. Chem.* **2006**, *49*, 6177−6196.

(29) Lyne, P. D.; Lamb, M. L.; Saeh, J. C. Accurate Prediction of the Relative Potencies of Members of a Series of Kinase Inhibitors Using Molecular Docking and MM-GBSA Scoring. *J. Med. Chem.* **2006**, *49*, 4805−4808.

(30) Genheden, S.; Ryde, U. The Mm/Pbsa and MM/GBSA Methods to Estimate Ligand-Binding Affinities. *Expert Opin. Drug Discovery* **2015**, *10*, 449−461.

(31) Bajusz, D.; Rácz, A.; Héberger, K. Why Is Tanimoto Index an Appropriate Choice for Fingerprint-Based Similarity Calculations? *J. Cheminf.* **2015**, *7*, 20.

(32) Cao, Y.; Charisi, A.; Cheng, L.-C.; Jiang, T.; Girke, T. ChemmineR: A Compound Mining Framework for R. *Bioinformatics* **2008**, *24*, 1733−1734.

(33) Bowers, K. J.; Chow, D. E.; Xu, H.; Dror, R. O.; Eastwood, M. P.; Gregersen, B. A.; Klepeis, J. L.; Kolossvary, I.; Moraes, M. A.; Sacerdoti, F. D. Scalable Algorithms for Molecular Dynamics Simulations on Commodity Clusters. *Proceedings of the 2006 ACM/ IEEE Conference on Supercomputing*, New York, 2006; pp 43−43.

(34) Mark, P.; Nilsson, L. Structure and Dynamics of the TIP3P, SPC, and SPC/E Water Models at 298 K. *J. Phys. Chem. A* **2001**, *105*, 9954−9960.

(35) Abdul Samad, F.; Suliman, B. A.; Basha, S. H.; Manivasagam, T.; Essa, M. M. A Comprehensive *In Silico* Analysis on the Structural and Functional Impact of SNPs in the Congenital Heart Defects Associated with NKX2−5 Gene—a Molecular Dynamic Simulation Approach. *PLoS One* **2016**, *11*, e0153999.

(36) Shivakumar, D.; Williams, J.; Wu, Y.; Damm, W.; Shelley, J.; Sherman, W. Prediction of Absolute Solvation Free Energies Using Molecular Dynamics Free Energy Perturbation and the OPLS Force Field. *J. Chem. Theory Comput.* **2010**, *6*, 1509−1519.

(37) Nosé, S. A Unified Formulation of the Constant Temperature Molecular Dynamics Methods. *J. Chem. Phys.* **1984**, *81*, 511−519.

(38) Evans, D. J.; Holian, B. L. The Nose-Hoover Thermostat. *J. Chem. Phys.* **1985**, *83*, 4069−4074.

(39) Zhang, L.; Lin, D.; Sun, X.; Curth, U.; Drosten, C.; Sauerhering, L.; Becker, S.; Rox, K.; Hilgenfeld, R. Crystal Structure of SARS-CoV-2 Main Protease Provides a Basis for Design of Improved α-Ketoamide Inhibitors. *Science* **2020**, *368*, 409−412.

(40) Graziano, V.; McGrath, W. J.; Yang, L.; Mangel, W. F. Sars Cov Main Proteinase: The Monomer- Dimer Equilibrium Dissociation Constant. *Biochemistry* **2006**, *45*, 14632−14641.

(41) Barrila, J.; Bacha, U.; Freire, E. Long-Range Cooperative Interactions Modulate Dimerization in SARS 3CL[pro]. *Biochemistry* **2006**, *45*, 14908−14916.

(42) Tan, J.; Verschueren, K. H.; Anand, K.; Shen, J.; Yang, M.; Xu, Y.; Rao, Z.; Bigalke, J.; Heisen, B.; Mesters, J. R.; et al. pH-dependent Conformational Flexibility of the SARS-CoV Main Proteinase (M[pro]) Dimer: Molecular Dynamics Simulations and Multiple X-ray Structure Analyses. *J. Mol. Biol.* **2005**, *354*, 25−40.

(43) Kiemer, L.; Lund, O.; Brunak, S.; Blom, N. Coronavirus 3CL[pro] Proteinase Cleavage Sites: Possible Relevance to SARS Virus Pathology. *BMC Bioinf.* **2004**, *5*, 72.

(44) Fan, K.; Ma, L.; Han, X.; Liang, H.; Wei, P.; Liu, Y.; Lai, L. The Substrate Specificity of SARS Coronavirus 3C-like Proteinase. *Biochem. Biophys. Res. Commun.* **2005**, *329*, 934−940.

(45) Jeffrey, G. A. *An Introduction to Hydrogen Bonding*; Oxford University Press: New York, 1997; Vol. *12*.

(46) Lim, L.; Shi, J.; Mu, Y.; Song, J. Dynamically-Driven Enhancement of the Catalytic Machinery of the SARS 3C-like Protease by the S284-T285-I286/A Mutations on the Extra Domain. *PLoS One* **2014**, *9*, e101941.

(47) Beg, M.; Thakur, S. C.; Meena, L. S. Structural Prediction and Mutational Analysis of Rv3906c Gene of *Mycobacterium tuberculosis* H$_{37}$Rv to Determine Its Essentiality in Survival. *Adv. Bioinf.* **2018**, *2018*, 1.

(48) Yamamoto, N.; Yang, R.; Yoshinaka, Y.; Amari, S.; Nakano, T.; Cinatl, J.; Rabenau, H.; Doerr, H. W.; Hunsmann, G.; Otaka, A.; et al. HIV Protease Inhibitor Nelfinavir Inhibits Replication of SARS-Associated Coronavirus. *Biochem. Biophys. Res. Commun.* **2004**, *318*, 719−725.

(49) Tahir ul Qamar, M.; Alqahtani, S. M.; Alamri, M. A.; Chen, L.-L. Structural Basis of SARS-CoV-2 3CL[pro] and Anti-Covid-19 Drug Discovery from Medicinal Plants. *J. Pharm. Anal.* **2020**, 1 DOI: 10.1016/j.jpha.2020.03.009.

(50) Enmozhi, S. K.; Raja, K.; Sebastine, I.; Joseph, J. Andrographolide as a Potential Inhibitor of SARS-CoV-2 Main Protease: An *In Silico* Approach. *J. Biomol. Struct. Dyn.* **2020**, 1−10.

(51) Das, S.; Sarmah, S.; Lyndem, S.; Singha Roy, A. An Investigation into the Identification of Potential Inhibitors of SARS-CoV-2 Main Protease Using Molecular Docking Study. *J. Biomol. Struct. Dyn.* **2020**, 1−18.

M