PLOS ONE

RESEARCH ARTICLE

In silico identification of potential inhibitors of key SARS-CoV-2 3CL hydrolase (Mpro) *via* molecular docking, MMGBSA predictive binding energy calculations, and molecular dynamics simulation

M. Iqbal Choudhary[1,2]☯*, **Muniza Shaikh**[1]☯*, **Atia- tul-Wahab**[1]☯*, **Atta- ur-Rahman**[1,2]☯

1 Dr. Panjwani Center for Molecular Medicine and Drug Research, International Center for Chemical and Biological Sciences, University of Karachi, Karachi, Pakistan, **2** H. E. J. Research Institute of Chemistry, International Center for Chemical and Biological Sciences, University of Karachi, Karachi, Pakistan

☯ These authors contributed equally to this work.
* iqbal.choudhary@iccs.edu (MIC); shaikhmuniza2016@gmail.com (MS); atia.tulwahab@iccs.edu (ATW)

Abstract

The incidence of 2019 novel corona virus (SARS-CoV-2) has created a medical emergency throughout the world. Various efforts have been made to develop the vaccine or effective treatments against the disease. The discovery of crystal structure of SARS-CoV-2 main protease has made the *in silico* identification of its inhibitors possible. Based on its critical role in viral replication, the viral protease can prove to be a promising "target" for antiviral drug therapy. We have systematically screened an *in-house* library of 15,754 natural and synthetic compounds, established at International Center for Chemical and Biological Sciences, University of Karachi. The *in silico* search for potential viral protease inhibitors resulted in nine top ranked ligands (compounds **1**–**9**) against SARS-CoV-2 main protease (PDB ID: 6LU7) based on docking scores, and predictive binding energies. The *in silico* studies were updated *via* carrying out the docking, and predictive binding energy estimation, with a recently reported crystal structure of main protease (PDB ID: 6Y2F) at a better resolution *i.e.*, 1.95 Å. Compound **2** (molecular bank code AAA396) was found to have highest negative binding energy of −71.63 kcal/mol for 6LU7. While compound **3** (molecular bank code AAD146) exhibited highest negative binding energy of -81.92 kcal/mol for 6Y2F. The stability of the compounds- *in complex* with viral protease was analyzed by Molecular Dynamics simulation studies, and was found to be stable over the course of 20 ns simulation time. Compound **2,** and **3** were predicted to be the significant inhibitors of SARS-CoV-2 3CL hydrolase (Mpro) among the nine short listed compounds.

1. Introduction

The recent outbreak of series of pneumonia cases in Wuhan (named as COVID-19 by WHO) has created a medical emergency, unprecedented in recent history. The disease has clinical

presentation resembling viral pneumonia, and has emerged as an epidemic [1]. The incident was first reported in central China, in December 2019 [2, 3]. By January 2020, 41 patients were admitted to hospital from which 73% were male with a median age of 49 years. Among them 66% of the patients were found to be exposed to Wuhan seafood market. Common symptoms observed at the onset of illness were cough, fever, and fatigue. All of the 41 patients were positive for pneumonia with abnormal findings on chest CT scan. Acute respiratory distress syndrome, RNAemia, acute cardiac injury, and secondary infection were recorded as complications [4, 5]. Extensive sequencing analysis of the samples from lower respiratory tract identified a virus resembling SARS CoV, and named as novel corona virus 2019 (2019-nCoV) or SARS-CoV-2 [6].

As of 29th May 2020, the number of COVID-19 cases has been reached to 5909,029 with 362,081 deaths reported worldwide. It has been spread in more than 94 countries, including major outbreaks in South Korea, Iran, and Italy [7].

Corona virus belongs to family Coronaviridae, and order Nidovirales. They are enveloped positive sense RNA virus, widely distributed in mammals including humans [8]. Once the cell is infected with SARS-CoV-2, the existing molecular machinery of the host cell is taken over by the virus to translate its RNA into long chains of proteins, and to produce more copies. These long viral proteins are activated when cut into smaller pieces by proteases. Hence, the viral proteases have a critical role in the propagation of the virus. A large number of viral protease inhibitors (such as Amprenavir, atazanavir, darunavir, boceprevir, grazoprevir, etc.) have been approved as antiviral drugs by FDA for the treatment of viral diseases such as HIV, and hepatitis C.

The main protease from SARS-CoV2 has been reported as a heart shape protein, which consist of dimer of identical subunits. The protease activity is triggered by binding of molecules to specific sites. Since 2019-nCov emergence created an urgency for the development of vaccines or effective treatment, various drug repurposing, and virtual screening approaches are being employed in order to find out the possible therapeutic strategy as early as possible. For this purpose, the development of specific inhibitors of the COVID-19 main protease can be of great importance in terms of proposing the treatment regimen.

The gradual evolution of computer hardware, and software technologies have maximized the likelihood of finding the new drugs from huge libraries of small molecules. Through the help of computer-aided drug design techniques, various potential drug targets emerging from genomic and proteomic initiatives can be effectively used to reduce the cost and speed up the drug discovery process. Various studies have employed the use of docking, molecular dynamics simulation, and combination of different advanced *in silico* approaches for drug design [9–11].

In an attempt to address COVID-19 global epidemic challenge, we conducted an *in silico* based virtual screening of an *in-house* chemical library of more than 15,754 natural products and synthetic compounds against the crystal structure of 2019-nCov main protease, also known as 3CL hydrolase (Mpro) submitted in PDB (PDB ID: 6LU7). The enzyme is complexed with an inhibitor, which is peptide in nature. These screened compounds are part of our ICCBS molecular bank, a national chemical repository of Pakistan. The database consists of various natural and synthetic small molecules, having significant biological activities. During this study, we have carried out *in silico* identification of potential inhibitors of COVID-19 main protease (6LU7) *via* molecular docking analysis with the predictive binding energy estimation of the selected ligands. The methodology was also carried out for the recently reported crystal structure of main protease (6Y2F) released on PDB on 4th March 2020 [12] having a better resolution of 1.95 Å. The selected complexes of the viral protease-ligand were subjected to molecular dynamics simulation studies for investigating the stability of the complex formed.

Based upon the results we have obtained, these new inhibitors, particularly compound **2**, and **3**, can serve as lead molecules in the discovery of direly needed treatment of COVID-19.

2. Material and methods

2.1 Protein preparation

For docking studies, the recently submitted crystal structures of SARS-CoV-2 main protease (6LU7 by Jin Z. et al. 2020, and 6Y2F by Zhang L. et al. 2020) in PDB were used. Before commencing docking, the protease structures were subjected to preparation by *Protein Preparation Wizard* in Maestro Schrödinger 12 [13, 14]. During preparation, the missing hydrogens were added, and partial charges were assigned using OPLS-3e force field. Hydrogens and heavy atoms were optimized by restrained minimization.

2.2. Ligand preparation

The 2D structures of 15,754 compounds from molecular bank converted to 3D structure, *via Ligprep* module in Maestro Schrödinger 12 [15]. *Ligprep* corrects the protonation, and ionization states of the compounds, and assigned proper bond orders. Afterwards, the tautomeric and ionization states were created for each ligand.

2.3. Receptor grid generation and molecular docking

The grid box was defined by selecting the co-crystallized peptide inhibitors (N$_3$ and α-ketoamide of 6LU7, and 6Y2F, respectively) to keep the center of each docked ligand with same dimensions of binding box. Rigid receptor docking protocol was run in standard precision (SP) mode of Glide based on OPLS-3e force field [16–18]. During the process of docking, the protein was fixed, while ligands were flexible.

Molecular mechanics-generalized Born surface area (MM-GBSA) method in Prime was used for rescoring the docked pose of ligand [19, 20]. These poses were taken as inputs for the energy minimization of the protein-ligand complexes (E$_{complex}$), the free protein (E$_{protein}$), and the free ligands (E$_{ligand}$). The binding free energy ΔG$_{bind}$ was determined according to the following equation:

$$\Delta Gbind = E\_complex\ (minimized) - E\_ligand\ (minimized) - E\_receptor\ (minimized)$$

2.4. ADME properties prediction

The ADME properties, and drug likeness of selected compounds were determined in terms of distribution, absorption, metabolism, excretion, etc., *via* QikProp module of Maestro Schrodinger [21].

2.5. Molecular dynamics simulation

MD simulations [22, 23] were performed with Desmond, as implemented in the Schrödinger package. The selected ligand-protein complexes were first immersed into SPC (simple point charge) water box, extending 10 Å beyond any of the complex's atoms. Counter ions (33 Na$^+$, and 29 Cl$^-$ ions) were added to neutralize charges. Salt concentration was set to 0.15 M sodium, and chloride ions to approximate physiological condition. The MD was performed in the NPT ensemble at temperature of 300 K and 1.63 bar pressure over 20 ns with recording intervals of 1.2 ps for energy and 20 ps for trajectory. Simulations were run with the OPLS-3e force field. Plots and figures were sketched with Desmond simulation interaction diagram tool of Maestro.

3. Results and discussion

3.1. Protein and ligand preparation

The conversion of 2D structures to 3D, tautomerization and ionization *via LigPrep* yielded 28,000 minimized 3D molecular structures. These energy minimized 3D structures were used for docking with crystal structure of the 3CL hydrolase (Mpro). Preparation of the viral protease (6LU7, and 6Y2F) by protein preparation wizard tool optimized the H-bond network, and minimized the geometry. The assurance of assigning the proper formal charge, and force field treatment was achieved by adding missing hydrogens and correct ionization states (Fig 1).

3.2. Molecular docking studies

After defining the grid box in the prepared viral protease *via Receptor Grid Generation* tool of *Glide* in Maestro, the prepared 3D molecular structures were docked into the co-crystallized inhibitor binding site of viral protease. Fig 2 shows the result of nine docked ligands chosen on the basis of most negative docking score from 15,754 diverse compounds. These scores represent the best-bound ligand conformations, and relative binding affinities. The molecule having the Molecular Bank code AAB492 (compound **1**), and AAD139 (compound **6**), exhibited the highly negative docking score of -8.897, and -8.892 kcal/mol, respectively, while compounds AAA396 (compound **2**), AAD308 (compound **7**), and AAD019 (compound **9**) exhibited a comparable docking score, (ranging from -8.417- -8.04 kcal/mol), and hence inferred to have similar relative binding affinity. AAD146 (compound **3**) and AAA127 (compound **8**), and AAA198 (compound **4**) and AAA210 (compound **5**) were found to have a similar docking score.

The docking analysis was updated in terms of re-docking the selected nine compounds with the recently released crystal structure of 3CL pro hydrolase (PDB ID: 6Y2F) at better resolution of 1.95 Å. Analysis of the docking score of the nine compounds with the α-ketoamide binding site of 2019-nCoV main protease (PDB ID 6Y2F) revealed different docking score, hence binding affinities (Fig 2). These differences can be attributed to the difference in the grid formation due to the presence of α-ketoamide inhibitor in the binding site of 6Y2F. Compounds **1**, **6**, **7**, and **9** exhibited similar docking score as that of 6LU7. Compounds **2**, **3**, **4**, and **5** showed different docking score from docking with 6LU7. However, compound **3** exhibited highly negative docking score of -9.920 kcal/mol against 6Y2F.

Literature survey for the selected compounds was conducted in order to find out their reported activities, mentioned in Table 1.

Docking of molecule in protein binding site is a robust approach to elucidate the correct binding pose among several predicted poses of a compound. However, the approach still lacks the ability to correctly rank the affinities of the small molecules to the target protein. In order to have a better ranking of the ligands, and determination of predictive binding energies, MMGBSA calculations were performed. The values obtained after calculation were approximate free energies of binding with a more negative value indicating a stronger binding. Fig 2 also depicts the predictive binding energies of ligand in the binding site of the viral protease. Compound **2** (AAA396) exhibited highest negative binding energy of approximately −71.63 kcal/mole with in the N_3 binding site of 6LU7 however, compound **3** (AAD146) manifested a predictive binding energy of -81.92 kcal/mole in the α-ketoamide binding site of 6Y2F.

Analysis of the co-crystallized peptide inhibitors (N_3, and α-ketoamide) revealed their molecular interaction with the binding site of viral proteases (6LU7, and 6Y2F, respectively). The peptide inhibitor N_3 interacted through hydrogen bonds with residues Thr190, Gln189, Glu166, His164, Phe140, and Gly143 (Fig 3A) of 6LU7 with the depiction of binding site
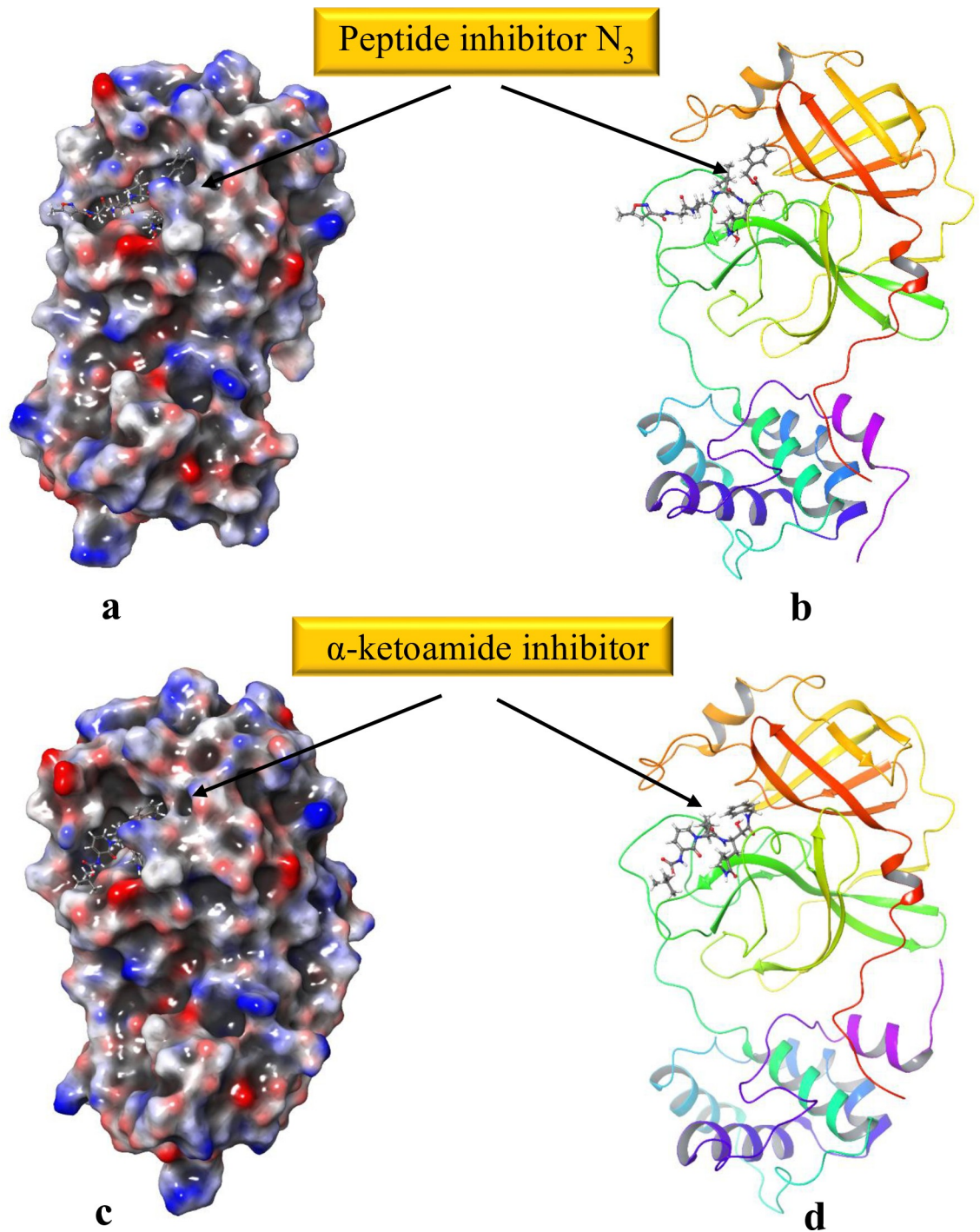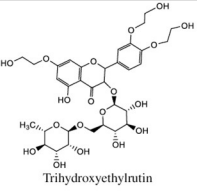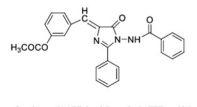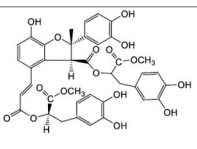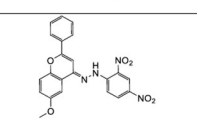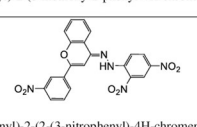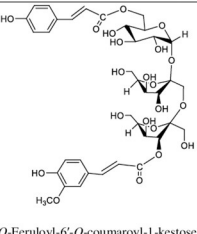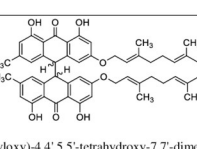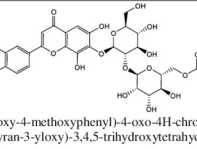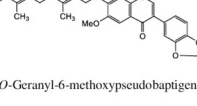
Fig 1. 3CL pro hydrolase prepared via *Protein Preparation Wizard*, Maestro. H-bond network optimized and geometry minimized structures of 2019-nCOV 3CL hydrolase (Mpro) PDB ID: **(a)**, and **(b)** 6LU7, **(c)**, and **(d)** 6Y2F represented as 3D molecular surface and ribbon structure.

occupancy in 3D molecular surface diagram (Fig 3B). The α-ketoamide inhibitor also inter-acted through hydrogen bonds in the binding site of 6Y2F with residues His41, Gly143, Cys145, His163, His164, and Glu166 (Fig 3C and 3D).

Compound No.	Structure	Molecular Weight	Source	For 6LU7		For 6Y2F	
				Docking score (kcal/mol)	MMGBSA score (kcal/mol)	Docking score (kcal/mol)	MMGBSA score (kcal/mol)
1 (AAB492)	Trihydroxyethylrutin	742.68	S	-8.897	-54.71	-8.076	-73.72
2 (AAA396)	(Z)-3-((1-Benzamido-5-oxo-2-phenyl-1H-imidazol-4(5H)-ylidene)methyl)phenyl acetate	425.4	S	-8.041	-71.63	-6.208	-57.81
3 (AAD146)	Dimethyl lithospermate	746.6	N	-6.603	-50.16	-9.920	-81.92
4 (AAA198)	(E)-1-(2,4-Dinitrophenyl)-2-(6-methoxy-2-phenyl-4H-chromen-4-ylidene)hydrazine	432.386	S New Compound	-7.881	-69.94	-5.977	-58.11
5 (AAA210)	(E)-1-(2,4-Dinitrophenyl)-2-(2-(3-nitrophenyl)-4H-chromen-4-ylidene)hydrazine	447.357	S	-7.591	-62.82	-6.490	-59.12
6 (AAD139)	O-Feruloyl-6'-O-coumaroyl-1-kestose	826.749	N	-8.892	-60.37	-8.450	-52.17
7 (AAD308)	2,2'-Bis((E)-3,7-dimethylocta-2,6-dienyloxy)-4,4',5,5'-tetrahydroxy-7,7'-dimethyl-9,9'-bianthracene-10,10'(9H,9'H)-dione	782.959	N	-8.417	-65.16	-8.322	-73.73
8 (AAA127)	(6-(2-(6,8-Dihydroxy-2-(3-hydroxy-4-methoxyphenyl)-4-oxo-4H-chromen-7-yloxy)-4,5-dihydroxy-6-(hydroxymethyl)tetrahydro-2H-pyran-3-yloxy)-3,4,5-trihydroxytetrahydro-2H-pyran-2-yl)methyl acetate	682.58	N New compound	-6.173	-38.58	-9.383	-63.34
9 (AAD019)	7-O-Geranyl-6-methoxypseudobaptigenin	448.508	N	-8.130	-61.40	-7.392	-54.52
	Remdesivir	602.6	Antiviral drug	-7.72	-58.16	-6.587	-49.9
	Indinavir	613.79	Antiviral drug	-7.59	-36.15	-7.150	-55.57

N= Natural
S= Synthet

Fig 2. Ligands identified against SARS-CoV-2 3CL protease (6LU7 and 6Y2F) from *in-house* chemical library *via in silico* screening.

Table 1. **Sources of the compounds and reported biological activities of *in silico* identified ligands of SARS-CoV-2 3CL-pro main proteases.**

Compound	Source	Reported activities
1 (AAB492)	Synthetic compound	Antioxidant [24], mild inhibitory activity against lipid peroxidation [25], and antiviral activity against vesicular stromatitis virus (VSV) [26]. Used in the treatment of lupus erythematosus disseminatus [27]
2 (AAA396)	Synthetic compound	Antioxidant activities [28].
3 (AAD146)	Found in the roots of *Salvia miltiorrhiza*, *Salvia chinensis*, *Salvia yunnanensis*,*Clerodendranthusspicatus*,and *Salvia przewalskii*	Anti-inflammatory activity [29], suppressesarrhythmogenesis [30], and liver fibrosis/cirrhosis [31], and also act as a Na$^+$ channel agonist [32]
4 (AAA198)	New* synthetic compound	–
5 (AAA210)	New* synthetic compound	–
6 (AAD139)	Natural product derived from *Lindelofiastylosa*	*In vitro* immunomodulatory activities [33].
7 (AAD308)	Natural product extracted from *Psorospermumaurantiacum*, *Cratoxylumformosum*,*Pruniflorum*, *Vismiaguineensis*, *Vismiaorientalis*, *Psorospermumtenuifolium*, and *Psorospermumfebrifugum*.	antiprotozoal activity [34]
8 (AAA127)	New* compound of natural origin extracted from *Stachys parviflora*	–
9 (AAD019)	Natural compound derived from *Millettiaconraui*	α-glucosidase inhibitory activity [35]

*The compound has been identified as new

Molecular recognition of viral protease-ligand complex at atomic level was also determined by analyzing the bound conformation of the ligands. The finally selected pose of the ligands, based on docking score and predicted binding energy, were studied to decipher the molecular interactions of viral protease (6LU7, and 6Y2F) with docked ligands.

Compound **1** (AAB492) interacted through hydrogen bonds (Fig 4A) with the binding site residues of SARS-CoV-2 main protease (6LU7). The binding site residues Thr45, Leu141, Gly143, Glu166, and Arg188, of viral protease exhibited hydrogen bonding with the various hydroxyl groups of compound **1**. Compound **2** (AAA396) showed interactions with residues Gln189, Glu166, and Thr26 *via* hydrogen bonds (Fig 4C). Compound **3** (AAD146) adopted a different binding pose, the interactions made between viral protease binding site, and compound **3** (AAD146) were through similar residues as that of N$_3$ peptide inhibitor, *i.e.*, hydrogen bonds with residues Thr26, Phe140, Glu166, and Thr190 (Fig 4E). The occupancy of the binding site can be observed in 3D molecular surface diagrams (Fig 4B, 4D and 4F). The compounds well accommodated the binding site with fully occupying the binding pocket. The docking analysis led to the prediction that the compounds are interacting with the SARS-CoV-2 hydrolase binding site through same residues as that of co-crystallized inhibitor.

Molecular interactions and binding site occupancy of compound **1**, **2**, and **3** in the α-ketoamide binding site of 6Y2F were also analysed. Compound **1** (AAB492) made molecular interactions with residues Thr26, Phe140, Gly143, Gln189, and Thr26 through hydrogen bonding (Fig 5A and 5B). Compound **2** (AAA396) showed hydrogen bonding with Thr190, Gln192, Gln189, and Glu166. In addition to this a π-π interaction with His41 was also observed (Fig 5C and 5D). Compound **3** (AAD146) exhibited hydrogen bonding with residues Thr26, Cys44, Ser46, Glu166, Phe140, and Gly143

Comparison of the molecular docking results of reported antivirals for COVID-19 *i.e.*, remdisivir, and indinavir with 6LU7, and 6Y2F was also carried out. The analysis revealed comparable docking score (Fig 2) as that of 9 selected compounds. However, the predicted binding energies were found to be lower as compared to the selected hits.
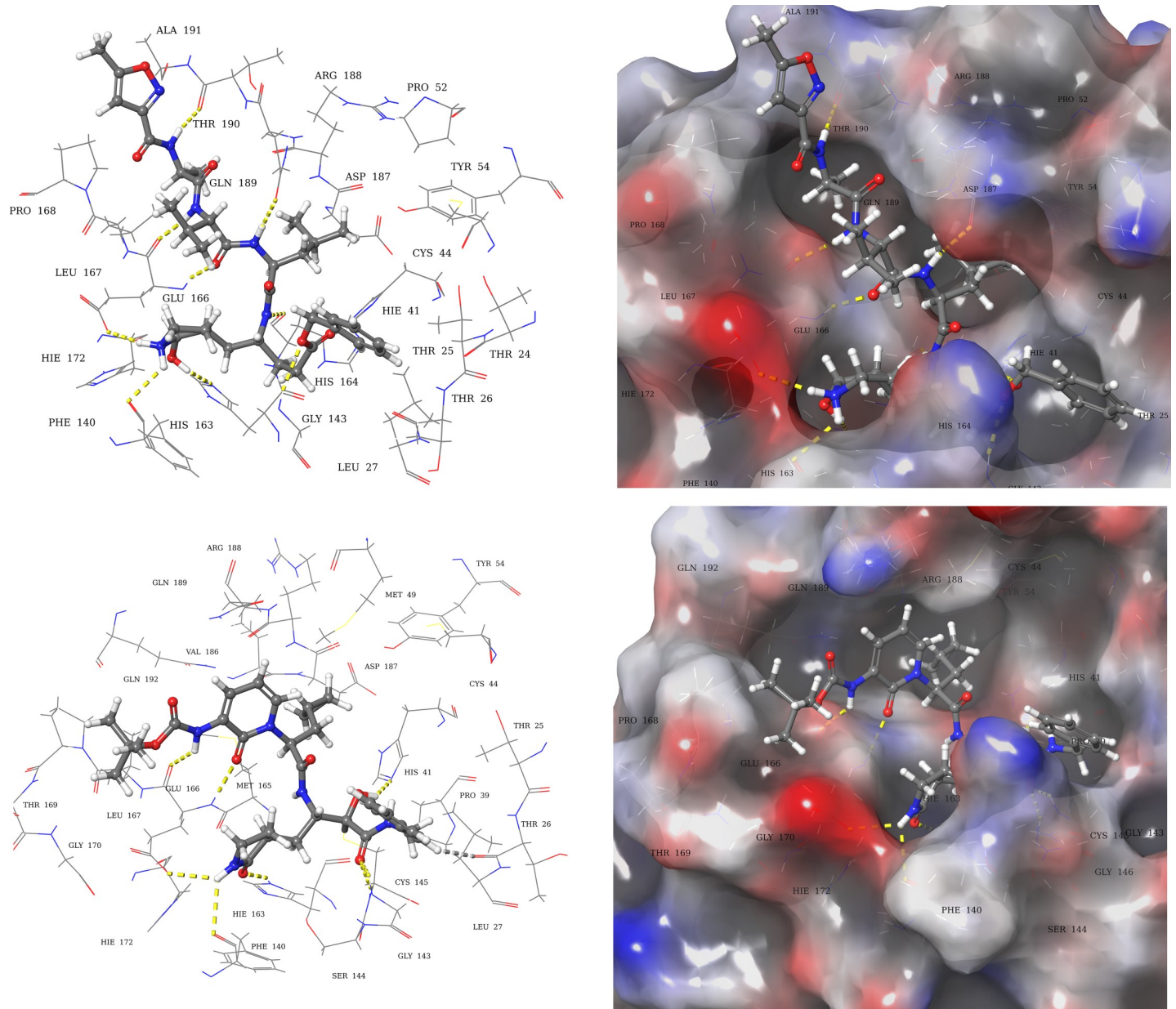
Fig 3. Molecular interactions of 3CL pro hydrolase with co-crystallized inhibitors. (**a**) 3D-Ligand interaction and (**b**) 3D molecular surface diagrams of co-crystallized peptide inhibitor in the binding site of 6LU7. (**c**) 3D-Ligand interaction and (**d**) 3D molecular surface diagrams of co-crystallized α-ketoamide inhibitor in the binding site of 6Y2F. The hydrogen bonds are represented as yellow dotted lines.

A summary of type of contacts between the nine ligands and biding site residues of 6LU7, and 6Y2F has been given as S1 Table in supporting information.

3.3. *In silico* ADME properties of selected ligands

Table 2 entails the ADME (absorption, distribution, metabolism, and excretion) properties predicted *via* QikProp. The ADME analysis determines the physicochemical properties and biological functions as well as drug likeness of the compound. This is important in terms of assessing the efficacy of drug molecules. The descriptors such as molecular weight, drug-
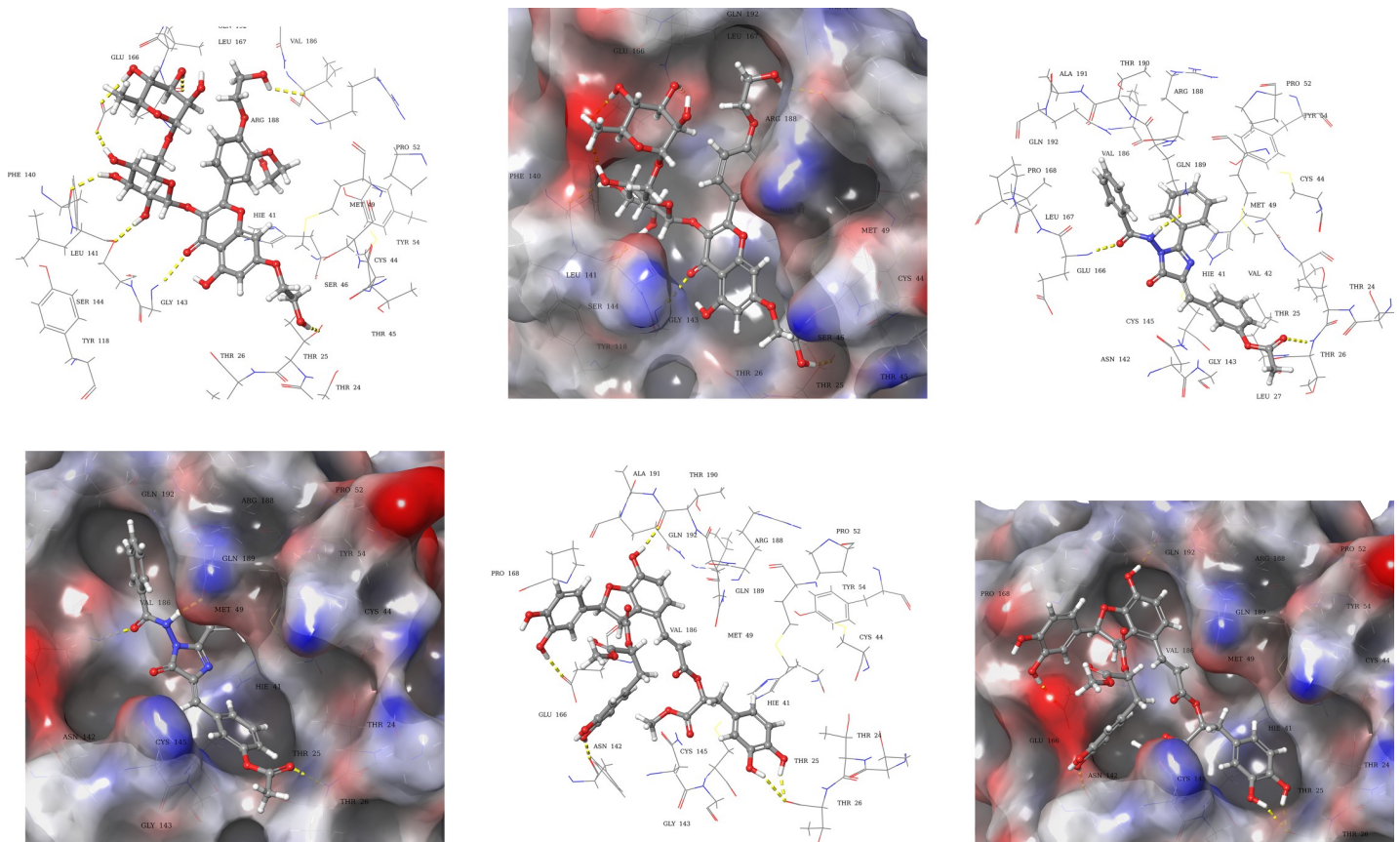
Fig 4. Docked poses and molecular interactions of compound 1, 2, and 3 in binding site of 6LU7. 3D-Ligand interaction and 3D molecular surface diagrams of compounds **1** (**a**) and (**b**), compound **2** (**c**) and (**d**) compound **3** (**e**) and (**f**), showing the docked pose in the binding site of 2019-nCOV 3CL hydrolase (Mpro). The hydrogen bonds are represented as yellow dotted lines.

likeness, dipole moment, solvent accessible surface area, hydrogen bond donor, and acceptor traits, octanol/water coefficient, aqueous solubility, binding to human serum albumin, number of likely metabolic reactions, brain/blood partition coefficient and human oral absorption were predicted for compound **2**, and **3** and compared with reported antiviral drugs indinavir, and remdesivir. The values obtained for compound **2** are in recommended range. Compound **2** exhibited better pharmacokinetic descriptors as compared to clinically approved antiviral drugs in terms of molecular weight drug-likeness, and human oral absorption. However the ADME analysis of compound **3** suggests that further derivatization in molecular structure can be done to improve the molecular descriptors such as molecular weight, acceptor hydrogen bond, number of likely reactions, brain/blood partition coefficient and % human oral absorption. This will eventually lead to an improvement in the drug-likeness score (# star) of the compound **3**.

3.4. Molecular dynamics simulation with Desmond package

Since docking is a static view of the binding pose of molecule in the active site of protein, Molecular dynamics (MD) simulation tend to compute the atom movements with time by integrating the Newton's classical equation of motion [36]. In MD, the dynamic behavior of the molecular system is simulated to assess the stability of protein ligand complex. Therefore, the docked conformers of compound **2** (AAA396), and **3** (AAD146) with highest predictive
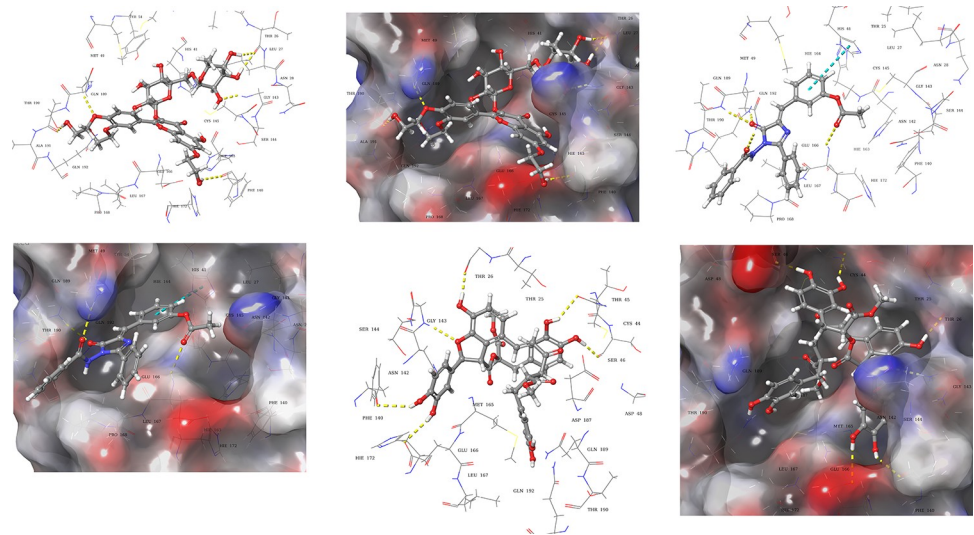
Fig 5. Docked poses and molecular interactions of compound 1, 2, and 3 in binding site of 6Y2F. 3D-Ligand interaction and 3D molecular surface diagrams of compounds **1** (**a**) and (**b**), compound **2** (**c**) and (**d**) compound **3** (**e**) and (**f**), showing the docked pose in the binding site of 2019-nCOV 3CL hydrolase (Mpro). The hydrogen bonds, and π-π interaction are represented as yellow and blue dotted lines.

binding energy of −71.63 and -81.92 kcal/mol, repectively, were used for molecular dynamics study with OPLS3e force field.

The root mean square deviation (RMSD) plot in Fig 6A indicates that the compound **2**-SARS-CoV-2 main protease (PDB ID 6LU7) complex stabilized shortly after commencing the simulation *i.e.*, 2.5 ns, with respect to the reference frame at time 0 ns. The plot showed the

Table 2. *In silico* predicted ADME properties of compounds 2, and 3.

Compound	Mol. Wt.	#Stars	Dipole	SASA	Donor H-bond	Acceptor H-bond	QPlogPo/w	QPlogS	QPlogkhsa	No. of Metabolites	QplogBB	% Human oral Absorption
2	425.4	2	6.56	748	0	8	3.7	-5.7	0.252	0	-1.28	100
3	746.6	11	9.749	103	7	14	2.643	-6.110	0.254	13	-5.723	0.5
Indinavir	613.7	6	5.549	1017	4	13	2.787	-3.866	-0.235	12	-0.724	59.6
Remdesivir	602.5	5	12.706	884	5	16	1.271	-4.856	-0.590	6	-3.278	34.504
Recommen-ded range	130–725	0–5	1.0–12.5	300–1000	0–6.0	2.0–20.0	-2-6.5	-6.5–0.5	-1.5–1.5	1–8	-3-1.2	>80% is high <25% is poor

Key:

#Stars: Number of property or descriptor values that fall outside the 95% range of similar values for known drugs. A large number of stars suggests that a molecule is less drug-like than molecules with few stars

Dipole: Computed dipole moment of the molecule.

SASA: Total solvent accessible surface area (SASA) in square angstroms using a probe with a 1.4 Å radius.

Donor H-bond: Estimated number of hydrogen bonds that would be donated by the solute to water molecules in an aqueous solution.

Acceptor H-bond: Estimated number of hydrogen bonds that would be accepted by the solute from water molecules in an aqueous solution.

QPlogPo/w: Predicted octanol/water partition coefficient.

QPlogS: Predicted aqueous solubility, log S.

QPlogkhsa: Prediction of binding to human serum albumin

No. of Metabolites: Number of likely metabolic reactions.

QplogBB: Predicted brain/blood partition coefficient.

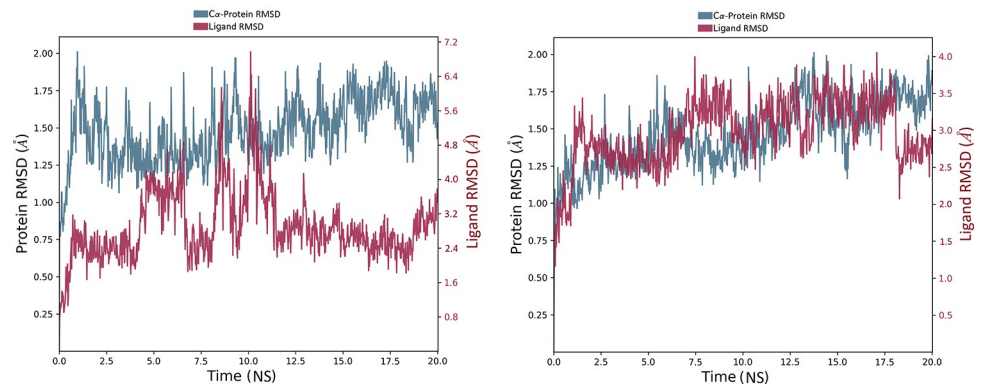**% Human Oral Absorption:** Predicted human oral absorption on 0 to 100% scale.

Fig 6. RMSD analysis of MD simulation trajectory. The RMSD plot obtained for **(a)** compound **2**-SARS-CoV-2 main protease complex (PDB ID 6LU7), and **(b)** compound **3**-SARS-CoV-2 main protease complex (PDB ID 6Y2F). The simulation time of 20 ns showing the formation of stable complex without any significant conformational changes in protein structure.

RMSD of viral protease on left Y-axis, while the Y-axis on right shows the ligand RMSD profile aligned on protein backbone. All the frames from 20 ns trajectory are aligned on reference frame backbone. In case of ligand RMSD, the fluctuation was observed during first 2 ns of trajectory. However, the variations were in order of 0.15 Å showing that the complex has not undergone large conformational changes. Furthermore, after 7 ns the simulation starts showing fluctuations in ligand RMSD of order of 2.8 Å for up to 12 ns. The ligand RMSD stabilized again after 12 ns which remain consistent till the end of simulation time of 20 ns.

The protein RMSD tend to be stable throughout the simulation with insignificant fluctuations ranging between 0.5–1.0 Å implying that the protease has not undergone large conformational changes. The overall RMSD plot of compound **2**-SARS-CoV-2 main protease complex shows fluctuations of 1.45 Å suggesting that the ligand is stably bound to protease binding site and has not diffused away from the bound position.

The RMSD plot in Fig 6B indicates the compound **3**-SARS-CoV-2 main protease (PDB ID 6Y2F) complex MD trajectory of 20 ns. The complex tends to be stabilized during the course of simulation with respect to reference frame at time 0 ns. A slight divergence can be seen towards the end of the simulation *i.e.*, 18 ns. However, the fluctuation lies under the permissible range of 1–3 Å, hence, can be considered as non-significant. Since the RMSD plots of compound **3** and protein backbone were lying over each other, formation of a stable complex can be inferred.

Fig 7A shows the residue interactions of viral protease with compound **2** (AAA396). Interactions that last more than 30% of the simulation time were considered. It shows that all of the interactions of docked pose were retained during the simulation time of 20 ns *i.e.* molecular interactions with residues Glu166, Gln189, and Thr26. Moreover, a hydrophobic contact in the form of π-π interaction was also established between the His41 and benzene moiety of compound **2**. Fig 7B shows the viral protease-compound **2** (AAA396) contacts in the form of stacked bar charts which are normalized over the course of 20 ns trajectory. These contacts are categorized into hydrogen bonds, hydrophobic, and ionic interactions, and water bridges. The bar chart shows that hydrogen bonds, hydrophobic contacts, and water bridges prevailed during the course of simulation. The important hydrogen bonds observed in the initial docked pose of compound **2** (Thr26, Glu166, Gln189) did not change during the MD simulation. Hydrogen bonding with residue Gln189 retained for more than 100% of the simulation time. This might be due to the formation of multiple hydrogen bonds of same subtypes *i.e*, backbone
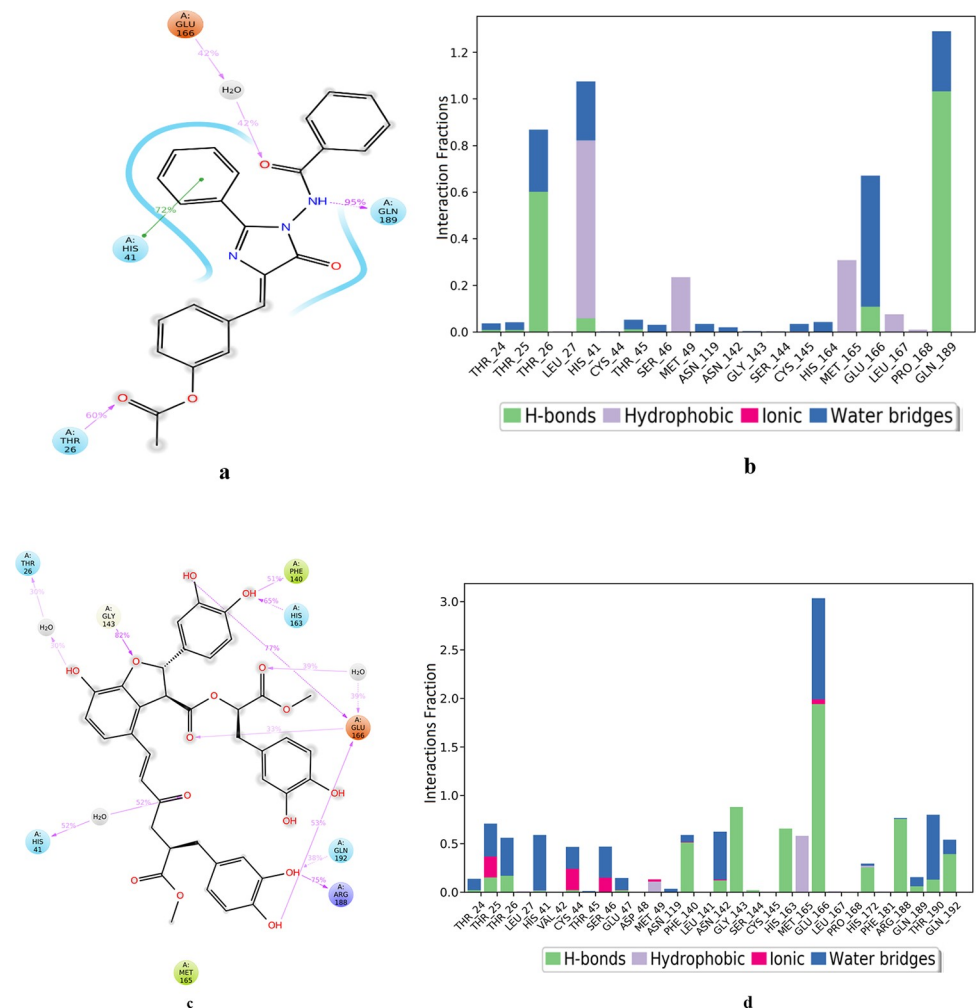
Fig 7. Analysis of molecular interaction and type of contacts with 3CL pro hydrolase after MD simulation. Detailed schematic interaction of **(a)** compound **2** (AAA396), and **(c)** compound **3** atoms with binding site residue of hydrolase crystal structures 6LU7, and 6Y2F, respectively. Interaction happening more than 30% of the simulation time are shown. Normalized stacked bar chart of viral protease binding site residues interacting with **(b)** compound **2** and **(d)** compound **3** *via* hydrogen bond, hydrophobic and ionic interactions, and water bridges.

acceptor; backbone donor; side-chain acceptor; side-chain donor. The same residues were also involved in forming water bridged interactions. Hydrophobic interactions were also observed with residues His 41, Met149, and Met165, which lasted for more than 30% of the simulation time. Water bridged interactions were also observed with residues Thr 24, Thr 25, Thr 26, Thr 45, Thr 46, Asn 119, Asn 142, Cys145, His 164, Glu 166, and Gln 189. However, the water bridges with Thr26, and Glu 166 last for 30% of the simulation. Hence, the four residues Thr 26, His 41, Glu 166, and Gln 189 were identified as the key residues involved in formation of 6LU7-compound **2** complex.

The fractional residue interactions of the α-ketoamide binding site of 6Y2F with compound **3** is presented in Fig 7C showing the interactions that persist for at least 30% of the simulation time. The interactions with residue Thr26, Glu166, Phe 140, and Gly143 were retained from the docked pose. However, new interactions were also established with residues His41, His163, Arg188, and Gln192. Fig 7D shows the types of viral protease-compound **3** (AAD146) contacts

that were formed during 20 ns MD simulation. It can be seen that the interactions were of various types *i.e.*, hydrogen bonds, water bridges, ionic, and few of them were hydrophobic. The cumulative fractional time for interactions made by most of the residues was 30–50% of total simulation time of 20 ns. However Glu166 showed interaction for more than 100% of the simulation time due to the simultaneous formation of interactions of same subtypes. This can be attributed to the predicted ADME properties of compound **3** depicted in Table 1. According to ADME analysis the compound **3** is capable of donating 7, and accepting 14 hydrogen bonds with the water molecules of the aqueous solution. Compound **3** also exhibited a very high solubility profile evident from its QPlogS value of -6.511. Glu166, Asn142, Arg188, Thr190 and Gly143 of 6Y2F were found to be critically important since they interacted with compound **3** for more than 70% of the total simulation time.

4. Conclusion

From the present study, it can be inferred that the computational methods can efficiently identify the inhibitors of a key enzyme/ protein of SARS-CoV-2, which is currently a global epidemic. The results need to be subjected to *in vitro* studies for further validation. The study identified nine compounds from an *in-house* chemical library of natural and synthetic compounds, chosen on the basis of docking score and predicted binding energies. Compound **2** (AAA396) and compound **3** (AAD146) are predicted to be the potential inhibitors of 2019-nCoV main protease. However, derivatization of compound **3** is recommended to improve its ADME profile. In view of the fact that no treatment remedy has been devised so far for corona virus infection, the data presented here can be helpful in identifying the lead molecule for future *in vitro* and *in vivo* studies and for the drug discovery and development against COVID-19.

Supporting information

S1 Table. Summary of the contacts and interacting residue of SARS-CoV-2 3CL protease (6LU& and 6Y2F) with the docked compounds.
(DOCX)

Acknowledgments

Author Contributions

Conceptualization: M. Iqbal Choudhary, Atia- tul-Wahab, Atta- ur-Rahman.

Formal analysis: Atta- ur-Rahman.

Investigation: Muniza Shaikh.

Methodology: Muniza Shaikh.

Software: Muniza Shaikh.

Supervision: M. Iqbal Choudhary, Atia- tul-Wahab.

Writing – original draft: Muniza Shaikh.

Writing – review & editing: M. Iqbal Choudhary, Atia- tul-Wahab, Atta- ur-Rahman.

References

1. WHO. Novel coronavirus–China. Jan 12, 2020. http://www.who.int/csr/doon/12-january-2020-novel-coronavirus-china/en/ (accessed Jan 19, 2020).

2. Zhu N, Zhang D, Wang W, Li X, Yang B, Song J, et al. A novel coronavirus from patients with pneumonia in China, 2019. N Engl J Med. 2020; 382: 727–733. https://doi.org/10.1056/NEJMoa2001017 PMID: 31978945

3. Lu H, Stratton CW, Tang YW. Outbreak of Pneumonia of Unknown Etiology in Wuhan China: the Mystery and the Miracle. J Med Virol. 2020; 92: 401–402. https://doi.org/10.1002/jmv.25678 PMID: 31950516

4. Huang C, Wang Y, Li X, Ren L, Zhao J, Hu Y, et al. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. The Lancet. 2020 Feb 15; 395(10223): 497–506.

5. Hui DS, I Azhar E, Madani TA, Ntoumi F, Kock R, Dar O, et al. The continuing 2019-nCoV epidemic threat of novel coronaviruses to global health—the latest 2019 novel coronavirus outbreak in Wuhan, China. Int J Infect Dis. 2020; 91: 264–6. https://doi.org/10.1016/j.ijid.2020.01.009 PMID: 31953166

6. National Health Commission of the People's Republic of China. Update on the novel coronavirus pneumonia outbreak. Beijing, China: National Health Commission of the People's Republic of China, 2020. http://en.nhc.gov.cn/2020-02/13/c_76512.htm. Accessed 13 February 2020.

7. https://www.worldometers.info/coronavirus/? Accessed 29 May 2020.

8. Richman DD, Whitley RJ, and Hayden FG. (Eds.). (2016). Clinical Virology. John Wiley & Sons.

9. Ahmad S, Navid A, Akhtar AS, Azam SS, Wadood A, Pérez-Sánchez H. Subtractive genomics, molecular docking and molecular dynamics simulation revealed LpxC as a potential drug target against multidrug resistant *Klebsiella pneumoniae*. Interdiscipl Sci Comput Life Sci. 2019 Sep 1; 11(3):508–26.

10. Hashemzadeh S, Ramezani F, Rafii-Tabar H. Study of molecular mechanism of the interaction between MEK1/2 and Trametinib with docking and molecular dynamic simulation. Interdiscipl Sci Comput Life Sci. 2019 Mar 1; 11(1):115–24.

11. Hassan M, Ashraf Z, Abbas Q, Raza H, Seo SY. Exploration of novel human tyrosinase inhibitors by molecular modeling, docking and simulation studies. Interdiscipl Sci Comput Life Sci. 2018 Mar 1; 10 (1):68–80.

12. Zhang L, Lin D, Sun X, Curth U, Drosten C, Sauerhering L, Hilgenfeld R. Crystal structure of SARS-CoV-2 main protease provides a basis for design of improved α-ketoamide inhibitors. Science. 2020.

13. Schrödinger Release 2020–1: Protein Preparation Wizard; Epik, Schrödinger, LLC, New York, NY, 2020; Impact, Schrödinger, LLC, New York, NY, 2020; Prime, Schrödinger, LLC, New York, NY, 2020.

14. Sastry GM, Adzhigirey M, Day T, Annabhimoju R, Sherman W. Protein and ligand preparation: parameters, protocols, and influence on virtual screening enrichments. J Comput Aided Mol Des. 2013; 27(3): 221–234. https://doi.org/10.1007/s10822-013-9644-8 PMID: 23579614

15. Schrödinger Release 2020–1: LigPrep, Schrödinger, LLC, New York, NY, 2020.

16. Schrödinger Release 2020–1: Glide, Schrödinger, LLC, New York, NY, 2020.

17. Halgren TA, Murphy RB, Friesner RA, Beard HS, Frye LL, Pollard WT, et al. Glide: a new approach for rapid, accurate docking and scoring. 2. Enrichment factors in database screening. J Med Chem. 2004; 47(7):1750–9. https://doi.org/10.1021/jm030644s PMID: 15027866

18. Friesner RA, Banks JL, Murphy RB, Halgren TA, Klicic JJ, Mainz DT, et al. Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. J Med Chem. 2004; 47(7):1739–49. https://doi.org/10.1021/jm0306430 PMID: 15027865

19. Schrödinger Release 2020–1: Prime, Schrödinger, LLC, New York, NY, 2020

20. Hayes JM, Archontis G. MM-GB (PB) SA calculations of protein-ligand binding free energies. In: Wang L, editor. Molecular Dynamics-Studies of Synthetic and Biological Macromolecules. UK: IntechOpen; 2012. pp. 171–90.

21. Schrödinger Release 2020–2: QikProp, Schrödinger, LLC, New York, NY, 2020.

22. Schrödinger Release 2020–1: Desmond Molecular Dynamics System, D. E. Shaw Research, New York, NY, 2020. Maestro-Desmond Interoperability Tools, Schrödinger, New York, NY, 2020.

23. Bowers KJ, Chow DE, Xu H, Dror RO, Eastwood MP, Gregersen BA, et al. Scalable algorithms for molecular dynamics simulations on commodity clusters. InSC'06: Proceedings of the 2006 ACM/IEEE Conference on Supercomputing 2006 Nov 11 (pp. 43–43). IEEE.

24. Vinothkumar R, Kumar RV, Karthikkumar V, Viswanathan P, Kabalimoorthy J, Nalini N. Oral supplementation with troxerutin (trihydroxyethylrutin), modulates lipid peroxidation and antioxidant status in 1, 2-dimethylhydrazine-induced rat colon carcinogenesis. Environ Toxicol Pharmacol. 2014; 37(1): 174–84. https://doi.org/10.1016/j.etap.2013.11.022 PMID: 24355798

25. Affany A, Salvayre R, Dousté-Blazy L. Comparison of the protective effect of various flavonoids against lipid peroxidation of erythrocyte membranes (induced by cumene hydroperoxide). Fund Clin Pharmacol. 1987; 1(6): 451–7.

26. Wacker A, Eilmes HG. Antiviral activity of plant components. 1st communication: Flavonoids (author's transl). Arzneimittel-Forsch. 1978; 28(3): 347–50.

27. Lamberton JN. Treatment of an extensive lupus erythematosus disseminatus by trihydroxyethylrutin. (1st personal trial). Bull Soc Fr Dermatol Syphiligr. 1965; 72(6): 808. PMID: 5882714

28. Khan KM, Mughal UR, Omar I, Choudhary MI. Microwaves-assisted syntheses of imidazolylbenzamides and their antioxidant activities. Lett Drug Des Discov. 2008; 5(3): 152–7.

29. Choi HG, Tran PT, Lee JH, Min BS, Kim JA. Anti-inflammatory activity of caffeic acid derivatives isolated from the roots of Salvia miltiorrhiza Bunge. Arch Pharm Res. 2018; 41(1): 64–70. https://doi.org/10.1007/s12272-017-0983-1 PMID: 29124660

30. Fish JM, Welchons DR, Kim YS, Lee SH, Ho WK, Antzelevitch C. Dimethyl lithospermate B, an extract of Danshen, suppresses arrhythmogenesis associated with the Brugada syndrome. Circulation. 2006; 113(11): 1393–400. https://doi.org/10.1161/CIRCULATIONAHA.105.601690 PMID: 16534004

31. Oh SH, Cho KH, Yang BS, Roh YK. Natural compounds from Danshen suppress the activity of hepatic stellate cells. Arch Pharm Res. 2006; 29(9): 762–7. https://doi.org/10.1007/BF02974077 PMID: 17024850

32. Yoon JY, Ahn SH, Oh H, Kim YS, Ryu SY, Ho WK, et al. A novel Na+ channel agonist, dimethyl lithospermate B, slows Na+ current inactivation and increases action potential duration in isolated rat ventricular myocytes. Br J Pharmacol. 2004; 143(6): 765–73. https://doi.org/10.1038/sj.bjp.0705969 PMID: 15504759

33. Mesaik MA, Jabeen A, Halim SA, Begum A, Khalid AS, Asif M, et al. In silico and in vitro immunomodulatory studies on compounds of Lindelofia stylosa. Chem Biol Drug Des. 2012; 79(3): 290–9. https://doi.org/10.1111/j.1747-0285.2011.01310.x PMID: 22181857

34. Mbwambo ZH, Apers S, Moshi MJ, Kapingu MC, Van Miert S, Claeys M, et al. Anthranoid compounds with antiprotozoal activity from Vismia orientalis. Planta Med. 2004; 70(08):706–10.

35. Tchinda AT, Khan SN, Fuendjiep V, Ngandeu F, Ngane AN, Choudhary MI. α-Glucosidase Inhibitors from Millettia conraui. Chem Pharm Bull. 2007; 55(9): 1402–3. https://doi.org/10.1248/cpb.55.1402 PMID: 17827772

36. Adcock SA, McCammon JA. Molecular dynamics: survey of methods for simulating the activity of proteins. Chem Rev. 2006; 106(5): 1589–615. https://doi.org/10.1021/cr040426m PMID: 16683746