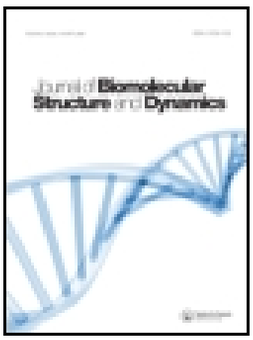


Taylor & Francis
Taylor & Francis Group

Journal of Biomolecular Structure and Dynamics

ISSN: 0739-1102 (Print) 1538-0254 (Online) Journal homepage: https://www.tandfonline.com/loi/tbsd20

Identification of chymotrypsin-like protease inhibitors of SARS-CoV-2 *via* integrated computational approach

Salman Ali Khan, Komal Zia, Sajda Ashraf, Reaz Uddin & Zaheer Ul-Haq

To cite this article: Salman Ali Khan, Komal Zia, Sajda Ashraf, Reaz Uddin & Zaheer Ul-Haq (2020): Identification of chymotrypsin-like protease inhibitors of SARS-CoV-2 *via* integrated computational approach, Journal of Biomolecular Structure and Dynamics, DOI: 10.1080/07391102.2020.1751298

To link to this article: https://doi.org/10.1080/07391102.2020.1751298

JOURNAL OF BIOMOLECULAR STRUCTURE AND DYNAMICS
https://doi.org/10.1080/07391102.2020.1751298

Identification of chymotrypsin-like protease inhibitors of SARS-CoV-2 *via* integrated computational approach

Salman Ali Khan[a], Komal Zia[a], Sajda Ashraf[a], Reaz Uddin[a] and Zaheer Ul-Haq[a,b]

[a]Dr. Panjwani Center for Molecular Medicine and Drug Research, International Center for Chemical and Biological Sciences, University of Karachi, Karachi, Pakistan; [b]H.E.J Research Institute of Chemistry, International Center for Chemical and Biological Sciences, University of Karachi, Karachi, Pakistan

Communicated by Ramaswamy H. Sarma

ABSTRACT

Recently, the world has witnessed outbreak of a novel Coronavirus (SARS-CoV-2), the virus which initially emerged in Wuhan, China has now made its way to a large part of the world, resulting in a public emergency of international concern. The functional importance of Chymotrypsin-like protease ($3CL^{pro}$) in viral replication and maturation turns it into an attractive target for the development of effective antiviral drugs against SARS and other coronaviruses. At present, there is no standard drug regime nor any vaccine available against the infection. The rapid development and identification of efficient interventions against SARS-CoV-2 remains a major challenge. Based on the available knowledge of closely related coronavirus and their safety profiles, repurposing of existing antiviral drugs and screening of available databases is considered a near term strategic and economic way to contain the SARS-CoV-2 pandemic. Herein, we applied computational drug design methods to identify Chymotrypsin-like protease inhibitors from FDA approved antiviral drugs and our in-house database of natural and drug-like compounds of synthetic origin. As a result three FDA approved drugs (Remdesivir, Saquinavir and Darunavir) and two natural compounds (. flavone and coumarine derivatives) were identified as promising hits. Further, MD simulation and binding free energy calculations were performed to evaluate the dynamic behavior, stability of protein-ligand contact, and binding affinity of the hit compounds. Our results indicate that the identified compounds can inhibit the function of Chymotrypsin-like protease ($3CL^{pro}$) of Coronavirus. Considering the severity of the spread of coronavirus, the current study is in-line with the concept of finding the new inhibitors against the vital pathway of the corona virus to expedite the process of drug discovery.

Abbreviations: $3CL^{pro}$: Chymotrypsin-like protease; CSG: Coronavirus Study Group; ICTV: International Committee on Taxonomy of Viruses; HCoV-229E: Human coronavirus 229E; HCoV-HKU1: Human coronavirus HKU1; HCoV-NL63: Human coronavirus NL63; HCoV-OC4: Human coronavirus OC34; MERS-CoV: Middle East Respiratory Syndrome Coronavirus; PLIF: Protein-Ligand Interactions Profile; SARS-CoV: Severe Acute Respiratory Syndrome Coronavirus; WHO: World Health Organization

ARTICLE HISTORY
Received 16 March 2020
Accepted 30 March 2020

KEYWORDS
$3CL^{pro}$; drug repurposing; MD simulation; remdesivir; SARS-CoV-2; virtual screening; corona virus

1. Introduction

Newly emerged coronavirus outbreak of severe asymptomatic pneumonia has geographically been associated with the seafood market of south Chinese Wuhan city, primarily designated as 2019 novel coronavirus (Salata, Calistri, Parolin, & Palù, 2019; Seah & Agrawal, 2020; Hemida & Ba Abduallah, 2020; Chan et al. 2020). The Coronavirus Study Group (CSG) taxonomists working under the umbrella of International Committee on Taxonomy of Viruses (ICTV) have named the new virus, Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) based on the novelty and comparative genomic analysis (Coronaviridae Study Group of the International Committee on Taxonomy of Viruses, 2020). We have witnessed the rapid spread of COVID-19; the official name of diseases caused by SARS-CoV-2, from China to Europe; the current epicenter, America and the Middle East (World Health Organization), the death curve is noticeably higher in many European countries, with the worst hit reported in Italy and Spain. The situation of the South-East Asian countries is also alarming. Being low income countries, most of these countries lack basic health care facilities required to combat the epidemic. With a national tally of 10 As of March 27th, Pakistan with 1057 confirmed cases, Pakistan is the most severely affected South-East Asian Country.

The SARS-COV-2 is the third most highly virulent human coronavirus of the 21st century followed by the SARS-COV and MERS-COV, expressing the highest fatality rate (Zhou et al. 2020; Wang et al., 2020). On 31 Jan 2020, the WHO declared COVID-19 as a public health emergency of international concern (Jiang et al., 2020). Coronaviruses are large

CONTACT Zaheer Ul-Haq zaheer.qasmi@iccs.edu Dr. Panjwani Center for Molecular Medicine and Drug Research, International Center for Chemical and Biological Sciences, University of Karachi, Karachi, 75270, Pakistan

enveloped positive sense single strand-RNA viruses recognized as highly evolving viruses, with high frequency of genomic recombination and mutation (Chen, Liu, & Guo, 2020; Fehr & Perlman, 2015). There are six known species of human coronavirus (HCoV-NL63, HCoV-229E, HCoV-OC34, HCoV-HKU1, SARS-CoV and MERS-CoV) associated with several respiratory tract diseases (Arden et al., 2005; Woo et al. 2005; de Wit et al. 2016; Su et al. 2016) . The new SARS-COV-2 is the seventh strain of human coronavirus which is taxonomically placed in the Genre *Betacoronavirus* exhibiting 89.1% and 60% nucleotide sequence similarity with SARS and MERS coronaviruses, respectively (Wu et al. 2020).

The SARS-COV-2 genome comprises 29,903 nucleotides, with 10 Open Reading Frames (ORFs), while the 3′ terminal regions encode structural viral proteins; spike, membrane envelope and nucleocapsid proteins (Figure 1A). Whereas, the 5′ terminal ORF1ab encodes two viral replicase polyproteins pp1a and pp1b. The proteolytic cleavage of pp1a and pp1b produces 16 non-structural proteins (nsp1 to nsp16) (Figure 1B). Among these non-structural proteins, nsp5 ($3CL^{pro}$) exist in the pp1a, play a key role in the life cycle of coronavirus replication and maturation (Wu et al., 2020). The functional importance of $3CL^{pro}$ turns it into an attractive target for the development of effective antiviral drugs against SARS and other coronaviruses (Anand et al., 2005; Yang et al., 2006).

The 3 D structure of $3CL^{pro}$ of SARS-Cov-2, like other coronaviruses, consists of three domains. The domain I (8-101 amino acid residues) and II (102-184 amino acid residues) are essentially beta-barrels and bear a resemblance to the chymotrypsin, whereas Domain III (201-306 amino acid residues) mainly comprised of alpha-helices. The substrate binding region, located at the cleft of Domain I and II, consists of conserved His41 and Cys145 catalytic dyad, in which Cys act as nucleophile while His acts as a proton acceptor (Figure 1C). In addition to the catalytic center there are two other deeply buried subsites (S1 & S2) and three shallow subsites (S3-S5). The S1 subsite consists of His163, Glu166, Cys145, Gly143, His172, and Phe140 while S2 consist of Cys145, His41, and Thr25 amino acid residues; mainly involved in hydrophobic and electrostatic interactions. The Shallow subsites S3-S5 comprises Met49, His41, Met165, Glu166 and Gln189 amino acid residues (Figure 2). These shallow subsites can tolerate different functionalities (Lu et al. 2006; Jin et al., 2020; Yang et al. 2005).

The FDA approved antiviral drugs remdesivir, lopinavir, ritonavir, oseltamivir and fapilavir have been reported to be effective against SARS-CoV-2. Remdesivir, a nucleotide analogue has presented antiviral activity (EC $50 = 0.77\ \mu M$) against SARS-Cov-2 (Wang et al., 2020). The approved HIV drug Kaletra (combination of lopinavir/ritonavir) has also been studied in combination with flu approved drug Oseltamivir against COVID-19. On Feb 18, 2020 it was reported that a Chinese woman who received this combination recovered after suffering severe SARS-CoV-2 infection. The Fapilavir, is an antiviral drug that has been approved for clinical testing against novel coronavirus. It has been found to be effective in an on-going clinical trial in Shenzhen, Guangdong province (https://www.coronavirustoday.com/covid-19-treatments). These studies provided the rationale for drug repurposing with the hope to discover antiviral drugs to treat COVID-19.

Inhibition of main protease or $3CL^{pro}$ activity is a promising strategy to control the emerging infection of SARS-CoV-2. In this study, we used an approach for target based-virtual screening of an in-house database containing synthetic and natural compounds with the aim of obtaining structurally diverse and potential inhibitors targeting the $3CL^{pro}$. Furthermore, this study is focused on repurposing the existing antiviral drugs based on their target interaction profile. As a result, five compounds were prioritized and proceeded to a detailed study through MD simulation and binding free energy calculation. This study is in-line with the aim of expediting the discovery of treatment against the COVID-19.

2. Material and methods

2.1. Database preparation

For structure-based virtual screening, an in-house database of natural and synthetic molecules along with 16 FDA approved drugs against viral protease was utilized. All the molecules were allowed to adjust hydrogens and lone pairs, where required charges were applied using MMFF94x force field. After protonation, the molecules were subjected to energy minimization in order to optimize the geometry using an RMS gradient of 0.1 kcal/mol/Å. The optimized molecules were then saved in mdb (MOE database) format for further processing.

2.2. Structure-based virtual screening

The MOE software suite [] was used to carry out molecular docking simulations of the designed database. At present, there is a single record of $3CL^{pro}$ of SARS-CoV-2 in the RCSB Protein Data Bank (Berman et al., 2000). The structure has been recently submitted to the protein data bank with the PDB ID: 6LU7 (Jin et al., 2020) The crystal structure of the target protein was protonated using the Protonate 3 D method (Labute, 2009), followed by energy minimization using Amber99 force field implemented in MOE. The grid was set on cognate ligands in the PDB to define the active site exclusively. The placement method Triangular Matcher Algorithm was implemented during the docking along with the London dG scoring functions and GBVI/WSA dG as a rescoring function. All the prepared compounds were docked in the defined binding site. Later on, the binding interactions of crucial residues were fingerprinted between protein and compounds using the Protein Ligand Interaction Fingerprinting (PLIF) module in MOE. Further interactions and binding poses were analyzed visually using Chimera (Pettersen et al., 2004). On the basis of the binding profiles, three FDA approved drugs (Remdesivir, Daraunvir and Saquinavir) and two compounds flavone and coumarin derivatives) from our in-house database were subjected to MD simulation studies and binding free energy calculation using the MM (GB/PB)SA method implemented in AMBER18 (Gohlke, Kiel, & Case, 2003; Case et al., 2005, 2018).

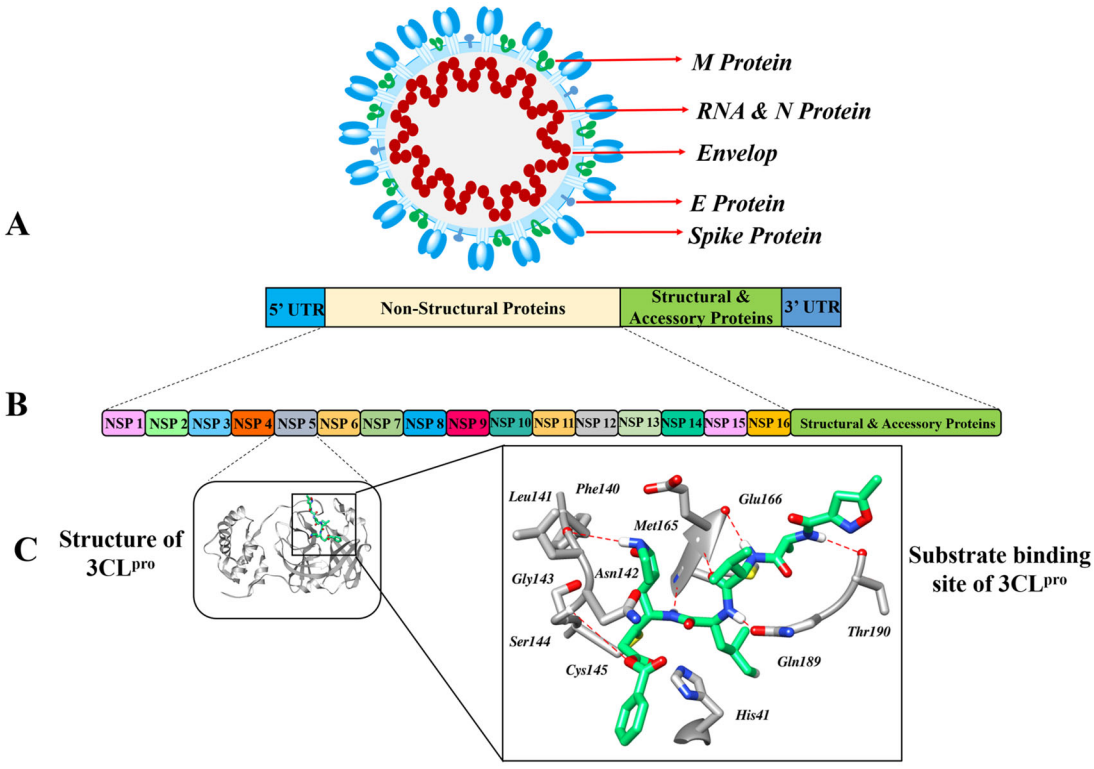


Figure 1. **A**) Diagram of coronavirus structure showing S (Spike) protein, M (membrane) protein, E (envelope) protein, N (nucleocapsid) protein & RNA. **B**) Schematic representation of genome sequence of SARS-CoV-2. Each colored box represented the product of protein (nsp1-nsp16 and structural and accessory protein). **C**) Structure of $3CL^{pro}$ in complex with N3 while the zoom region represented the substrate binding site.

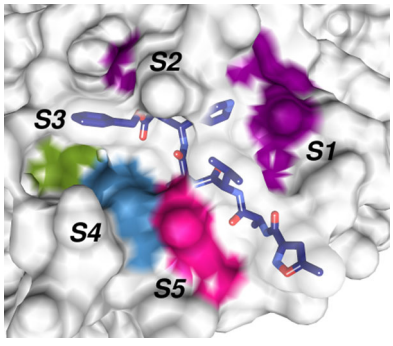


Figure 2. The surface of $3CL^{pro}$ showing substrate binding subsites, color-coded as follows: purple site S1 and S2, olive green site S3, blue site S4, pink site S5.

2.3. Molecular dynamics simulation

Molecular dynamics simulation studies were performed using the pmemd.cuda module in AMBER 18. The antechamber (Wang et al. 2006) and tleap modules were used to prepare the parameters for five systems. The Generalized Amber Force Field (GAFF) was used for the parameterization of ligands (Wang et al., 2004) and the FF14SB force field was used for protein preparation (Maier et al. 2015). A box of 8 Å was generated and solvated by the TIP3P water model (Jorgensen et al., 1983). The systems were neutralized by adding counter ions and periodic boundary conditions were applied. The Particle Mesh Ewald method was used to calculate the long-range electrostatic interactions while the SHAKE algorithm was used to constrain the hydrogen bonds (Essmann et al., 1995; Kräutler et al., 2001).

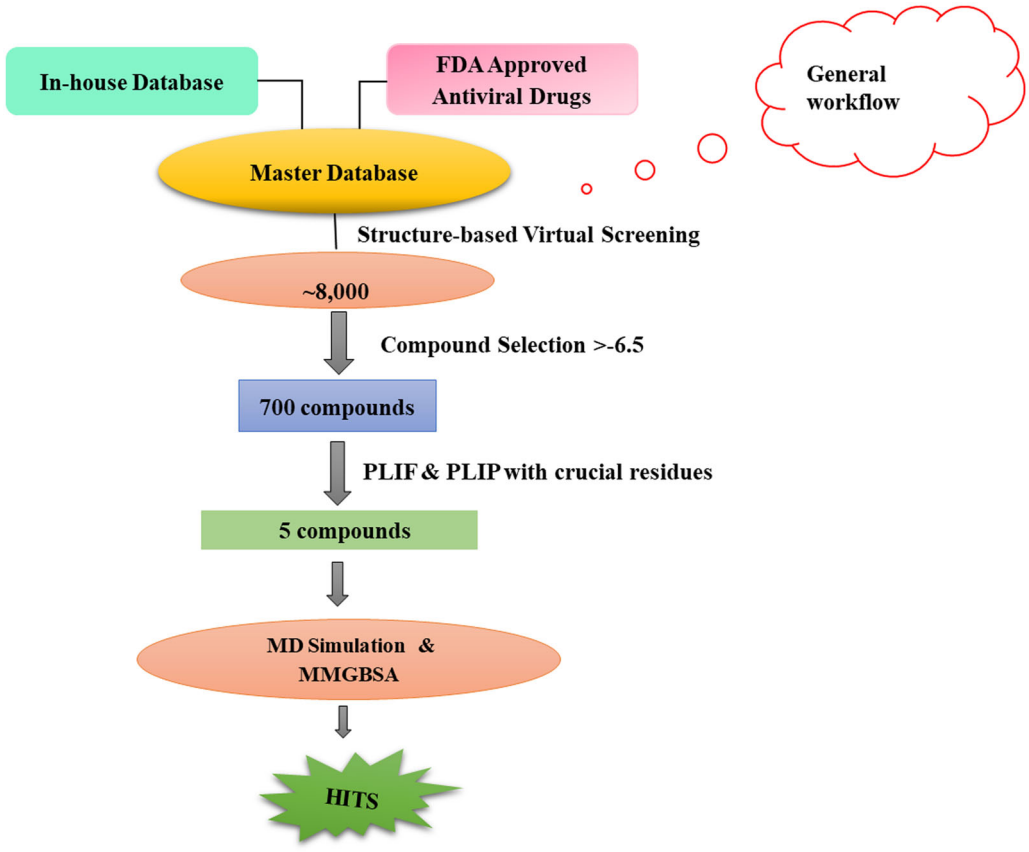
The prepared systems were energy minimized to correct the geometry and steric clashes. The first 2500 steps were carried out using the steepest descent algorithm, while the remaining 2500 steps were carried out by using conjugate gradient algorithms (Fletcher & Powell, 1963). The systems were then gradually heated from 0 to 300 K within 500 ps. A further 500 ps of equilibration was carried out at 300 K to obtain a stable system. Subsequently, the NPT equilibration for the unconstrained system was carried out for 1000 ps at 1 atm pressure and 300 K temperature. The final MD production run of 20 ns was performed with an integration time step of 2.0 fs. The trajectories were collected after every 1 ps and analyzed using the CPPTRAJ module within the Amber software (Roe & Cheatham, 2013). The root mean square deviation, root mean square fluctuation, and radius of gyration were calculated for all the five simulated systems to predict the reasons attributable to their stability. The graphs were plotted using the gnuplot (Racine, 2006).

2.4. Binding free energies

The binding free energy calculation is a promising approach to estimate binding affinities of small molecules with the target protein in explicit solvent (Genheden & Ryde, 2015). In the study Molecular Mechanics-combined with Poisson–Boltzmann or Generalized Born surface (MMGB/PBSA) method was used to evaluate binding free energies of five simulated systems using the following equation:

$$\begin{aligned}\Delta Gbind &= Gcomplex - Gprotein - Gligand \\ &= \Delta E_{MM} + \Delta G_{GB} + \Delta G_{SA} - T\Delta S \\ &= \Delta E_{vdw} + \Delta E_{ele} + \Delta G_{GB} + \Delta G_{SA} - T\Delta S\end{aligned}$$

Where ΔE_{MM} represent the gas-phase interaction energy between protein-ligand complex including van der Waals



Scheme 1. Schematic workflow of structure based virtual screening.

energy contribution (ΔE_{vdw}) and electrostatic energy contribution (ΔE_{ele}), ΔG_G represents the polar while ΔG_{SA} represents the nonpolar components of desolvation free energy. $-T\Delta S$ is the entropy contribution at temperature T. Herein, ΔG_{SA} and ΔG_{GB} were estimated using the accessible surface area (SASA) model with the LCPO method: $\Delta GSA = 0.0072 \times \Delta SASA$ and generalized Born (GB) model, respectively. For each protein complex the binding free energy was calculated from 1000 snapshots extracted from the last 5 ns simulation trajectories.

3. Results and discussion

3.1. Structure-Based virtual screening

In the present study, a structure-based virtual screening was performed using an in-house database and FDA approved drugs against viral proteases using the crystal structure of Coronavirus main proteinase (3CLPro) of SARS-CoV-2. The approach is based on computationally fitting molecules in the target protein using 3 D structure of the active site, followed by ranking of these compounds on the basis of their interaction profile. Recently, the crystal structure of SARS-CoV-2 $3CL^{pro}$ has been submitted by Jin *et al* in the RCSB Protein Data Bank with PDB ID code 6LU7. The crystal structure was used as a starting point for computer aided drug design because of its functional importance in virus survival. The binding site of $3CL^{s}$ consists of conserved catalytic dyad i.e. Cys145 and His41 with other crucial residues i.e. Phe140, Leu141 Asn142, Gly143, Ser144, Cys145, Met165, Glu166, Gln189 and Thr190.

The prepared database was filtered using the docking-based procedure to obtain the potential chemical probes with significant binding affinity towards aforementioned active site residues. The grid was placed on the cognate ligand (N3) to cover all the active site residues. Approximately 8,000 compounds were docked whereas 700 compounds were prioritized with significant affinities or docking scores. The interaction between the active site residue and the selected compounds were carefully analyzed using the PLIF module in MOE. These compounds were then ranked according to the following criteria: (i) low S-score (lower the S-score, stronger the binding affinity) and (ii) interaction with the residues of the binding pocket.

The collection of compounds that resulted mostly showed hydrogen bond interactions with the catalytic dyad. Thus among 700 compounds 5 hits were selected on the basis of significant interaction with catalytic dyad and other hotspot residues of the active site (Scheme 1).

The docking results revealed that among selected drugs Saquinavir showed highest binding affinity compared to other compounds. All 5 selected compounds exhibit significant interactions with the catalytic dyad (Cys145 and His41) in addition to S1-S5 subsite pocket residues. The selected hits were subjected to MD simulation studies followed by binding energy calculation to estimate the stability of these compounds.

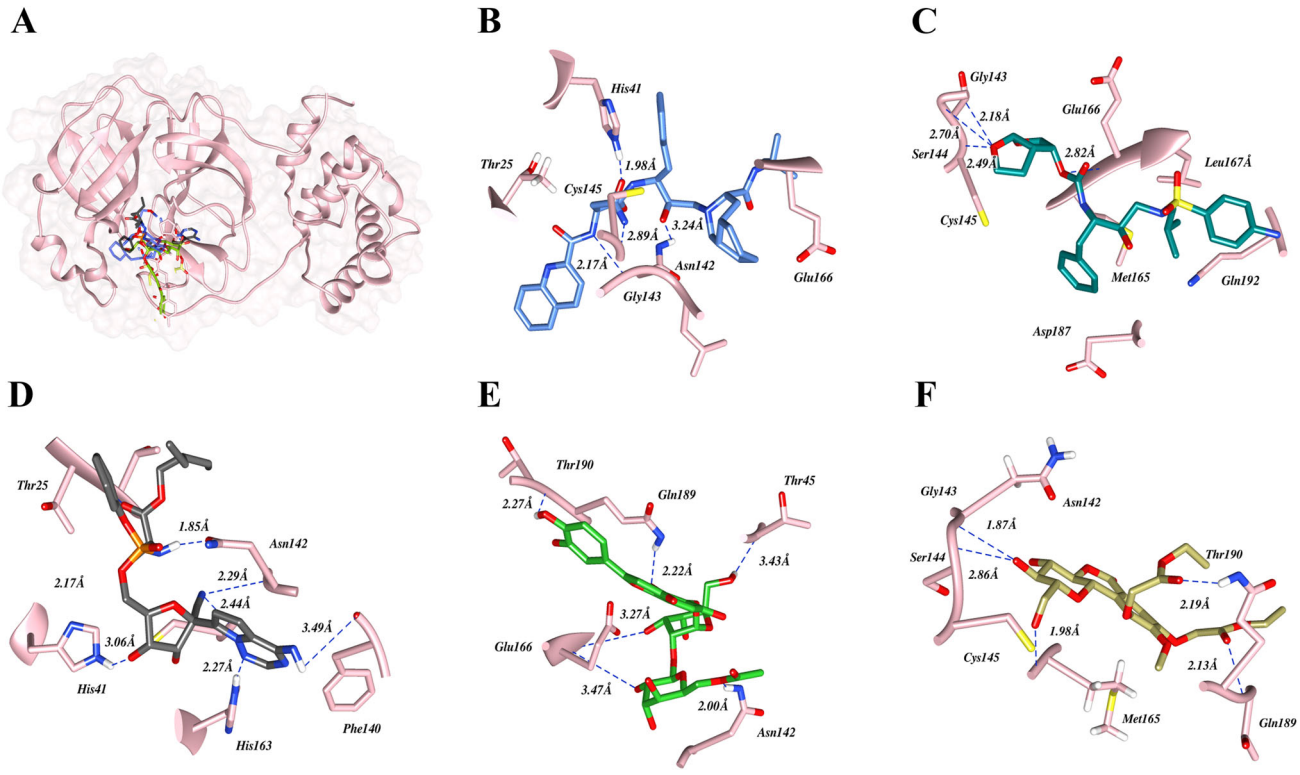


Figure 3. A) Binding mode of all virtual hits in the active site of 3CL^{pro}. B-F) Binding interactions of Saquinavir, Darunavir, Remdesivir, Nat-1 & Syn-16 with the active site residues, respectively. Protein residues are shown in pink stick while blue dash lines depict hydrogen bonds.

3.2. Molecular dynamics simulation

Molecular Dynamics (MD) simulation is an attractive approach to explore the real-time dynamics and conformational stability of a protein upon binding of a ligand. Simulation studies of 20 ns for five selected systems (Saqunavir, Remdesivir, Darunavir, Syn-16, and Nat-1) were carried out. The Root Mean Square Deviation (RMSD) from the initial crystal structure, Root Mean Square Fluctuation (RMSF) of the active site residues and protein-ligand binding contacts in terms of hydrogen bonding and electrostatic interactions were estimated to get an insight into the overall stability of the system. To determine the compactness of the system during simulation, the radial distance of mass weighted atoms from their center of mass (radius of gyration) was also estimated. Because of the vital role of catalytic dyad of 3CL^{pro} in viral replication, His41 and Cys145 are considered to be essential for the activity of viruses. Moreover, in addition to catalytic center, amino acid residues of subsites designated as S1-S5 (Thr25, His163, Glu166, Cys145, Gly143, His172, Phe140, Met49, His41, Met165, Asn142 and Gln189) play a key role in the substrate binding. The visual analysis of MD simulation trajectories suggests that all the virtual hits engaged in significant binding interactions with the hotspot residues of the target protein (Figure 3A).

Upon examining the molecular interaction, it was observed that Saquinavir exhibits a number of significant interactions with the active site residues as depicted in Figure 3B. The amide group filled the S1 and S2 subsite and formed a hydrogen bond with the nitrogen of the imidazole ring of His41. The linker amide group mediated the three hydrogen bonds with the backbone of Asn142, Gly143 and Cys145. Moreover, Thr25 and Glu166 prompted to have alkyl-alkyl interaction in the hydrophobic S2 and S4 pocket. Similarly, all these interactions were highly conserved during the course of simulation with high occupancies and accommodating the ligand to be held at the binding pocket firmly.

The binding mode of Darunavir demonstrated that oxygen of furan ring mediates three hydrogen bonds with the backbone of Gly143, Ser144 and Cys145 in S1 pocket. The fourth hydrogen bond was mediated by a carbonyl group with the backbone of Glu166 in the S4 subsite pocket. Furthermore, aromatic ring in the ligand have prompted π-alkyl interaction with Met165 and Asp187 while alkyl-alkyl interaction observed with Leu167 and Gln192 in the shallow hydrophobic S4 and S5 subsite pockets, which projected ligand firmly in the binding pocket as shown Figure 3C.

The binding mode of Remdesivir illustrated in Figure 3D, suggested that the triazine ring mediated two hydrogen bonds with backbone of Phe140 and nitrogen of imidazole ring of His163 in the S1 pocket. Similarly, the hydroxyl group of the furan ring formed a strong hydrogen bond with His41, and two hydrogen bonds with Thr25 and Asn142 in the S2 subsite. Additionally, alkyl-alkyl interactions were also observed with Thr25 and Thr26. All these interactions were conserved during the simulation with significant occupancies and nestling the ligand in the binding pocket.

The binding mode of the polar ligand Nat-1 depicted in Figure 3E, suggested that aromatic ring in the ligand mediated a π-alkyl interaction with Gln189 in the S5 subsite while the rest of the interactions were hydrogen bonds. The side

Table 1. Comprehensive interactions between protein and selected hits.

System	Alkyl-Alkyl and π-Alkyl Interactions	Hydrogen Bonding	Docking Score
Saquinavir	His41, Thr26, Glu166, Asn142	Gly143, Ser144, Cys145	−8.5
Darunavir	Met165, Leu167, Asp187, Gln192	Gly143, Ser144, Cys145, Glu166	−6.9
Remdesivir	Thr25, Thr26	Thr26, His41, Phe140, Asn142, Cys145, His163	−7.7
Nat-1	Gln189	Gly143, Ser144, Cys145, Glu166, Gln189, Thr190	−8.0
Syn-16	Gln189, Cys145	Asn142, Glu166, Gln189, Thr190	−7.3

Table 2. Distance and occupancy of H Bond contacts between protein and selected hits.

System	Donor	Donor H	Acceptor	Distance (Å)	% Occupancy
Saqunavir	His41:NE2	His41:HE2	Lig:O34	1.98	85
	Asn142:ND2	Asn142:HD21	Lig:O49	3.23	98
	Gly143:N	Gly143:H	Lig:N36	2.16	95
	Cys145:N	Cys145:H	Lig:N35	2.95	99
Darunavir	Gly143:N	Gly143:H	Lig:O27	2.18	95
	Ser144:N	Ser144:H	Lig:O27	2.70	99
	Cys145:N	Cys145:H	Lig:O27	2.49	98
	Glu166:N	Glu166:H	Lig:O19	2.82	98
Remdesivir	His41:NE2	His41:HE2	Lig:O35	3.06	97
	Thr26:N	Thr26:H	Lig:O9	2.17	75
	Lig:N33	Li:H27	Phe140:O	3.86	79
	Asn142:N12	Asn142:H18	Lig:OD1	1.85	95
	Gly143:N	Gly143:H	Lig:N23	2.29	88
	Cys145:N	Cys145:H	Lig:N23	2.44	91
	His163:NE2	His163:HE2	Lig:N29	2.27	96
Nat-1	Lig:O10	Lig:H18	Thr45:N	3.43	68
	Asn142:ND2	Asn142:HD22	Lig:O16	2.00	95
	Glu166:N	Glu166:H	Lig:O8	3.27	97
	Glu166:N	Glu166:H	Lig:O13	3.47	98
	Gln189:NE2	Gln189:HE22	Lig:O1	2.22	50
	Lig:O2	Lig:H6	Thr190:N	2.27	65
Syn-16	Gly143:N	Gly143:H	Lig:O6	1.87	75
	Ser144:N	Ser144:H	Lig:O6	2.86	95
	Met165:N	Met165:H	Lig:O7	1.98	70
	Gln189:NE2	Gln189:HE22	Lig:O9	2.19	98
	Thr190:N	Thr190:H	Lig:O11	2.13	50

chain of Asn142 and backbone of Glu166 and Thr45 in the S1 subsite established hydrogen bonds with different hydroxyl groups. Similarly, the fourth and fifth hydrogen bonds were mediated by Gln189 and Thr190 in the S5 subsite pocket with the hydroxyl group. All these interactions were shown to exhibit more than 50% occupancy during the course of simulation.

The analysis of molecular interactions of Syn-16 after simulation depicted in Figure 3F, suggested that it stabilizes the protein by interacting with the S1, S2 and S5 pocket residues. Three different hydroxyl groups of ligand exhibit hydrogen bonds with the backbone of Met165 and Thr190 and with the side chain of Gln189 in the S5 pocket. Further, three other hydrogen bonds were observed with Gly143, Ser144 and Cys145, which may further stabilize the ligand in the S1 and S2 pocket. All the five selected hits are found to mediate strong hydrogen bonds along with hydrophobic interactions with the hotspot residues of $3CL^{Pro}$ of SARS-CoV-2. Table 1 illustrates the detail of interactions between protein and selected hits whereas hydrogen bonding distances between heavy atoms and their occupancies are provided in Table 2.

3.2.1. Rmsd analysis

The Root Mean Square Deviation (RMSD) imparts the information of overall stability of the protein complex in terms of calculating deviation from the initial structure. As depicted in Figure 4 all the five systems were significantly stable with variable deviation. All the systems have projected RMSD around 3 Å except Nat-1, which shows the RMSD up to 4 Å. Upon calculating the average deviation, Remdesivir presented an average RMSD of 2.45 Å with slight fluctuation in Cα backbone around 13000ps and was stabilized in the remaining simulation. Similarly, in the case of Saquinavir the analysis of RMSD of Cα backbone has prompted an average RMSD of 2.72 Å. However, a streak of continuous drop was observed around 11000ps followed by inconsiderable variation. For Darunavir bound complex, the average RMSD was found to be 2.59 Å with the overall RMSD of around 3 Å suggesting its stability. A negligible fluctuation was observed in the starting 5000ps and around 10000ps and stable thereafter throughout simulation. In case of Nat-1 bound to target protein, the overall RMSD of around 4 Å was observed which was the highest among the five selected systems. However, RMSD was significantly stable throughout the simulation with negligible fluctuation as compared with other systems. Similarly, the average RMSD of 2.53 Å was observed for Syn-16 bound to target protein with inconsiderable fluctuation till 8000ps followed by a stable RMSD till the end of the simulation, indicating the convergence of the system. The data indicates that all the systems showed stable internal motion.

3.2.2. RMSF analysis

The Root Mean Square Fluctuation (RMSF) was also calculated that imparts knowledge on the flexibility of the protein residues. As depicted in Figure 4, all the systems showed

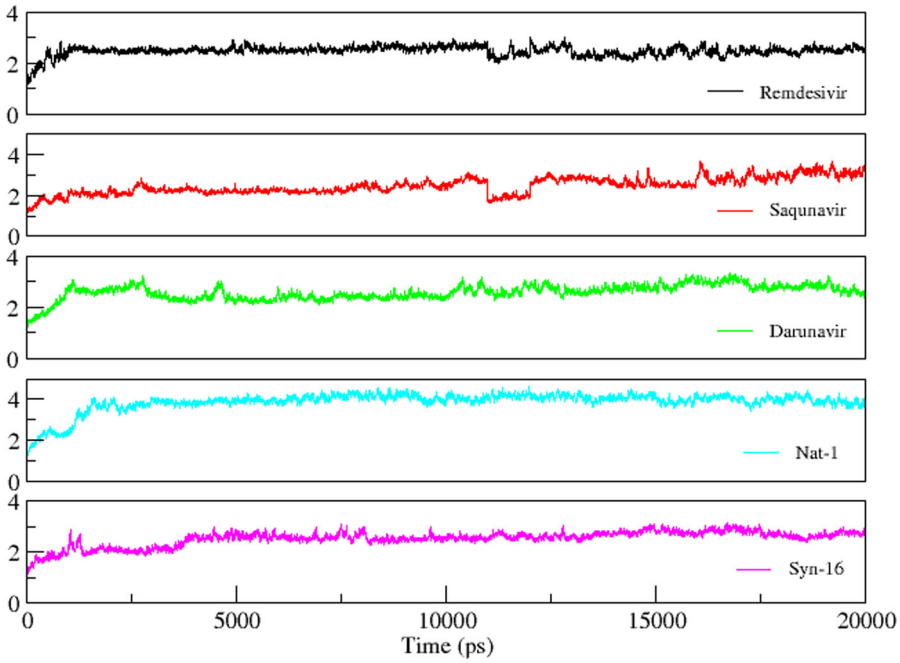


Figure 4. RMSD of the simulated systems calculated as a function of time.

almost similar patterns. The overall value of RMSF for all the five systems was around 3 Å as clearly seen in Figure 5. The loop region was slightly fluctuated in all the systems while regions of active site residues were quite stable throughout the course of simulation. The results here indicate that binding of all five selected hits stabilizes the target protein.

3.2.3. Radius of gyration

Radius of Gyration (RoG) imparts the knowledge of folding and unfolding of protein structure upon binding of the ligands. Therefore, RoG was calculated to determine the compactness of the system with the time. Higher RoG values explain less compactness (more unfolded) with high conformational entropy while low RoG values show high compactness and more stability in the structure (more folded). As evident from Figure 6, all the systems have projected the gyration scores between 21 ± 0.2 to 23 ± 0.2 Å. In the case of Remdesivir, the average gyration score was found to be 22.25 ± 0.1 Å while Saquinavir and Darunavir prompted to have average scores of 22.32 ± 0.4 Å and 22.29 ± 0.6 Å, respectively. Similarly, in case Nat-1 and Syn-16 bound to the target protein the average gyration score was found to be 21.91 ± 0.2 and 22.21 ± 0.7 Å, respectively. The data reveal that all the systems were compact throughout the simulation, which indicate that the systems are well converged.

3.3. Binding free energy

Energy calculation was carried out of all five systems to determine their binding affinities using MM (GB/PB)SA method. To evaluate the binding free energy, 1000 conformations were extracted from the last 5 ns of the MD production run. The calculated ΔG binding energy for Remdesivir, Saqunavir, Darunavir, Nat-1 and Syn-16 with target protein was found to be -45.5240, −36.3026, −48.1041, −41.2565 and -31.5581, respectively. By comparing the binding free energy, it was found that Darunavir bound more tightly to the $3CL^{pro}$ than others. In order to understand the impact of individual energy terms in the binding process, total free energy decomposed into the E_{vdw}, E_{ele}, E_{GB} and ESURF energy components. The results indicated that the major favorable contributors were electrostatic (ΔG_{ele}) and van der Waals interactions (ΔG_{vdw}) while the polar component of solvation (ΔG polar) contributed unfavorably to the binding of all five compounds (Table 3).

Examining the results of MM (GB/PB)SA suggest that the van der Waals and electrostatic interactions play a dominant role in the binding of the ligands. To gain further insight into the protein-ligand interaction or the key residues involved in the binding process, the total binding free energies were decomposed into the individual residue energy contribution (Figure 7). The residues that contributed more than 1.0 kcal/mol in the binding energy, considered as hot-spot residues. It is observed that Phe140, Leu141, Asn142, Gly143, Ser144, Cys145 Met164, Glu166 and Gln189 residues contribute most significantly to the binding energy. For all the five systems the major binding contribution came from catalytic dyad (His41 and Cys145) in the binding affinity.

In Remdesivir, Asn142 of the electrostatic S1 subsite contributes -6.58 kcal/mol to the binding free energy. Similarly, Phe140, Leu141, Gly143, Ser144, Glu166 and Gln189 of the S1 and S5 subsite pockets, contributed between -1.10 to -1.49 kcal/mol to the binding free energy by mediating strong hydrophobic and electrostatic interactions. In case of Saquinavir, the residues of S1 and S4 pocket (His41, Asn142, and Met165) contribute more significantly with binding energies ranging from -1.16 to -2.20 kcal/mol as compared to other active site residues by establishing electrostatic and hydrophobic interactions. The binding of Darunavir to the

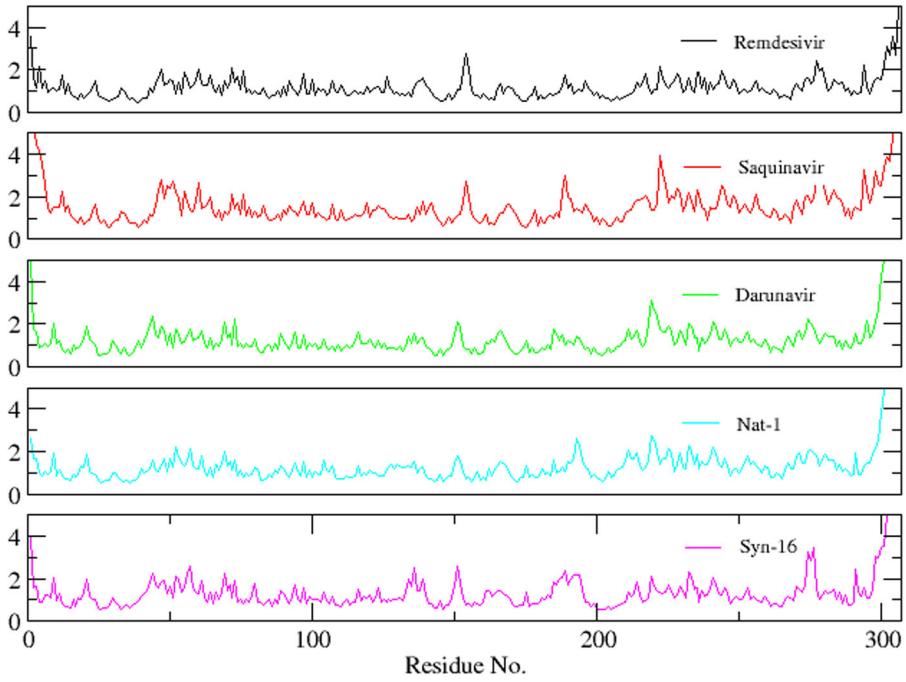


Figure 5. RMSF of amino acid residues of the simulated systems calculated as a function of time.

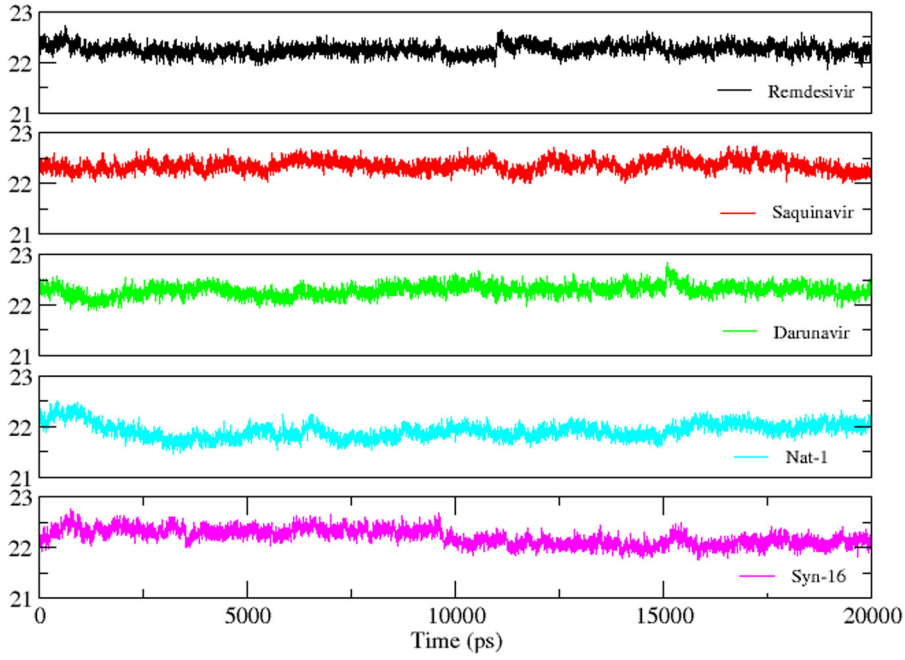


Figure 6. Radius of Gyration of the simulated systems calculated as a function of time.

Table 3. Contribution of each energy component in the binding of all five simulated systems.

System	E_{vdw}	E_{ele}	E_{GB}	ESURF	ΔGB bind	ΔGP Bind
Remdesivir	−55.5595	−40.7738	57.1003	−6.2910	−45.5240	−38.2654
Saquinavir	-53.6286	−20.0359	44.2065	−6.8447	−36.3026	−34.3634
Darunavir	−55.7290	−15.8202	29.8952	−6.4501	−48.1041	−32.6737
Nat-1	−49.8898	−48.5398	63.5133	−6.3401	−41.2565	−30.3325
Syn-16	−45.4254	−86.6787	107.0819	−6.5359	−31.5581	−21.6986

target protein was mediated strongly by the residues of S1, S2 and S5 pockets (His41, Gly143, Ser144, Cys145, Met164, and Gln189) that contribute between -1.21 to -3.71 kcal/mol by establishing hydrophobic contacts with the aromatic rings, and through electrostatic interactions. However, in case of Nat-1, Gln189 contributed most significantly (binding energy of -4.26 kcal/mol) by mediating strong electrostatic interactions. Similarly, the residues of S1 and S5 pocket (His41, Asn142, Met165, and Glu166) majorly contribute to the binding of Nat-1. In the case of Syn-16 the major

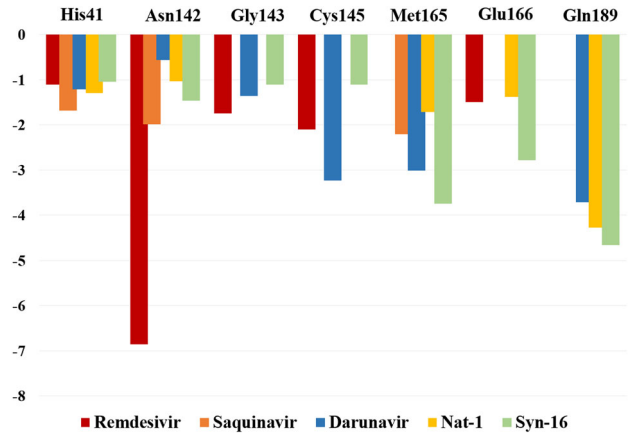


Figure 7. Graphical representation of each residue contribution in binding energy.

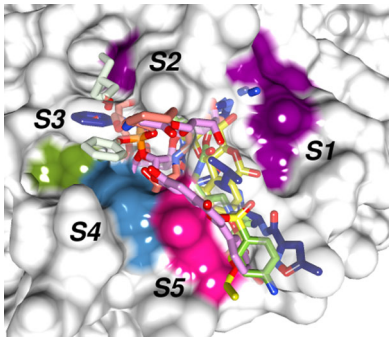


Figure 8. Short-listed candidates potentially interact with S1, S2 and S5 sub sites for $3CL^{pro}$ inhibition.

contribution came from Met165, Glu166 and Gln189 in the range of -2 to -4 kcal/mol mainly by establishing electrostatic interactions. However, Asn142, Gly143, also favorably contribute to the binding affinity, in addition to the catalytic dyad.

4. Conclusion

3C-like protease ($3CL^{pro}$) is essential for the initiation of the viral replication cycle and thus considered as a validated drug target for the treatment of SARS-COV-2 infections. In this study, we identified three FDA-approved drugs (Remdesivir, Saquinavir and Darunavir) and two small molecules (flavone and coumarin derivatives) as potential inhibitors of $3CL^{pro}$ by using an *in silico* approach. Primarily structure-based screening was performed using an in-house database of natural and synthetic molecules along with existing antiviral FDA-approved drugs. Candidates with good docking scores were selected for binding mode analysis. Among them, five compounds were shortlisted for MD simulation and binding free energy calculation to get deep insight in their inhibitory mechanism. Detailed analysis of per-residue energy contribution with short-listed compounds revealed functionally important S1, S2 and S5 subsites for $3CL^{pro}$ inhibition (Figure 8).

Our findings address an urgent need, as currently there is no drug available against this target. As demonstrated in this combined approach, screening and repurposing of FDA-approved antiviral drugs to inhibit $3CL^{pro}$ may provide an alternative fast-track approach for identifying and developing new treatments for SARS-COV-2 infection.

Acknowledgements

Authors gratefully acknowledge the valuable suggestions provided by Prof. Dr. Sarah Harris, University of Leeds, UK.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

ORCID

Zaheer Ul-Haq http://orcid.org/0000-0002-8530-8711

References

Anand, K., Yang, H., Bartlam, M., Rao, Z., & R. Hilgenfeld. (2005). Coronavirus main proteinase: Target for antiviral drug therapy. In A. Schmidt, O. Weber, & M. H. Wolff (Eds.), *Coronaviruses with special emphasis on first insights concerning SARS* (pp. 173–199). Birkhäuser Basel. 10.1007/3-7643-7339-3_9.

Arden, K. E., Nissen, M. D., Sloots, T. P., & Mackay, I. M. (2005). New human coronavirus, HCoV-NL63, associated with severe lower respiratory tract disease in Australia. *Journal of Medical Virology*, *75*(3), 455–462. https://onlinelibrary.wiley.com/doi/abs/10.1002/jmv.20288. doi:10.1002/jmv.20288

Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N., & Bourne, P. E. (2000). The protein data bank. *Nucleic Acids Research*, *28*(1), 235–242. no doi:10.1093/nar/28.1.235

Case, D. A., Ben-Shalom, I. Y., Brozell, S. R., Cerutti, D. S., Cheatham, T. E., III, Cruzeiro, V. W. D., & Darden, T. A. (2018). *AMBER 2018*. San Francisco: University of California.

Case, D. A., Cheatham, T. E., III,Darden, T., Gohlke, H., Luo, R., Merz, K. M., Jr.,Onufriev, A., Simmerling, C., Wang, B., & Woods, R. J. (2005). The amber biomolecular simulation programs. *Journal of Computational Chemistry*, *26*(16), 1668–1688. doi:10.1002/jcc.20290

Chan, J. F.-W., Yuan, S., Kok, K.-H., To, K. K.-W., Chu, H., Yang, J., Xing, F., Liu, J., Yip, C. C.-Y., Poon, R. W.-S., Tsoi, H.-W., Lo, S. K.-F., Chan, K.-H., Poon, V. K.-M., Chan, W.-M., Ip, J. D., Cai, J.-P., Cheng, V. C.-C., Chen, H., Hui, C. K.-M., & Yuen, K.-Y. (2020, February 15). A familial cluster of pneumonia associated with the 2019 novel coronavirus indicating person-to-person transmission: A study of a family cluster. *The Lancet*, *395*(10223), 514–523. doi:10.1016/S0140-6736(20)30154-9

Chen, Y., Liu, Q., & Guo, D. (2020, April). Emerging coronaviruses: Genome structure, replication, and pathogenesis. *Journal of Medical Virology*, *92*(4), 418–423. doi:10.1002/jmv.25681

Coronaviridae Study Group of the International Committee on Taxonomy of Viruses. (2020, March 2). The species severe acute respiratory syndrome-related coronavirus: Classifying 2019-nCoV and naming it SARS-CoV-2. *Nature Microbiology*, 1. 10.1038/s41564-020-0695-z.

de Wit, van Doremalen, N., Falzarano, D., & Munster, V. J. (2016, August). SARS and MERS: Recent insights into emerging coronaviruses. *Nature Reviews Microbiology*, *14*(8), 523–534. doi:10.1038/nrmicro.2016.81

Essmann, U., Perera, L., Berkowitz, M. L., Darden, T., Lee, H., & Pedersen, L. G. (1995). A smooth particle mesh ewald method. *The Journal of Chemical Physics*, *103*(19), 8577–8593. doi:10.1063/1.470117

Fehr, A. R., & Perlman, S. (2015). Coronaviruses: An overview of their replication and pathogenesis. *Methods in Molecular Biology*, *1282*, 1–23. 10.1007/978-1-4939-2438-7_1.

Fletcher, R., & Powell, M. J. D. (1963, August). A rapidly convergent descent method for minimization. *The Computer Journal*, *6*(2), 163–168. https://academic.oup.com/comjnl/article-abstract/6/2/163/364776. doi:10.1093/comjnl/6.2.163

Genheden, S., & Ryde, U. (2015, May). The MM/PBSA and MM/GBSA methods to estimate ligand-binding affinities. *Expert Opinion on Drug Discovery*, *10*(5), 449–461. doi:10.1517/17460441.2015.1032936

Gohlke, H., Kiel, C., & Case, D. A. (2003, July 18). Insights into protein-protein binding by binding free energy calculation and free energy decomposition for the Ras-Raf and Ras-RalGDS complexes. *Journal of Molecular Biology*, *330*(4), 891–913. doi:10.1016/S0022-2836(03)00610-7

Hemida, M. G., & Ba Abduallah, M. M. (2020, March 16). The SARS-CoV-2 outbreak from a one health perspective. *One Health*, 100127. http://www.sciencedirect.com/science/article/pii/S2352771420300185. doi:10.1016/j.onehlt.2020.100127

Jiang, F., Deng, L., Zhang, L., Cai, Y., Cheung, C. W., & Xia, Z. (2020, March 4). Review of the clinical characteristics of coronavirus disease 2019 (COVID-19). *Journal of General Internal Medicine*. 10.1007/s11606-020-05762-w

Jin, Z., Du, X., Xu, Y., Deng, Y., Liu, M., Zhao, Y., & Zhang, B. (2020). Structure of Mpro from COVID-19 virus and discovery of its inhibitors. *bioRxiv*. https://www.biorxiv.org/content/10.1101/2020.02.26.964882v2

Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W., & Klein, M. L. (1983). Comparison of simple potential functions for simulating liquid water. *The Journal of Chemical Physics*, *79*(2), 926–935. doi:10.1063/1.445869

Kräutler, V., van Gunsteren, W. F., & Hünenberger, P. H. (2001). A fast SHAKE algorithm to solve distance constraint equations for small molecules in molecular dynamics simulations. *Journal of Computational Chemistry*, *22*(5), 501–508. https://onlinelibrary.wiley.com/doi/abs/10.1002/1096-987X(20010415)22:5%3C501::AID-JCC1021%3E3.0.CO;2-V. doi:10.1002/1096-987X(20010415)22:5<501::AID-JCC1021>3.0.CO;2-V

Labute, P. (2009, April). Protonate3D: Assignment of ionization states and hydrogen coordinates to macromolecular structures. *Proteins: Structure, Function, and Bioinformatics*, *75*(1), 187–205. doi:10.1002/prot.22234

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. (2006, August 24). Structure-based drug design and structural biology study of novel nonpeptide inhibitors of severe acute respiratory syndrome coronavirus main protease. *Journal of Medicinal Chemistry*, *49*(17), 5154–5161. doi:10.1021/jm060207o

Maier, J. A., Martinez, C., Kasavajhala, K., Wickstrom, L., Hauser, K. E., & Simmerling, C. (2015, August 11). ff14SB: Improving the accuracy of protein side chain and backbone parameters from ff99SB. *Journal of Chemical Theory and Computation*, *11*(8), 3696–3713. doi:10.1021/acs.jctc.5b00255

Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., & Ferrin, T. E. (2004). UCSF chimera—A visualization system for exploratory research and analysis. *Journal of Computational Chemistry*, *25*(13), 1605–1612. doi:10.1002/jcc.20084

Racine, J. (2006, January). Gnuplot 4.0: A portable interactive plotting utility. *Journal of Applied Econometrics*, *21*(1), 133–141. http://doi.wiley.com/10.1002/jae.885. doi:10.1002/jae.885

Roe, D. R., & Cheatham, T. E. III. (2013). PTRAJ and CPPTRAJ: Software for processing and analysis of molecular dynamics trajectory data. *Journal of Chemical Theory and Computation*, *9*(7), 3084–3095. doi:10.1021/ct400341p

Salata, C., Calistri, A., Parolin, C., & Palù, G. (2019, December 1). Coronaviruses: A paradigm of new emerging zoonotic diseases. *Pathogens and Disease*, *77*(9). doi:10.1093/femspd/ftaa006

Seah, I., & Agrawal, R. (2020, March 16). Can the coronavirus disease 2019 (COVID-19) affect the eyes? A review of coronaviruses and ocular implications in humans and animals. *Ocular Immunology and Inflammation*, 1–5. 10.1080/09273948.2020.173850

Su, S., Wong, G., Shi, W., Liu, J., Lai, A. C. K., Zhou, J., Liu, W., Bi, Y., & Gao, G. F. (2016, June). Epidemiology, genetic recombination, and pathogenesis of coronaviruses. *Trends in Microbiology*, *24*(6), 490–502. doi:10.1016/j.tim.2016.03.003

Wang, C., Horby, P. W., Hayden, F. G., & Gao, G. F. (2020). A novel coronavirus outbreak of global health concern. *The Lancet*, *395*(10223), 470–473. thelancet.com. doi:10.1016/S0140-6736(20)30185-9

Wang, J., Wang, W., Kollman, P. A., & Case, D. A. (2006, October). Automatic atom type and bond type perception in molecular mechanical calculations. *Journal of Molecular Graphics and Modelling*, *25*(2), 247–260. doi:10.1016/j.jmgm.2005.12.005

Wang, J., Wolf, R. M., Caldwell, J. W., Kollman, P. A., & Case, D. A. (2004). Development and testing of a general amber force field. *Journal of Computational Chemistry*, *25*(9), 1157–1174. doi:10.1002/jcc.20035

Woo, P. C. Y., Lau, S. K. P., Chu, C-m., Chan, K-h., Tsoi, H-w., Huang, Y., Wong, B. H. L., Poon, R. W. S., Cai, J. J., Luk, W-k., Poon, L. L. M., Wong, S. S. Y., Guan, Y., Peiris, J. S. M., & Yuen, K-Y. (2005, January). Characterization and complete genome sequence of a novel coronavirus, coronavirus HKU1, from patients with pneumonia. *Journal of Virology*, *79*(2), 884–895. doi:10.1128/JVI.79.2.884-895.2005

World Health Organization. (2020). *Naming the coronavirus disease (covid-19) and the virus that causes it*.

Wu, F., Zhao, S., Yu, B., Chen, Y.-M., Wang, W., Song, Z.-G., Hu, Y., Tao, Z.-W., Tian, J.-H., Pei, Y.-Y., Yuan, M.-L., Zhang, Y.-L., Dai, F.-H., Liu, Y., Wang, Q.-M., Zheng, J.-J., Xu, L., Holmes, E. C., & Zhang, Y.-Z. (2020, March). A new coronavirus associated with human respiratory disease in China. *Nature*, *579*(7798), 265–269. doi:10.1038/s41586-020-2008-3

Yang, H., Bartlam, M., & Rao, Z. (2006). Drug design targeting the main protease, the Achilles' Heel of coronaviruses. *Current Pharmaceutical Design*, *12*(35), 4573–4590. doi:10.2174/138161206779010369

Yang, H., Xie, W., Xue, X., Yang, K., Ma, J., Liang, W., Zhao, Q., Zhou, Z., Pei, D., Ziebuhr, J., Hilgenfeld, R., Yuen, K. Y., Wong, L., Gao, G., Chen, S., Chen, Z., Ma, D., Bartlam, M., & Rao, Z. (2005, October). Design of wide-spectrum inhibitors targeting coronavirus main proteases. *PLoS Biology*, *3*(10), e324. doi:10.1371/journal.pbio.0030324

Zhou, Y., Hou, Y., Shen, J., Huang, Y., Martin, W., & Cheng, F. (2020, March 16). Network-based drug repurposing for novel coronavirus 2019-nCoV/SARS-CoV-2. *Cell Discovery*, *6*(1), 14. doi:10.1038/s41421-020-0153-3