Drug Repurposing Studies Targeting SARS-nCoV2: An Ensemble Docking Approach on Drug Target 3C-like Protease (3CLpro)

Shruti Koulgi, Vinod Jani, Mallikarjunachari Uppuladinne, Uddhavesh Sonavane, Asheet Kumar Nath, Hemant Darbari, Rajendra Joshi

Submitted date: 01/05/2020 • Posted date: 04/05/2020
Licence: CC BY-NC-ND 4.0

Citation information: Koulgi, Shruti; Jani, Vinod; Uppuladinne, Mallikarjunachari; Sonavane, Uddhavesh; Nath, Asheet Kumar; Darbari, Hemant; et al. (2020): Drug Repurposing Studies Targeting SARS-nCoV2: An Ensemble Docking Approach on Drug Target 3C-like Protease (3CLpro). ChemRxiv. Preprint.
https://doi.org/10.26434/chemrxiv.12228831.v1

The COVID-19 pandemic has been responsible for several deaths worldwide. The causative agent behind this disease is the Severe Acute Respiratory Syndrome – novel Coronavirus 2 (SARS-nCoV2). SARS-nCoV2 belongs to the category of RNA viruses. The main protease, responsible for the cleavage of the viral polyprotein is considered as one of the hot targets for treating COVID-19. Earlier reports suggest the use of HIV anti-viral drugs for targeting the main protease of SARS-CoV, which caused SARS in the year 2002-03. Hence, drug repurposing approach may prove to be useful in targeting the main protease of SARS-nCoV2.

The high-resolution crystal structure of 3CL\textsuperscript{pro} (main protease) of SARS-nCoV2 (PDB ID: 6LU7) was used as the target. The Food and Drug Administration (FDA) approved and SWEETLEAD database of drug molecules were screened. The apo form of the main protease was simulated for a cumulative of 150 ns and 10 μs open source simulation data was used, to obtain conformations for ensemble docking. The representative structures for docking were selected using RMSD-based clustering and Markov State Modeling analysis. This ensemble docking approach for main protease helped in exploring the conformational variation in the drug binding site of the main protease leading to efficient binding of more relevant drug molecules. The drugs obtained as best hits from the ensemble docking possessed anti-bacterial and anti-viral properties. Small molecules with these properties may prove to be useful to treat symptoms exhibited in COVID-19. This in-silico ensemble docking approach would support identification of potential candidates for repurposing against COVID-19.

File list (2)

| File Name                                      | Size   |
|------------------------------------------------|--------|
| Covid19_Manscript_30April_v3.pdf              | 1.70 MiB |
| Supplementary_Data_30April.pdf                | 1.49 MiB |
Drug repurposing studies targeting SARS-nCoV2: An ensemble docking approach on drug target 3C-like protease (3CL\textsuperscript{pro})

Shruti Koulgi\textsuperscript{+}, Vinod Jani\textsuperscript{+}, Mallikarjunachari Uppuladinne V N\textsuperscript{±}, Uddhavesh Sonavane, A. K. Nath, Hemant Darbari and Rajendra Joshi*  

Centre for Development of Advanced Computing (C-DAC), Panchwati, Pashan, Pune, India  

*corresponding author: rajendra@cdac.in  

\textsuperscript{±}joint first authorship

Abstract

The COVID-19 pandemic has been responsible for several deaths worldwide. The causative agent behind this disease is the Severe Acute Respiratory Syndrome – novel Coronavirus 2 (SARS-nCoV2). SARS-nCoV2 belongs to the category of RNA viruses. The main protease, responsible for the cleavage of the viral polyprotein is considered as one of the hot targets for treating COVID-19. Earlier reports suggest the use of HIV anti-viral drugs for targeting the main protease of SARS-CoV, which caused SARS in the year 2002-03. Hence, drug repurposing approach may prove to be useful in targeting the main protease of SARS-nCoV2. The high-resolution crystal structure of main protease of SARS-nCoV2 (PDB ID: 6LU7) was used as the target. The Food and Drug Administration (FDA) approved and SWEETLEAD database of drug molecules were screened. The apo form of the main protease was simulated for a cumulative of 150 ns and 10 μs open source simulation data was used, to obtain conformations for ensemble docking. The representative structures for docking were selected using RMSD-based clustering and Markov State Modeling
analysis. This ensemble docking approach for main protease helped in exploring the conformational variation in the drug binding site of the main protease leading to efficient binding of more relevant drug molecules. The drugs obtained as best hits from the ensemble docking possessed anti-bacterial and anti-viral properties. Small molecules with these properties may prove to be useful to treat symptoms exhibited in COVID-19. This *in-silico* ensemble docking approach would support identification of potential candidates for repurposing against COVID-19.

**Keywords:** COVID-19, SARS-nCoV2, main protease, repurposing, cryptic pockets, 3CL\textsuperscript{pro}
Introduction

The COVID-19 pandemic caused by the SARS-nCoV2 is known to spread quite rapidly. The first incidence of this disease was found in Wuhan, China in December 2019. The World Health Organization (WHO) has reported over 2.8 million affected individuals and 1.9 million deaths by the end of April 2020 [1]. In a very short span of time COVID-19 has spread all over the globe. The SARS-nCoV2, has an RNA genome of around 30K nucleotides. This genome is known to code for the entire viral proteome. The entire coding RNA can be divided into three regions, non-structural protein (nsp) coding region, structural protein coding region and accessory protein coding region [2,3,4]. The non-structural protein region consisting of ORF1a and ORF1b codes for the polyprotein pp1a and pp1b. These polyproteins are further cleaved to form 16 nsp. The structural protein region codes for the Spike(S)-glycoprotein, envelope protein, membrane protein and nucleoprotein [4]. These proteins are responsible for the viral replication, viral functioning and viral-host interaction. Hence, therapeutic studies targeting these proteins have gained importance in the drug industry [5]. Few of the major proteins which have been considered as potential drug targets based on the earlier therapeutics developed against old coronaviruses include, viral proteases, RNA dependent RNA polymerase and viral surface Spike protein [6,7]. The need of the hour lies in development of fast therapeutics for combating the SARS-nCoV2. Drug repurposing is one such strategy which is being extensively used worldwide to design vaccines against this coronavirus [8-11]. To understand how similar is the SARS-nCoV2 in comparison to the earlier coronaviruses sequence comparison studies have been performed to understand the variation in the sequence of these potential target proteins [12-18]. These studies would help in repurposing and repositioning the drugs used earlier to be targeted against this novel coronavirus. Drug repurposing studies are being performed through experimental as well as
computational techniques [19-30]. All the three drug targets mentioned have been extensively studied using drug repurposing [20,21,33,34]. The viral main protease also known as 3Chymotrypsin-like protease (3CL<sub>p</sub>) or main protease is formed by the autocleavage of the polyprotein and further responsible for cleavage of various other nsp [4]. 3CL<sub>p</sub> itself is the nsp5 amongst the 16 nsp. The high-resolution crystal structure of 3CL<sub>p</sub> was elucidated in February 2020 [35]. Before the availability of this structure, modeling studies were performed for the SARS-nCoV2 3CL<sub>p</sub> and it was found to be similar to the SARS-CoV and MERS main protease [24]. Drug repurposing studies on this model using in-silico approaches revealed the role of previously used antivirals as a potential drug for COVID19 [24]. The inhibitor interactions with 3CL<sub>p</sub> has been well explored using electron density maps, the residues crucial to initiate the inhibitory effect upon interacting with the drugs have been identified in these experimental studies [36]. There are reports on protein-protein interaction networks, where the entire viral proteome has been studied to find its interactions with the human proteins [37]. This study reveals the role of 3CL<sub>p</sub> in obstructing the inflammatory and interferon pathway in humans by inhibiting the nuclear transport of epigenetic regulatory proteins [37]. There are experimental studies that reveal the inhibitor binding site of the 3CLpro is divided into various sub-sites [38]. Interaction with the residues in these sites would help in designing inhibitors for the same [38].

This current article describes the scope of repurposing drugs for COVID-19 considering the conformational variation in the inhibitory binding site of 3CL<sub>p</sub> of SARS-nCoV2. The states of the protein than remain unexplored in experimental techniques such as X-ray crystallography were witnessed, which further helped in identifying even better potential candidates for drug
repurposing. The crystal structure with PDB ID 6LU7 shows the presence of covalently bound inhibitor N3 in its active site (Figure 1) [35]. This structure has been widely used by the research community for performing molecular docking and simulation studies [39-42]. The 3D coordinates for the apo form of 3CL\textsuperscript{pro} were obtained from the PDB ID 6LU7 [35]. This SARS-nCoV2 drug target was screened for probable drug candidates against FDA approved and SWEETLEAD drug database [43,44]. Two approaches were used for drug repurposing studies against the drug target 3CL\textsuperscript{pro} (Figure 2). The first approach involved docking of the FDA approved and SWEETLEAD drug database against the crystal structure 6LU7 (Figure 2A). This approach has been referred as “Direct Docking” further. The second approach involved docking of the two mentioned databases against an ensemble of structures obtained from molecular dynamics (MD) simulations of the apo form of 3CL\textsuperscript{pro} (Figure 2B). This approach has been referred as “Ensemble docking” further. MD simulations were performed for a cumulative of 150 ns for the apo 3CL\textsuperscript{pro}. An open source 10 μs MD simulation data of 3CL\textsuperscript{pro} dimer was obtained from the simulations performed on MDGRAPE-4A supercomputing cluster located at RIKEN BDR, Japan [41]. An ensemble of conformations to perform high throughput docking were obtained using Root Mean Square Deviation (RMSD) based clustering and Markov State Modelling (MSM) analysis. A total of 16 conformations of 3CL\textsuperscript{pro} were obtained from RMSD-based clustering and MSM analysis. These, 16 conformations
were docked against the FDA approved and SWEETLEAD drug database. The best ranked drugs obtained from both these approaches belonged to the class of anti-bacterial and anti-viral drugs. The docking scores obtained for the ensemble docking revealed to be better than those obtained through direct docking. The conformation variation in the drug binding site of 3CL\textsuperscript{pro} was observed, as a greater number of target protein residues interacted with the drug molecule. This lacked when the drugs were docked state against the crystal structure. The drug binding pockets were found to be more accessible in case of the ensemble docking. The findings obtained through the direct docking and ensemble docking of 3CL\textsuperscript{pro} of SARS-nCoV2 have been discussed further in this article. These observations may prove to be useful \textit{in-silico} approaches in designing/repurposing drugs against COVID-19.
Methodology

The high-resolution crystal structure of 3CL\textsuperscript{pro} was retrieved from the Protein Data Bank with PDB ID: 6LU7 [35]. This PDB file was cleaned by removing the ligand coordinates. This PDB in the apo form was further considered for the molecular dynamics simulations and docking studies. The detailed protocol followed was been explained in the Figure 2. Direct docking was performed on the 3CL\textsuperscript{pro} protein against the FDA approved drug database and the SWEETLEAD database. This docking was performed using DOCK 6 [45]. The receptor preparation was done using UCSF Chimera and further the active site pocket identification and docking was performed using DOCK 6 (Figure 2A) [45,46]. In case of ensemble docking the coordinates for 3CL\textsuperscript{pro} were obtained from molecular dynamics simulations (Figure 2 B). The simulations were performed using the AMBER 16 simulation package [47]. The AMBER FF14SB force field was used for generation of the parameters. The system was neutralized by Na+ ions and solvate using the TIP3P water model.

![Figure 2: Two approaches used for the drug repurposing and docking studies viz. Direct docking (A) and Ensemble docking (B)](image-url)
The minimization was performed for 20000 steps using the steepest descent and the conjugate gradient method. The system gradually heated to 300 K using the Langevian thermostat. The SHAKE algorithm was employed for dealing with the hydrogen restraints. The equilibration was performed for 1ns at NPT with temperature being 300 K and pressure being 1 atm. The production run was performed for 50 ns using the NPT ensemble. Three parallel runs of 50 ns each were performed based on the explained MD protocol, hence a cumulative of 150 ns data was generated. A 10 µs simulation data on dimer of 3CL\textsuperscript{pro} was obtained MDGRAPE-4A, at RIKEN BDR, JAPAN [41]. However, only monomer data was used from this 10 µs as the simulations performed in-house belonged to the monomer. A cumulative data of 10.15 µs was subjected to RMSD based clustering and Markov State Modelling (MSM) analysis. The RMSD cut-off of 1.7 Å was used for the clustering using the \textit{dbscan} method of \textit{cpptraj} module of AmberTools 17 [48]. A total of 12 representative conformations were obtained through RMSD-based clustering. The MSM analysis was performed using the PyEmma software [49]. The backbone dihedral angles were used as the collective variable for performing the MSM analysis. The complete details of the MSM analysis performed to obtained different states of 3CL\textsubscript{pro} has been given in the supplementary data as SI1. MSM analysis is widely used for significant sampling of the MD simulation data [50-54]. Similar methodology for identifying significant states from the simulation data was used in one of the earlier works reported by our group [50]. A total of 4 representative conformations for 3CL\textsuperscript{pro} were obtained from the MSM analysis. Hence, a total of 16 conformations exploring the ensemble of 3CL\textsuperscript{pro} were considered for ensemble docking approach described in Figure 1B. The FDA approved and SWEETLEAD drug database were screened against these 16 ensemble structures.

Results and Discussion

Direct Docking
The direct docking of the 3CL\textsuperscript{pro} against the FDA approved drug database was performed using DOCK6 [45]. The drugs were screened and ranked based on their grid score. The drugs that’s ranked in the top five were ceftazidime (PubChem CID: 5481173), quetiapine (PubChem CID: 5002), cabergoline (PubChem CID: 54746), enoxolone (PubChem CID: 10114) and apremilast (PubChem CID: 11561674) in the order of decreasing rank. These drugs and their current usage for treating different diseases has been listed in the Table 1. The information on the current usage of these drugs was obtained from PubChem. The grid scores obtained for these drugs have been shown in Figure 3A. The best ranked drug ceftazidime is known to be used as an antibacterial in respiratory tract infections. Enoxolone, which ranked fourth as per the grid score has glycyrrhetinic acid as a sub-component, which is a known plant derivative and is also known to possess antibacterial and anti-viral properties. The drugs, quetiapine and cabergoline, that ranked as second and third respectively, are known to be effective in treating neurological diseases such as bipolar disorder and Parkison’s disease. Apremilast, which ranked as fifth is known to be used in treating psoriasis. The goal of docking against the FDA approved drug database was to find previously known drugs which would be effective in treating the symptoms of the current disease in investigation. The patient suffering from COVID-19 are known to have pneumonia and severe lung infections, hence drugs possessing antibacterial and antiviral properties may prove to have the potential of being repurposed. The results obtained from direct docking infer that out of the
best five docked drugs, two of them appear to carry antibacterial and antiviral properties which were also found to be specific to the respiratory tract infections.

The direct docking of 3CL\textsuperscript{pro} was performed against the SWEETLEAD drug database using DOCK 6. The SWEETLEAD drug hosts around more than 10K drug molecules which include the approved drugs, rejected drugs and molecules isolated from traditional medicinal herbs [44]. One of the docking studies by Smith. et. al. on the SARS-nCoV2 spike protein mentioned the use of SWEETLEAD for identifying few molecules with potential inhibitory activity against the viral spike protein [33]. The docking of 3CL\textsuperscript{pro} against this database lead to the identification of few small molecules which may prove to be potential candidates for drug development against the COVID-19 disease [33]. The drugs that ranked in the top five were dibekacin (PubChem CID: 470999), micronomicin (PubChem CID: 3037206), catalposide (PubChem CID: 93039), dihydro-alpha-ergocryptine (PubChem CID: 114948) and itopride (PubChem CID: 3792) in the decreasing order of their ranks. The details about the current usage of these drugs for treating various diseases and their respective grid scores has been given in Table 2. The grid scores of each of these drugs has been shown in Figure 3B. The drugs that were observed to be ranked in the top two were dibekacin and micronomicin which belonged to class of aminoglycoside antibiotics. Catalposide which ranked third is known to be a plant derivative and it used as an anti-inflammatory agent.
The remaining two drugs dihydro-alpha-ergocryptine and itopride are used for treating Parkinson’s disease and gastrointestinal ailments, respectively. Two of the drugs amongst the top five ranked drugs are known to have antibacterial activity viz. dibekacin and micrionicin.

**Ensemble generation**

The ensemble generation for docking was prepared based on two clustering methods. The first method employed was RMSD-based clustering. The simulation data of 10.15 μs was clustered based on the all atom RMSD of all the residues of the 3CL\textsuperscript{pro} system. The cpptraj module of AMBERTOOLS 17 was used for performing this clustering using the dbscan method \[48\]. The RMSD cut-off used here was of 1.7 Å for the in-house as well as the open source simulation data.

Total three clusters were obtained from the in-house simulation data and nine clusters were obtained from the open source simulation data. The representatives of these 12 clusters were considered for the further docking studies. Figure 4 shows the RMSD values of these cluster representatives against the experimental structure of 3CL\textsuperscript{pro} i.e. PDB ID: 6LU7. The structures represented in blue are the ones that were obtained from RMSD based clustering. Ten of the representative structures...
from the ensemble had an RMSD of below 3 Å. In order, to have more significant conformations from the ensemble a second approach of MSM analysis was performed on the entire simulation data. The collective variable used for the MSM analysis was the backbone torsion angle. Based on this CV, four significant states were obtained for the entire simulation data (Supplementary Info SI 1). The RMSD values of these four representative structures obtained from MSM analysis have been given in the figure. The representative structures shown in purple were obtained from the MSM analysis (Figure 4). All the four representative structures had an RMSD of less than 3 Å, whereas three of them showed an RMSD below 2 Å. These 16 structures represented the ensemble covered by the protein. The varying RMSD values infer that the dynamics of the protein helped in surfacing out conformations that differ from the experimentally derived static conformation of the protein. The flexibility of the protein was captured in these representative structures which further helped in docking of few other small molecules. Considering these ensemble structures, helped to explore wider range of drug molecules that would bind to the target protein, in this case the 3CL\textsuperscript{pro} protein. There are studies where the role of molecular dynamics in exploring the different conformations of binding site also referred to as cryptic pockets helps in computer aided drug discovery [55]. Hence, the identification of different states of 3CL\textsuperscript{pro} through MD simulations helped in visiting different conformations of the binding site. The information on varying binding site may also lead to the identification of more significant drug molecules, further increasing the scope of therapeutics through drug repurposing.

**Ensemble Docking**

A total of 16 representatives were selected from the MD simulation data using RMSD-based clustering and MSM analysis. The FDA approved and SWEETLEAD drug database was screened to obtain the best docked drugs against these 16 representative structures. Figure 5 explains the
name of the best ranked drugs against these ensemble representatives and their corresponding grid scores obtained from DOCK 6. The details about these drugs obtained from FDA approved and SWEETLEAD drug database mentioning their earlier purpose has been described in Table 3 and 4 respectively. Figure 5A explains the screening of the FDA approved drug database against the ensemble representative structures. The drug with the best grid score was indinavir (PubChem CID: 5362440), which is a known HIV protease inhibitor. This was followed by ceftin (PubChem CID: 6321416) which a cephalosporin-derivative and is used as an antibiotic to fight bacterial infection. The third best score was observed for the drug ivermectin (PubChem CID: 6321424), which is used in treating head lice, and is known to possess anti-parasitic property. In-vitro studies approve of the use of ivermectin as a repurposed drug against COVID-19 [32]. However, three more drugs that belong to cephalosporin-derivatives viz. cefiofur (PubChem CID: 6328657), cefazedone (PubChem CID: 71736) and ceftizoxime (PubChem CID: 6533629) were also ranked in the ensemble docking. All these drugs are known to possess anti-bacterial property and are used to treat severe bacterial infections. The top ranked drug obtained through direct docking, ceftazidime, is also a cephalosporin derivative (Figure 3A). Amprenavir (PubChem CID: 65016),
which is also a known HIV protease inhibitor docked with the best grid score to one of the 16 representative ensemble structures. However, the value of the grid score was comparatively higher than the other best hit drug molecules.

Figure 5B explains the top ranked drugs obtained on screening the SWEETLEAD drug database using the ensemble docking approach. Amongst, the 16 representative structures used for doing, three structures were observed to show the best grid score for the anti-bacterial drug Neomycin (PubChem CID: 8378). The drug with the best grid score was Neomycin. The drug with second best grid score was vasopressin tannate (PubChem CID: 8230) which is known to be used as an anti-diuretic drug to treat diabetes insipidus. The drug with third best grid score was again neomycin. However, amikacin which is also known for its antibacterial property was ranked as fourth best according to the grid score. In comparison to the direct docking approach, 10 of the 16 ensemble structures screened drug molecules possessing anti-bacterial activity viz. tobramycin.

![Figure 5: The top ranked drugs from the FDA approved (A) and SWEETLEAD (B) drug database obtained through ensemble docking approach](image-url)
(PubChem CID: 36294), lenapenem (PubChem CID: 216262), neomycin, gentamicin (PubChem CID: 3467), ribostamycin (PubChem CID: 33042), amikacin (PubChem CID: 37768) and netilmicin (PubChem CID: 441306). Neomycin and tobramycin appeared as the best ranked for three and two of the 16 representative structures respectively.

**Discussion**

**Drug-3CL\textsuperscript{pro} interactions: Direct Docking**

The active site of the main protease shows the presence of a few polar residues viz. histidine, asparagine, glutamate and glutamine. These residues were observed to be involved in hydrogen bonding and hydrophobic interactions with the drug molecules. The LigPlus and PLIP were used to calculated the various interactions between the drug molecule and the 3CL\textsuperscript{pro} \cite{51,52}. Figure S1 shows the residues of 3CL\textsuperscript{pro} present in the vicinity of the FDA approved ligand molecules and the residues that are involved in hydrogen bonding. The drug ceftazidime which showed the best dock score was observed to form hydrogen bonds with GLY 143 and HIS 164 (Figure 6). Ceftazidime was also involved in forming hydrophobic interactions with GLU 166 and π-π interactions with HIS 41 (Figure 6). Quetiapine formed hydrogen bonds with GLY 143 and THR 26. It showed hydrophobic interactions with ASP 187 and GLU 189 and π-π interactions with HIS 41. Cabergoline had no hydrogen bonding interactions however, showed hydrophobic interactions with PHE 140, GLU 166 and GLN 189. π-π interaction with HIS 41 was
observed for cabergoline too. Enoxolone was observed to form hydrogen bonds with GLU 166 and GLU 192 and hydrophobic interactions with THR 25, ASN 142, MET 165 and GLU 189. Apremilast was observed to form hydrogen bonding interactions with THR 26, GLY 143, ASN 142 and SER 144 and hydrophobic interactions with ASN 142 and MET 165. It was observed that HIS 41, GLY 143, ASN 142 and GLU 166 were involved interaction with the three of the drug molecules amongst the top five. However, π-π interactions with HIS 41 were observed in the top three molecules. The region around CYS 145 of the main protease is known to interact with human proteins viz. human deacteylase 2 (HDAC2) tRNA-methyl transferase 1 (TRMT1) [37]. Both these proteins are known epigenetic regulators and their nuclear localization is blocked by this viral protease [37]. The docking studies performed here revealed that the residues ASN142 and GLY143 which are neighboring to CYS145 were known to be involved in interacting with the drug molecules. This may suggest their crucial role in inhibiting the protein-protein interaction between the main protease and the human epigenetic regulatory proteins.

Figure S2 shows the residues of 3CL\textsuperscript{pro} present in the vicinity of the SWEETLEAD drug molecules and the residues that are involved in hydrogen bonding. The best ranked drug, dibekacin was observed to form five hydrogen bonds with THR 25, THR 26, SER 46, CYS 145, GLU 166 and GLN 189 (Figure 7). Micronomicin, which ranked second in terms of the grid score was observed to form hydrogen bonds with THR 24, SER 46 and ASN 142. The next drug in top five ranked drugs was catalposide which showed hydrogen bonding interactions with THR 24,
THR 26, GLY 143 and GLU 166. It also showed π- π interactions with HIS 41 and hydrophobic interactions with MET 165 and GLN 189. Dihydro-alpha ergocryptine, which ranked fourth in terms of the grid score was observed to form hydrogen bonds with ASN 142 and GLY 143. It was also involved in hydrophobic interactions with GLU 166. Itopride which ranked last amongst the top five drug molecules did not show any significant interactions with the residues of the protein.

**Drug-3CLpro interactions: Ensemble docking**

The residues interacting with the best ranked drugs from the FDA approved and SWEETLEAD database for all the 16 representative clusters have been shown in Figure S3 and S4 respectively. The top three ranked drugs from the FDA approved drug database included, Indinavir, CEFTIN and Ivermectin which showed interactions with THR 24, LEU 27, VAL 42, THR 45, ARG 60, LYS 61, ASN 142, GLU 166 and GLN 189 (Figure 8). Figure S3 depicts the hydrogen bonding interactions of all the best ranked drug molecules of the 16 representative structures. It was observed that apart from the residues mentioned above HIS 41, PHE 140, HIS163 and GLN 192 were also responsible for forming hydrogen bonding interactions with the other best ranked drug molecules. The top four ranked drugs from the SWEETLEAD drug database included Neomycin at the first and third position, Vasopressin tannate and Amikacin. Neomycin, was observed to be the best ranked drug in case of three ensemble structures. The conformational variability in the

![Figure 8: Hydrogen bonding interactions of Indinavir (A), CEFTIN (B) and Ivermectin (C) with the residues of 3CLpro](image-url)
ensemble structures was clearly visible on observing the number of interactions of the drug Neomycin in the three different 3CL\textsuperscript{pro} states that were captured (Figure 9). GLU 166 was involved in formation of strong hydrogen bonding with the atoms of the neomycin in all the three states. GLU 166 formed 3 (Figure 9A), 1 (Figure 9B) and 2 (Figure 9C) hydrogen bonds in the three representative structures of the ensemble. The other residues involved in hydrogen bonding were THR 24, SER 46, HIS 164, MET 165, PRO 168 and GLN 189. Vasopressin tannate, drug with second best grid score, formed hydrogen bonds with THR 24, THR 25, HIS 41 and SER 46 residues of the 3CL\textsuperscript{pro}. Amikacin, the drug obtained with fourth best grid score formed hydrogen bonds with CYS 145, ASN 142 and MET 165.

Experimental studies performed to elucidate the crystal structure of 3CLpro suggests that the inhibitor binding site of this protein is divided into sub sites \[38\]. The S1 sub-site consists of PHE 140, ASN 142, GLU 166, HIS 163 and HIS 172. Whereas, the S2 sub-site consists of the hydrophobic pocket made by the residues viz. HIS 41, MET 49, TYR 54 and MET 165. The CYS 145 is involved in covalent bond with the inhibitor N3 (Figure 1) \[35,38\]. It was observed that most of these residues which play a crucial role in interacting with the inhibitor showed similar
results for the drugs obtained through direct and ensemble docking. However, obtaining the same
drug as the best docked in case of different ensemble representative states which varying
interactions suggests conformational variability in the inhibitor binding site of the 3CL\textsuperscript{pro}.

**Conclusion**

The high throughput docking and ensemble docking studies of 3C-like protease reveal few
potential drugs that can be considered for repurposing. The docking against the FDA approved and
SWEETLEAD drug database helped to enlist few antibacterial and antiviral drugs that may be
used as candidates for repurposing studies against 3C-like protease. Indinavir, ivermectin,
cephalosporin-derivatives, neomycin and amprenavir were few of the drugs which may prove to
be effective against the symptoms seen in COVID-19. As, the earlier purpose of indinavir and
amprenavir states inhibition of the HIV protease. Similarly, ivermectin, cephalosporin-derivatives
and neomycin are used against treating anti-parasitic and anti-bacterial infections especially the
respiratory tract infections. In support to these findings, these drugs were also observed to show
better docking scores in comparison to other drugs. The ensemble docking approach helped to
explore the conformational variability of the inhibitor binding site of 3C-like protease. The
conformations captured through effective sampling methods like Markov State Modelling analysis
reveal a more accessible region for inhibitors to bind to the 3C-like protease. The docking scores
for the ligands when docked against these structures were observed to be better in comparison to
the other. The drug residue interactions also complement the role of crucial residues that were
earlier defined by the electron-density studies of 3C-like protease and its inhibitors [36]. The
ensemble docking approach coupled with a strong sampling technique would help to explore the
more accessible conformations of the drug target which would further help in designing a better
drug as an inhibitor.
Acknowledgements

The authors would like to acknowledge the National Supercomputing Mission (NSM), Ministry of electronics and information technology (MeitY), Government of India for funding this work. The authors would like to acknowledge the PARAM supercomputing facility and the Bioinformatics Resources and Applications Facility (BRAF) at Centre for Development of Advanced Computing (C-DAC), Pune for providing the computing infrastructure. The authors would like to acknowledge S R Rajesh Kumar, Saurabh Patil and Akash Khade for their timing administrative support and services.
Table 1: Best five ranked drugs from the FDA approved database obtained through direct docking of 3CLpro of SARS-nCoV2

| PubChem CID | Name of the drug | Earlier Purpose | Structure | Grid Score (kcal/mole) |
|-------------|------------------|----------------|-----------|-----------------------|
| 5481173     | Ceftazidime      | Antibacterial, used in pneumonia | ![Structure](image1) | -46.05 |
| 5002        | Quetiapine       | Bipolar disorder, used as a treatment against Schizophrenia | ![Structure](image2) | -42.26 |
| 54746       | Cabergoline      | Dopamine agonists, used in Parkinson’s disease | ![Structure](image3) | -41.81 |
| 1011        | Enoxolone        | Consists of Glycyrrhetinic acid, a plant derivative, used as anti-allergitic, anti-bacterial and anti-viral | ![Structure](image4) | -40.94 |
| 11561674    | Apremilast       | Psoriasis       | ![Structure](image5) | -40.81 |
Table 2: Best five ranked drugs from the SWEETLEAD database obtained through direct docking of 3CLpro of SARS-nCoV2

| PubChem CID | Name of the drug          | Earlier purpose                                      | Structure                           | Grid Score (kcal/mole) |
|-------------|---------------------------|------------------------------------------------------|-------------------------------------|------------------------|
| 470999      | Dibekacin                 | Aminoglycoside, Antibiotic                          | ![Structure](image1)               | -54.018                |
| 3037206     | Micronomicin              | Aminoglycoside, Antibiotic                          | ![Structure](image2)               | -45.963                |
| 93039       | Catalposide               | Plant derivative, anti-inflammatory effect           | ![Structure](image3)               | -45.831                |
| 114948      | Dihydro-alpha-ergocryptine| Parkinson’s diseases                                 | ![Structure](image4)               | -45.045                |
| 3792        | Itopride                  | Functional dyspepsia, gastrointestinal drug, Blocks dopamine receptor | ![Structure](image5)               | -44.98                 |
Table 3: Best ranked drugs from the FDA approved drug database obtained through ensemble docking of 3CLpro of SARS-nCoV2

| 3CL<sub>pro</sub> Ensemble Representative | Name of the drug (PubChem CID) | Earlier Purpose | Structure | Grid Score (kcal/mole) |
|-----------------------------------------|--------------------------------|----------------|-----------|------------------------|
| RMSD-Based clustering                    |                                |                |           |                        |
| 1                                       | Hespiridin (3594)              | Bioflavonoid, anti-oxidant, anti-inflammatory | ![Structure](image1.png) | -41.402 |
| 2                                       | Etoposide (36462)              | Chemotherapy drug, used in lung cancer too | ![Structure](image2.png) | -43.659 |
| 3                                       | Pranlukast (4887)              | Anti-asthamatic, reduces bronchospasm | ![Structure](image3.png) | -42.833 |
| 4                                       | Azelnidipine (65948)           | Treats hypertension, calcium channel blocker | ![Structure](image4.png) | -43.362 |
|   | Chemical Name          | Description                              | Molecular Structure | Value   |
|---|------------------------|------------------------------------------|---------------------|---------|
| 5 | Epicatechin gallate    | Flavonoid, treated for pre-diabetes      | ![Molecule Image](image) | -31.073 |
| 6 | Brinzolamide           | Ocular Hypertension                      | ![Molecule Image](image) | -34.4   |
| 7 | Ceftiofur              | Anti-bacterial, veterinary drug           | ![Molecule Image](image) | -37.726 |
| 8 | Artesunate             | Treats malaria, combination therapy      | ![Molecule Image](image) | -37.925 |
| 9 | Ivermectin             | Treats parasitic infections              | ![Molecule Image](image) | -44.703 |
| 10| Peimine                | Anti-inflammatory                         | ![Molecule Image](image) | -38.504 |
| 11| Empagliflozin          | Treats Type2-Diabetes                    | ![Molecule Image](image) | -38.504 |
|   | Name                    | Category          | Description                                                                 | Score  |
|---|-------------------------|-------------------|-----------------------------------------------------------------------------|--------|
| 12| Agenerase/Amprenavir   | Antiviral         | inhibits HIV protease                                                       | -37.565|
|   | (65016)                 |                   |                                                                             |        |
| 13 | Cefazedone              | Antibacterial     |                                                                             | -43.024|
|   | (71736)                 |                   |                                                                             |        |
| 14 | Indinavir               | Antiviral         | inhibits HIV protease                                                       | -53.089|
|   | (5362440)               |                   |                                                                             |        |
| 15 | Ceftin                  | Antibacterial     | used against pneumonia                                                       | -49.33 |
|   | (6321416)               |                   |                                                                             |        |
| 16 | Ceftizoxime             | Antibiotic        | used against life threatening bacterial infections                           | -42.027|
|   | (6533629)               |                   |                                                                             |        |
Table 4: Best ranked drugs from the SWEETLEAD drug database obtained through ensemble docking of 3CLpro of SARS-nCoV2

| 3CLpro Ensemble Representative | Name of the drug (PubChem CID) | Earlier Purpose | Structure | Grid Score (kcal/mole) |
|--------------------------------|--------------------------------|-----------------|-----------|-----------------------|
| RMSD-based clustering           | Tobramycin (36294)             | Antibiotic, antibacterial activity | ![Tobramycin Structure](image) | -48.33 |
|                                 | Lanreotide acetate (71349)     | Used to treat Acromegaly, inhibits the growth hormone | ![Lanreotide Acetate Structure](image) | -52.94 |
|                                 | Lenapenem (216262)             | carbapenem antibiotic with bactericidal activity, penicillin binding protein | ![Lenapenem Structure](image) | -41.91 |
|   | Drug Name               | Description                                                                 | Code  |
|---|------------------------|-----------------------------------------------------------------------------|-------|
| 4 | Neomycin (8378)        | Aminoglycoside antibiotic                                                  | -53.316 |
| 5 | Riboflavin tetrabutyrate (92140) | One component of the multi-vitamin drugs                                   | -35.101 |
| 6 | Sennosides (5199)      | Stimulant laxative                                                          | -42.306 |
| 7 | Gentamicin (3467)      | Antibacterial, used against pneumonia                                       | -45.604 |
| 8 | Terlipressin (72081)   | Vasoactive drug, used to manage low blood pressure                         | -46.807 |
| 9 | Ribostamycin (33042)   | Aminoglycoside-aminocyclitol antibiotic                                     | -52.346 |
|   | Name                  | Description                                                                 | Value    |
|---|-----------------------|-----------------------------------------------------------------------------|----------|
|10 | Tobramycin (36294)    | Antibiotic, antibacterial activity                                          | -47.518  |
|11 | Neomycin (8378)       | Aminoglycoside antibiotic                                                   | -59.469  |
|12 | Neomycin (8378)       | Aminoglycoside antibiotic                                                   | -65.975  |
|13 | Lypressin (644076)    | Used against diabetes insipidus                                             | -46.047  |
|14 | Amikacin (37768)      | Antibiotic, Multi-drug resistant tuberculosis                               | -58.178  |
|15 | Vasopressin tannate (8230) | Antidiuretic drug, used against diabetes insipidus                          | -63.335  |
|   | Netilmicin (441306) | Antibiotic, treatment against severe bacterial infections | -57.513 |
References:

1. Data obtained from https://covid19.who.int/ (World Health Organization)

2. Wang H, Li X, Li T, Zhang S, Wang L, Wu X, Liu J. The genetic sequence, origin, and diagnosis of SARS-CoV-2. Eur J Clin Microbiol Infect Dis. 2020 Apr 24. doi: 10.1007/s10096-020-03899-4. [Epub ahead of print] Review. PubMed PMID:32333222.

3. Andersen KG, Rambaut A, Lipkin WI, Holmes EC, Garry RF. The proximal origin of SARS-CoV-2. Nat Med. 2020 Apr;26(4):450-452. doi: 10.1038/s41591-020-0820-9. PubMed PMID: 32284615.

4. Kim D, Lee JY, Yang JS, Kim JW, Kim VN, Chang H. The Architecture of SARS-CoV-2 Transcriptome. Cell. 2020 Apr 18. pii: S0092-8674(20)30406-2. doi: 10.1016/j.cell.2020.04.011. [Epub ahead of print] PubMed PMID: 32330414; PubMed Central PMCID: PMC7179501.

5. Li G, De Clercq E. Therapeutic options for the 2019 novel coronavirus (2019-nCoV). Nat Rev Drug Discov. 2020 Mar;19(3):149-150. doi: 10.1038/d41573-020-00016-0. PubMed PMID: 32127666.

6. Prajapat M, Sarma P, Shekhar N, Avti P, Sinha S, Kaur H, Kumar S, Bhattacharyya A, Kumar H, Bansal S, Medhi B. Drug targets for corona virus: A systematic review. Indian J Pharmacol. 2020 Jan-Feb;52(1):56-65. doi: 10.4103/ijp.IJP_115_20. Epub 2020 Mar 11. PubMed PMID: 32201449; PubMed Central PMCID: PMC7074424.

7. Nadeem MS, Zamzami MA, Choudhry H, Murtaza BN, Kazmi I, Ahmad H, Shakoori AR. Origin, Potential Therapeutic Targets and Treatment for Coronavirus Disease
8. Tu YF, Chien CS, Yarmishyn AA, Lin YY, Luo YH, Lin YT, Lai WY, Yang DM, Chou SJ, Yang YP, Wang ML, Chiou SH. A Review of SARS-CoV-2 and the Ongoing Clinical Trials. Int J Mol Sci. 2020 Apr 10;21(7). pii: E2657. doi: 10.3390/ijms21072657. Review. PubMed PMID: 32290293.

9. Cherian SS, Agrawa M, Basu A, Abraham P, Gangakhedkar RR, Bhargava B. Perspectives for repurposing drugs for the coronavirus disease 2019. Indian J Med Res. 2020 Apr 10. doi: 10.4103/ijmr.IJMR_585_20. [Epub ahead of print] Review. PubMed PMID: 32317408.

10. Asai A, Konno M, Ozaki M, Otsuka C, Vecchione A, Arai T, Kitagawa T, Ofusa K, Yabumoto M, Hirotsu T, Taniguchi M, Eguchi H, Doki Y, Ishii H. COVID-19 Drug Discovery Using Intensive Approaches. Int J Mol Sci. 2020 Apr 18;21(8). pii: E2839. doi: 10.3390/ijms21082839. Review. PubMed PMID: 32325767.

11. Pawar AY. Combating Devastating COVID-19 by Drug Repurposing. Int J Antimicrob Agents. 2020 Apr 17:105984. doi: 10.1016/j.ijantimicag.2020.105984. [Epub ahead of print] PubMed PMID: 32305589; PubMed Central PMCID: PMC7162749.

12. Wang C, Liu Z, Chen Z, Huang X, Xu M, He T, Zhang Z. The establishment of reference sequence for SARS-CoV-2 and variation analysis. J Med Virol. 2020 Mar 13. doi: 10.1002/jmv.25762. [Epub ahead of print] PubMed PMID: 32167180.

13. Yadav PD, Potdar VA, Choudhary ML, Nyayanit DA, Agrawal M, Jadhav SM, Majumdar TD, Shete-Aich A, Basu A, Abraham P, Cherian SS. Full-genome sequences
of the first two SARS-CoV-2 viruses from India. Indian J Med Res. 2020 Mar 27. doi: 10.4103/ijmr.IJMR_663_20. [Epub ahead of print] PubMed PMID: 32242873.

14. Bauer DC, Tay AP, Wilson LOW, Reti D, Hosking C, McAuley AJ, Pharo E, Todd S, Stevens V, Neave MJ, Tachedjian M, Drew TW, Vasan SS. Supporting pandemic response using genomics and bioinformatics: a case study on the emergent SARS-CoV-2 outbreak. Transbound Emerg Dis. 2020 Apr 19. doi: 10.1111/tbed.13588. [Epub ahead of print] PubMed PMID: 32306500.

15. Forster P, Forster L, Renfrew C, Forster M. Phylogenetic network analysis of SARS-CoV-2 genomes. Proc Natl Acad Sci U S A. 2020 Apr 8. pii: 202004999. doi: 10.1073/pnas.2004999117. [Epub ahead of print] PubMed PMID: 32269081.

16. Stefanelli P, Faggioni G, Lo Presti A, Fiore S, Marchi A, Benedetti E, Fabiani C, Anselmo A, Ciammaruconi A, Fortunato A, De Santis R, Fillo S, Capobianchi MR, Gismondo MR, Ciervo A, Rezza G, Castrucci MR, Lista F, On Behalf Of Iss Covid-Study Group. Whole genome and phylogenetic analysis of two SARS-CoV-2 strains isolated in Italy in January and February 2020: additional clues on multiple introductions and further circulation in Europe. Euro Surveill. 2020 Apr;25(13). doi: 10.2807/1560-7917.ES.2020.25.13.2000305. PubMed PMID: 32265007; PubMed Central PMCID: PMC7140597.

17. Li C, Yang Y, Ren L. Genetic evolution analysis of 2019 novel coronavirus and coronavirus from other species. Infect Genet Evol. 2020 Mar 10;82:104285. doi:10.1016/j.meegid.2020.104285. [Epub ahead of print] PubMed PMID: 32169673.

18. Wu F, Zhao S, Yu B, Chen YM, Wang W, Song ZG, Hu Y, Tao ZW, Tian JH, Pei YY, Yuan ML, Zhang YL, Dai FH, Liu Y, Wang QM, Zheng JJ, Xu L, Holmes EC, Zhang
YZ. Author Correction: A new coronavirus associated with human respiratory disease in China. Nature. 2020 Apr;580(7803):E7. doi: 10.1038/s41586-020-2202-3. PubMed PMID: 32296181.

19. Serafin MB, Bottega A, Foletto VS, da Rosa TF, Hörner A, Hörner R. Drug repositioning is an alternative for the treatment of coronavirus COVID-19. Int J Antimicrob Agents. 2020 Apr 9;105969. doi: 10.1016/j.ijantimicag.2020.105969. [Epub ahead of print] PubMed PMID: 32278811.

20. Shah B, Modi P, Sagar SR. In silico studies on therapeutic agents for COVID-19: Drug repurposing approach. Life Sci. 2020 Apr 9;252:117652. doi: 10.1016/j.lfs.2020.117652. [Epub ahead of print] PubMed PMID: 32278693.

21. Muralidharan N, Sakthivel R, Velmurugan D, Gromiha MM. Computational studies of drug repurposing and synergism of lopinavir, oseltamivir and ritonavir binding with SARS-CoV-2 protease against COVID-19. J Biomol Struct Dyn. 2020 Apr 16;1-6. doi: 10.1080/07391102.2020.1752802. [Epub ahead of print] PubMed PMID: 32248766.

22. Rosa SGV, Santos WC. Clinical trials on drug repositioning for COVID-19 treatment. Rev Panam Salud Publica. 2020 Mar 20;44:e40. doi: 10.26633/RPSP.2020.40. eCollection 2020. PubMed PMID: 32256547; PubMed Central PMCID: PMC7105280.

23. Liu S, Zheng Q, Wang Z. Potential covalent drugs targeting the main protease of the SARS-CoV-2 coronavirus. Bioinformatics. 2020 Apr 1. pii: btaa224. doi: 10.1093/bioinformatics/btaa224. [Epub ahead of print] PubMed PMID: 32239142.

24. Phadke M, Saunik S. COVID-19 treatment by repurposing drugs until the vaccine is in sight. Drug Dev Res. 2020 Mar 29. doi: 10.1002/ddr.21666. [Epub ahead of print] PubMed PMID: 32227357.
25. Elfiky AA. Ribavirin, Remdesivir, Sofosbuvir, Galidesivir, and Tenofovir against SARS-CoV-2 RNA dependent RNA polymerase (RdRp): A molecular docking study. Life Sci. 2020 Mar 25;253:117592. doi: 10.1016/j.lfs.2020.117592. [Epub ahead of print] PubMed PMID: 32222463; PubMed Central PMCID: PMC7102646.

26. Chen YW, Yiu CB, Wong KY. Prediction of the SARS-CoV-2 (2019-nCoV) 3C-like protease (3CLpro) structure: virtual screening reveals velpatasvir, ledipasvir, and other drug repurposing candidates. F1000Res. 2020 Feb 21;9:129. doi: 10.12688/f1000research.22457.2. PMID: 32194944; PMCID: PMC7062204.

27. Baron SA, Devaux C, Colson P, Raoult D, Rolain JM. Teicoplanin: an alternative drug for the treatment of COVID-19? Int J Antimicrob Agents. 2020 Apr;55(4):105944. doi: 10.1016/j.ijantimicag.2020.105944. Epub 2020 Mar 13. PubMed PMID: 32179150; PubMed Central PMCID: PMC7102624.

28. Fan HH, Wang LQ, Liu WL, An XP, Liu ZD, He XQ, Song LH, Tong YG. Repurposing of clinically approved drugs for treatment of coronavirus disease 2019 in a 2019-novel coronavirus (2019-nCoV) related coronavirus model. Chin Med J (Engl). 2020 Mar 6. doi: 10.1097/CM9.0000000000000797. [Epub ahead of print] PubMed PMID: 32149769; PubMed Central PMCID: PMC7147283.

29. Elfiky AA. Anti-HCV, nucleotide inhibitors, repurposing against COVID-19. Life Sci. 2020 May 1;248:117477. doi: 10.1016/j.lfs.2020.117477. Epub 2020 Feb 28. PubMed PMID: 32119961; PubMed Central PMCID: PMC7089605.

30. Kruse RL. Therapeutic strategies in an outbreak scenario to treat the novel coronavirus originating in Wuhan, China. F1000Res. 2020 Jan 31;9:72. doi:
31. Gao J, Tian Z, Yang X. Breakthrough: Chloroquine phosphate has shown apparent efficacy in treatment of COVID-19 associated pneumonia in clinical studies. Biosci Trends. 2020 Mar 16;14(1):72-73. doi: 10.5582/bst.2020.01047. Epub 2020 Feb 19. PubMed PMID: 32074550.

32. Caly L, Druce JD, Catton MG, Jans DA, Wagstaff KM. The FDA-approved drug ivermectin inhibits the replication of SARS-CoV-2 in vitro. Antiviral Res. 2020 Apr 3;178:104787. doi: 10.1016/j.antiviral.2020.104787. [Epub ahead of print] PubMed PMID: 32251768; PubMed Central PMCID: PMC7129059.

33. Smith, Micholas; Smith, Jeremy C. (2020): Repurposing Therapeutics for COVID-19: Supercomputer-Based Docking to the SARS-CoV-2 Viral Spike Protein and Viral Spike Protein-Human ACE2 Interface. ChemRxiv. Preprint. https://doi.org/10.26434/chemrxiv.11871402.v4

34. Khan SA, Zia K, Ashraf S, Uddin R, Ul-Haq Z. Identification of chymotrypsin-like protease inhibitors of SARS-CoV-2 via integrated computational approach. J Biomol Struct Dyn. 2020 Apr 13:1-10. doi: 10.1080/07391102.2020.1751298. [Epub ahead of print] PubMed PMID: 32238094.

35. Jin Z, Du X, Xu Y, Deng Y, Liu M, Zhao Y, Zhang B, Li X, Zhang L, Peng C, Duan Y, Yu J, Wang L, Yang K, Liu F, Jiang R, Yang X, You T, Liu X, Yang X, Bai F, Liu H, Liu X, Guddat LW, Xu W, Xiao G, Qin C, Shi Z, Jiang H, Rao Z, Yang H. Structure of M(pro) from COVID-19 virus and discovery of its inhibitors. Nature. 2020 Apr 9. doi: 10.1038/s41586-020-2223-y. [Epub ahead of print] PubMed PMID:32272481.
36. Fearon, D., Owen, C.D., Douangamath, A., Lukacik, P., Powell, A.J., Strain-Damerell, C.M., Resnick, E., Krojer, T., Gehrtz, P., Wild, C., Aimon, A., Brandao-Neto, J., Carbery, A., Dunnett, L., Skyner, R., Snee, M., London, N., Walsh, M.A., von Delft, F. PanDDA analysis group deposition of SARS-CoV-2 main protease fragment screen. RCSB-PDB, To be published

37. A SARS-CoV-2-Human Protein-Protein Interaction Map Reveals Drug Targets and Potential Drug-Repurposing. David E. Gordon, Gwendolyn M. Jang, Mehdi Bouhaddou, Jiewei Xu, Kirsten Obernier, Matthew J. O’Meara, Jeffrey Z. Guo, Danielle L. Swaney, Tia A. Tummino, Ruth Huettenhain, et. al bioRxiv 2020.03.22.002386; doi: https://doi.org/10.1101/2020.03.22.002386

38. 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 Apr 24;368(6489):409-412. doi: 10.1126/science.abb3405. Epub 2020 Mar 20. PubMed PMID: 32198291; PubMed Central PMCID: PMC7164518.

39. Kandeel M, Al-Nazawi M. Virtual screening and repurposing of FDA approved drugs against COVID-19 main protease. Life Sci. 2020 Jun 15;251:117627. doi: 10.1016/j.lfs.2020.117627. Epub 2020 Apr 3. PubMed PMID: 32251634.

40. Ortega JT, Serrano ML, Pujol FH, Rangel HR. Unrevealing sequence and structural features of novel coronavirus using in silico approaches: The main protease as molecular target. EXCLI J. 2020 Mar 17;19:400-409. doi: 10.17179/excli2020-1189. eCollection 2020. PubMed PMID: 32210741; PubMed Central PMCID: PMC7081067
41. Komatsu, Teruhisa S.; Koyama, Yohei M.; Okimoto, Noriaki; Morimoto, Gentaro; Ohno, Yousuke; Taiji, Makoto (2020), COVID-19 related trajectory data of 10 microseconds all atom molecular dynamics simulation of SARS-CoV-2 dimeric main protease, Mendeley Data, v1 http://dx.doi.org/10.17632/vpps4vhryg.1

42. Hall DC Jr, Ji HF. A search for medications to treat COVID-19 via in silico molecular docking models of the SARS-CoV-2 spike glycoprotein and 3CL protease. Travel Med Infect Dis. 2020 Apr 12:101646. doi: 10.1016/j.tmaid.2020.101646. [Epub ahead of print] PubMed PMID: 32294562; PubMed Central PMCID: PMC7152904.

43. Center for Drug Evaluation and Research (U.S.). (2004). Drugs@FDA. Washington D.C: U.S. Food and Drug Administration, Center for Drug and Evaluation Research.

44. Novick PA, Ortiz OF, Poelman J, Abdulhay AY, Pande VS. SWEETLEAD: an in silico database of approved drugs, regulated chemicals, and herbal isolates for computer-aided drug discovery. PLoS One. 2013 Nov 1;8(11):e79568. doi: 10.1371/journal.pone.0079568. eCollection 2013. PubMed PMID: 24223973; PubMed Central PMCID: PMC3815155.

45. Allen WJ, Balius TE, Mukherjee S, Brozell SR, Moustakas DT, Lang PT, Case DA, Kuntz ID, Rizzo RC. DOCK 6: Impact of new features and current docking performance. J Comput Chem. 2015 Jun 5;36(15):1132-56. doi: 10.1002/jcc.23905. PubMed PMID: 25914306; PubMed Central PMCID: PMC4469538.

46. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE. UCSF Chimera--a visualization system for exploratory research and analysis. J Comput Chem. 2004 Oct;25(13):1605-12. PubMed PMID: 15264254.
47. D.A. Case, R.M. Betz, D.S. Cerutti, T.E. Cheatham, III, T.A. Darden, R.E. Duke, T.J.
   Giese, H. Gohlke, A.W. Goetz, N. Homeyer, S. Izadi, P. Janowski, J. Kaus, A.
   Kovalenko, T.S. Lee, S. LeGrand, P. Li, C. Lin, T. Luchko, R. Luo, B. Madej, D.
   Mermelstein, K.M. Merz, G. Monard, H. Nguyen, H.T. Nguyen, I. Omelyan, A.
   Onufriev, D.R. Roe, A. Roitberg, C. Sagui, C.L. Simmerling, W.M. Botello-Smith, J.
   Swails, R.C. Walker, J. Wang, R.M. Wolf, X. Wu, L. Xiao and P.A. Kollman (2016),
   AMBER 2016, University of California, San Francisco.

48. Ester, M.; Kriegel, H.; Sander, J.; Xu, X. A Density-Based Algorithm for Discovering
   Clusters in Large Spatial Databases with Noise. Proceedings of the Second
   International Conference on Knowledge Discovery and Data Mining (KDD-96), 1996,
   pp 226–231.

49. Scherer MK, Trendelkamp-Schroer B, Paul F, Pérez-Hernández G, Hoffmann M,
   Plattner N, Wehmeyer C, Prinz JH, Noé F. PyEMMA 2: A Software Package for
   Estimation, Validation, and Analysis of Markov Models. J Chem Theory Comput. 2015
   Nov 10;11(11):5525-42. doi: 10.1021/acs.jctc.5b00743. Epub 2015 Oct 14. PubMed
   PMID: 26574340.

50. Jani, Vinod & Sonavane, Uddhavesh & Joshi, Rajendra. (2019). Detecting early stage
    structural changes in wild type, pathogenic and non-pathogenic prion variants using
    Markov state model. RSC Advances. 9. 14567-14579. 10.1039/C9RA01507H.

51. Prinz JH, Wu H, Sarich M, Keller B, Senne M, Held M, Chodera JD, Schütte C, Noé
    F. Markov models of molecular kinetics: generation and validation. J Chem Phys. 2011
    May 7;134(17):174105. doi: 10.1063/1.3565032. PubMed PMID: 21548671.
52. Chodera JD, Noé F. Markov state models of biomolecular conformational dynamics. Curr Opin Struct Biol. 2014 Apr;25:135-44. doi: 10.1016/j.sbi.2014.04.002. Epub 2014 May 16. Review. PubMed PMID: 24836551; PubMed Central PMCID: PMC4124001.

53. Chodera JD, Singhal N, Pande VS, Dill KA, Swope WC. Automatic discovery of metastable states for the construction of Markov models of macromolecular conformational dynamics. J Chem Phys. 2007 Apr 21;126(15):155101. PubMed PMID: 17461665.

54. Sirur A, De Sancho D, Best RB. Markov state models of protein misfolding. J Chem Phys. 2016 Feb 21;144(7):075101. doi: 10.1063/1.4941579. PubMed PMID: 26897000; PubMed Central PMCID: PMC4760979.

55. Kuzmanic A, Bowman GR, Juarez-Jimenez J, Michel J, Gervasio FL. Investigating Cryptic Binding Sites by Molecular Dynamics Simulations. Acc Chem Res. 2020 Mar 17;53(3):654-661. doi: 10.1021/acs.accounts.9b00613. Epub 2020 Mar 5. PubMed PMID: 32134250.

56. Laskowski RA, Swindells MB. LigPlot+: multiple ligand-protein interaction diagrams for drug discovery. J Chem Inf Model. 2011 Oct 24;51(10):2778-86. doi: 10.1021/ci200227u. Epub 2011 Oct 5. PubMed PMID: 21919503.

57. Salentin S, Schreiber S, Haupt VJ, Adasme MF, Schroeder M. PLIP: fully automated protein-ligand interaction profiler. Nucleic Acids Res. 2015 Jul 1;43(W1):W443-7. doi: 10.1093/nar/gkv315. Epub 2015 Apr 14. PMID: 25873628; PMCID: PMC4489249.
Supplementary Data: Drug repurposing studies targeting SARS-nCoV2: An ensemble docking approach on drug target 3C-like protease (3CLpro)

(Supplementary Information and Supplementary Figures)

Supplementary Information SI1:

Markov State Modelling Analysis

Markov State Modelling Analysis helps in exploring the different states visited by the protein in a molecular dynamics by significant sampling the simulation data along the slowest changing variable. These states would depend on the collective variable (CV) that is selected to sample the simulation data. These collective variables in case of protein simulations can consists of one or many conformational parameters. In the present study, the backbone torsion angles were selected as the collective variable. Apart from collective variable (CV), the choice of lag time plays an important role in MSM analysis [50-54]. Lag time states the time duration after the minimum time after which a variation in the data may be seen. It order to achieve the sampling, the dimensionality reduction is achieved by performing the time independent component analysis (tICA). Figure SI.1 shows the behavior of tICA 1 (upper panel) and tICA 2 (lower panel) for the complete 10.15 μs simulation data of 3CLpro. The plots show that two transitions were observed one captured by tICA 1 and the other captured by tICA 2.

After dimensionality reduction the data was clustered using the K-means clustering where the value for k was set to 250. These clusters obtained are termed as the microstates. This step was followed by the choosing the lag time based on the implied time scale. Figure SI.3 depicts the implied time scale and it can be inferred that the lag time of 100 would be selected for obtained the Markov states. The Chapman-Kolmogorov test was performed to check the markovian behavior and validate the Markov state models that would be generated. The results obtained through the Chapman-Kolmogorov test suggested that the predicted Markov state models where in significant agreement with the MSM estimation with a confidence score of 95 %. Figure SI.4 shows the comparison of probability of the

Figure SI.1: Behavior of independent component 1 (upper panel) and independent component 2 (lower panel)

Figure SI.2: Microstates obtained through k-means clustering (k=250)
estimated and predicted values of each of the Markov states. The microstates were clustered together using the spectral clustering method PCCA+ to obtain the macrostates. Figure SI.5 shows the four macrostates obtained, all the conformations of 3CLpro obtained in each of these clusters were extracted. The average structure of each of the cluster has been shown in the Figure SI.5. These average structures were further used for ensemble docking.

Figure SI.3: Implied time scales for choosing the lag time

Figure SI.4: Chapman-Kolmogorov test comparing the predicted and estimated probabilities for the Markov states obtained

Figure SI.5: The four states obtained through MSM analysis and further considered for ensemble docking
Supplementary Figures:

**Figure S1:** Hydrogen bonding and residues of $3\text{CL}^{\text{pro}}$ in the vicinity of the top five drugs viz. ceftazidime (A), quetiapine (B), cabergoline (C), enoxolone (D) and apremilast (E) obtained from direct docking of FDA approved drug database.

![Figure S1](image1)

**Figure S2:** Hydrogen bonding and residues of $3\text{CL}^{\text{pro}}$ in the vicinity of the top five drugs viz. dibekacin (A), micronomicin (B), catalposide (C), dihydro-alpha ergocryptine (D) and itopride (E) obtained from direct docking of SWEETLEADS drug database.

![Figure S2](image2)
Figure S3: Hydrogen bonding and residues of 3CL\textsuperscript{pro} in the vicinity of the best ranked drugs obtained from ensemble docking of FDA approved drug database on the 16 representative structures
Figure S4: Hydrogen bonding and residues of $3\text{CL}^{\text{pro}}$ in the vicinity of the best ranked drugs obtained from ensemble docking of SWEETLEADS drug database on the 16 representative structures.
