Drug repurposing against SARS-CoV-2 using E-pharmacophore based virtual screening and molecular docking with main protease as the target.

Arun KG¹, Sharanya CS¹, Abhithaj J¹, Dileep Francis², Sadasivan C*¹

¹Department of Biotechnology & Microbiology, Kannur University, Kerala, India.
²Department of Life sciences, Kristu Jayanti College, Bengaluru, Karnataka, India.

*Corresponding Author; e-mail: csadasivan@gmail.com
Abstract:
Since its first report in December 2019 from China the COVID-19 pandemic caused by the beta-coronavirus SARS-CoV-2 has spread at an alarming pace infecting about 26 lakh, and claiming the lives of more than 1.8 lakh individuals across the globe. Although social quarantine measures have succeeded in containing the spread of the virus to some extent, the lack of a clinically approved vaccine or drug remains the biggest bottleneck in combating the pandemic. Drug repurposing can expedite the process of drug development by identifying known drugs which are effective against SARS-CoV-2. The SARS-CoV-2 main protease is a promising drug target due to its indispensable role in viral multiplication inside the host. In the present study an E-pharmacophore hypothesis was generated using the crystal structure of the viral protease in complex with an imidazole carbamimide inhibitor as the drug target. Drugs available in the superDRUG2 database were used to identify candidate drugs for repurposing. The hits were further screened using a structure based approach involving molecular docking at different precisions. The most promising drugs were subjected to binding free energy estimation using MM-GBSA. Among the 4600 drugs screened 17 drugs were identified as candidate inhibitors of the viral protease based on the glide scores obtained from molecular docking. Binding free energy calculation showed that six drugs viz, Binifibrate, Macimorelin acetate, Bamifylline, Rilmazafon, Afatinib and Ezetimibe can act as potential inhibitors of the viral protease.

Key words: SARS-CoV-2, COVID-19, Main protease, Drug repurposing, E-pharmacophore, Molecular docking
1. Introduction

COVID-19, a severe viral pneumonia was first reported on December 31, 2019 in the city of Wuhan in the Hubei province of China by the Chinese Centre for Disease Control (CDC, China). The causative virus was shortly identified as a novel betacoronavirus, dubbed SARS-CoV-2. The virus belongs to the order Nidovirales of the Coronaviridae family comprising of the alpha- and beta-coronaviruses. These are enveloped, positive sense RNA viruses with comparatively large genomes among known RNA viruses (26.4–31.7 kb) [1,2]. Six members of the family are previously known to infect humans including SARS-CoV and MERS-CoV, which are known to cause severe respiratory ailments in the host [2, 3]. SARS-CoV-2 is the latest addition to the group and has presented itself as a potent human respiratory pathogen due to a mutation in the Receptor Binding Domain (RBD) of its spike protein that enables high affinity binding to the ACE2 receptor in humans and a polybasic furin cleavage site at the junction of the S1 and S2 subunits of the spike protein [4]. Since the first report on the virus, it has spread across continents inflicting a global health-care and economic emergency. In view of the global spread of the outbreak the World Health Organization (WHO) declared it as a pandemic in January 2020. On the date of writing this paper 2,672,260 infections and 186,933 deaths have been reported across 210 countries of the world. The number of infections and the death toll is increasing relentlessly despite concerted efforts to contain the spread of the virus using rigorous diagnostic testing, isolation of positive cases and tracing of contacts. The scenario is further made grim by the fact that at present there are no specific drugs or vaccines against the virus. The current treatments focus on symptom management and supportive therapy [5]. Government agencies, pharmaceutical companies and research institutes across the globe has taken up the formidable challenge of inventing a specific, viable and validated therapeutic agent against SARS-CoV-2 as it is probably the only solution to the ongoing crisis.

Drug repurposing refers to the identification of novel applications/targets for an approved or investigational drug outside the premise of its medical indication [6]. At present, this strategy would be a logical choice for developing a therapy for COVID-19 considering the substantial time-scales and attrition rates associated with new drug discovery and the trial-based validation of its safety and efficacy. The major advantage lies in the fact that the repurposed drug has been already evaluated for safety in animal and human trials, which would save significant amounts of time and money [7], a priority concern in SARS-CoV-2 drug discovery. Indeed, most of the drugs currently under investigation for efficacy against SARS-CoV-2 are
repurposed, known medicines. Drugs that are either under development or prescribed off-label against COVID-19 include Ribavirin, interferon (IFN) -α, mycophenolic acid ritonavir, lopinavir, oseltamivir, remdesivir, and chloroquine [5, 8-10]. Among these hydroxychloroquine, an approved anti-malarial drug, and two known antivirals, ritonavir and remdesivir, have been reported to be effective against SARS-CoV-2 in vitro [11].

The aim of the present study is to identify clinically approved drugs that can be targeted to the main protease, M\(^{\text{pro}}\) which also called 3CL\(^{\text{pro}}\) because it has similar structural folds and cleavage site specificity as that of the picornavirus3C protease [12]. The main protease is an attractive and well characterised drug target in corona viruses owing to the pivotal role it plays in the propagation of the virus inside the host cells [13]. The non-structural proteins of the virus (n=16) is encoded in the ORF1a/b of the RNA genome and gets transcribed and translated into two polyproteins (PP1a and PP1ab). Proteolytic cleavage of the PPs into its components is required to derive functionally active proteins. After its auto-cleavage from PP1a and PP1ab, M\(^{\text{pro}}\) cleaves PP1ab at about 11 sites. PP1ab contains the subunits of the replicase complex including the RNA dependant RNA polymerase (RdRP) and hence the cleavage becomes an essential requirement for viral replication [13, 14]. Thus, the inhibition of the main protease would effectively stop viral spread by preventing its replication. Since human proteases with the same cleavage specificity as the SARS-CoV-2 protease (Leu-Gln↓Ser,Ala,Gly) are not known, it is unlikely that an inhibitor would cross react with a human protease [13]. In 2020, Jin and others reported the structure of the SARS-CoV-2 M\(^{\text{pro}}\)in complex with a potent inhibitor [15]. The protein is 306 amino-acid long and has a molecular weight 33.8 KDa. In-order to identify clinically approved drugs that would bind to the catalytic site of M\(^{\text{pro}}\), an E-pharmacophore model based virtual screening was performed on a chemical library of known drugs from the SuperDRUG2 database. SuperDRUG2 contains more than 4600 active pharmaceuticals which are marketed/approved [16]. A subset of the drugs selected based on pharmacophore screening was further screened using molecular docking to identify potential drug leads. The free energy of binding for the identified drugs were calculated using MM-GBSA
2. Methods

All computational studies like E-pharmacophore hypothesis generation, virtual screening, molecular docking and MM-GBSA were carried out using Maestro version 11.4 from Schrodinger Inc. Visualisation of molecular interactions were performed using PyMol.

2.1. Generation of E-pharmacophore model

The crystal structure of M\textsuperscript{pro} of SARS-CoV-2 bound to a non-covalent inhibitor X77 at a resolution of 2.1 Å was used to generate an energy-optimised pharmacophore hypothesis (E-pharmacophore). The structural coordinates of M\textsuperscript{pro}-X77 complex was downloaded from the PDB (ID: 6W63). The structure of the protein-ligand complex was pre-processed and water molecules within 5 Å distance from the ligand were eliminated. Missing hydrogens and loops were added and the structure was subjected to energy minimisation using restrained minimisation by the OPLS3 force field [17]. These steps were performed using the protein preparation wizard of the Schrodinger suite [18]. The E-pharmacophore model was developed using the ‘Develop Pharmacophore from protein-ligand complex’ option in the Phase module [19]. For this, the prepared protein-ligand complex was imported to the workspace and default pharmacophore features such as hydrogen bond accepter (A), hydrogen bond donor (D), aromatic ring (R) and hydrophobicity (H) were mapped.

2.2. E-pharmacophore based virtual screening

E-pharmacophore based virtual screening was performed using the chemical structures of 4600 drugs (ligands) from the SuperDRUG2 database. Prior to the screening, the ligands were structurally optimised at near neutral pH (7±1). All plausible tautomers and stereoisomers were generated and protonation states were assigned. The ligands were subjected to energy minimisation with OPLS3 force field using the ligprep module of Schrodinger suite. In order to generate a subset of drugs with the desired molecular features for optimal binding to M\textsuperscript{pro}, as mapped by the E-pharmacophore model, a pharmacophore based virtual screening was carried out using the phase module of Schrodinger suite. The fitness scores were used to select the best hits.

2.3. Structure based virtual screening

The initial screening using the E-pharmacophore model enabled the selection of 1000 drugs with a potential to make energetically favourable interactions with the active site of M\textsuperscript{pro}. 
Further, to identify the most promising candidate drugs from this subset, a structure-based screening was performed on the selected drugs using Molecular Docking. The GLIDE (Grid-based Ligand Docking with Energetics) module of Schrodinger suite was used to perform all the molecular docking studies. A receptor grid was generated by keeping the crystallographic ligand (X77) as the centroid of the grid box. The size of the box was set to 15×15×15 Å. GLIDE scores (g-scores) were used to rank the drugs based on binding affinity [20]. The initial screening was performed using the High-Throughput Virtual Screening (HTVS) module of glide and the top scoring compounds were subjected to standard precision (SP) docking. Finally, extra precision (XP) docking method was used to identify the best hits.

2.3.1. Validation of docking procedure: The docking procedure was validated by a control study. For this the bound ligand in the crystal structure was re-docked to the pre-pre-processed and prepared protein keeping the same grid box. The glide score for this docking was used as a standard value against which the scores for the drugs were compared. The control docking was performed in the XP mode.

2.4. Estimation of Binding Free Energy

The theoretical binding free energy of the potent inhibitors of M\textsuperscript{pro} identified using combined E-pharmacophore and structure-based screening were calculated using the prime module of Schrodinger suit [21]. MM-GBSA is a popular method to calculate binding energy, which uses energy properties of free ligand, free receptor and receptor – ligand complex for binding affinity calculation. Binding energies were estimated for the 40 drugs selected based on the glide scores of XP docking, using the MM-GBSA method.

3. Results and discussion

3.1 E-pharmacophore hypothesis

Pharmacophore is defined as “an ensemble of steric and electronic features that is necessary to ensure the optimal supramolecular interactions with a specific biological target and to trigger (or block) its biological response” [22]. Pharmacophore model based screening has evolved as a key tool in computer aided drug discovery because of its ability to screen large libraries for potent hits within a short period of time and minimal computational capacities. Energy optimised pharmacophore models tries to combine the stereo-electronic features of the ligand with the energetics of its interactions with the protein structure [23]. In the present study an E-pharmacophore hypothesis was generated to screen for the inhibitors of the M\textsuperscript{pro} protein of
SARS-CoV-2 using its crystal structure in complex with a strong, broad spectrum non-covalent inhibitor. The bound inhibitor, an imidazole carboximide derivative dubbed X77 interacts strongly with the active site amino acids. Based on the ligand-protein complex an energy optimised five-featured pharmacophore hypothesis, AARRR was obtained. The generated E-pharmacophore model contains three aromatic rings (R) and two hydrogen bond accepters (A) (Figure 1). Figure 2 shows the planar representation of the pharmacophore hypothesis with distance between the features. The hypothesis AARRR was used as a 3D search query to screen 4600 drugs from the SuperDRUG2 database to identify drugs with comparable pharmacophore features. During the screening, the phase module analyses the fitness of compounds with the query hypothesis and ranks the search results based on a fitness score. Based on the fitness score 1000 compounds were selected for a structure based screening based on molecular docking.

Figure 1: E-pharmacophore model for M\textsuperscript{pro}-X77 complex mapped to the bound inhibitor X77: The left panel shows the bound inhibitor X77 (ball and stick) in the active site of M\textsuperscript{pro}(ribbon). The E-pharmacophore features on the inhibitor are shown in red. The right panel is a zoomed in image of the inhibitor, X77 with the pharmacophore marked in red.
3.2. Structure Based Virtual Screening

The best hits obtained in the E-pharmacophore based screening (n=1000) was further screened using molecular docking. The docking study analyses the molecular interactions of the different plausible geometries of the drugs (poses) with the active site residues of the SARS-CoV-2 M\textsuperscript{pro} and ranks them based on binding properties. Most promising inhibitors were identified through docking using HTVS, SP and XP methods; by filtering the outputs after each stage based on glide scores. The docking studies also revealed the interactions of the drugs with active site of the protease. Re-docking the crystallographic ligand X77 (N-(4-tert-butylphenyl)-N-[(1R)-2-(cyclohexylamino)-2-oxo-1-(pyridin-3-yl) ethyl]-1H imidazole-4-carboxamide) to the active site of the protease using the same protocol resulted in the binding of the ligand in the same position and orientation (Figure 3) and this revealed that the docking parameters and the grid box chosen were optimal. Analysis of the crystal structure of M\textsuperscript{pro}-X77 complex showed that the ligand interactions were stabilized through three hydrogen bonds with the active site residues, viz, Gly 143, His 163 and Glu 166. The docked structure also showed these three hydrogen bonds (Figure 4).
Figure 3: The docked pose of X77 (red) superimposed on the crystal structure (green).

Figure 4: Interactions between standard inhibitor X77 and Mpro active site. (a) X77-Mpro interactions in the crystal structure. (b) X77-Mpro interactions in the docked structure.

Based on HTVS and SP docking, 40 drugs were selected for XP docking (Supplementary data, S1). The g-score calculated for the crystallographic ligand, X77 was -8.243 Kcal/mol.
Seventeen of the 40 drugs used for XP docking showed g-scores comparable to that of the standard inhibitor used in the study (g-scores above -7.0 Kcal/mol) (Figure 5). Six drugs showed g-scores better than that of X77. Three highest scoring drugs, viz, Hidrosmin (-12.689 Kcal/mol), Diosmin (-11.409 Kcal/mol) and Monoxerutin (-10.745 Kcal/mol) are flavanoids with similar pharmacological properties. They are used as vaso-protectives and capillary stabilising agents. Remikirin (-9.429 Kcal/mol) is an interesting hit because it is a well-known inhibitor of Renin, an aspartyl endoprotease which acts as the primary enzyme in the renin-angiotensin system [24]. Remikirin may thus serve as a potential starting point for a drug against SARS-CoV-2. Doxorubisin (-9.16 Kcal/mol), an anthracyne class antineoplastic used as anti-cancer drug and fluvastatin (-8.346 Kcal/mol), an inhibitor of hydroxymethylglutaryl-coenzyme A (HMG-CoA) reductase used as antilipemic agent also showed high binding affinity to the protease. Of note, doxorubisin and its derivatives had been previously shown to be effective in-vitro against viruses like HIV, HSV, Dengue virus, Yellow Fever Virus, Rauscher leukemia virus, and avian myeloblastosis virus [25-27]. Although the drug targets in each case vary, the drugs were able to reduce virus replication in vitro. Statins, in general, are known to inhibit the replication of many enveloped viruses by the inhibition of cholesterol/isoprenoid pathway [28]. Fluvastatin was earlier shown to have an inhibitory effect on Heamophilus influenzae replication in-vitro [29]. Thus, the present study identifies the main protease of SARS-CoV-2 as a novel targets for the known antivirals, doxorubisin and fluvastatin. Fluvastatin is also an interesting hit considering the fact that the SARS-CoV-2 is also an enveloped virus. The drug might have a cumulative inhibitory effect on the propagation of the virus, if it can inhibit both the main protease and the cholesterol synthesis pathway.
3.3. Estimation of Binding energy

Drugs which showed high binding score was further subjected to binding energy prediction using the MM-GBSA method. The binding energy of crystallographic ligand, X77 as determined by MMGBSA was -73.68 Kcal/mol. Six drugs from the pool of 40 drugs identified using molecular docking showed binding free energies better than -60 Kcal/mol. These drugs were Binifibrate, Macimorelin acetate, Bamifylline, Rilmazafon, Afatinib and Ezetimibe. Except Bamyfilline (g-score; 6.61Kcal/mol) all of these drugs had shown g-scores better than -7 Kcal/mol in the docking studies. However few drugs that showed glide scores comparable to the standard inhibitor showed lower binding energies. Binifibrate, a hypolipidaemic drug showed a very high binding energy (-69.04Kcal/mol) similar to that of the standard inhibitor, X77. Binifibrate forms hydrogen bonds with three active site residues, viz, Gly 143, His 163, and Glu 166. The binding energies and interacting residues of the drugs that showed the highest binding free energies are shown in Table 1. The primary targets/activity for the drugs as are also given. From the results presented in Table 1 it can be seen that the hydrogen bonding interactions with Gly 143, His 163, and Glu 166, three polar amino-acid residues in the active site of the protease is critical in the binding of the drugs to the protein with high affinity. Figure 6 shows the interactions of the promising hits with the active site of the protease.

Figure 5: Drugs which showed comparable glide scores to that of X77, the standard inhibitor.
Table 1: Drugs with the potential to be repurposed against SARS-CoV-2 M\textsuperscript{pro}

<table>
<thead>
<tr>
<th>Drug</th>
<th>Binding Free Energy (Kcal/mol)</th>
<th>Fitness score</th>
<th>H-bonded residues</th>
<th>Primary Target/Activity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Binifibrate</td>
<td>-69.04</td>
<td>1.3</td>
<td>Gly143, His163, Glu166</td>
<td>Peroxisome proliferator-activated receptor agonist/ Hypolipidimic agent</td>
</tr>
<tr>
<td>Macimorelin acetate</td>
<td>-64.25</td>
<td>1.296</td>
<td>Gly 143, His 164, Glu 166, Thr 190</td>
<td>Agonist of Growth Hormone (GH) secretagogue receptor</td>
</tr>
<tr>
<td>Bamifylline</td>
<td>-63.19</td>
<td>1.534</td>
<td>Gly 143, Glu 166, Gln 189</td>
<td>Selective A1 adenosine receptor antagonist.</td>
</tr>
<tr>
<td>Rilmazafone</td>
<td>-61.37</td>
<td>1.409</td>
<td>Thr 26, Gly143, His163, Glu 166</td>
<td>GABA-A receptor agonist/ Non-benzodiazepine sedative</td>
</tr>
<tr>
<td>Afatinib</td>
<td>-60.89</td>
<td>1.169</td>
<td>Gly 143, Glu 166</td>
<td>Tyrosine kinase inhibitor</td>
</tr>
<tr>
<td>Ezetimibe</td>
<td>-60.21</td>
<td>1.175</td>
<td>Glu 166</td>
<td>Inhibits intestinal cholesterol absorption by physical interactions with Niemann-Pick C1-Like 1 (NPC1L1) transporter</td>
</tr>
</tbody>
</table>
Figure 6: Interactions of the top-scoring drugs with the active site of SARS-CoV-2 M\textsuperscript{pro}: (a) Afatinib, (b) Bamifylline, (c) Ezetimibe, (d) Binifibrate, (e) Macimorelin acetate, (f) Rilmazafone
4. Conclusion

Drug repurposing is perhaps the best way to combat the medical emergency poised by the SARS-CoV-2 infections that grows in magnitude with the passing of each day. Repurposing involves screening and identification of known bio-actives against specific therapeutic targets in SARS-CoV-2. Repurposed drugs gets to the market at relatively lesser time periods and costs compared to novel drugs. The main protease of SARS-CoV-2 is involved in the proteolytic processing of viral polyproteins to form key non-structural components involved in viral multiplication and hence is an attractive target for drug development. In the present study using a combination of E-pharmacophore and structure based virtual screening followed by binding energy estimation; a subset of known drugs from the superDRUG database are repurposed as putative drug leads against COVID-19. Of the 4600 drugs from the database six drugs were shown to bind to the main protease active site with high binding free energies. Previously known drugs Binifibrate, Macimorelin acetate, Bamifylline, Rilmazafon, Afatinib and Ezetimibe are proposed as potential inhibitors of the SARS-CoV-2 main protease. However, these drugs need to be further evaluated using in-vitro studies to confirm their inhibitory activity before they can be adapted to a drug development pipeline.

Conflicts of Interest:

The authors declare no conflicts of interest

Acknowledgements:

Arun KG and Abhithaj J acknowledge the Indian Council for Medical Research (ICMR), New Delhi, India for financial support. The authors would like to thank the Bioinformatics Infrastructure facility (BIF) at the Department of Biotechnology & Microbiology, Kannur University, supported by Department of Biotechnology (DBT), Government of India for computational facilities.
References


