

Repurposing Simeprevir, Calpain inhibitor IV and a cathepsin F inhibitor against SARS-CoV-2: A study using *in silico* pharmacophore modeling and docking methods.

Abhithaj J¹, Dileep Francis², Sharanya C.S¹, Arun K.G¹, Sadasivan C¹ and E Jayadevi Variyar^{*1}

¹Department of Biotechnology and Microbiology, Kannur University, Thalassery Campus, Kannur, Kerala-670661, India.

²Department of Life sciences, Kristu Jayanti College (Autonomous), Bengaluru, Karnataka-560077, India.

*Corresponding Author; e-mail: ejayadevi@gmail.com.

Abstract

The world has come to a sudden halt due to the incessant spread of a viral pneumonia dubbed COVID-19 caused by the beta-coronavirus, SARS-CoV-2. The pandemic spread of the virus has already claimed lakhs of valuable lives and has infected millions of people across the globe. The situation is further worsened by the fact that there is no approved therapeutics currently available for the treatment of the disease. The only way to handle the crisis is the rapid development of a therapeutic strategy to combat the virus. Computational biology offers resources to rapidly identify novel drug leads and to repurpose existing drugs at the expense of minimal resources and time. The main protease of SARS-CoV-2 is key to the replication and propagation of the virus in the host cells. Inhibiting the protease blocks replication and hence is an attractive therapeutic target in the virus. The crystal structures of the protein in complex with inhibitors are available in public databases. Here we describe the screening of the DrugBank database, a public repository for small molecule therapeutics, to identify approved or experimental phase drugs that can be repurposed against the main protease of SARS-CoV-2. The initial screening was performed on more than 13,000 drug entries in the target database using an energy optimised pharmacophore hypothesis AARRR. A sub-set of the molecules selected based on the fitness score was further screened using molecular docking by sequentially filtering the molecules through the high throughput virtual screening, extra precision and standard precision docking modalities. The most promising hits were subjected to binding free energy estimation using the MMGBSA method. Approved drugs viz, Cobicistat, Larotrectinib and Simeprevir were identified as potential candidates for repurposing. Drugs in the discovery phase identified as inhibitors include the known cysteine protease inhibitors, Calpain inhibitor IV and an experimental cathepsin F inhibitor.

Key words: Covid-19, SARS-CoV-2, Main protease, E-pharmacophore, Molecular docking, Simeprevir, Protease inhibitors

1. Introduction:

Covid-19, a novel viral pneumonia, was first reported from China in the late December of 2019 and has since then spread across the globe at an alarming rate. The etiological agent of the disease is an enveloped, non-segmented positive sense RNA virus belonging to the Nidovirales order of the Coronaviridae family and is named SARS-CoV-2 [[1], [2]]. The virus shows high sequence similarity with corona viruses previously known to infect humans including the SARS-CoV (82% identity) and the MERS-CoV (50% identity) which inflicts severe respiratory ailments in humans [3]. Mutations in the surface glycoprotein of the virus called as the spike protein facilitates effective binding with the human trans-membrane peptidase, angiotensin converting enzyme-2 (ACE2) and subsequent fusion of the viral membrane with host membrane, enabling the rapid development of infections and strong potential for human-to-human transmission seen with SARS-CoV-2 [[2], [3]]. Clinical manifestations of severe infections include acute respiratory distress syndrome (ARDS) and septic shock that in many cases eventually lead to multiple organ dysfunctions [4]. The outbreak which was declared a pandemic by the World health Organisation (WHO) in January 2020 has resulted in 3,660,471 infections and 252,680 deaths as on 5th April 2020.

Attempts to ease the pace of the spread of the infection through measures like patient quarantining, tracing and surveillance of patient contacts and social lockdown has contained the outbreak to some extent, but the risk still looms large. The crisis is aggravated by the fact that there are no approved drugs or prophylactics that is known to be effective against SARS-CoV-2. At present, the treatment for the infections relies on supportive therapy and some off-label drugs [5]. The only way, perhaps, to overcome the public health and economic crisis posed due to Covid-19 is to discover at a therapeutic moiety effective against the virus. However, novel drug discovery is a time and cost consuming process and has high uncertainty associated with it due to high attrition rates and regulatory constraints. In the fight against Covid-19, however both time and money is sparse to spare. Hence drug repurposing is envisaged as a promising approach to drug discovery. Drug repurposing or repositioning involves the identification of a novel application or target for a drug which is already present in the market or is in an advanced stage of pharmaceutical investigation [6]. Such drugs, having completed the regulatory requirements regarding the assessment of safety in humans can hit the market at relatively shorter periods of time and at lesser costs [7]. Indeed, most of the drugs under development against SARS-CoV-2 and many of the off-label drugs used against SARS-CoV-2 are repurposed known drugs [[8], [9]].

Corona virus genomes are among the largest in the RNA virus clade [10]. The SARS-CoV-2 genome is approximately 30 kb long and is composed of 12 full length open reading frames. The ORF1a and 1b encodes non-structural proteins and covers two-thirds of the total genome length. They are translated to form poly-proteins (pps) which are composite structures composed of many functionally distinct proteins [11]. Poly-proteins 1a and 1ab contains proteins that are required to build the replicase complex, a multi-protein complex containing the RNA dependant RNA polymerase, the enzyme that is essential for the replication of the virus. Before the complex can assemble the poly-proteins needs to be cleaved at specific locations to release the component proteins. This process which is critical for viral replication and propagation in the host cells is carried out by two viral proteases, viz, the papain-like protease (PL^{pro}) and the main protease (M^{pro}). PL^{pro} cleaves the polyproteins 1a and 1ab at three sites, while M^{pro} cleaves it at eleven sites [[12], [13]]. Considering their role in viral replication, these proteins are attractive drug targets against SARS-CoV-2. The 3D structure of SARS-CoV-2 M^{pro}, a 309 residue long protein with a molecular weight of 34 kDa determined using X-ray crystallography was first reported in March 2020 [[14], [15]]. Subsequently many groups have reported the 3D structures of the protein in complex with different inhibitors. The protein recognises the amino acid sequence, Leu-Gln↓ (Ser,Ala,Gly) and cleaves at the site denoted by the arrow [14]. This cleavage site specificity is not displayed by any known human proteases. Hence, it is less likely that a drug against SARS- CoV-2 M^{pro} would cross-react with a human protease. Using M^{pro} as the target, our group had recently reported a subset of drugs from the SuperDrug2 database which showed potential to be used against SARS-CoV-2 [16]. In the present study, we report the virtual screening of the DrugBank database, a larger database of drugs using the method of E-pharmacophore modeling and docking to identify compounds that would inhibit M^{pro} and hence could be repurposed against SARS-CoV-2.

The DrugBank database (version 5.1.6, released 2020-04-22) contains around 13571 drug entries including approved small molecules, experimental (discovery-phase) drugs, and nutraceuticals [17]. An E-pharmacophore hypothesis was developed based on the crystallographic structure of M^{pro} in complex with a non-covalent inhibitor. This hypothesis was then used to screen the database for drugs possessing similar pharmacophoric features. The hits from the pharmacophore based screen were further screened using molecular docking to select the most promising drug candidates. The binding free energy of the potential hits in complex with M^{pro} was determined using MMGBSA, to identify candidates for repurposing.

2. Methods:

Maestro version 12.2 from Schrodinger Inc. was used to perform all the *in silico* methods including the generation of the E-pharmacophore hypothesis, virtual screening based on E-pharmacophore and molecular docking and calculation of binding free energy. Visualisations of molecular structures were done using Pymol.

2. 1. Generation of the E-pharmacophore hypothesis:

An energy optimized pharmacophore hypothesis (E-pharmacophore) was built based on the crystal structure of SARS-CoV-2 3CL^{pro} in complex with a non-covalent inhibitor X77 at a resolution of 2.1 Å (unpublished data). The atomic co-ordinates of the 3CL^{pro}-X77 complex were retrieved from Protein Data Bank (PDB ID: 6W63). The protein structure with the bound inhibitor was pre-processed by removing the water molecules in the vicinity of the ligand binding pocket and by adding the missing hydrogens. After pre-processing the structure underwent restrained energy minimisation using OPLS 2005 force field [18]. The protein preparation wizard of Schrodinger suite was employed to perform these steps [19]. After processing, the structure was imported to the maestro workspace and fed to the ‘Develop pharmacophore from protein-ligand complex’ option available in the phase module of Schrodinger suite to generate the E- pharmacophore hypothesis [[20], [21]]. The default parameters were maintained allowing the program to map the pharmacophore features like hydrogen bond acceptor (A), hydrogen bond donor (D), aromatic ring (R) and hydrophobicity (H).

2. 2. Virtual Screening based on the E-pharmacophore hypothesis:

The E-pharmacophore hypothesis generated was utilised to screen the DrugBank database for pharmaceutical molecules that possess the pharmacophoric features presented by the hypothesis. The molecular structures of 13563 drugs from the database were subjected to an optimisation process at the near physiological pH (7 ± 1) by generating all possible tautomeric and stereo-isomeric conformations and assigning the correct protonation states. The optimised structures were energy minimised using OPLS 2005 force field. These steps were executed using the Ligprep module of Schrodinger suite. The prepared ligands were then subjected to E-pharmacophore based virtual screening using the phase module of the Schrodinger suite [20]. The program assigns a fitness score to the screened molecules based on their degree of relatedness to the E-pharmacophore hypothesis. The fitness score was used to select the best hits from the database.

2. 3. Virtual Screening using Molecular Docking:

The best hits obtained from the previous step were then subjected to further screening using molecular docking. The grid based ligand docking (GLIDE) module offered by the Schrodinger suite was used for the screening [22]. A glide grid box, which confines the search space for the ligand while docking, was generated on the protein by keeping the crystallographic ligand, X77 as the centroid of the grid box. The dimensions of the grid were set to 20×20×20 Å along the X, Y and Z axes. The docking program expresses the binding affinity of the ligands to the active site of the protein in terms of glide scores (g score) [23]. The molecular docking flow involved an initial screening at the lowest docking precision using High Throughput Virtual Screening (HTVS), followed by the standard precision (SP) docking and finally the extra precision (XP) docking. The ligand pool used for docking with each precision was filtered and narrowed down based on the g scores.

2. 3. 1. Quality control for docking parameters

In order to validate the docking parameters used in the study, the high affinity, non covalent ligand bound to M^{pro} in the crystallographic structure was extracted from the complex, and subjected to optimisation and energy minimisation under the OPLS 2005 force field. The ligand was then re-docked to the pre-processed and minimised protein using the same grid box that was used for the screening of the candidate drugs. The glide score obtained for this docking was used as a reference value to which the g scores obtained for the docking based screening of drugs were compared. The re-docking procedure was carried out in XP mode.

2. 4. Estimation of Binding Free Energy

The best hits obtained from XP docking were subjected to binding free energy estimation. This was carried out using the MMGBSA method in the prime module of the Schrodinger suite [24]. As a control exercise, the binding free energy for the re-docked crystallographic ligand X77 in complex with M^{pro} was also calculated. The value obtained was used as a reference to compare the values obtained with the candidate drugs.

3. Results and Discussion

3.1. The E-pharmacophore hypothesis

A pharmacophore represents the minimum conformational and electronic features that should be present in a molecule to bring about a desired biological effect by binding to target. E-pharmacophore combines steric and electronic features of a pharmacophore with the energy parameters associated with the optimal interaction with the target [25]. Given the crystallographic structure of a protein-ligand complex, the minimum stereo-electronic ensemble on the ligand required to bind to the protein along with optimal energy parameters can be mapped to generate an energy optimised pharmacophore hypothesis. The hypothesis can then be used to rapidly screen large libraries of molecules to identify the ones that possess similar pharmacophoric features using lower computational capacities and time as compared to molecular docking approaches. In the present study the crystal structure of M^{pro} from SARS-CoV-2 in complex with a non-covalent inhibitor was used to develop the E-pharmacophore hypothesis. The generated hypothesis had 2 hydrogen bond acceptors (designated as A) and 3 aromatic rings (designated as R) (**Figure 1**). The hypothesis lacked hydrogen bond donors and hydrophobic moieties. The hypothesis can be represented in short as AARRR. This hypothesis was provided as a search query in the phase module of Schrodinger suite to search against the DrugBank database for molecules with similar features. From 13, 563 drugs available in the DrugBank, 1000 drugs were selected based on the screening for further analysis. The drugs were selected as the best hits because they ranked highest in terms of the fitness score, a measure of the relatedness of the subject molecule with the original hypothesis. The advantage of an E-pharmacophore based screening is evident from the approximately 13 fold decrease in the number of candidates to be screened using Molecular docking in the subsequent step. The computational time and power required for docking all the 13, 563 molecules against M^{pro} would have been significantly higher compared to what was required for the E-pharmacophore based screening.

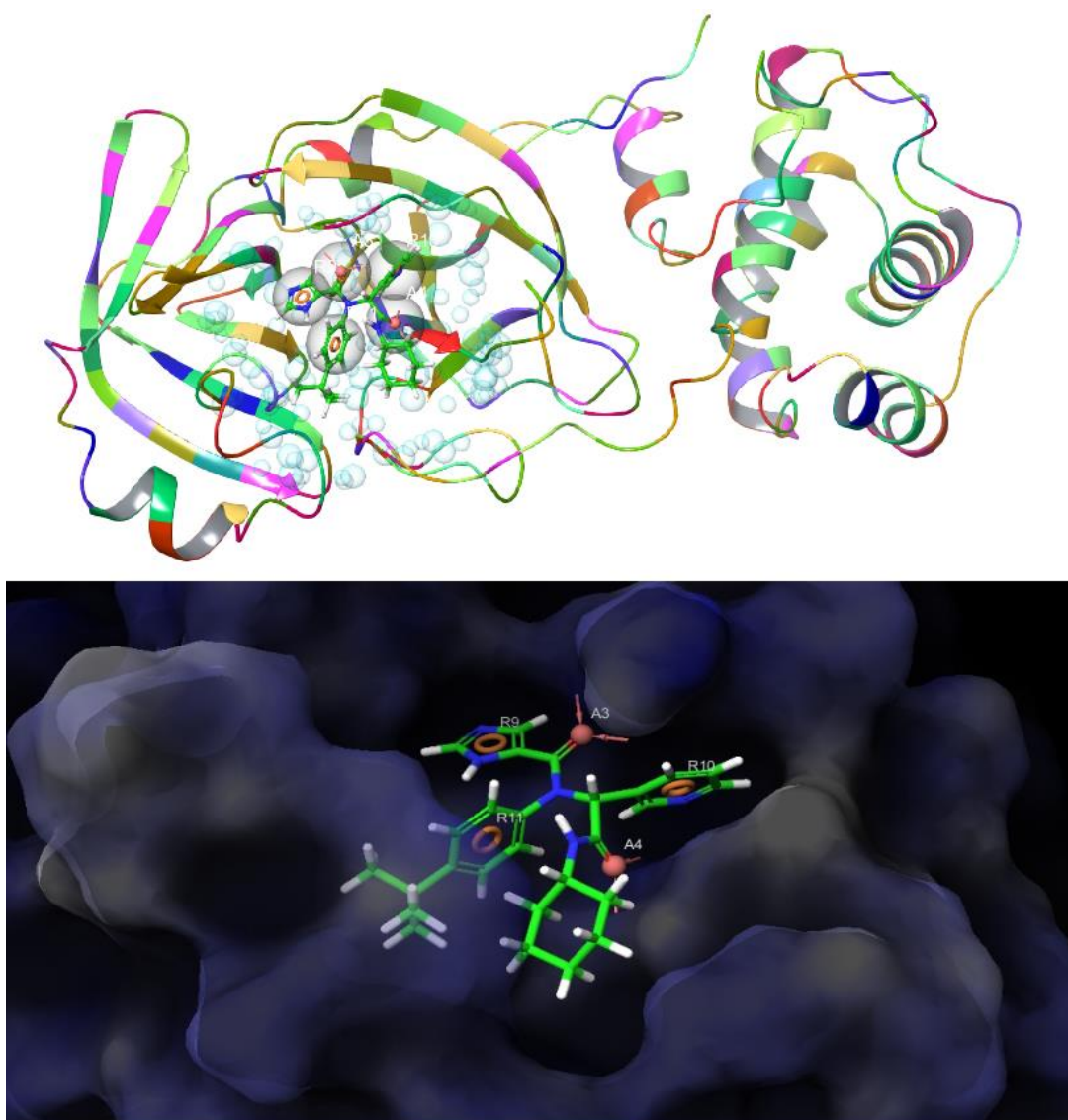


Figure 1: The E-pharmacophore hypothesis AAARR: (A) The E- pharmacophore hypotheis mapped on the non-covalent inhibitor X77 (ball and stick) bound to the active site of M^{pro} (ribbon). The E-pharmacophore features are shown in red and enclosed in spheres. (B) Surface view of the protein active site with the bound ligand. The E-pharmacophore features are shown in red.

3.2. Docking based virtual screening

The top scoring hits obtained from the E-pharmacophore based screening was further screened using molecular docking using the GLIDE module of the Schrodinger suite. Molecular docking employs a structure based strategy to estimate the binding affinity of a drug to the active site of a protein. Different conformations of a ligand are generated and docked on to the binding site to estimate to the binding affinity in terms of a scoring function called as the glide score. The glide score is based on the energetically favourable, non covalent interactions established by a given pose of a ligand. In addition docking also reveals the molecular interactions that stabilise the binding of a ligand to the active site. The thousand hits obtained from the E-pharmacophore based screening was further screened using HTVS, SP and XP docking modules by filtering and removing low-scoring hits at each stage. The selection of drugs was based on the g scores. Based on the control experiment using the non-covalent inhibitor X77 (N-(4-tert-butylphenyl)-N-[(1R)-2-(cyclohexylamino)-2-oxo-1-(pyridin-3-yl) ethyl]-1H imidazole-4-carboxamide), a reference value for analysing the results from the docking based screening was derived. Re-docking of the crystallographic ligand to the active site of M^{pro} revealed that the docking parameters chosen were optimal for the screening as the docked structure engaged the active site through the same type of interactions as observed in the crystallographic complex between X77 and M^{pro}. Both the docked and the crystallographic structure of M^{pro}-X77 complex had three hydrogen bonds between the ligand and the protein involving the active site residues Gly 143, His 163 and Glu 166. Interaction with these residues are assumed to be critical in binding of drug candidates to M^{pro}. The conformation adopted by the ligand in the docked structure and the crystallographic structure were also found to be similar which further improved the confidence in the docking protocol (**Figure 2**). X77 docked in the active site of M^{pro} with a glide score of -4.9 Kcal/mol.

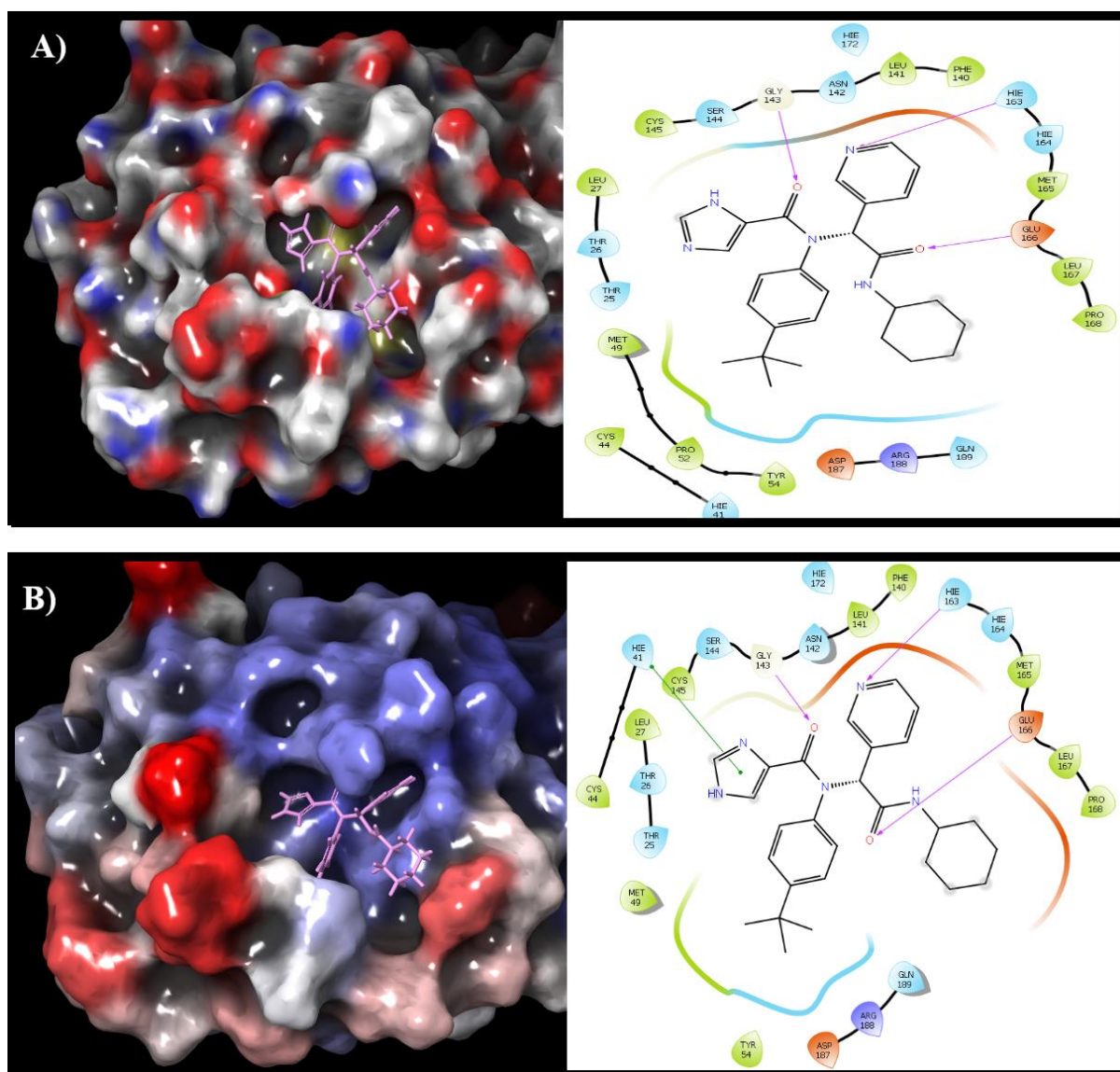


Figure 2: Re-docking of the crystallographic ligand X77 to M^{pro} (A) Binding pose of X77 in the active site of the energy minimized 3CL^{pro}-X77 complex in the crystallographic structure. The right panel shows the 2D ligand interaction diagram. The H-bonding interactions of the ligand with Gly 143, His 163 and Glu 166 are shown. (B) Binding pose of re-docked X77 in the active site of 3CL^{pro}. The 2D interaction diagram shown to the right displays the same H-bonding interactions seen in the crystallographic structure as well as an additional π - π stacking interaction with His 41.

Forty-five drugs were selected and subjected to XP docking based on the scores obtained from HTVS and SP docking. Thirty drugs from this pool docked M^{pro} with glides scores comparable to the reference value (g score above -4.5 Kcal/mol). Of the thirty drugs, seven are approved/investigational drugs (**Figure 3**) while the rest are experimental small molecules (**Table 1**). The approved drugs that are identified as potential inhibitors of SARS-CoV-2

M^{pro} include, Ezetimibe, Larotrectinib, Cobicistat, Regadenoson, Capmatinib, Alogliptin and Simeprevir. Ezetimibe is a hypolipidemic agent that inhibits intestinal and hepatic cholesterol absorption through interactions with the sterol transporter Niemann-Pick C1-Like 1 (NPC1L1) [26]. Inhibition of cellular lipid synthesis or absorption pathway is a common strategy for viral entry inhibition and Ezetimibe is previously known to block HCV and HBV infection through this mechanism [27]. The present study identifies M^{pro} of SARS-CoV-2 as a novel antiviral target for Ezetimibe.

Larotrectinib and Capmatinib are anti-neoplastics. The former inhibits the tropomyosin receptor kinase (Trk) resulting in apoptosis and inhibition of cell growth in tumors [28] while the latter is approved for the treatment of metastatic non-small cell lung cancer [29]. Cobicistat is pharmacokinetic enhancer used for the treatment of HIV. Interestingly it neither inhibits the virus directly nor does it have a therapeutic target in HIV. Instead the drug inhibits the cytochrome P450 3A (CYP3A) enzymes which are involved in the oxidative biotransformation of useful drugs. The co-administration of Cobicistat along with anti-retrovirals like atazanavir or darunavir results in increased systemic exposure of the drugs, thereby decreasing the required drug dosage and side effects [30]. Present study shows that the drug may directly inhibit the main protease of SARS-CoV-2. Regadenoson is an agonist of the A2A adenosine receptor that causes a rapid increase in the coronary blood flow and is used as a pharmacological stress agent for myocardial perfusion imaging [31]. Alogliptin is a dipeptidyl peptidase-4 (DPP-4) inhibitor used for the clinical management of type 2 diabetes mellitus due to its potential to stimulate insulin secretion and reduce glucagon secretion [32]. DPP-4 is a serine exo-peptidase that cleaves and inactivates peptide hormones like glucagon-like peptide-1 [GLP-1] and glucose-dependent insulintropic polypeptide [GIP]) that regulate glucose homeostasis. It has to be noted that the structural fold of the SARS-CoV-2 main protease is similar to serine proteases but uses a cysteine and a nearby histidine to cut proteins [33]. Thus alogliptin is a candidate that be further explored to be repurposed against the M^{pro} of SARS-CoV-2.

Simeprevir is the most interesting hit obtained in the study as it is a ‘directly acting antiviral’ (DAA) used in the treatment of chronic Hepatitis C Virus (HCV) infection [34]. Another recent molecular docking based virtual screening on a different data set also had identified the drug as a potent inhibitor of the SARS-CoV-2 M^{pro} [35]. Despite a lower glide score compared to the other hits (but still similar to the standard inhibitor) the binding free energy for simeprevir-M^{pro} complex is comparable to that of the standard inhibitor. Docking results shows that the

interaction of Simeprevir with the 3CL^{pro} active site is stabilised through 3 hydrogen bonds with HIS41, CYS44 and GLY143 in the active site. Simeprevir is a known inhibitor of the NS3/4A protease of HCV, a membrane targeted serine protease that proteolytically processes the ~3000 amino acid long poly-protein encoded by the virus to generate the non-structural proteins NS3, NS4A, NS4B, NS5A and NS5B by cleaving at 4 non-structural (NS) sites [36]. M^{pro}, the therapeutic target used in this study performs a function analogous to that of the NS3/4A protease of HCV in SARS-CoV-2 by proteolytically cleaving the polyproteins pp1a and pp1ab. As mentioned earlier the catalytic fold of M^{pro} is structurally similar to that of serine proteases, the class of protease to which NS3/4A of HCV belongs to. Thus it is safe to reason that the predicted affinity of Simeprevir for the catalytic domain of M^{pro} may be due to the structural similarities between the catalytic folds of SARS-CoV-2 M^{pro} and HCV nsp3/4A. In addition, Simeprevir has displayed remarkable pharmacokinetics when used against HCV and gets extensively distributed in the intestinal tract and liver with a bioavailability of 44%, upon administration of a single oral dose [34]. When used in conjunction with therapeutics like sofosbuvir or ribavirin and peginterferon-alfa the cure rates are 80-90% in HCV infections [37]. It was also shown that the drug is effective against HCV/HIV co-infection [38]. A recent study demonstrated that the drug can be extended for therapy against the positive sense RNA viruses, Zika virus (ZIKV), enterovirus A71 (EV-A71) as well as the DNA virus Herpes simplex virus type 1 (HSV-1) [39]. In these viruses, Simeprevir was shown to up-regulate the expression of key components of the antiviral innate immune signals including interferon beta (IFN-beta) as well as the interferon stimulated gene 15 (ISG15), which are generally inhibited during a viral infection as part of immune evasion mechanisms elaborated by viruses. Taken together Simeprevir is a high potential candidate for drug repurposing against the M^{pro} of SARS-CoV-2.

The highest docking scores were shown by experimental molecules in the DrugBank database. These molecules do not fit into the general scheme of drug repurposing, however they may be considered as drug leads for novel drug development against SARS-CoV-2 M^{pro}. The experimental drug Beta-1, 2, 3, 4, 6 Penta-O-Galloyl-D-Glucopyranose (Drug bank ID: DB03208) showed the highest glide score of -17.49 Kcal/mol. The high glide score is due to eight hydrogen bonding interactions it establishes with the binding groove of M^{pro}. The compound is under review for a patent for its inhibitory effect on the proliferation of HepG2 cancer cells in combination with 5-fluorouracil. With a glide score of -12.708 Kcal/mol, Coenzyme F420 was the second best hit. In nature, this heterocyclic molecule mediates redox

transformations in bacteria and archaea involved in microbial persistence, antibiotic biosynthesis, pro-drug activation and methanogenesis [40]. However the compound did not show an impressive binding free energy. A substituted flavin-adenine-dinucleotide (FAD) named 8-demethyl-8-dimethylamino-flavin-adenine-dinucleotide (DB03482) with a glide score of -10.636 Kcal/mol establishes 6 hydrogen bonds with the active site including two with Glu 166, a critical residue involved in the binding of X77 to M^{pro}.

Table 1: Experimental drugs that showed glide scores better than -4.5 Kcal/mol

Drug Bank ID	Name of the Drug	Glide score (Kcal/mol)
DB03208	Beta-1,2,3,4,6-Penta-O-Galloyl-D-Glucopyranose	-17.491
DB03913	Coenzyme F420	-12.708
DB03482	8-demethyl-8-dimethylamino-flavin-adenine-dinucleotide	-10.636
DB01940	Balanol Analog 2	-8.85
DB03960	Cp-Coeleneterazine	-8.421
DB03541	10-Propargyl-5,8-Dideazafolic Acid	-8.38
DB04653	N-[(benzyloxy)carbonyl]-L-leucyl-N-[(1S)-3-fluoro-1-(4-hydroxybenzyl)-2-oxopropyl]-L-leucinamide	-8.342
DB04049	Coelenteramide	-8.195
DB07471	5-[5,6-bis(methyloxy)-1h-benzimidazol-1-yl]-3-[[1-(2-chlorophenyl)ethyl]oxy]-2-thiophenecarboxamide	-8.17
DB03141	3-[[[(5R,6R)-5-Benzyl-6-hydroxy-2,4-bis(4-hydroxybenzyl)-3-oxo-1,2,4-triazepan-1-yl]sulfonyl]benzonitril	-8.163
DB01772	3-[3-(2,3-Dihydroxy-Propylamino)-Phenyl]-4-(5-Fluoro-1-Methyl-1h-Indol-3-Yl)-Pyrrole-2,5-Dione	-8.009
DB08229	[N-(3-dibenzylcarbamoxy)-oxiranecarbonyl]-hydrazino]-acetic acid	-7.701
DB02830	FR236913	-7.32
DB02243	4-Morpholin-4-Yl-Piperidine-1-Carboxylic Acid [1-(3-Benzenesulfonyl-1-Propyl-Allylcarbamoxy)-2-Phenylethyl]-Amide	-7.281
DB08701	2-(3-bromophenyl)-6-[(2-hydroxyethyl)amino]-1h-benzo[de]isoquinoline-1,3(2h)-dione	-7.237
DB07785	1-[(1r,2s)-2-hydroxy-1-[2-(2-naphthyloxy)ethyl]propyl]-1h-imidazole-4-carboxamide	-7.194
DB07006	9-hydroxy-6-(3-hydroxypropyl)-4-(2-methoxyphenyl)pyrrolo[3,4-c]carbazole-1,3(2h,6h)-dione	-7.147
DB03572	FR230513	-6.851
DB07265	3-(9-hydroxy-1,3-dioxo-4-phenyl-2,3-dihydropyrrolo[3,4-c]carbazol-6(1h)-yl)propanoic acid	-6.739

DB02483	Etheno-NAD	-6.419
DB04601	4-[(4-o-sulfamoylbenzyl)(4-cyanophenyl)amino]-4h-[1,2,4]-triazole	-6.39
DB03363	3-Acetylpyridine Adenine Dinucleotide	-6.02
DB01946	3-[1-(3-Aminopropyl)-1h-Indol-3-Yl]-4-(1-Methyl-1h-Indol-3-Yl)-1h-Pyrrole-2,5-Dione	-6.01

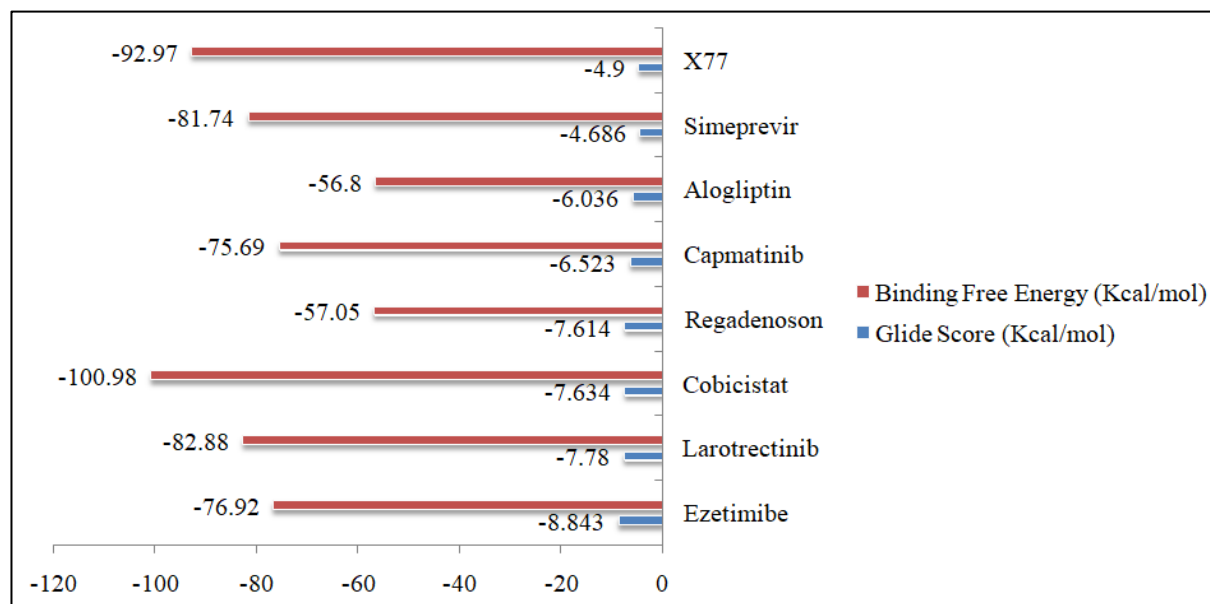


Figure 3: Approved/ investigational drugs that glide scores comparable to that of the reference inhibitor X77. The binding free energy of the drugs is also shown.

3.3. Estimation of binding free energy

The best hits from the docking based screening were subjected to the estimation binding free energy using MMGBSA method. The binding free energy of the M^{pro}-X77 complex was -92.97 Kcal/mol. This value was used as a reference for selecting the best inhibitors of M^{pro} from the database. Among the 30 drugs identified using the molecular docking studies 8 had binding energies comparable to that of X77 (better than -80 kcal/mol) (**Table 2**). (The molecular structures of these drugs are shown in **Figure 4**). Overall, there was good correlation between the g scores and the predicted binding free energies. However Simeprevir which showed the lowest glide score (-4.686 kcal/mol) when compared other hits interacted with M^{pro} with a binding free energy of -81.74 Kcal/mol. Simeprevir, for reasons already discussed shows high promise as a drug against SARS-CoV-2. Cobicistat (-100.98) and Larotrectinib (-82.88 Kcal/mol), both are approved drugs, had good glide scores as well as binding free energy and hence are potential candidates for repurposing. DB03208, DB03482, DB04653, DB07471, and DB02243 are the experimental drugs that showed the highest binding energies. DB03208 (g

score -17.491 Kcal/mol) and DB03482 (g score -10.636 Kcal/mol) had shown impressive glide scores too. DB03913 (g score -12.708), despite having high glide scores showed lower binding energy (-74.64 Kcal/mol). DB04653 and DB02243 are remarkable hits as they showed both high glide scores and binding free energy. In addition, both of these experimental drugs are cysteine protease inhibitors. It is important to note that the target protein, M^{pro} of SARS-CoV-2 also belongs to the cysteine protease family. DB04653 (Refer Table 1 for IUPAC name) also called Calpain inhibitor IV is a known inhibitor of Calpain, a calcium dependant cysteine protease [41]. DB02243 is an inhibitor of Cathepsin F, a protease belonging to the papain family of cysteine proteases. Although not available for immediate development through repurposing, these molecules show promise as novel drug candidates against SARS-CoV-2. Another experimental drug that binds M^{pro} with high affinity is DB07471 a benzimidazole that targets cyclin A2 and cyclin dependant kinase 2. **Figure 5** shows the non-covalent interactions that stabilise the binding of the 8 drugs which showed the highest binding energies and hence identified as promising candidate drug leads against SARS-CoV-M^{pro}.

Table 2: Candidate inhibitors of SARS-CoV-2 M^{pro} based on binding free energy estimation

Sl. No.	Drug/ Experimental Molecule	Binding Free Energy (Kcal/mol)	Active site interactions	Fitness score	Primary Target/Activity
1	DB02243	-101.52	GLU166	1.083	Cathepsin F
2	Cobicistat	-100.98	ASN142, HIS163	0.686	Cytochrome P450 3A (CYP3A) isoforms.
3	DB07471	-92.09	HIS41, CYS44, GLY143, GLU166	1.161	Cyclin- dependent kinase 2
4	DB03208	-89.12	HIS41, PHE140, GLN189, GLU166, THR190, GLN192	1.000	Receptor tyrosine kinase- like orphan receptor 1
5	DB03482	-89.05	HIS41, ASN142, GLU166, GLN189, GLN 192	1.320	p- hydroxybenzoat e hydroxylase
6	Larotrectinib	-82.88	GLY143, HIS163	0.968	Tropomyosin kinase receptors TrkA, TrkB, and TrkC.
7	Simeprevir	-81.74	HIS41, CYS44, GLY143	1.136	NS3/4 protease
8	DB04653	-81.28	HIS164, GLU166, THR190, GLN192	0.536	Calpain-1 catalytic subunit

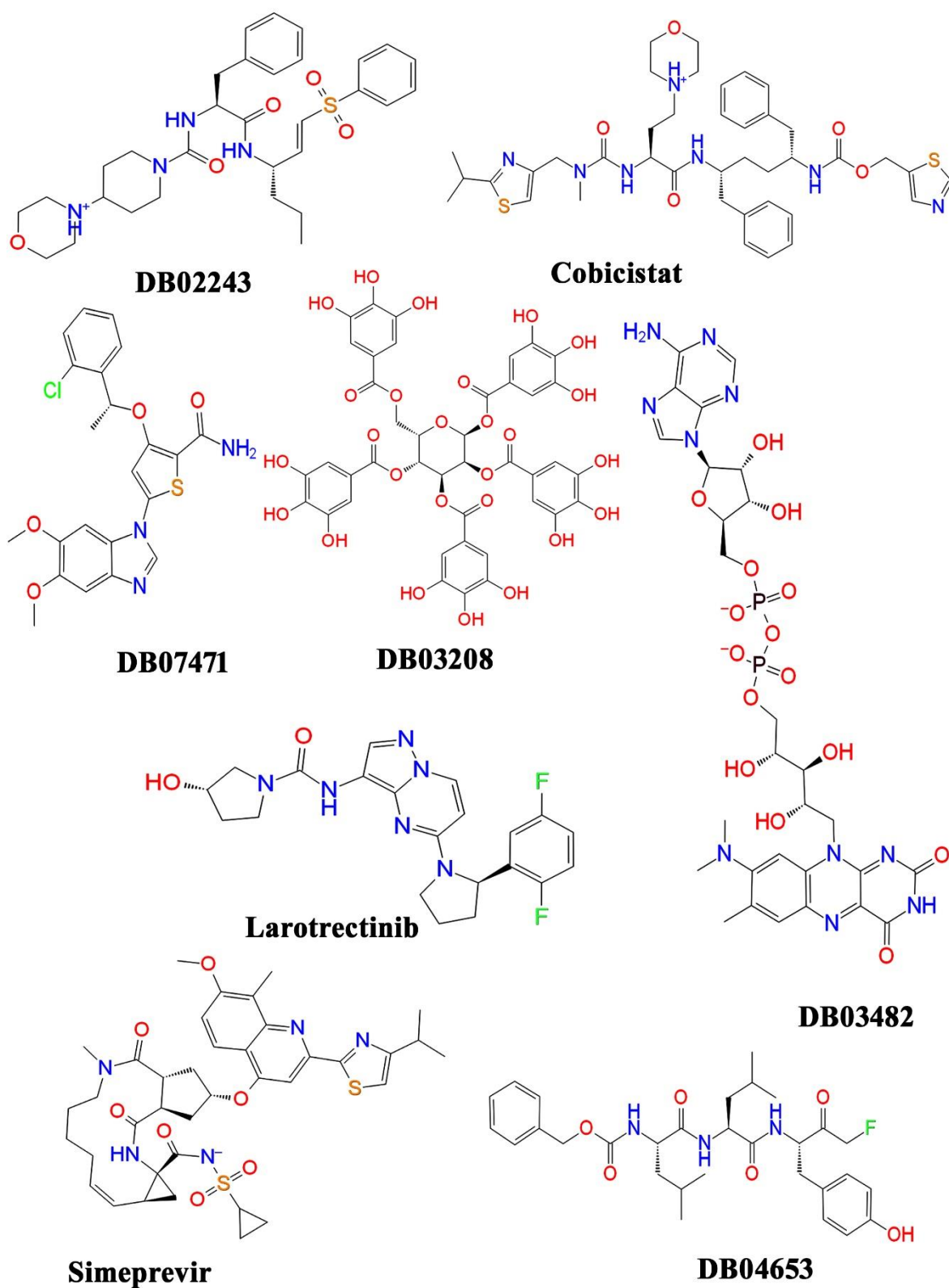
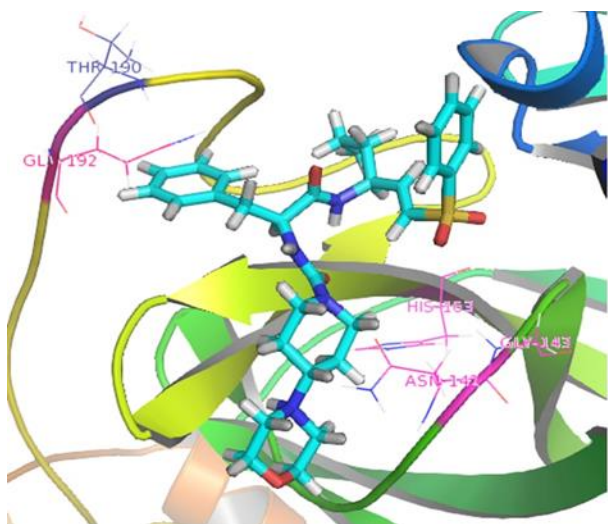
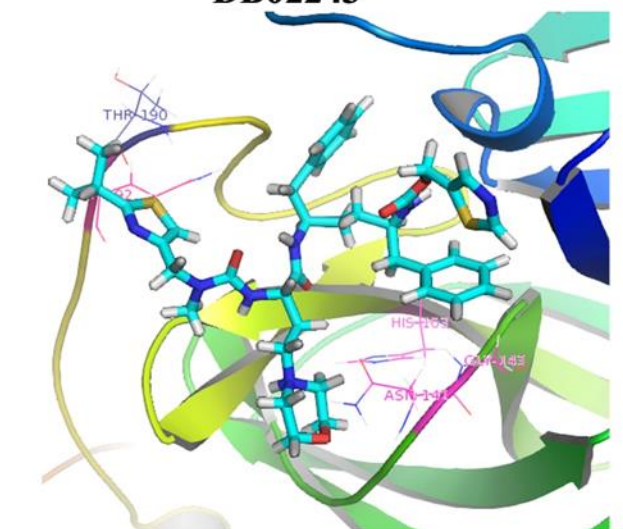
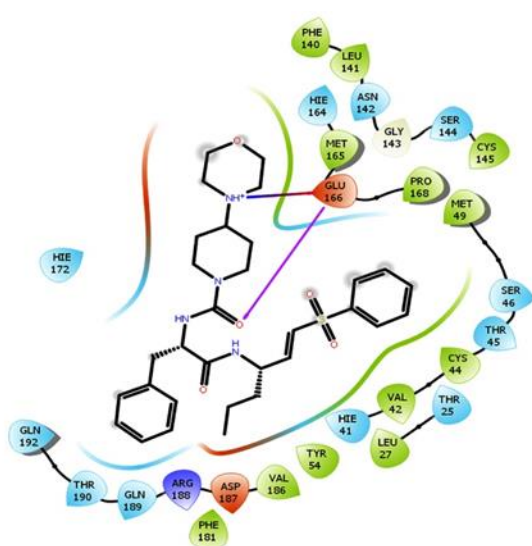


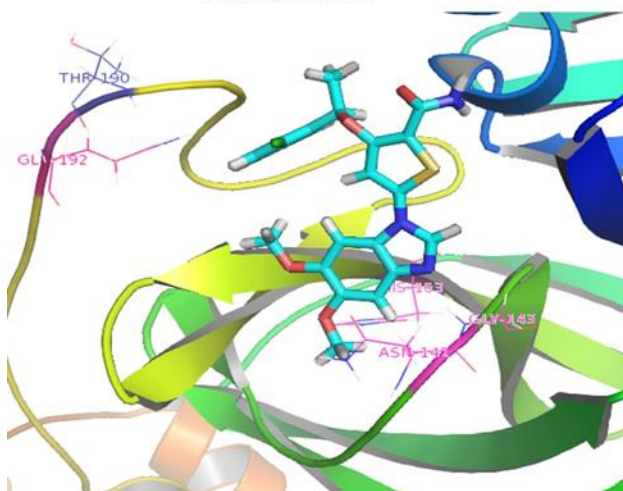
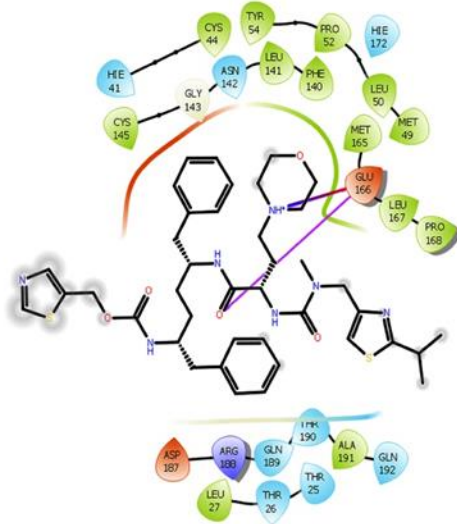
Figure 4: Structures of candidate inhibitors of M^{pro} of SARS-CoV-2 identified in the study. For experimental drugs the drug bank IDs are given. Refer to table for the IUPAC names.



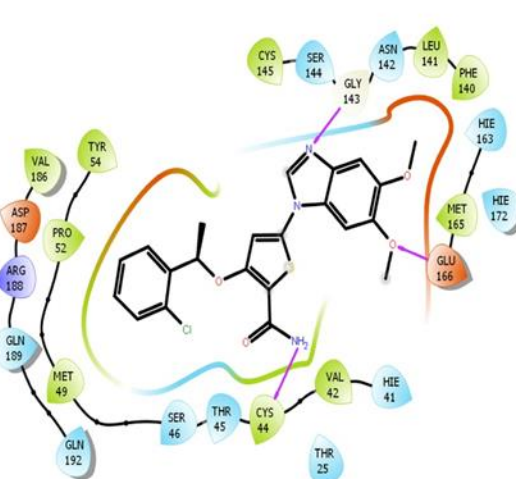
DB02243

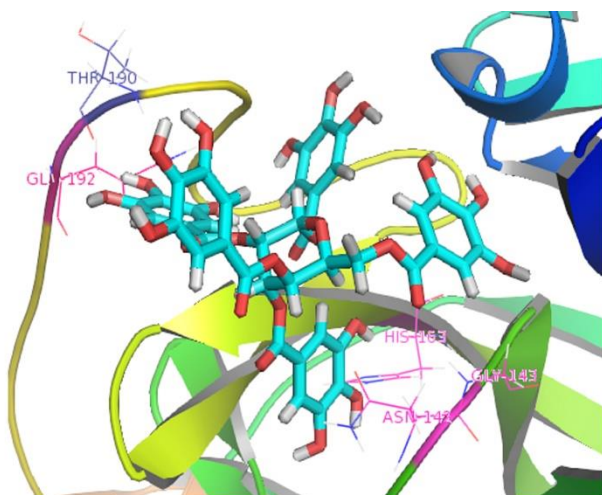


Cobicistat

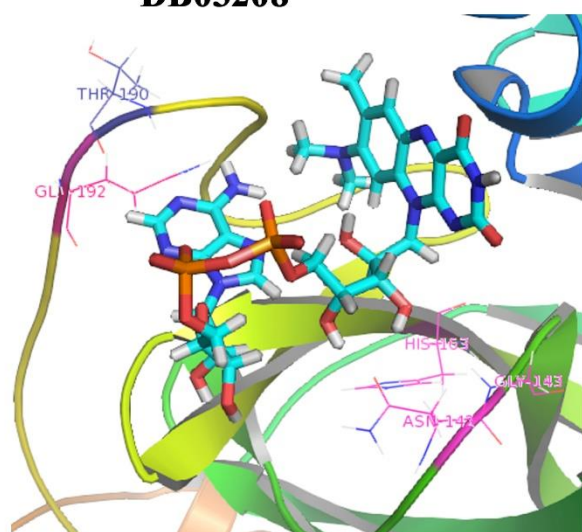
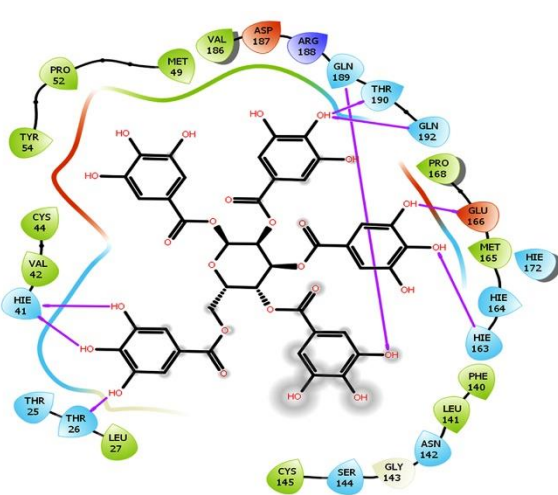


DB07471

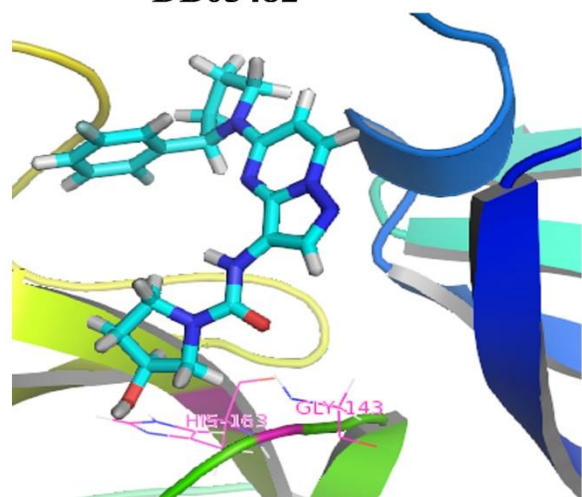
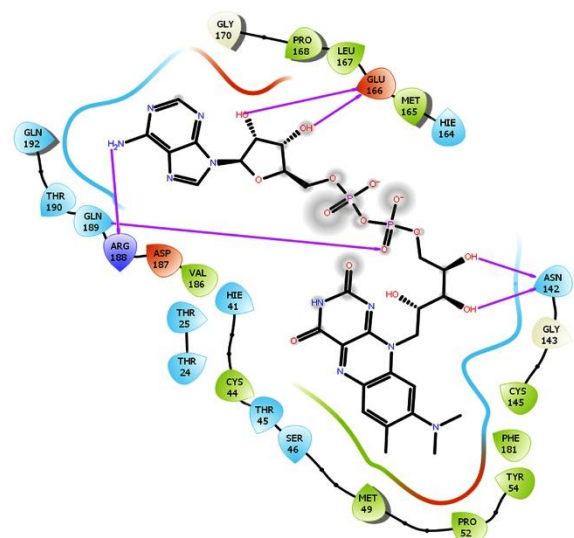




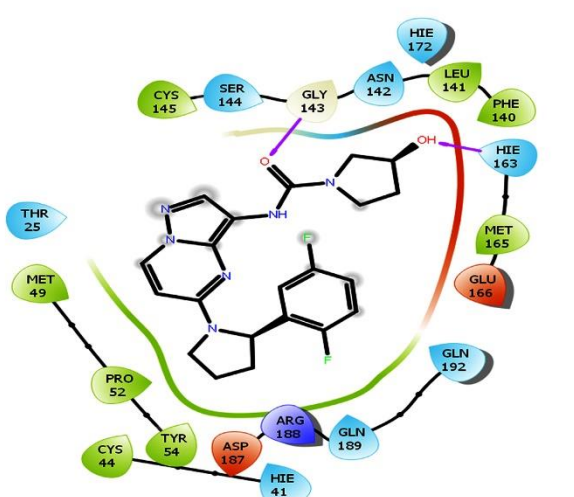
DB03208



DB03482



Larotrectinib



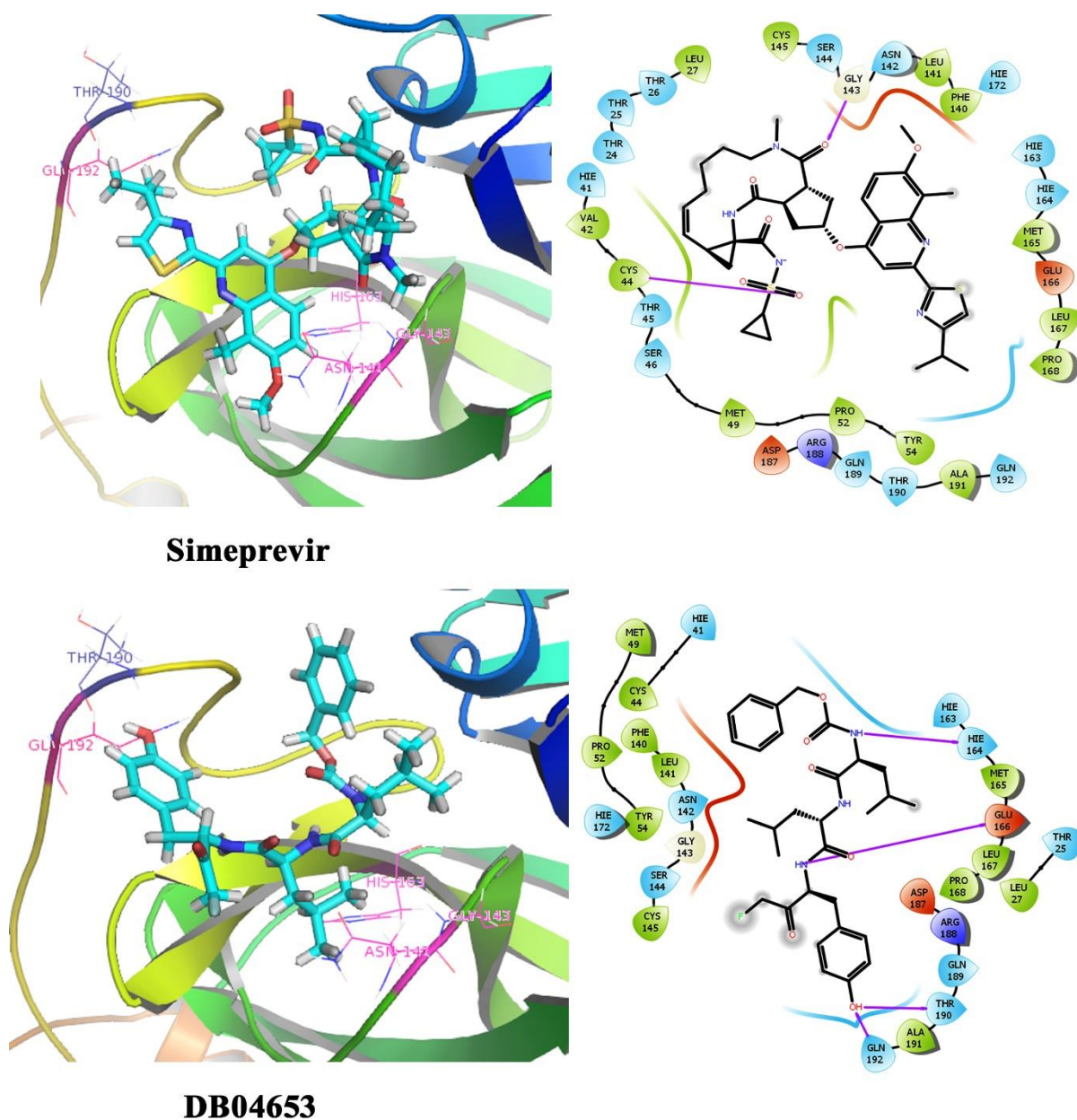


Figure 5: Active site interactions of the lead molecules identified in the study. The left panel shows the docked ligand interacting with the active site of M^{pro} . The 2D interaction diagrams are given in right panel.

4. Conclusion:

The only way to overcome the health care related and economic crisis inflicted by the global spread of SARS- CoV-2 is to develop a viable and approved therapeutic strategy against the virus in a short period of time. The sooner it is made possible, the lesser will be the fatalities and consequences the outbreak will have on humanity. The proteome of the SARS-CoV-2 contains about 29 proteins and offers many potential drug targets to develop therapeutics. Structures of some of these proteins are experimentally determined and made available in PDB.

Among them, the main protease, M^{pro} proteolytically processes the polyproteins 1a and 1ab synthesised by the virus during an infection to derive functional proteins serving critical functions in viral propagation. The proteins cleaved by M^{pro} include the RNA dependant RNA polymerase which is the key enzyme for viral genome replication. Evidently, inhibiting M^{pro} will block replication and spread of the virus through the host cells. Thus it is an attractive drug target in SARS-CoV-2. In the present study by employing a 3D experimental structure of M^{pro} in complex with a strong binding, non-covalent inhibitor, an energy optimised pharmacophore hypothesis was developed to screen ~13,000 small molecule bio-actives available in the DugBank database. From a subset of 1000 molecules identified based on the screen, 8 candidate drug leads were identified through molecular docking and binding free energy estimations. Of these drugs some are already known medicines and hence can be used for rapid drug development through Drug repurposing. Based on the study Cobicistat, Larotrectinib and Simeprevir are proposed as potential candidates for drug repurposing against the M^{pro} of SARS-CoV-2. Simeprevir is a known direct action antiviral, inhibiting a functionally similar protease in HCV and hence stands out as the most promising and relevant hit obtained in the study. Five experimental, molecules viz, DB03208, DB03482, DB04653, DB07471, and DB02243 were also identified as potential inhibitors of M^{pro}. However these drugs have not been approved for any medical indications so far and hence will have to be adopted to a novel drug discovery pipeline against SARS-CoV-2 and subsequently enter the development stage. Though this process which is time and resource consuming is not ideal to solve the current crisis, the molecules still are of value for future drug development. However it is important to evaluate both the known and experimental drugs identified in this study through *in-vitro* methods for confirming the inhibitory activity.

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 the financial support. Sharanya CS acknowledge CSIR-SRF for the 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:

- [1] F. Wu *et al.*, “A new coronavirus associated with human respiratory disease in China,” *Nature*, vol. 579, no. 7798, pp. 265–269, Mar. 2020, doi: 10.1038/s41586-020-2008-3.
- [2] P. Zhou *et al.*, “A pneumonia outbreak associated with a new coronavirus of probable bat origin,” *Nature*, vol. 579, no. 7798, pp. 270–273, 2020, doi: 10.1038/s41586-020-2012-7.
- [3] K. G. Andersen, A. Rambaut, W. I. Lipkin, E. C. Holmes, and R. F. Garry, “The proximal origin of SARS-CoV-2,” *Nature Medicine*, vol. 26, no. 4, pp. 450–452, Apr. 2020, doi: 10.1038/s41591-020-0820-9.
- [4] M. Xie and Q. Chen, “Insight into 2019 novel coronavirus — an updated interim review and lessons from SARS-CoV and MERS-CoV,” *International Journal of Infectious Diseases*, vol. 94, pp. 119–124, May 2020, doi: 10.1016/j.ijid.2020.03.071.
- [5] M. Prajapat *et al.*, “Drug targets for corona virus: A systematic review,” *Indian Journal of Pharmacology*, vol. 52, no. 1, p. 56, Jan. 2020, doi: 10.4103/ijp.IJP_115_20.
- [6] T. T. Ashburn and K. B. Thor, “Drug repositioning: identifying and developing new uses for existing drugs,” *Nature Reviews Drug Discovery*, vol. 3, no. 8, pp. 673–683, Aug. 2004, doi: 10.1038/nrd1468.
- [7] S. Pushpakom *et al.*, “Drug repurposing: progress, challenges and recommendations,” *Nature Reviews Drug Discovery*, vol. 18, no. 1, pp. 41–58, Jan. 2019, doi: 10.1038/nrd.2018.168.
- [8] O. Mitjà and B. Clotet, “Use of antiviral drugs to reduce COVID-19 transmission,” *The Lancet Global Health*, vol. 8, no. 5, pp. e639–e640, May 2020, doi: 10.1016/S2214-109X(20)30114-5.
- [9] L. R. Baden and E. J. Rubin, “Covid-19 — The Search for Effective Therapy,” *New England Journal of Medicine*, Mar. 2020, doi: 10.1056/NEJMe2005477.
- [10] V. C. C. Cheng, S. K. P. Lau, P. C. Y. Woo, and K. Y. Yuen, “Severe Acute Respiratory Syndrome Coronavirus as an Agent of Emerging and Reemerging Infection,” *Clinical Microbiology Reviews*, vol. 20, no. 4, pp. 660–694, Oct. 2007, doi: 10.1128/CMR.00023-07.
- [11] C. Wang *et al.*, “The establishment of reference sequence for SARS-CoV-2 and variation analysis,” *Journal of Medical Virology*, vol. 92, no. 6, pp. 667–674, 2020, doi: 10.1002/jmv.25762.
- [12] K. Anand, J. Ziebuhr, P. Wadhwani, J. R. Mesters, and R. Hilgenfeld, “Coronavirus Main Proteinase (3CLpro) Structure: Basis for Design of Anti-SARS Drugs,” *Science*, vol. 300, no. 5626, pp. 1763–1767, Jun. 2003, doi: 10.1126/science.1085658.
- [13] R. Hilgenfeld, “From SARS to MERS: crystallographic studies on coronaviral proteases enable antiviral drug design,” *The FEBS Journal*, vol. 281, no. 18, pp. 4085–4096, 2014, doi: 10.1111/febs.12936.
- [14] L. Zhang *et al.*, “Crystal structure of SARS-CoV-2 main protease provides a basis for design of improved α -ketoamide inhibitors,” *Science*, vol. 368, no. 6489, pp. 409–412, Apr. 2020, doi: 10.1126/science.abb3405.
- [15] Z. Jin *et al.*, “Structure of M pro from COVID-19 virus and discovery of its inhibitors,” *Nature*, pp. 1–9, Apr. 2020, doi: 10.1038/s41586-020-2223-y.

- [16] A. Kumar, S. C. S, A. J, D. Francis, and S. C, "Drug Repurposing Against SARS-CoV-2 Using E-Pharmacophore Based Virtual Screening and Molecular Docking with Main Protease as the Target," Apr. 2020, doi: 10.26434/chemrxiv.12199610.v1.
- [17] D. S. Wishart *et al.*, "DrugBank 5.0: a major update to the DrugBank database for 2018," *Nucleic Acids Res.*, vol. 46, no. D1, pp. D1074–D1082, 04 2018, doi: 10.1093/nar/gkx1037.
- [18] G. A. Kaminski, R. A. Friesner, J. Tirado-Rives, and W. L. Jorgensen, "Evaluation and Reparametrization of the OPLS-AA Force Field for Proteins via Comparison with Accurate Quantum Chemical Calculations on Peptides," *J. Phys. Chem. B*, vol. 105, no. 28, pp. 6474–6487, Jul. 2001, doi: 10.1021/jp003919d.
- [19] G. Madhavi Sastry, M. Adzhigirey, T. Day, R. Annabhimoju, and W. Sherman, "Protein and ligand preparation: parameters, protocols, and influence on virtual screening enrichments," *J Comput Aided Mol Des*, vol. 27, no. 3, pp. 221–234, Mar. 2013, doi: 10.1007/s10822-013-9644-8.
- [20] S. L. Dixon, A. M. Smondirev, E. H. Knoll, S. N. Rao, D. E. Shaw, and R. A. Friesner, "PHASE: a new engine for pharmacophore perception, 3D QSAR model development, and 3D database screening: 1. Methodology and preliminary results," *J Comput Aided Mol Des*, vol. 20, no. 10, pp. 647–671, Oct. 2006, doi: 10.1007/s10822-006-9087-6.
- [21] N. R. Tawari and M. S. Degani, "Pharmacophore mapping and electronic feature analysis for a series of nitroaromatic compounds with antitubercular activity," *J Comput Chem*, vol. 31, no. 4, pp. 739–751, Mar. 2010, doi: 10.1002/jcc.21371.
- [22] R. A. Friesner *et al.*, "Extra Precision Glide: Docking and Scoring Incorporating a Model of Hydrophobic Enclosure for Protein–Ligand Complexes," *Journal of Medicinal Chemistry*, vol. 49, no. 21, pp. 6177–6196, Oct. 2006, doi: 10.1021/jm051256o.
- [23] T. A. Halgren *et al.*, "Glide: A New Approach for Rapid, Accurate Docking and Scoring. 2. Enrichment Factors in Database Screening," *J. Med. Chem.*, vol. 47, no. 7, pp. 1750–1759, Mar. 2004, doi: 10.1021/jm030644s.
- [24] P. D. Lyne, M. L. Lamb, and J. C. Saeh, "Accurate Prediction of the Relative Potencies of Members of a Series of Kinase Inhibitors Using Molecular Docking and MM-GBSA Scoring," *J. Med. Chem.*, vol. 49, no. 16, pp. 4805–4808, Aug. 2006, doi: 10.1021/jm060522a.
- [25] H. Sun, "Pharmacophore-Based Virtual Screening," *CMC*, vol. 15, no. 10, pp. 1018–1024, Apr. 2008, doi: 10.2174/092986708784049630.
- [26] E. A. Nutescu and N. L. Shapiro, "Ezetimibe: A Selective Cholesterol Absorption Inhibitor," *Pharmacotherapy: The Journal of Human Pharmacology and Drug Therapy*, vol. 23, no. 11, pp. 1463–1474, 2003, doi: 10.1592/phco.23.14.1463.31942.
- [27] J. W. Schoggins and G. Randall, "Lipids in Innate Antiviral Defense," *Cell Host & Microbe*, vol. 14, no. 4, pp. 379–385, Oct. 2013, doi: 10.1016/j.chom.2013.09.010.
- [28] L. J. Scott, "Larotrectinib: First Global Approval," *Drugs*, vol. 79, no. 2, pp. 201–206, Feb. 2019, doi: 10.1007/s40265-018-1044-x.
- [29] J. F. Vansteenkiste, C. V. D. Kerkhove, E. Wauters, and P. V. Mol, "Capmatinib for the treatment of non-small cell lung cancer," *Expert Review of Anticancer Therapy*, vol. 19, no. 8, pp. 659–671, Aug. 2019, doi: 10.1080/14737140.2019.1643239.

- [30] E. D. Deeks, “Cobicistat: A Review of Its Use as a Pharmacokinetic Enhancer of Atazanavir and Darunavir in Patients with HIV-1 Infection,” *Drugs*, vol. 74, no. 2, pp. 195–206, Feb. 2014, doi: 10.1007/s40265-013-0160-x.
- [31] K. P. Garnock-Jones and M. P. Curran, “Regadenoson,” *Am J Cardiovasc Drugs*, vol. 10, no. 1, pp. 65–71, 2010, doi: 10.2165/10489040-000000000-00000.
- [32] C. I. Jarvis, A. Cabrera, and D. Charron, “Alogliptin: A New Dipeptidyl Peptidase-4 Inhibitor for Type 2 Diabetes Mellitus,” *Annals of Pharmacotherapy*, Nov. 2013, doi: 10.1177/1060028013504076.
- [33] H. Yang *et al.*, “The crystal structures of severe acute respiratory syndrome virus main protease and its complex with an inhibitor,” *PNAS*, vol. 100, no. 23, pp. 13190–13195, Nov. 2003, doi: 10.1073/pnas.1835675100.
- [34] X. Zhang, “Direct anti-HCV agents,” *Acta Pharmaceutica Sinica B*, vol. 6, no. 1, pp. 26–31, Jan. 2016, doi: 10.1016/j.apsb.2015.09.008.
- [35] F. S. Hosseini and M. Amanlou, “Simeprevir, Potential Candidate to Repurpose for Coronavirus Infection: Virtual Screening and Molecular Docking Study,” Feb. 2020, doi: 10.20944/preprints202002.0438.v1.
- [36] L. Chatel-Chaix, M. Baril, and D. Lamarre, “Hepatitis C Virus NS3/4A Protease Inhibitors: A Light at the End of the Tunnel,” *Viruses*, vol. 2, no. 8, pp. 1752–1765, Aug. 2010, doi: 10.3390/v2081752.
- [37] A. Majumdar, M. T. Kitson, and S. K. Roberts, “Systematic review: current concepts and challenges for the direct-acting antiviral era in hepatitis C cirrhosis,” *Alimentary Pharmacology & Therapeutics*, vol. 43, no. 12, pp. 1276–1292, 2016, doi: 10.1111/apt.13633.
- [38] S. Flanagan, A. Crawford-Jones, and C. Orkin, “Simeprevir for the treatment of hepatitis C and HIV/hepatitis C co-infection,” *Expert Review of Clinical Pharmacology*, vol. 7, no. 6, pp. 691–704, Nov. 2014, doi: 10.1586/17512433.2014.956091.
- [39] Z. Li *et al.*, “Antiviral effects of simeprevir on multiple viruses,” *Antiviral Research*, vol. 172, p. 104607, Dec. 2019, doi: 10.1016/j.antiviral.2019.104607.
- [40] G. Bashiri *et al.*, “A revised biosynthetic pathway for the cofactor F 420 in prokaryotes,” *Nature Communications*, vol. 10, no. 1, p. 1558, Apr. 2019, doi: 10.1038/s41467-019-09534-x.
- [41] N. Carragher, “Calpain Inhibition: A Therapeutic Strategy Targeting Multiple Disease States,” *CPD*, vol. 12, no. 5, pp. 615–638, Feb. 2006, doi: 10.2174/138161206775474314.