Identification of potential molecules against COVID-19 main protease through structure-guided virtual screening approach

Lovika Mittal[1,#], Anita Kumari[1,#], Mitul Srivastava[1,#], Mrityunjay Singh[1], Shailendra Asthana[1]*
[1] Translational Health Science and Technology Institute (THSTI), Haryana, 121001, India

[#] these authors contributed equally
[*] To whom the correspondence should be addressed.
[#] These authors contributed equally
[*] To whom the correspondence should be addressed.

Dr. Shailendra Asthana
Translational Health Science and Technology Institute (THSTI),
NCR Biotech Science Cluster, 3rd Milestone
Faridabad – Gurgaon Expressway, Haryana-121001, India
E-Mail: sasthana@thsti.res.in

Abstract

The epidemic caused by novel coronavirus disease 2019 (COVID-19) infecting around ~ 1 million populations worldwide and counting, has demanded quick and potential therapeutic strategies. Current approved drugs or molecules under clinical trials can be a good pool for repurposing through computer-aided drug design techniques to quickly identify promising drug repurposing candidates. The structural information of recently released crystal structure of COVID-19 proteases in complex with inhibitor, N3, and 13b molecules was utilized to conduct extensive virtual docking screening of aforementioned drug pool. Considering the recent success of HIV protease molecules, we also used protease molecules (from different viruses) for drug repurposing purposes. The top docking hits were further used for focused docking followed by binding free energy calculations using MM-GBSA. Interestingly, in our screening, several promising already documented drugs stand out as potential inhibitors of COVID-19 main protease. However, based on N3 and 13B molecules as control, we have identified six potential molecules, Leupeptin Hemisulphate, Pepstatin A, Nelfinavir, Birinapant, Lypression and Octeotide which have shown the docking energy > -8.0 kcal/mol and MMGBSA > -68.0 kcal/mol. The pharmacokinetic annotations and therapeutic importance of these molecules have suggested that they possess drug-like properties and pave their way for in vitro studies. Further insight shows that the molecules forms stable interactions with *hot-spot* residues, that are mainly conserved and can be targeted for structure and pharmacophore-based designing of more potent molecules.

Key-words: Virtual screening, COVID-19, SARS-CoV-2, M^{pro} protease, docking, binding free energy.

1. Introduction

Coronaviruses (CoVs) belong to a pathogenic group of viruses that causes respiratory tract infections in mammals and birds. It originated through zoonotic transmission between bats and humans and then transmitted from humans to humans spreading via saliva droplets or nasal discharge from an infected person (*Coronavirus*, n.d.). It belongs to a family of coronaviridae and they can be divided into four genera, i.e., alphacoronavirus, betacoronavirus, gammacoronavirus, and deltacoronavirus. The current outbreak of coronavirus disease (COVID-19; previously known as 2019-nCoV-2) was first reported from Wuhan, China, in late December 2019 (H. Chen et al., 2020), which has subsequently affected countries worldwide, reporting nearly half a million of confirmed cases of COVID-19 along with ~39,000 deaths and nearly 200,000 people have recovered as per data recorded in march 2020 (*Coronavirus Update (Live): 803,772 Cases and 39,070 Deaths from COVID-19 Virus Outbreak - Worldometer*, n.d.). Human coronaviruses were initially discovered in the 1960s (Kahn & McIntosh, 2005) with a bronchitis infection in chickens and in humans. SARS-CoV (in 2003), HCoV NL63 (2004), HKU1 (2005), MERS-CoV (2012) and SARS-CoV-2 (COVID-19) are members of its family that have been identified till now (Su et al., 2016). The COVID-19 is a new strain that has emerged due to which they are not familiar to our immune system causing infection in the respiratory tract. According to WHO, the symptoms include mild to moderate respiratory illness and in severe conditions may develop difficulty in breathing and pain in chest, but patients may recover without special treatment. Old age people with medical complications like cardiovascular disease, diabetes, chronic respiratory disease, and cancer are known to be more likely to develop serious illness.

Unfortunately, at present there is no well-defined treatment or therapeutics against COVID-19 is available but the preventive measures are being recommended worldwide which includes social distancing and isolation of infected individuals along with providing treatment as per the symptoms with mild disease and oxygen therapy/ventilator support for patients with severe disease. However, the clinical trials for already marketed drugs such as lopinavir, ritonavir, hydroxychloroquine, azithromycin, (Tirumalaraju, 2020c) chloroquine (*Chloroquine/ Hydroxychloroquine Prevention of Coronavirus Disease (COVID-19) in the Healthcare Setting - Full Text View - ClinicalTrials.gov*, n.d.), Remdesivir (Tirumalaraju, 2020b) etc. along with antibiotics are being evaluated to treat the secondary infections (www.clinicaltrials.gov). All of the drug options come from experience treating SARS, MERS or some other new influenza virus previously (Lu, 2020). These drugs would be helpful but the efficacy needs to be further confirmed. Few COVID-19 vaccines are also under clinical trials such as Moderna's mRNA-1273, first US clinical vaccine funded by NIH's NIAID (National Institute of Allergy and Infectious Diseases) (Tirumalaraju, 2020a). Thus, there is an unmet requirement for the specific anti-COVID-19 therapeutics to limit the severity of the deadly disease. Various clinicians and researchers are engaged in investigating and developing antivirals using different strategies combining experimental and *in-silico* approaches (Ahmed et al., n.d.; Hatada et al., n.d.; Nair & Narayanan, n.d.) with the goal of identifying novel, selective and potent therapeutic agents. The RCSB protein data bank (PDB) (www.rcsb.org) has collected ~104 crystal structures of COVID-19 proteins to allow the rational designing of small molecules. COVID-19 (or SARS-CoV-2) is an enveloped non-segmented large (+)ssRNA virus (~30kb) with 5′-cap structure and 3′-poly-A tail belonging to β-CoV category (Cascella et al., 2020; Su et al., 2016). It is round or elliptic or pleomorphic with diameter ranging in 60–140 nm (Cascella et al., 2020). It is presumed that like other CoVs, it is sensitive to ultraviolet rays and heat. Moreover, these viruses could be effectively inactivated by lipid solvents including ether (75%), ethanol, chlorine-

containing disinfectant, peroxyacetic acid and chloroform except for chlorhexidine(Cascella et al., 2020) Its RNA genome contains 29891 nucleotides, and encoding for ~9860 amino acids (Cascella et al., 2020). The genome codes for both structural (namely, spike (S), envelope (E), membrane (M) and nucleocapsid (N)) and non-structural proteins (NSPs 1-16) (Graham et al., 2008).

Its replicase gene encodes two overlapping polyproteins pp1a and pp1ab, that are required for viral replication and transcription. Its first ORF (orf1) is the largest, that encodes ~ 70% of its entire genome and 16 NSPs (Bzówka et al., n.d.; Y. Chen et al., 2020). Among these nsps, the main protease (Mpro, also known as a chymotrypsin-like cysteine protease 3CLpro), encoded by nsp5, has been found to play a fundamental role in viral gene expression and replication, thus it became an attractive target for anti-CoV drug design (Bzówka et al., n.d.; Y. Chen et al., 2020). The remaining ORFs are responsible to code for accessory and structural proteins. The functional polypeptides are released from the polyproteins by extensive proteolytic processing via 33.8-kDa main protease (M^{pro}). M^{pro} digests the polyprotein at 11 conserved sites, starting with the autolytic cleavage of this enzyme itself from pp1a and pp1ab5 (Anand et al., 2003; "The Molecular Biology of Coronaviruses," 2006) and that results in 16 NSPs (including itself)(Y. Chen et al., 2020) The present study focused on the main proteases in CoVs ($3CL^{pro}/M^{pro}$) (PDB ID 6LU7), as potential target proteins for COVID-19 treatment. M^{pro} in COVID-19 has been structured and repositioned in PDB and has been accessible by the public since early February 2020. Mpro/3CLpro is active in its dimer state but till now there is no crystal structure available for the dimer form. Its monomer inhabits the 306 amino acids including 3 domains, folded into helices and β-strands. The electron density map for the monomer protein is clearly visible (**Figure 1A**). The domain I (residues 1–101) and II (residues 102–184) includes an antiparallel β-barrel structure; and domain III (residues 201–303) includes five α-helices arranged into a largely antiparallel globular cluster, and is connected to domain II by means of a long loop region (residues 185–200). The catalytic dyad (H41 and C145) is responsible for the catalytic activity of SARS-COV-2 and is placed at the junction of domain I and domain II (**Figure 1A**). Recently, a crystal structure for monomeric Mpro complexed with N3 peptide like molecule (inhibitor) was crystallised (PDB-ID: 6LU7) and without any inhibitor (PDB-ID: 6M03). The M^{pro} of 2019-nCov shares 96% similarity with the M^{pro} of the SARS-CoV(Xu et al., n.d.; Zhavoronkov et al., n.d.). It is reported that 12 residues vary in both SARS-CoV-1 and SARS-CoV-2 but the residue S46 in SARS-CoV-2 (COVID-19) (corresponding residue A46 in SARS-CoV-1) is part of the binding site of the N3 molecule or active site (Bzówka et al., n.d.) . A fragment molecular orbital (FMO) based interaction analysis on a complex between the SARS-CoV-2 main protease (Mpro) and its peptide-like inhibitor N3 (PDB ID: 6LU7) is reported. The target inhibitor molecule was segmented into five fragments in order to capture site-specific interactions with amino acid residues of the protease. From this study, H41, H163, H164, and E166 were found to be the key interacting residue with the inhibitor, mainly due to hydrogen bonding (Hatada et al., n.d.). Another co-crystal (PDB-ID 6Y2F), α-ketoamide inhibitor (13b) is also reported recently, providing the structural and residue-based architecture of catalytic sites (L. Zhang et al., 2020). These co-crystals are paving the route for the application of virtual screening to get more efficacious molecules.

Considering the emergency situation, the drug repurposing approach is being widely applied to quickly identify therapeutic solutions due to availability of their pharmacokinetic, toxicological and manufacturing data. Drug repurposing includes drugs that are either FDA approved, investigational, withdrawn or shelved molecules. Although there are studies of the repurposing and marketed drugs which proposed several candidates for SARS-CoV-2

treatment (Contini, n.d.). With this aim, we have used a computer-aided drug design approach to identify the new molecules against COVID-19. To facilitate the rapid identification of potent molecules against COVID-19, we have performed the virtual screening of currently FDA approved drugs, repurpose and anti-protease inhibitors of different viruses. The findings from this study may provide opportunities to other researchers to identify and develop the right molecules to combat against COVID-19.

2. Materials and Methods

2.1 Protein Structure Preparation. The protein crystals of M^{pro} of Covid-19 were retrieved from Protein Data Bank (PDB ID: 6LU7, 6M03) (Jin et al., n.d.). The crystal (6LU7) is found to be bound with a peptide-like inhibitor 'N3' and another crystal (6M03) is without any ligand (APO). The structures were optimized and then minimized using the Protein Preparation Wizard module of Maestro (Anang et al., 2018; *Maestro, version 2017; Schrödinger, LLC: New York, 2017*, n.d.) in which OPLS3 force field was used (Jorgensen et al., 1996).

2.2 SiteMap Analysis. The SiteMap (*SiteMap, Version 4.2, Schrödinger, LLC, New York, NY, 2017*, n.d.) program of Schrodinger Suite was also used for calculating binding sites on crystal 6LU7. The method was implemented as an unbiased approach to undermine the presence of any secondary or allosteric binding site. SiteMap which is a ligand independent method, will also help in calculating the druggability of the identified site (Mattapally et al., 2018; Srivastava et al., 2018). The method identifies putative binding sites by implementing different parameters such as: *site score, size, exposure score, enclosure, hydrophobic/hydrophilic character, contact, and donor/acceptor character* are considered to highlight the potential binding site. As per Halgren analysis (Halgren, 2009), the average number of sites for sub-micromolar sites is 132, where lower exposure scores of 0.52 and Higher exposure scores of 0.76 average is considered better for sub-micromolar sites. For the donor/ acceptor character and site score, the average for the sub-micromolar sites is 0.76 and 1.01, respectively. Druggability of the site is denoted by Dscore. Dscore values provide a rough estimate of whether the site is druggable. These scores were derived by Halgren (Halgren, 2009) by executing the SiteMap program on a number of proteins that have inhibitors bound with potencies in the sub-micromolar range and performing statistical analyses to produce optimized scores (Halgren, 2009). The OPLS-2005 force field (Jorgensen et al., 1996) was employed, and a standard grid was used with 15 site points per reported site and cropped at 4.0 Å from the nearest site point.

2.3 Ligand Selection and Preparation. The ligand structures were taken from SELLEKCHEM database (http://www.selleckchem.com/), the DrugBank database (https://www.drugbank.ca/) and the Repurposing hub (https://clue.io/repurposing) (see Scheme 1). The ligands were prepared using Schrödinger's (version 2017-1), LIGPREP (*LigPrep, version 4.2; Schrödinger, LLC: New York, 2017*, n.d.), which generates tautomers, and possible ionization states at the pH range 7 ± 2 using Epik (*Epik Version 4.0, Schrödinger, LLC, New York, NY, 2017*, n.d.) and also generates all the stereoisomers of the ligand if necessary. The optimization was done using the OPLS3 force field (Jorgensen et al., 1996). The pharmacokinetic descriptors of the hits were curated from the above-mentioned databases.

2.4 Virtual Screening of virtual libraries on M^{pro}. The crystal used in this study was bound with a peptide-like inhibitor which provided the primary site for ligand docking and also SiteMap confirmed the primary site as the most druggable site. The grid was generated using the centroid of bound inhibitor by using the Receptor Grid Generation panel in Glide. Docking studies were carried out using the Virtual Screening Workflow (VSW) of Glide Schrodinger Suite (Friesner et al., 2006; Mittal et al., 2020). The ligands chosen from the database were passed through three stages of the screening workflow starting from high-throughput screening (50% filtered), followed by standard precision (30%) and finally extra precision (10%) stages. The final poses were processed using the Prime MM/GBSA panel at the end (Schrodinger suite, LLC, New York, NY, 2016-3) (Scheme 1). The results of the docking were then quantified on the basis of docking scores and MM/GBSA ΔG_{bind}.

2.5 Figures. All the images were generated using VMD and Schrodinger Suite (Asthana et al., 2014, 2015; Humphrey et al., 1996).

3. Results and discussions

3.1 Comparative structural analysis: The availability of co-crystal 6LU7 in which a peptide (N3: a covalently-bonded inhibitor) is bound and knowledge of interacting residues with molecule 13b (α-ketoamide inhibitor) made the understanding of protein interesting. Since the M^{pro} is known to function in dimer state and compound 13b is reported to be bound in dimer state, the crystal structure (coordinates file for PDB-ID: 6Y2F) for the same is not available in RCSB Protein Data Bank so far. The residues reported for compound 13b involve the same residues as for N3 molecule except for residue S1 from chain B. Therefore, only monomer crystal was used for further analysis. The architecture of binding site with any possible conformational changes was concurred after superimposition of the crystal structures COM (6LU7) and APO (6M03). It shows that overall structure of protein is well aligned (RMSD APO-vs.-COM is 0.4 Å) except C-terminal region (**Figure 1B**). We also observed some noticeable differences in the binding site architecture of COM and APO (**Figure 1C**). It was found that all the residues of the binding site are well aligned with that of APO except the residues T25, M49, R188, T190, M165 and Q189 that shows side chain conformational changes, while P168 shows the backbone movement also (Figure 1C). The inhibitor forms the hydrogen bond (HB) interactions with residues G143, H163, E166, Q189, and T190.

3.2 Exploring the druggable binding sites, docking and benchmark setup for compound screening: To bring in more robustness in confirming the final binding site before performing the Virtual Screening, we performed ligand independent binding site search on M^{Pro}. The top score of the SiteMap program also confirms the co-crystal site as the primary binding site (marked as Site 1) with the highest D_score of 1.096 (best druggability score) (**Figure 2 and Table 1**). The volume of the pocket is 287.09 $Å^3$. This binding site of M^{pro} is encompassed by domain I and II and the loop region connecting domains II and II. Sitemap result shows that this pocket is relatively smaller in size with a size score of 120 (reference value: 130), and more exposed to solvent with an exposure value of 0.614 (reference value: 0.52) (Halgren, 2009). Furthermore, we performed the focused docking on the Site 1 with the N3 and 13b molecules to set up the benchmark for VS (Virtual screening) execution based on two main criteria i.e docking score and MM-GBSA values. The docking score and MM-GBSA values for N3 is -10.6 kcal/mol and -64.32 kcal/mol, respectively, while for 13b it is -8.34 kcal/mol and -70.9 kcal/mol, respectively.

3.3 Virtual screening of two different libraries

3.3.1 Protease library on the Site1 of M^{pro} protein: It has been a proven and successful strategy to inhibit the viral proteases for the treatment of viral infections such as in the cases of human immunodeficiency virus (HIV) and hepatitis C virus (HCV) and hence screening of protease inhibitors could be a useful approach against COVID-19 M^{pro} (Fischer et al., n.d.; Ghosh et al., 2016; Yang et al., 2006). With this logic, 227 protease molecules from Selleckchem were curated and prepared for screening on Site1. Based on our filtering criteria of optimum MM/GBSA and docking scores, the top 30 initial hits were filtered out (**Figure 3A**). The docking scores, MM/GBSA, pharmacokinetic descriptors, their known targets, along with the structures of the hits are shown in **Table 2**. However, some of these molecules violate the pharmacokinetic properties but they can be used as a starting point for the experimental validation and hits optimization that can facilitate the identification of more potent drug-like molecules. For all selected molecules (top hits), the residue mapping based on interaction pattern was executed. The top three molecules found in our study are Leupeptin hemisulfate, pepstatin A and nelfinavir on the basis of MM/GBSA and docking scores and their ligand interaction diagrams are shown in **Figure 4A-4C** respectively. However, few of the molecules are observed to be common in our screening and some in-silico studies published targeting M^{pro} and they are Nelfinavir, Lopinavir, indinavir, ritonavir, darunavir (Adem et al., n.d.; Contini, n.d.; Farag et al., n.d.).

3.3.2 FDA/repurposing library on the Site1 of M^{pro} protein: Another library of FDA/repurposing molecule curated from different sources such as repurposing hub, Selleckchem and DrugBank were used to prepare for screening purpose. FDA approved molecules are the fastest route to identify the active molecules. From the multi-step route of VS, we filtered out the best 41 hits (**Figure 3B**). The docking scores, MM/GBSA, pharmacokinetic descriptors, their known targets, along with the structures of the hits are shown in **Table 3**. From this library, the best three molecules are Birinapant, Lypressin and Octreotide and their ligand interaction diagrams are shown in Figure 5**A-5C** respectively. Nevertheless, these molecules violate the pharmacokinetic properties but could be optimised to drug-like properties based on the experimental results. The finding of some common top molecules in this study are also reported in other screening studies on the same target, claiming the robustness of our protocols. These common molecules are Lypressin Octreotide, Mitoxantrone, Hesperidin, Echanoside, Pralmorelin, Epicatechin, Diosmin, Flavitan (flavin adenine dinucleotide), Curcumin, Saquinavir, Montelukast, Baicalin, Thymopentin (Adem et al., n.d.; Contini, n.d.; Farag et al., n.d.).

3.4 Characterization of binding site: The residues H41, M49, N142, C145, H164, M165, E166, D187, R188 and Q189 having highest occupancy (>90%) for all the hits obtained from protease library (**Figure 6**) with respect to the control compound N3. The residues G143, E166, and Q189 are observed to be majorly involved in HBs (hydrogen bonds) interaction for most of the hits while H41 is observed to form pi-pi stacking and pi-cation interaction in some cases (**Figure 6**). Similarly, for the hits obtained from FDA/repurpose library, the residue interaction mapping showed that the residues H41, M49, N142, C145, M165, E166, R188, and Q189 have higher occupancy (>90%) in the binding site (**Figure 7**). The HBs analysis found that residues T26, G143, E166, and Q189 are highly involved in the hydrogen bond interaction (**Figure 7**). From the residue interaction mapping analysis of hits, we observed that the static binding site of M^{pro} is composed of 44.7% of polar residues, 42.1 % of non-polar residues, 10.5% of acidic residues and 2.6% of basic residues which encompasses a diverse class of molecules. This indicates that the binding site (site1) of M^{pro}

is hydrophilic in nature and solvent exposed which is in concordance with our site map results. However, the flexibility and adaptability of the pocket towards the ligand can be further explored by molecular dynamics simulation studies. Interestingly, we find H41, M49, N142, C145, H164, M165, E166 and Q189 are the most conserved residues and their occupancy is 97-100% with all our top hits. As reported by Wang J. (Wang, n.d.) that MERS, SARS and COVID-19 have four (H41, H163, M165 and Q189) common residues, while HCV NS3/4A and COVID-19 have only one common hotspot residue (H41), confirming our findings that the hits having interaction with these residues might have broad spectrum value.

3.5 Top rated molecules for in vitro analysis: We identified six molecules with a lower free energy of binding combined with a higher theoretical drug discovery value compared to the co-crystallized ligand N3 and 13b (**Figure 8**). From the interaction pattern and other analysis, the identified set of molecules should be considered as early hit molecules, albeit we accept that no experimentally supported hit-optimization was conducted. Some molecules have shown a minor decrease in their binding energies, also some of them have presented a substantial improvement in comparison to the co-crystallized ligand. Different studies based on virtual screening (Contini, n.d.), artificial intelligence (H. Zhang et al., n.d.) followed by docking, neural network (Zhou et al., 2020) etc to discovery SARS-CoV-2 inhibitors we observed that their docking scores ranging from -7.0 to -9.0 kcal/mol, however, Some of our molecules have shown substantially good docking energy as well as their free energy is also considerably higher side. Our identified antivirals interacted with the protease with at least two HBs with an average of over four (**Figures 4 and 5**). From structure-based approaches the knowledge of hot-spot residues is critical.

Furthermore, the identified molecules have shown relevant pharmacokinetic descriptors with logP values, MW, PSA and hydrogen bond donors/acceptors in the range described by Lipinski. Though we are aware that there are many FDA approved and successful candidates who are documented that they are not following the standard Lipinski rule. It is interesting that proposed inhibitors have shown their availability of molecules at commercial suppliers such as MCULE and Selleckchem. The identified top six molecules presented comparatively better docking and MM/GBSA scores with respect to the control molecules and therefore represent excellent candidates for further investigation in vitro (**Figure 8**).

4. Conclusion

As the cases increase day by day there is an extremely urgent need of the designing small compound or peptide drugs to cure the 2019-nCoV. A structure based virtual screening of different known and existing molecule libraries was performed to identify useful molecules that have potential to bind and could be repurposed against M^{pro} of COVID-19. The protein-ligand interaction analysis is suitable to overcome the challenge of screening thousands of drugs in a short time in certain emergency situations, such as covid-19 outbreak. Considering the current emergency situation for the need of effective drugs, our paper is more focused on drug efficacy and their availability from different sources as they can be used immediately skipping various time taking stages in the drug discovery and development pipeline. The emergence of highly infectious SARS-CoV-2 is only possible to combat if we target this deadly virus from different approaches as designed multiple treatment plans within a time frame.

Computational drug repurposing study is known to provide treatment options in a short period of time. Docking, ab-initio calculation, molecular modeling, needs CPU-GPU based computational resources to compute and identify the potential possible molecule in target dependent manner. We can run the drug repurposing screenings within four to five days using a reliable HVS strategy. Depending upon the size of the M^{pro} and the type of molecules it can accomodate the screening time could be even optimised for it. The availability of high-quality crystals of drug target structures is very important that can shed light into the mechanistic insights of protein function. A high-resolution crystal structure of COVID-19 protease in complex with N3 and knowledge of key residues from 13b molecules (crystal is not yet released) is available on time, allowing us to conduct this drug repurpose screening. In absence of high-quality structure, one can rely on homology modeling technique, probably with a reduced success rate of identifying repurposing drugs. We compared the crystal structure of proteases of other SARS with COVID-19 and found that it is close to SARS-COV-1.

In conclusion, we obtained some useful hit molecules from the structure-based virtual screening having a binding energy of > -6.0 kcal/mol and MM/GBSA > -32.0 kcal/mol with relevant ADME properties. Currently, antineoplastic, immunomodulators, nucleotide inhibitors, antimalarial, ribonucleoside inhibitors, steroid hormones, protease inhibitors, antiretrovirals are being evaluated in clinical trials against COVID-19. Our screening molecules also belong to the similar category of molecules. ***The hits found in our study belong to different chemical classes and have the potential to accommodate inside the pocket that can be further validated by experimental assays.*** Our interaction mapping has identified key interacting residues H41, M49, N142, C145, E166 and Q189 and out of which H41 and C145 belong to the catalytic dyad. These hits or their scaffolds can be used as a good starting point for developing novel hits by optimization or can be used in combination treatments. The investigation of solvation sites or the role of water molecules in the binding site could be carried out to explore the druggability of this pocket that might help in rational designing of molecules. The potential candidate hits provided in our paper can aid in facilitating the hunt of anti-covid-19 M^{pro} drug discovery. We believe that the structure-based virtual screening approach can play a vital role in designing anti-coronaviral drugs.

Acknowledgements

The work has been supported by the facilities and support provided by THSTI.

Author's Contribution

S. Asthana proposed and designed the study. M. Srivastava, L. Mittal, A. Kumari and M. Singh performed experiments and analysis of results. M. Srivastava, L. Mittal, A. Kumari and S. Asthana wrote the manuscript.

Conflicts of Interest

The authors declare that they do not have conflict of interest.

References:

Adem, S., Eyupoglu, V., Sarfraz, I., Rasul, A., & Ali, M. (n.d.). *Identification of Potent COVID-19 Main Protease (Mpro) Inhibitors from Natural Polyphenols: An in Silico Strategy Unveils a Hope against CORONA*. https://doi.org/10.20944/preprints202003.0333.v1

Ahmed, M., Dwivedy, A., Mariadasse, R., Tiwari, S., Jeyakanthan, J., & Biswal, B. K. (n.d.). *Prediction of small molecule inhibitors targeting the novel coronavirus (SARS-CoV-2) RNA-dependent RNA polymerase*. https://doi.org/10.31219/osf.io/fjnzc

Anand, K., Ziebuhr, J., Wadhwani, P., Mesters, J. R., & Hilgenfeld, R. (2003). *Coronavirus Main Proteinase (3CLpro) Structure: Basis for Design of anti-SARS Drugs*. https://doi.org/10.2210/pdb1p9t/pdb

Anang, S., Kaushik, N., Hingane, S., Kumari, A., Gupta, J., Asthana, S., Shalimar, Nayak, B., Ranjith-Kumar, C. T., & Surjit, M. (2018). Potent Inhibition of Hepatitis E Virus Release by a Cyclic Peptide Inhibitor of the Interaction between Viral Open Reading Frame 3 Protein and Host Tumor Susceptibility Gene 101. *Journal of Virology*, *92*(20). https://doi.org/10.1128/JVI.00684-18

Asthana, S., Shukla, S., Ruggerone, P., & Vargiu, A. V. (2014). Molecular mechanism of viral resistance to a potent non-nucleoside inhibitor unveiled by molecular simulations. *Biochemistry*, *53*(44), 6941–6953. https://doi.org/10.1021/bi500490z

Asthana, S., Zucca, P., Vargiu, A. V., Sanjust, E., Ruggerone, P., & Rescigno, A. (2015). Structure–Activity Relationship Study of Hydroxycoumarins and Mushroom Tyrosinase. In *Journal of Agricultural and Food Chemistry* (Vol. 63, Issue 32, pp. 7236–7244). https://doi.org/10.1021/acs.jafc.5b02636

Bzówka, M., Mitusińska, K., Raczyńska, A., Samol, A., Tuszyński, J., & Góra, A. (n.d.). *Molecular Dynamics Simulations Indicate the SARS-CoV-2 Mpro Is Not a Viable Target for Small-Molecule Inhibitors Design*. https://doi.org/10.1101/2020.02.27.968008

Cascella, M., Rajnik, M., Cuomo, A., Dulebohn, S. C., & Di Napoli, R. (2020). Features, Evaluation and Treatment Coronavirus (COVID-19). In *StatPearls*. StatPearls Publishing. https://www.ncbi.nlm.nih.gov/pubmed/32150360

Chen, H., Guo, J., Wang, C., Luo, F., Yu, X., Zhang, W., Li, J., Zhao, D., Xu, D., Gong, Q., Liao, J., Yang, H., Hou, W., & Zhang, Y. (2020). Clinical characteristics and intrauterine vertical transmission potential of COVID-19 infection in nine pregnant women: a retrospective review of medical records. *The Lancet*, *395*(10226), 809–815. https://doi.org/10.1016/S0140-6736(20)30360-3

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

Chloroquine/ Hydroxychloroquine Prevention of Coronavirus Disease (COVID-19) in the Healthcare Setting - Full Text View - ClinicalTrials.gov. (n.d.). Retrieved March 28, 2020, from https://clinicaltrials.gov/ct2/show/NCT04303507

Contini, A. (n.d.). *Virtual Screening of an FDA Approved Drugs Database on Two COVID-19 Coronavirus Proteins*. https://doi.org/10.26434/chemrxiv.11847381

Coronavirus. (n.d.). Retrieved March 31, 2020, from https://www.who.int/health-topics/coronavirus#tab=tab_1

Coronavirus Update (Live): 803,772 Cases and 39,070 Deaths from COVID-19 Virus Outbreak - Worldometer. (n.d.). Retrieved March 31, 2020, from https://www.worldometers.info/coronavirus/

Epik Version 4.0, Schrödinger, LLC, New York, NY, 2017. (n.d.).

Farag, A., Wang, P., Ahmed, M., & Sadek, H. (n.d.). *Identification of FDA Approved Drugs Targeting COVID-19 Virus by Structure-Based Drug Repositioning*. https://doi.org/10.26434/chemrxiv.12003930.v1

Fischer, A., Sellner, M., Neranjan, S., Lill, M. A., & Smieško, M. (n.d.). *Inhibitors for Novel Coronavirus Protease Identified by Virtual Screening of 687 Million Compounds*. https://doi.org/10.26434/chemrxiv.11923239.v1

Friesner, R. A., Murphy, R. B., Repasky, M. P., Frye, L. L., Greenwood, J. R., Halgren, T. A., Sanschagrin, P. C., & Mainz, D. T. (2006). Extra precision glide: docking and scoring incorporating a model of hydrophobic enclosure for protein-ligand complexes. *Journal of Medicinal Chemistry*, *49*(21), 6177–6196. https://doi.org/10.1021/jm051256o

Ghosh, A. K., Osswald, H. L., & Prato, G. (2016). Recent Progress in the Development of HIV-1 Protease Inhibitors for the Treatment of HIV/AIDS. *Journal of Medicinal Chemistry*, *59*(11), 5172–5208. https://doi.org/10.1021/acs.jmedchem.5b01697

Graham, R. L., Sparks, J. S., Eckerle, L. D., Sims, A. C., & Denison, M. R. (2008). SARS coronavirus replicase proteins in pathogenesis. In *Virus Research* (Vol. 133, Issue 1, pp. 88–100). https://doi.org/10.1016/j.virusres.2007.02.017

Halgren, T. A. (2009). Identifying and characterizing binding sites and assessing druggability. *Journal of Chemical Information and Modeling*, *49*(2), 377–389. https://doi.org/10.1021/ci800324m

Hatada, R., Okuwaki, K., Mochizuki, Y., Fukuzawa, K., Komeiji, Y., Okiyama, Y., & Tanaka, S. (n.d.). *Fragment Molecular Orbital Based Interaction Analyses on COVID-19 Main Protease - Inhibitor N3 Complex (PDB ID:6LU7)*. https://doi.org/10.26434/chemrxiv.11988120

Humphrey, W., Dalke, A., & Schulten, K. (1996). VMD: Visual molecular dynamics. *Journal of Molecular Graphics*, *14*(1), 33–38. https://doi.org/10.1016/0263-7855(96)00018-5

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, H. (n.d.). *Structure of Mpro from COVID-19 virus and discovery of its inhibitors*. https://doi.org/10.1101/2020.02.26.964882

Jorgensen, W. L., Maxwell, D. S., & Tirado-Rives, J. (1996). Development and Testing of the OPLS All-Atom Force Field on Conformational Energetics and Properties of Organic Liquids. In *Journal of the American Chemical Society* (Vol. 118, Issue 45, pp. 11225–11236). https://doi.org/10.1021/ja9621760

Kahn, J. S., & McIntosh, K. (2005). History and Recent Advances in Coronavirus Discovery. In *The Pediatric Infectious Disease Journal* (Vol. 24, Issue Supplement, pp. S223–S227). https://doi.org/10.1097/01.inf.0000188166.17324.60

LigPrep, version 4.2; Schrödinger, LLC: New York, 2017. (n.d.).

Lu, H. (2020). Drug treatment options for the 2019-new coronavirus (2019-nCoV). *Bioscience Trends*. https://doi.org/10.5582/bst.2020.01020

Maestro, version 2017; Schrödinger, LLC: New York, 2017. (n.d.).

Mattapally, S., Singh, M., Murthy, K. S., Asthana, S., & Banerjee, S. K. (2018). Computational modeling suggests impaired interactions between NKX2.5 and GATA4 in individuals carrying a novel pathogenic D16N NKX2.5 mutation. *Oncotarget*, *9*(17), 13713–13732. https://doi.org/10.18632/oncotarget.24459

Mittal, L., Kumari, A., Suri, C., Bhattacharya, S., & Asthana, S. (2020). Insights into structural dynamics of allosteric binding sites in HCV RNA-dependent RNA polymerase. *Journal of Biomolecular Structure & Dynamics*, *38*(6), 1612–1625. https://doi.org/10.1080/07391102.2019.1614480

Nair, D. t., & Narayanan, N. (n.d.). *Vitamin B12 may inhibit RNA-dependent-RNA polymerase activity of nsp12 from the COVID-19 Virus*. https://doi.org/10.35543/osf.io/p48fa

SiteMap, Version 4.2, Schrödinger, LLC, New York, NY, 2017. (n.d.).

Srivastava, M., Suri, C., Singh, M., Mathur, R., & Asthana, S. (2018). Molecular dynamics simulation reveals the possible druggable of USP7. *Oncotarget*, *9*(76), 34289–34305. https://doi.org/10.18632/oncotarget.26136

Su, S., Wong, G., Shi, W., Liu, J., Lai, A. C. K., Zhou, J., Liu, W., Bi, Y., & Gao, G. F. (2016). Epidemiology, Genetic Recombination, and Pathogenesis of Coronaviruses. In *Trends in Microbiology* (Vol. 24, Issue 6, pp. 490–502). https://doi.org/10.1016/j.tim.2016.03.003

The Molecular Biology of Coronaviruses. (2006). In *Advances in Virus Research* (Vol. 66, pp. 193–292). Academic Press. https://doi.org/10.1016/S0065-3527(06)66005-3

Tirumalaraju, D. (2020a, March 17). *First US clinical trial of Covid-19 vaccine candidate begins*. Clinical Trials Arena. https://www.clinicaltrialsarena.com/news/first-us-covid-19-vaccine-trial-moderna/

Tirumalaraju, D. (2020b, March 23). *China begins Phase I clinical trial of Covid-19 vaccine*. Clinical Trials Arena. https://www.clinicaltrialsarena.com/news/china-covid-19-vaccine-trial-begins/

Tirumalaraju, D. (2020c, March 26). *Pfizer reports safety data of azithromycin in Covid-19 trial*. Clinical Trials Arena. https://www.clinicaltrialsarena.com/news/pfizer-data-azithromycin-covid-19-trial/

Wang, J. (n.d.). *Fast Identification of Possible Drug Treatment of Coronavirus Disease -19 (COVID-19) Through Computational Drug Repurposing Study*. https://doi.org/10.26434/chemrxiv.11875446.v1

Xu, Z., Peng, C., Shi, Y., Zhu, Z., Mu, K., Wang, X., & Zhu, W. (n.d.). *Nelfinavir was predicted to be a potential inhibitor of 2019-nCov main protease by an integrative approach combining homology modelling, molecular docking and binding free energy calculation*. https://doi.org/10.1101/2020.01.27.921627

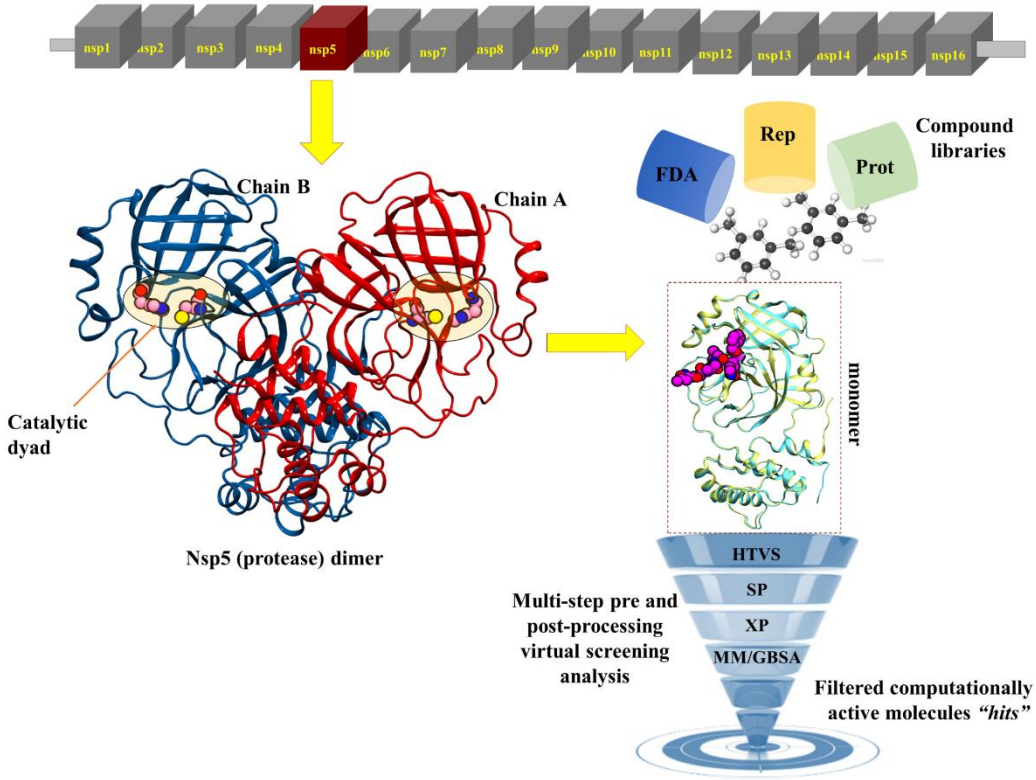
Yang, S., Chen, S.-J., Hsu, M.-F., Wu, J.-D., Tseng, C.-T. K., Liu, Y.-F., Chen, H.-C., Kuo, C.-W., Wu, C.-S., Chang, L.-W., Chen, W.-C., Liao, S.-Y., Chang, T.-Y., Hung, H.-H., Shr, H.-L., Liu, C.-Y., Huang, Y.-A., Chang, L.-Y., Hsu, J.-C., … Hsu, M.-C. (2006). Synthesis, Crystal Structure, Structure−Activity Relationships, and Antiviral Activity of a Potent SARS Coronavirus 3CL Protease Inhibitor. In *Journal of Medicinal Chemistry* (Vol. 49, Issue 16, pp. 4971–4980). https://doi.org/10.1021/jm0603926

Zhang, H., Saravanan, K. M., Yang, Y., Hossain, M. T., Li, J., Ren, X., & Wei, Y. (n.d.). *Deep Learning Based Drug Screening for Novel Coronavirus 2019-nCov*. https://doi.org/10.20944/preprints202002.0061.v1

Zhang, L., Lin, D., Sun, X., Curth, U., Drosten, C., Sauerhering, L., Becker, S., Rox, K., & Hilgenfeld, R. (2020). Crystal structure of SARS-CoV-2 main protease provides a basis for design of improved α-ketoamide inhibitors. *Science*. https://doi.org/10.1126/science.abb3405

Zhavoronkov, A., Aladinskiy, V., Zhebrak, A., Zagribelnyy, B., Terentiev, V., Bezrukov, D. S., Polykovskiy, D., Shayakhmetov, R., Filimonov, A., Orekhov, P., Yan, Y., Popova, O., Vanhaelen, Q., Aliper, A., & Ivanenkov, Y. (n.d.). *Potential 2019-nCoV 3C-like Protease Inhibitors Designed Using Generative Deep Learning Approaches*. https://doi.org/10.26434/chemrxiv.11829102.v1

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



Scheme1: The schematic representation of non-structural proteins of SARS-CoV-2. The targeted protein Mpro (main protease) is shown in red block. The modelled dimer was created. The crystal monomer 6LU7 was used for virtual screening and other post-processing analysis. **Rep** and **Prot** are repurpose and protease libraries.

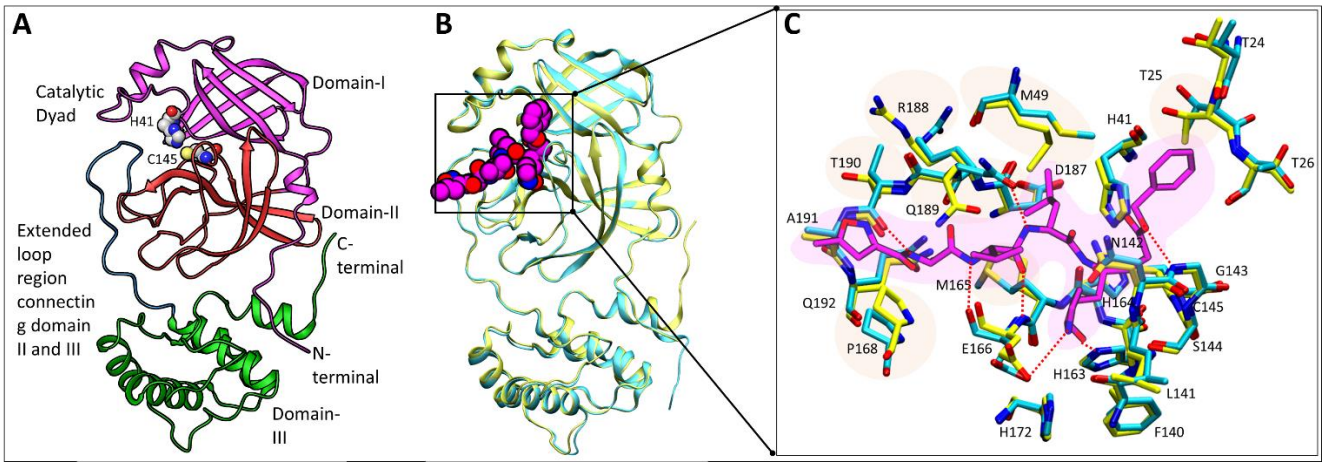


Figure 1: Assessment of APO and COM structures of SARS-COV-2. (A) Overview of the APO structure (PDB-ID: 6M03), (B) Superimposition of APO (Yellow) and COM (Cyan) structure with peptide-like inhibitor represented in VdW and coloured in Purple, (C) Binding site alignment highlights the conformational difference in the residues. The residues are shown in licorice representation along with inhibitors shown in purple.

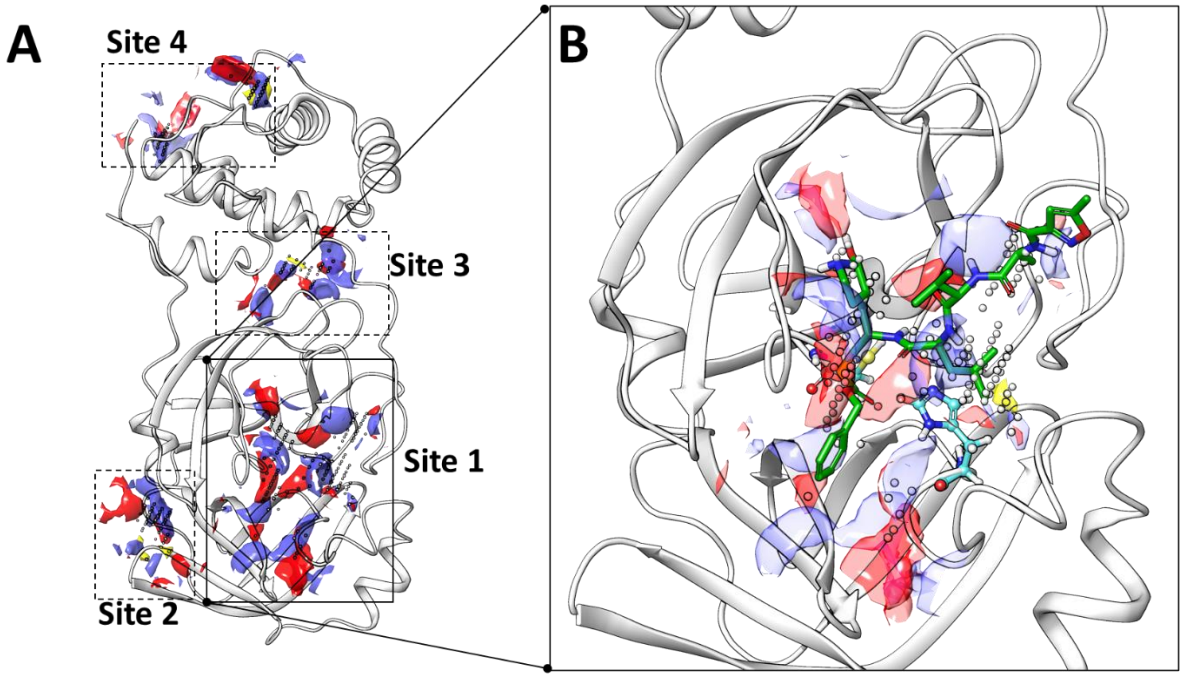


Figure 2. Possible binding sites and poses found by SiteMap. The yellow, red, and blue regions indicating the hydrophobic, ligand acceptor and ligand donor sites, respectively.

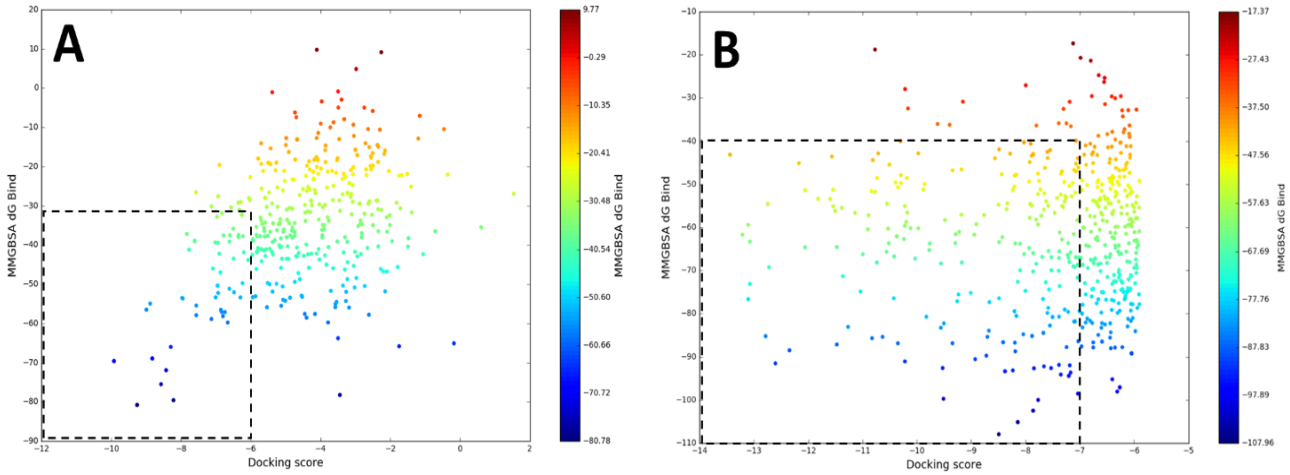


Figure 3: Scatter plot to calculate the docking scores (x-axis) and predicted MM-GBSA binding free energies (kcal/mol) (y-axis) for hits obtained from virtual screening of (A) protease library (B) FDA/repurposed molecules library. The points are coloured by MM-GBSA ΔG values. In plot (A) the total conformers obtained after docking from 227 molecules protease library: 399 and in plot (B) total conformers reported after virtual screening from 13947 FDA/repurposed library: 531

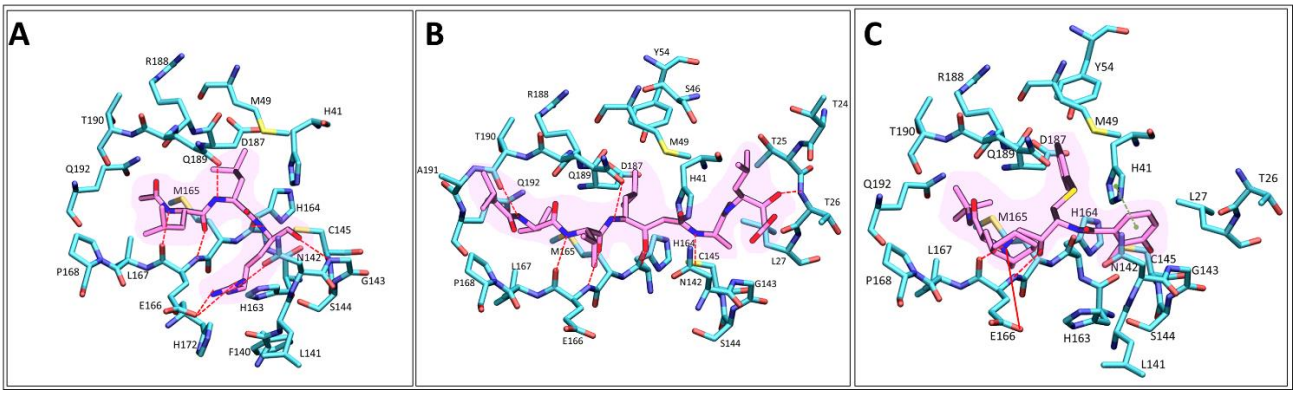


Figure 4: (A-C) the docked poses of Leupeptin Hemisulfate, Pepstatin A, Nelfinavir respectively within the binding site. Residues lining the pocket (Cyan) under 4.0 Å and its respective inhibitors (purple) are shown in licorice representation. Red dotted line shows the hydrogen bond interaction.

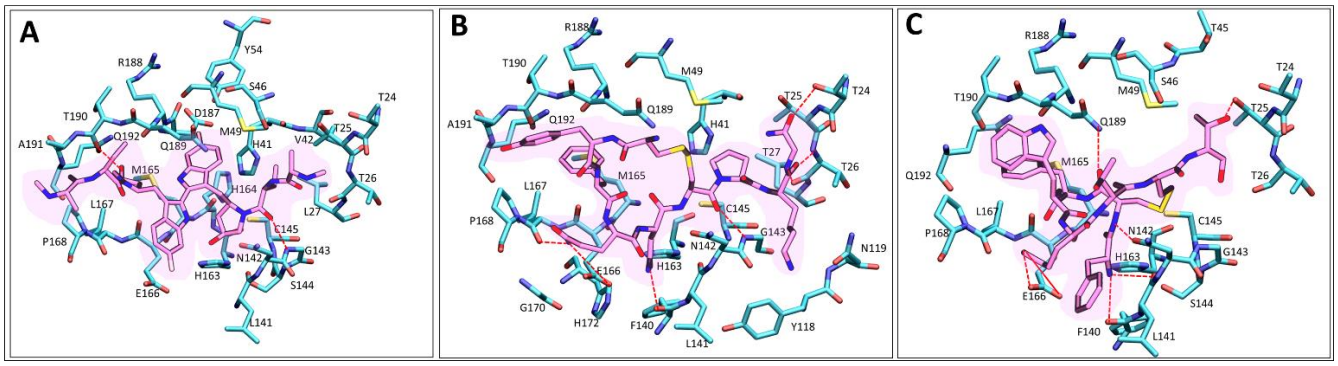


Figure 5: (A-C) The docked poses of Birinapant, Lypressin, Octreotide respectively within the binding site. Residues lining the pocket (cyan) under 4.0 Å and its respective inhibitors (purple) are shown in licorice representation. Red dotted line shows the hydrogen bond interactions.

	T24	T25	T26	L27	H41	V42	C44	T45	S46	M49	L50	P52	Y54	S139	F140	L141	N142	G143	S144	C145	H163	H164	M165	E166	L167	P168	T169	G170	H172	F181	V186	D187	R188	Q189	T190	A191	Q192
6LU7-lig	1	1	1	1	1	0	0	0	0	1	0	0	1	0	1	1	1	1	1	1	1	1	1	1	1	1	0	0	1	0	0	1	1	1	1	1	1
Leupeptin Hemisulfate	0	0	0	0	1	0	0	0	0	1	0	0	1	0	1	1	1	1	1	1	1	1	1	1	1	1	0	0	1	0	0	1	1	1	1	0	1
Pepstatin A	1	1	1	1	1	0	0	1	1	1	0	0	1	0	0	0	1	1	1	1	0	1	1	1	1	1	0	0	0	0	0	1	1	1	1	1	1
Nelfinavir	0	0	1	1	1	0	1	0	0	1	0	1	1	0	0	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	1	1	1	1	1	1
Bortezomib	0	0	0	0	1	0	0	0	0	1	0	1	1	0	1	1	1	1	1	1	1	1	1	1	1	0	0	0	1	0	0	1	1	1	1	0	0
Ixazomib	0	0	0	0	1	0	1	0	0	1	0	1	1	0	1	1	1	1	1	1	1	1	1	1	0	0	0	0	1	0	0	1	1	1	0	0	0
MG-101	0	0	0	0	1	0	0	0	0	1	0	0	1	0	1	1	1	1	1	1	1	1	1	1	1	1	0	0	1	0	0	1	1	1	1	1	1
Carfilzomib	0	1	1	1	1	0	0	1	1	1	0	0	1	0	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	0	0	1	1	1	1	1	1
L-685,458	0	1	1	1	1	0	0	0	0	1	0	0	1	0	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	0	0	1	1	1	1	1	1
Calpeptin	0	0	0	0	1	0	0	0	0	1	0	0	1	0	1	1	1	1	1	1	1	1	1	1	1	1	0	0	1	0	0	1	1	1	1	1	1
Z-FA-FMK	0	0	0	0	1	0	0	0	0	1	0	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	0	0	1	0	0	1	1	1	1	1	1
Atazanavir	0	0	0	0	1	0	1	0	0	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	0	0	1	1	1	1	0	1
ITF2357	0	0	0	0	1	0	0	0	0	1	0	0	0	0	1	1	1	0	1	1	1	1	1	1	1	1	0	0	1	0	0	1	1	1	1	0	1
Indinavir	0	0	0	0	1	0	0	0	0	1	0	0	1	0	1	1	1	0	0	1	0	1	1	1	1	1	1	1	1	0	0	1	1	1	1	1	1
LAQ824	0	1	1	1	1	0	0	0	0	1	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	0	0	1	0	0	1	1	1	1	1	1
Anagliptin	0	1	0	1	1	0	1	0	0	1	0	1	1	0	1	1	1	1	1	1	1	1	1	1	0	0	0	0	1	0	0	1	1	1	0	0	0
Aloin	0	0	0	0	1	0	0	0	0	1	0	0	0	0	1	1	1	1	1	1	1	1	1	1	0	0	0	0	1	0	0	1	1	1	0	0	1
LY411575	0	0	0	0	1	0	0	0	0	1	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	0	0	1	0	0	1	1	1	1	1	1
RG2833	0	0	0	1	1	0	0	0	0	1	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	0	0	1	0	0	0	0	1	1	1	1
Ritonavir	0	1	1	1	1	0	1	0	0	1	0	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	0	0	1	0	0	1	1	1	1	1	1
E-64	0	0	0	0	1	0	0	0	0	1	0	0	1	0	1	1	1	1	1	1	1	1	1	1	0	0	0	0	1	0	0	1	1	1	1	0	0
GI254023X	0	0	0	0	1	0	0	0	0	1	0	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	1	1	1	1	0	0
Isorhamnetin	0	0	0	0	1	0	0	0	0	1	0	0	1	0	0	0	0	0	0	1	0	1	1	1	1	1	0	0	0	0	0	1	1	1	1	1	1
TAPI-1	0	1	1	1	1	0	0	0	0	1	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	0	0	1	0	0	1	1	1	1	0	1
Trelagliptin	0	0	0	0	1	0	1	0	0	1	0	0	1	0	0	0	0	0	0	0	0	1	1	1	1	1	0	0	0	1	0	1	1	1	1	0	1
Amprenavir	0	1	1	1	1	1	1	0	0	1	0	1	1	0	1	1	1	1	1	1	1	1	1	1	1	0	0	0	1	0	0	1	1	1	1	0	0
Lopinavir	1	1	1	1	1	1	0	0	0	1	0	0	1	0	1	1	1	1	1	1	1	1	1	1	1	1	0	0	1	0	0	1	1	1	1	1	1
Delanzomib	0	1	1	1	1	0	0	0	0	1	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	0	0	1	0	0	1	1	1	1	1	1
Telaprivir	0	1	1	1	1	0	0	0	0	1	0	1	1	0	1	1	1	1	1	1	1	1	1	1	1	0	0	1	1	0	0	1	1	1	1	1	1
Lycorine	0	0	0	0	1	0	0	0	0	1	0	0	1	0	0	1	1	0	1	1	1	1	1	1	1	0	0	0	0	0	0	1	1	1	0	0	0
Darunavir	0	1	1	1	1	0	1	0	0	1	0	1	1	0	0	0	1	1	0	1	0	1	1	1	1	1	0	0	0	0	0	1	1	1	1	0	1
Occupancy (%)	9.7	42	42	48	100	6.5	26	6.5	6.5	100	0	35	77	3.23	80.6	87.1	93.5	83.9	87.1	96.8	83.9	100	100	100	87.1	74.2	3.23	16.1	77.4	3.23	0	96.8	96.8	100	87.1	54.8	77.4

Hydrogen bond Salt bridge Hydrogen bond + salt bridge Pi-pi stacking Pi-cation

Figure 6: Residue sampling for hits obtained from protease library at site 1 residues involved in the interaction site are within 4.0 Å. The presence and absence of residues are marked as 1 and 0 respectively. The residues are coloured on the basis of polar (blue), hydrophobic (green), acidic (red) and basic (purple) properties. The residues which are showing > 90% occupancy are considered as key residues.

	T24	T25	T26	L27	H41	V42	C44	T45	S46	E47	M49	L50	P52	Y54	Y118	N119	S139	F140	L141	N142	G143	S144	C145	H163	H164	M165	E166	L167	P168	T169	G170	H172	D177	V186	D187	R188	Q189	T190	A191	Q192
6LU7-lig	1	1	1	1	1	0	0	0	0	0	1	0	0	1	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	0	0	1	0	0	1	1	1	1	1	1
Birinapant	0	1	1	1	1	1	0	0	1	0	1	0	0	1	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	1	1	1	1	1	1
Lypressin	1	1	1	1	1	0	0	0	0	0	1	0	0	0	1	1	0	1	1	1	1	1	1	1	0	1	1	1	1	0	1	1	0	0	0	1	1	1	1	1
Octreotide	1	1	1	0	0	0	1	1	1	0	1	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	0	0	1	0	0	0	1	1	1	1	1
PD-173212	1	1	1	1	1	0	0	0	0	0	1	0	1	1	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	1	1	1	1	1	1
Mitoxantrone	0	1	1	1	1	1	0	0	0	0	1	0	0	0	0	0	0	1	1	1	1	1	1	1	0	1	1	1	1	0	0	1	0	0	0	1	1	1	1	1
Dipyridamole	0	1	1	1	1	1	1	0	0	0	1	0	1	1	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	0	0	1	0	0	1	1	1	1	0	1
Amikacin	0	0	0	0	1	0	0	0	0	0	1	0	0	1	0	0	0	1	1	1	0	1	1	1	1	1	1	1	1	0	0	1	0	0	1	1	1	1	1	1
Hesperidin	0	1	1	1	1	1	1	0	0	0	1	0	1	1	0	0	0	0	1	1	1	0	1	0	0	1	1	1	1	0	0	0	0	0	1	1	1	1	1	0
Echinacoside	1	1	1	1	1	0	0	0	1	0	1	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	0	0	1	0	0	0	1	1	1	1	1
Ginsenoside Rg1	0	1	1	1	1	0	0	0	1	1	1	1	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	1	1	0	0
Pralmorelin	0	1	1	1	1	1	0	0	0	0	1	0	0	1	0	0	0	1	1	1	1	0	1	1	0	1	1	1	1	0	1	1	0	0	1	1	1	1	1	1
PSI-7976	1	1	1	1	1	0	1	0	1	0	1	0	1	1	0	0	0	1	1	1	1	1	1	1	1	1	1	0	0	0	0	1	1	0	0	1	1	0	0	0
Boceprevir	0	0	0	0	1	0	0	0	0	0	1	0	0	1	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	1	1	1	1	1	1
Acteoside	1	1	1	1	1	0	0	0	1	0	1	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	0	0	1	0	1	0	1	1	1	1	1
TMC353121	0	0	0	0	1	0	0	0	0	0	1	0	0	1	0	0	1	1	1	0	0	1	1	1	1	1	1	1	1	0	1	1	0	0	1	1	1	1	1	1
TXA127	0	1	1	1	1	0	0	0	0	0	1	0	1	1	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	1	1	1	1	1	1
Evista	0	1	1	1	1	0	0	0	0	0	1	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1	0	0	0	0	1	0	0	1	1	1	0	0	0
Sunitinib	0	0	0	0	1	0	1	0	0	0	1	0	1	1	0	0	0	0	0	0	0	0	1	0	1	1	1	1	1	1	1	0	0	0	1	1	1	1	1	1
Rutin	0	1	1	1	1	0	1	0	0	0	1	0	1	1	0	0	0	1	1	1	1	0	1	0	1	1	1	1	1	0	0	0	0	0	1	1	1	1	1	1
MM-102	0	1	1	1	1	0	1	0	0	0	1	1	1	1	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	0	1	0	0	0	1	1	1	1	1	1
Felypressin	0	1	1	1	1	1	0	0	0	0	1	1	0	1	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	1	1	1	1	1	1
Epicatechin	0	1	1	1	1	1	0	0	0	0	1	0	0	1	0	0	0	1	1	1	1	1	1	1	1	1	1	0	0	0	0	1	0	0	1	1	1	0	0	0
Quercetin 3-rutinoside	0	1	1	1	1	0	1	0	0	0	1	0	1	1	0	0	0	0	1	1	1	0	1	0	1	1	1	1	1	0	0	0	0	0	1	1	1	1	1	1
Lanreotide	0	1	1	1	1	1	0	0	1	0	1	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	1	1	1	1	1
Diosmin	1	1	1	1	1	0	0	0	1	0	1	0	0	0	0	0	0	0	0	1	1	1	1	0	0	1	1	1	1	0	0	0	0	0	1	1	1	1	1	1
KD025	0	1	1	1	1	0	1	0	1	0	1	0	1	1	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	1	1	1	1	1	1
Cefoperazone	1	1	1	1	1	0	1	1	1	0	1	0	1	1	0	0	0	0	0	1	1	0	1	1	1	1	1	1	1	0	0	0	0	0	1	1	1	1	1	1
Pilaralisib	0	0	0	0	1	0	1	0	0	0	1	0	0	1	0	0	0	1	1	1	0	1	1	1	0	1	1	1	1	0	0	1	0	0	1	1	1	1	1	1
Troxerutin	0	1	1	1	1	0	0	0	0	0	1	1	0	0	0	0	0	1	1	1	1	0	1	1	0	1	1	1	1	0	1	0	0	0	1	1	1	1	1	1
Vitexin	0	1	1	1	1	1	0	0	0	0	1	0	0	0	0	0	0	1	1	1	1	1	1	1	0	1	1	0	0	0	0	1	0	0	1	1	1	0	0	0
Tenofovir alafenamide	0	1	1	1	1	0	0	0	0	0	1	0	0	1	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	0	0	1	0	0	1	1	1	1	1	1
Argatroban Monohydrate	0	1	1	1	1	1	1	0	0	0	1	0	1	1	0	0	0	1	1	1	1	1	1	1	1	1	1	0	0	0	0	1	0	0	1	1	1	1	0	0
Flavitan	1	1	1	0	1	0	0	1	1	0	1	0	0	0	0	1	0	0	0	1	1	0	0	0	1	1	1	1	1	0	0	0	0	0	1	1	1	1	1	1
Curcumin	0	1	1	1	1	1	1	0	0	0	1	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	1	1	1	1	1	1
Sequinavir	0	1	1	1	1	0	0	0	0	0	1	0	0	1	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	1	1	1	1	1	1
Montelukast	0	1	1	1	1	0	0	0	0	0	1	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	1	1	1	1	1	1
Dapagliflozin	0	0	0	0	1	0	0	0	0	0	1	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	0	0	1	0	0	1	1	1	1	0	1
Baicalein	0	1	1	1	1	1	1	0	0	0	1	0	1	1	0	0	0	0	0	1	1	0	1	0	1	1	0	0	0	0	0	0	0	0	1	1	1	0	0	0
Pitavastatin	0	1	1	1	1	1	1	0	0	0	1	0	1	1	0	0	0	0	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	1	1	1	1	0	0
Thymopentin	0	0	0	0	1	0	1	0	0	0	1	0	1	1	0	0	0	1	1	1	0	0	1	0	1	1	1	1	1	1	1	1	0	0	1	1	1	1	0	0
Occupancy (%)	24	83	83	78	98	32	39	7.3	27	2.4	100	9.8	37	63	2.44	4.88	2.44	73.2	87.8	95.1	87.8	75.6	97.6	80.5	80.5	100	97.6	80.5	80.5	12.2	24.39	51.2	2.44	2.44	80.5	97.6	100	87.8	73.2	75.61

Hydrogen bond Salt bridge Hydrogen bond + salt bridge Pi-pi stacking

Figure 7: Residue sampling for hits obtained from FDA and Repurposing library at site 1 residue involved in the interaction site are within 4.0 Å. The presence and absence of residue are marked as 1 and 0 respectively. The residues are coloured on the basis of polar (blue), hydrophobic (green), acidic (red) and basic (purple) properties. The residues which are showing > 90% occupancy are considered as key residues.

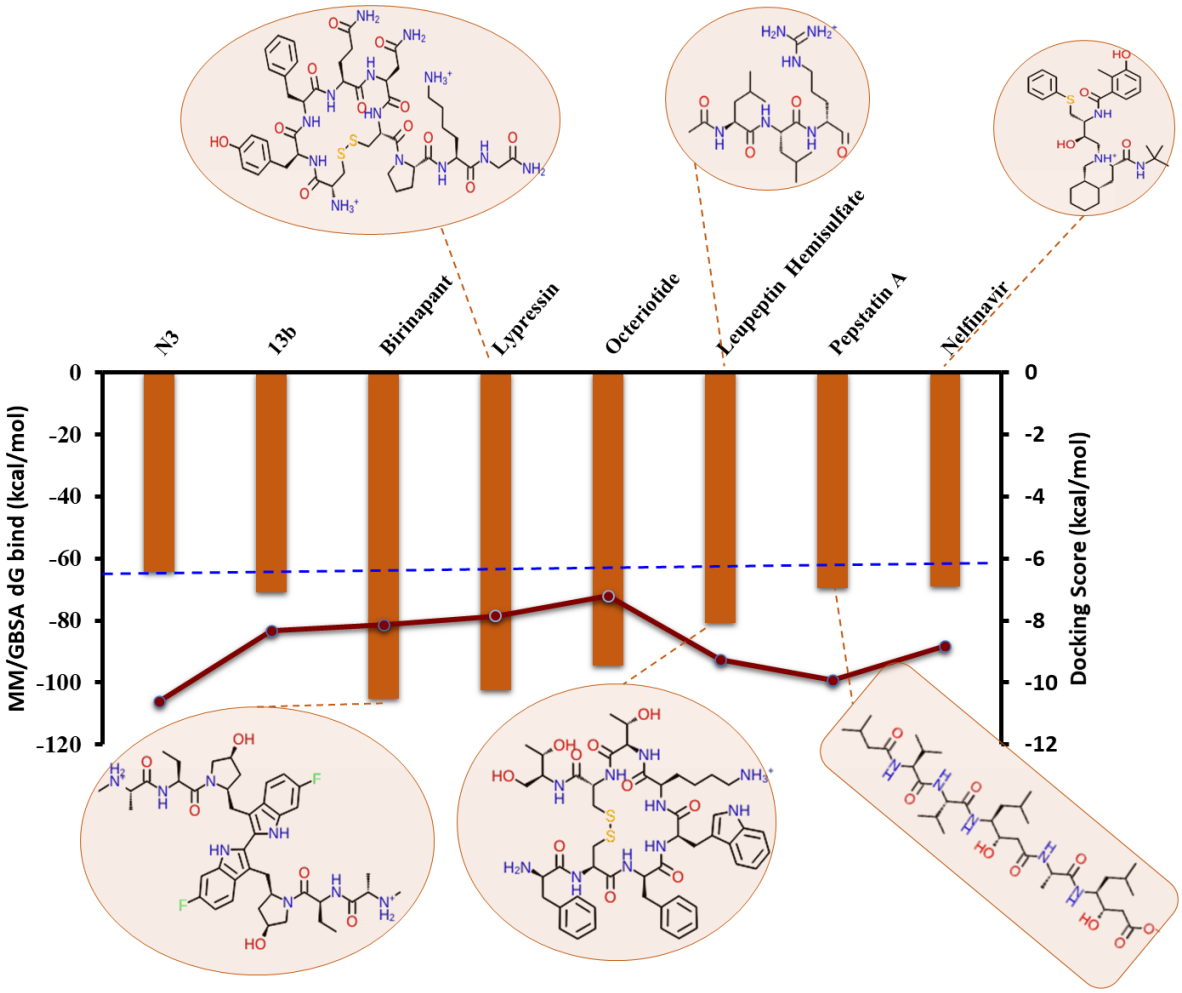
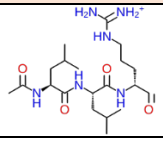
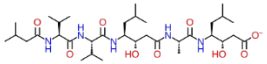


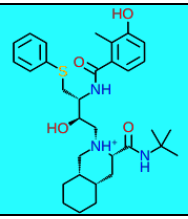
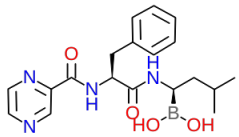
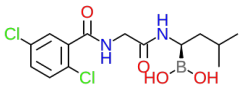
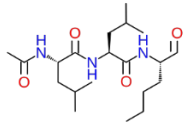
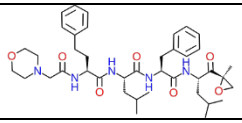
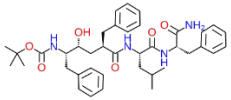
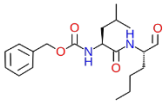
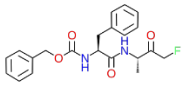
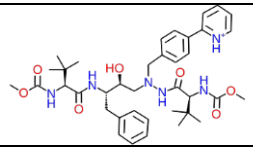
Figure 8: A plot of the MM/GBSA values (primary y-axis) and docking score (secondary y-axis) for the control molecules and 6 finalized potential hits along with their 2D structures.

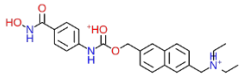
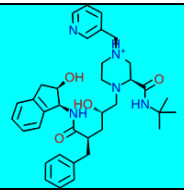
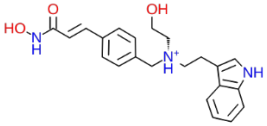
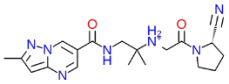
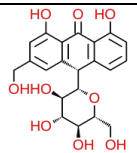
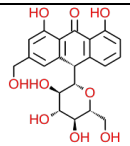
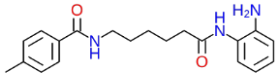
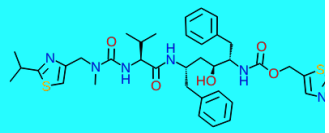
Table 1: SiteMap analysis on M^{pro} monomer.

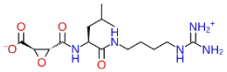
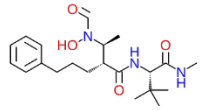
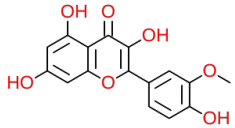
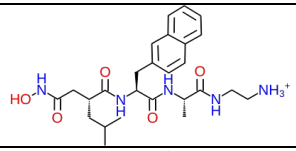
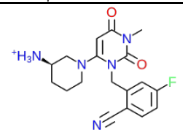
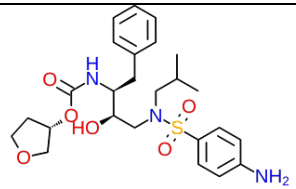
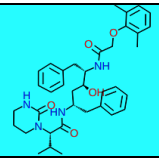
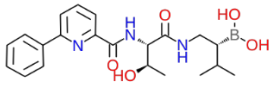
Title	SiteScore	size	D_score	Volume (Å^3)	exposure	enclosure	contact	phobic	philic	balance	don/acc
site_1	1.02	120	1.09	287.09	0.61	0.65	0.87	1.21	0.71	1.70	0.86
site_2	0.64	41	0.59	116.62	0.76	0.55	0.78	0.27	1.03	0.27	0.60
site_3	0.65	30	0.45	106.33	0.63	0.68	1.00	0.17	1.44	0.12	0.56
site_4	0.61	25	0.56	73.75	0.75	0.59	0.72	0.73	0.79	0.91	4.17

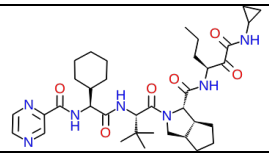
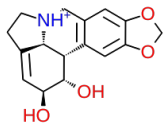
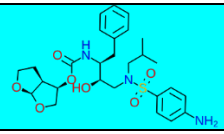
Table 2: Hits selected from proteases library along with their pharmacokinetic parameters and structures.

Name	Docking score	MMGBSA ΔG Bind	[a]Mol Weight (g/mol)	Target	Developmental Phase	[b]HBA	[c]HBD	Rotatable bonds	[d]PSA (Å^2)	[e]logP	[f]logS	Structure
Leupeptin Hemisulfate	-9.257	-80.784	426.554	Serine Protease	Experimental	5	6	18	166.27	0.58	-3.72	
Pepstatin A	-9.919	-69.603	685.892	HIV Protease	Experimental	9	8	22	223	2.46	-4.19	

Nelfinavir	-8.822	-68.943	567.8	HIV Protease	Approved	5	4	10	101.9	4.61	-5.5	
Bortezomib (Velcade)	-8.291	-65.989	384.237	Proteasome	Approved	6	4	9	124.44	0.89	-3.9	
Ixazomib (MLN2238)	-6.658	-59.781	361.029	Proteasome	Approved, Investigational	4	4	7	98.66	2.57	-4.5	
MG-101 (ALLN) (calpain)	-7.126	-58.882	383.525	Cysteine Protease	Investigational	4	3	13	104	2.62	-3.07	
Carfilzomib (PR-171)	-6.795	-58.155	719.91	Proteasome	Approved, Investigational	8	4	20	158.47	4.2	-5.2	
L-685,458	-7.56	-57.905	672.853	Gamma-secretase	Experimental	6	5	22	159.85	4.76	-6.57	
Calpeptin	-6.74	-57.113	362.463	Cysteine Protease	Investigational	4	2	12	84.5	3.23	-3.77	
Z-FA-FMK	-6.829	-57.077	386.417	Cysteine Protease	Experimental	5	2	12	84.5	2.46	-3.3	
Atazanavir	-6.853	-56.572	704.855	HIV Protease	Approved, Investigational	7	5	18	171.22	4.54	-5.3	

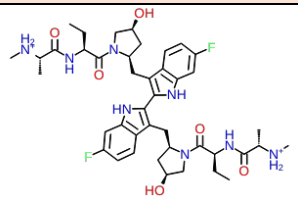
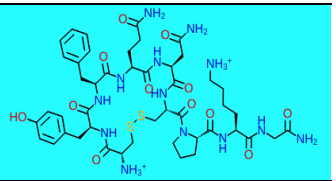
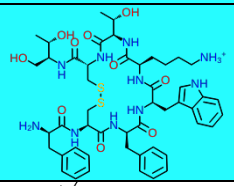
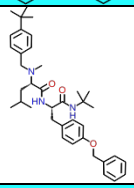
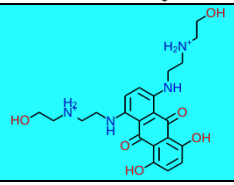
ITF2357 (Givinostat)	-8.991	-56.502	475.965	HDAC	Investigational	5	3	9	90.9	3.51	-4.9	
Indinavir	-7.946	-53.636	613.8	HIV Protease	Approved	7	4	12	118.03	2.81	-4.1	
LAQ824 (Dacinostat)	-7.965	-53.51	379.452	HDAC	Phase I	4	4	10	88.59	2.52	-3.42	
Anagliptin	-6.295	-53.377	383.447	DPP-4	Investigational	6	2	6	115.42	-0.54	-3.3	
Aloin (Barbaloin)	-6.933	-52.231	418.394	Tyrosinase	Withdrawn	9	7	3	167.91	-0.14	-2.46	
LY411575	-6.477	-52.07	479.475	Gamma-secretase	Experimental	6	3	5	98.7	2.99	-4.63	
RG2833 (RGFP109)	-6.61	-51.173	339.431	HDAC	Experimental	3	3	8	84.2	3.04	-3.38	
Ritonavir	-6.949	-50.393	720.944	HIV Protease	Approved	6	4	18	145.78	5.22	-5.8	

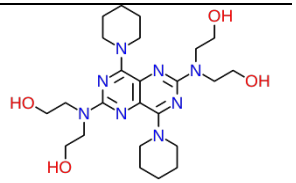
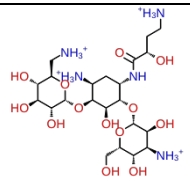
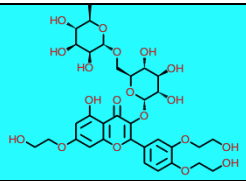
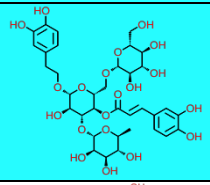
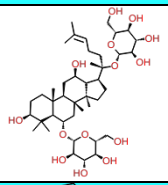
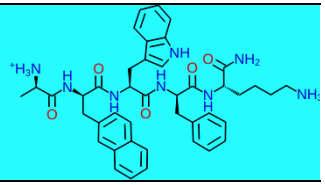
E-64	-7.085	-46.011	357.405	Cysteine Protease	Experimental	6	5	11	172	-0.43	-0.8	
GI254023X	-6.492	-43.484	391.504	Immunology & Inflammation Related	NF*	4	3	10	98.7	2.65	-3.41	
Isorhamnetin	-6.901	-42.236	316.262	tyrosinase inhibitor	Experimental	7	4	2	116	1.65	-3.36	
TAPI-1	-7.768	-41.196	499.602	TNF-alpha protease inhibitor I	Experimental	6	6	13	163	1.75	-2.87	
Trelagliptin	-6.936	-39.477	357.382	DPP-4	Investigational	5	1	3	93.67	1.3	-3.2	
Amprenavir (Agenerase)	-6.61	-39.138	505.627	HIV Protease	Approved	6	3	11	131.19	2.43	-4	
Lopinavir (ABT-378)	-7.607	-38.997	628.801	HIV Protease	Approved	5	4	15	120	4.69	-5.5	
Delanzomib (CEP-18770)	-6.352	-37.975	413.275	Proteasome	Investigational	6	5	9	131.78	2.57	-3.8	

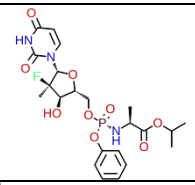
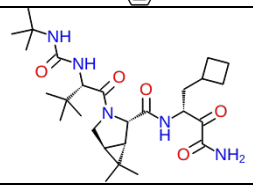
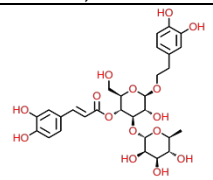
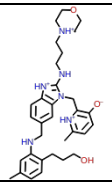
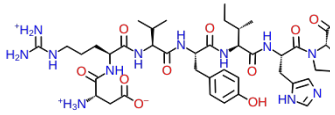
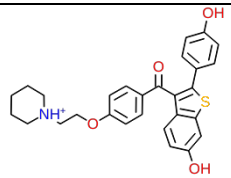
Telaprivir	-6.78	-37.58	679.85	HCV proteases	Approved, Withdrawn	8	4	14	179.56	2.58	-4.3	
Lycorine	-6.502	-35.836	287.31	HCV proteases	Experimental	5	3	0	62.2	0.8	-2.54	
Darunavir	-6.9	-32.78	547.66	HIV Protease	Approved	7	3	11	140.42	2.82	-3.9	

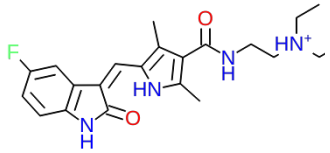
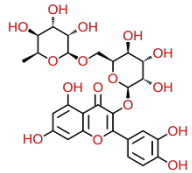
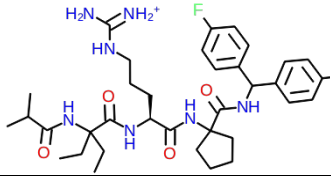
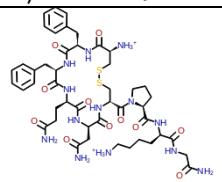
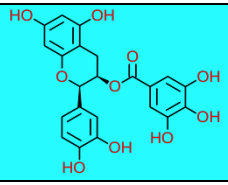
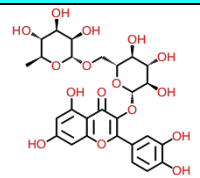
The general recommended ranges are as follows:
[a]Molecular weight, <500.
[b]Number of hydrogen bond acceptors, <10.
[c]Number of hydrogen bond donors, <5.
[d]Polar surface area, <140 $Å^2$,
[d]Predicted octanol/water partition coefficient, -0.4 to +5.6.
[e]Predicted aqueous solubility, <-5.0.
*NF, not found

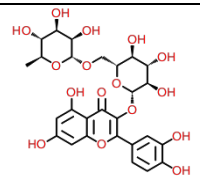
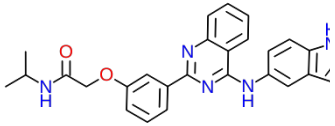
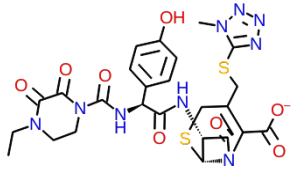
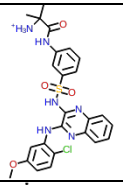
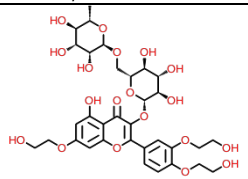
Table 3: Hits selected from proteases library along with their pharmacokinetic parameters and structures.

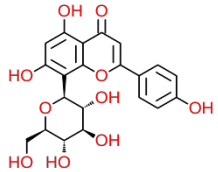
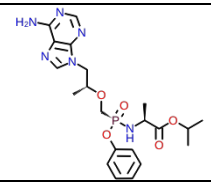
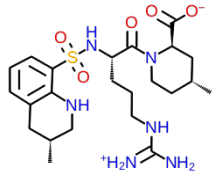
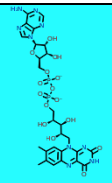
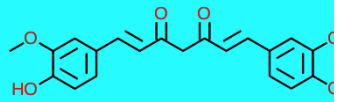
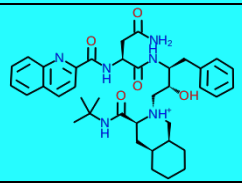
Title	Docking score	MMGBSA ΔG Bind	Phase	Source	[a]MW	[b]RB	[c]PSA (Å2)	[d]HBD	[e]HBA	[f]logP	[g]logS	Pharmacological activity	Structures
Birinapant	-8.141	-105.15	Approved	drugbank	806.9	15	194.2	8	10	2.1	-4.6	peptidomimetic activator of SMAC and inhibitor of IAP; potential antineoplastic activity.	
Lypressin	-7.859	-102.499	Approved	drugbank	1056.2	19	452.9	16	24	2.2	-4.3	Antidiuretic hormone	
Octreotide	-7.202	-94.415	Approved	drugbank	1019.2	17	332.3	13	12	0.4	-4.9	potent inhibitor of growth hormone, glucagon, and insulin	
PD-173212	-7.168	-93.65	pre-clinical	repurphub	599.85	15	65.72	1	7	7.7	-9.08	calcium channel blocker	
Mitoxantrone	-8.232	-93.159	Approved	drugbank	444.5	12	163.1	8	10	0.9	-2.8	antineoplastic activity	

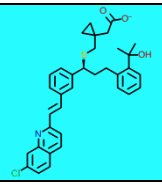
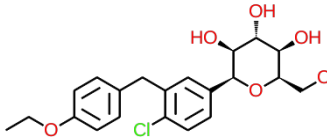
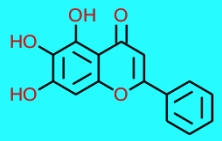
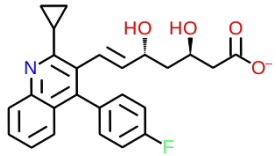
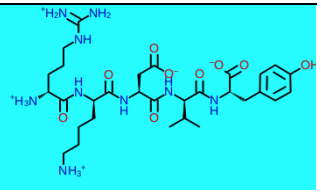
Dipyridamole (Persantine)	-7.184	-91.983	Approved	drugbank	504.6	12	145.4	4	12	1.5	-2.7	nucleoside transport inhibitor and a PDE3 inhibitor; inhibits blood clot formation	
Amikacin	-8.187	-88.745	Approved	drugbank	585.607	10	331.9	13	17	-8.6	-1.1	aminoglycoside antibacterial agent	
Hesperidin	-12.344	-88.5	Approved	drugbank	610.5	7	234.2	8	15	-0.2	-2.4	neurological conditions, antioxidant and anti-inflammatory effects	
Echinacoside	-11.473	-87.149	investigational	drugbank	786.7	14	324.4	12	19	-0.9	-2.3	Treatment of Neurological and other Disorders	
Ginsenoside Rg1	-10.349	-86.87	approved	repurposing hub	444.7	4	40.4	2	2	6.3	-6.4	antitumor, anti-inflammatory, antioxidation, and inhibition of cell apoptosis.	
Pralmorelin (GHRP-2)	-9.344	-86.865	phase II	repurposing hub	818	21	256	9	8	2.4	-5	Acts to endogenously increase growth hormone release from the pituitary.	

PSI-7976	-8.19	-86.73	pre-clinical	repurp hub	529.4 5	10	170.8 45	3	14	1.07	-3.71	HCV inhibitor	
Boceprevir	-7.028	-86.398	Approved, withdrawn	drugba nk	519.6	10	150.7	4	5	1.9	-4.4	HCV protease inhibitor (genotype 1).	
Acteoside	-10.631	-85.38	investigati onal	drugba nk	624.5	11	245.2	9	14	1	-2.8	a neuroprotective agent, an antileishmanial agent, an anti-inflammatory agent, a plant metabolite and an antibacterial agent.	
TMC353121	-11.266	-83.08	Approved	repurpo sing hub	558.7 27	13	107.7	4	6	4.09	-5.45	potent respiratory syncytial virus (RSV) fusion inhibitor	
TXA127	-9.5	-82.253	phase II	repurp hub	899.0 2	28	421.1 53	10	20	-2.93	-2.12	angiotensin receptor agonist	
Evista (Raloxifene Hydrochlori de)	-7.086	-81.815	approved, investigati onal	drugba nk	473.5	7	70	2	5	5.4	-6	Treatment of osteoporosis in postmenopausal women and those on glucocorticoids	

Sunitinib	-7.381	-80.705	approved	drugbank	398.4	7	77.2	3	3	3.2	-4.1	Used to treat certain types of advanced or progressive tumors of the stomach, intestines, esophagus, pancreas, or kidneys.	
Rutin	-13.104	-76.635	approved	drugbank	610.5	6	267	11	17	1.7	-3.1	radioprotective and antiplatelet activity. Rutin reduces hepatic and blood cholesterol levels.	
MM-102	-10.424	-76.45	pre-clinical	repurposing hub	669.8	21	178.3	7	7	2.85	-5.49	Peptidomimetic MLL1 inhibitor	
Felypressin	-8.486	-76.45	pre-clinical	drugbank	669.8	19	264.7	12	13	-1.1	-4.4	Non-catecholamine vasoconstrictor	
Epicatechin	-11.036	-74.878	investigational	drugbank	290.2	1	110.2	5	6	1	-2.6	Prevent the onset of type II diabetes and many cardiovascular diseases	
Quercetin 3-Rutinoside	-11.362	-71.34	Approved	drugbank	610.5	6	265.6	10	16	0.1	-2.2	antiallergic, anti-inflammatory, antiproliferative, and anticarcinogenic	

												properties	
Lanreotide	-8.097	-69.96	Approved	drugbank	1096.3	17	355	13	12	1.87	-5.3	management of acromegaly and symptoms caused by neuroendocrine tumors,	
Diosmin	-10.032	-67.23	Approved,investigational	drugbank	608.5	7	234.2	8	15	-0.4	-2.6	venous disease	
KD025 (Slx-2119)	-7.023	-66.999	phase II	repurposing hub	608.5	8	104.82	3	5	2.95	-5.7	Use idiopathic pulmonary fibrosis (IPF), selective ROCK2 inhibitor.	
Cefoperazone	-7.074	-66.95	Approved,investigational	drugbank	645.6	9	220.2	4	11	-0.1	-3.4	Effective against Pseudomonas infection and various bacterial infections.	
Pilaralisib	-7.691	-66.86	investigational	drugbank	541	7	148.3	4	8	3.98	-5	potential antineoplastic activity	
Troxerutin	-10.993	-66.24	investigational	drugbank	742.6	15	293.2	10	19	-0.5	-2.7	chronic venous insufficiency; vasoprotective	

Vitexin	-7.858	-65.995	investigational	drugbank	432.4	3	177	7	10	1.38	-2.8	platelet aggregation inhibitor, an antineoplastic agent and a plant metabolite.	
Tenofovir Alafenamide	-7.379	-65.61	Approved	drugbank	476.4	12	143.8	2	7	1.49	-3.3	treat chronic hepatitis B, and prevent HIV-1 infections	
Argatroban Monohydrate	-7.932	-64.8	Approved, investigational	drugbank	508.6	8	177.7	6	9	-0.9	-3.5	synthetic direct thrombin inhibitor derived from L-arginine	
Flavitan	-10.237	-62.71	Approved	drugbank	785.5	13	356	9	20	-2.7	-3.2	Vitamin B2 deficiency or metabolic disorder including stomatitis, eczema, etc	
Curcumin	-8.941	-61.71	Approved, Experimental, Investigational	drugbank	368.3	8	93	2	6	3	-4.3	chemopreventive and anticancer agents.	
Saquinavir	-7.461	-56.66	Approved, Investigational	drugbank	670.8	13	166.7	5	7	4	-5.4	anti-HIV protease inhibitor	

Montelukast	-7.418	-53.21	Approved	drugbank	586.1	12	70.4	2	4	7.2	-7.8	leukotriene antagonist, an anti-asthmatic drug and an anti-arrhythmia drug	
Dapagliflozin	-8.205	-50.56	Approved	drugbank	408.8	6	99.3	4	6	2.5	-3.4	odium-glucose cotransporter 2 inhibitor	
Baicalein	-8.169	-47.81	pre-clinical	repurposing hub	270.2	1	87	3	5	1.7	-3.3	anti-inflammatory, antioxidant, antiviral, and antitumor effects, lipoxygenase inhibitor	
Pitavastatin	-8.435	-47.71	Approved	drugbank	421	8	90.6	3	5	3.75	-5	novel statin that induces plaque regression and elevates HDL-cholesterol levels	
Thymopentin	-8.026	-40.28	investigational	drugbank	679.9	22	327.6	11	14	-3.3	-4	treatment of rheumatoid arthritis, AIDS, and other primary immunodeficiencies.	

The general recommended ranges are as follows:
[a]Molecular weight, <500.
[b]Rotatable bonds
[c]Polar surface area, <140 Å^2,
dNumber of hydrogen bond donors, <5.
[d]Number of hydrogen bond acceptors, <10.
[e]Predicted octanol/water partition coefficient, -0.4 to +5.6.
[f]Predicted aqueous solubility, < -5.0.
*NF, not found

Blue highlighted cells are the molecules found common in screening results of other papers.