

Computational Studies to Identify Potential Main Protease Inhibitors for SARS-CoV-19

Authors: M. Elizabeth Sobhia ^{*1}, Ketan Ghosh¹, Srikanth Sivangula¹, Siva Kumar¹, Harmanpreet Singh¹

Affiliations: ¹National Institute of Pharmaceutical Education and Research (NIPER), Sector 67, SAS Nagar, Punjab.

*Correspondence: mesophia@niper.ac.in

Abstract

The Coronavirus pandemic has put the entire humanity in total shock and has forced the world to go under total lockdown. It is time for the entire scientific community across the globe to find a solution for this deadly and unseen enemy. *In silico* studies play a vital role in situations like this, as experimental studies are not feasible by all researchers particularly with relevance to BSL4 procedures. In this study, using the high resolution crystal structure of SARS-CoV-2 main protease (PDB: 5R82), we have identified molecules which can potentially inhibit the main protease (M^{pro}). We used a three-tier docking protocol making use of three different databases. We analysed the residues which are lying near the ligand binding pocket of the main protease structure and it shows a wide cavity, which can accommodate chemically diverse ligands, occupying different sub-pockets. Using the small fragment bound in the 5R82, we have identified several larger molecules whose functional groups make interactions with the active site residues covering. This study also presumably steers the structure determination of many ligand-main protease complexes using x- ray diffraction methods. These molecules can be used as ‘*in silico* leads’ and further be explored in the development of SARS-CoV-2 drugs.

Keywords: SARS-COVID-19, Main Protease, Molecular Docking, MMGBSA

Introduction

Inchoate pathogens are becoming worldwide threats to public health in recent times. In December 2019, a cluster of patients with pneumonia of unknown origin were reported in Wuhan, China. The causative agent was later found to be a virus of the corona virus family. The world health organization named this infectious disease as Coronavirus disease-2019 (Covid-19). The virus is now named as Severe Acute Respiratory Syndrome Coronavirus-2 (SARS-CoV-2). Though the earlier outbreaks of severe acute respiratory syndrome coronavirus (SARS CoV) in China and middle east respiratory syndrome coronavirus (MERS CoV) in the Middle east, posed a significant threat, with the Covid-19 numbers increasing exponentially, the global public health has been under serious stress [1, 2]. As of 22 April, WHO reports 24,71,136 confirmed cases and 1,69,006 deaths [3].

Coronaviruses (CoVs) are large, enveloped, positive single stranded RNA viruses with 26 to 32 kilo bases belonging to the *Coronaviridae* family. The glycoprotein spikes on its envelop and gives them their crown like appearance. There are four subtypes α , β , δ and γ corona viruses. While α and β are known to affect humans, the δ and γ are found in various animals [1]. The large RNA genome of CoVs provides the flexibility for host adaptation and genome modification [4]. The identification of a variety of CoVs from bats proves that they may be the reservoir of these viruses. They generally infect humans through an intermediate mammal host like civet cats and dromedary camels [5]. There are seven CoVs known to infect humans (HCoVs). They are HCoV-229E, HCoV-NL63, HCoV-OC43, HCoV-HKU1, SARSCoV, MERSCoV and the current SARS-Cov-2 [6]. They lead to various respiratory, gastrointestinal and neurological diseases in human hosts [7]. The virus uses the surface spike (S) glycoprotein for binding to host cell receptor, Angiotensin converting enzyme2 (ACE2) in case of SARS-CoV-2. Once the RNA enters the cell, it uses the host cell ribosomes to synthesize polyproteins pp1a/pp1ab and these polyproteins undergo proteolysis by the virus coded main protease (Mpro) also called the chymotrypsin-like protease (3CLpro), as well as the papain-like proteases to produce 16 non-structural proteins necessary for the replication process of the virus. Remaining accessory proteins and structural proteins viz., spike (S) protein, membrane (M) protein, envelop (E) protein, nucleocapsid (N) protein, hemagglutinin esterase (HE) glycoprotein are further synthesized from sub genomic RNAs [4] [8].

With no approved drugs or vaccines for Covid-19, there is lot of research on various therapeutic strategies and a number of clinical trials are in progress. Considerable work on development of drugs for other corona viruses such as SARS CoV and MERS CoV has enabled rapid identification of drug targets for SARS-CoV-2 [9]. The

essential role of SARS-CoV-2 main protease in viral replication makes it an attractive target for drug design and development. The human proteases not matching with the SARS-CoV-2 main protease is an added advantage to target this protein [10]. The antivirals Lopinavir and Ritonavir known to target the main protease (M^{pro}) of SARS CoV are used in clinical trials for SARS CoV-2 [11], [12].

The aim of the present study is to use the fragment binding to the active site of M^{pro} structure (5R82) of SARS-CoV-2 and to explore the ways to use other sub pockets of binding site to obtain potential molecules for the inhibition of main protease of SARS-CoV-2. We have performed a three-tier docking protocol using three databases to come to a conclusion.

Structural information on M^{pro} of SARS-CoV-2

Table S1 contain the details of all the PDB (11) crystal structures of the main protease. The main protease helps in replication and transcription of novel corona virus. The resolution of these structures having 306 residues ranges from 2.20Å to 1.31Å. Also it has CYS-HIS catalytic dyad and substrate binding pocket placed in extended conformation [13].

Based on the this analysis, 5R82 (main protease) was selected as the target for our study as it has the best resolution (1.31Å) amongst all the PDB structures available. This protein is crystallized with a fragment molecule bound to the substrate binding site. This structure provides important information regarding the possible features of molecular fragments that could have good affinity for binding and using this aspect we could design possible substitutions of fragments.

The 3D-structural analysis of 5R82; shows the binding cavity is bigger than the space occupied by ligand in 5R82. From their analysis, the binding cavity is divided into 5 sub-pocket regions. The first sub pocket S1, residues 139-148 (green), a core region of the binding site in M^{pro} protein. This sub pocket is essential to inhibit the protein; all the amino acids in the region are capable of forming hydrogen bonds. The second sub pocket S2, residues from 163-171 (marine) which are little away from the S1. This region can also anchor small molecules by forming hydrogen bonds. The third sub-pocket S3, residues 41-49 (olive) which lies on the other side of S1. This region is comparatively smaller but is helpful in interacting with the side chain portions of inhibitors designed either S1 or S2 sub-pocket. The fourth sub-pocket region S4 lie in between residues 187-189 (purple) is tiny; it can also form hydrophobic interactions with the designed inhibitors. A final sub-pocket S5, between residues 21-28 (brick red) has the capability to form both hydrogen bonds and hydrophobic interactions. Below fig 1 shows the co-crystallized ligand, and sub pocket division in clock wise direction of 5R82.

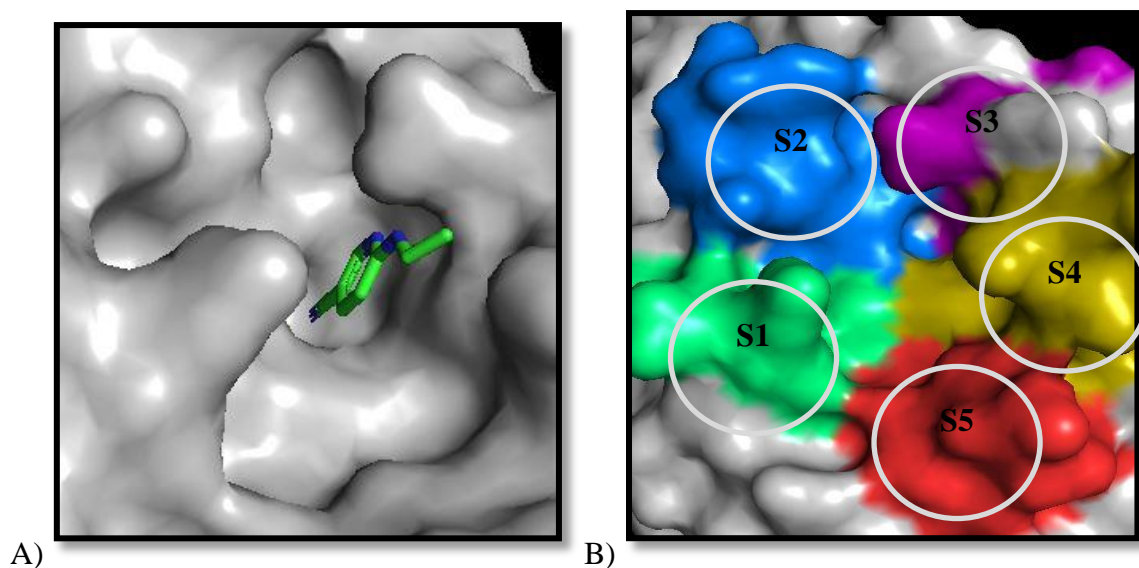


Fig. 1: A) Surface view of 5R82 with ligand (RZS) bound in 5R82; B) Sub-pockets S1-S5 in the binding cavity (clock-wise).

Table 1: Residues making up the cavity for binding

Residue Type	Residue no. and 3 letter name
Polar residues	Asn28, Thr24, Thr25, Thr21, Gln19, Asn119, Ser144, Asn142, Ser46, Thr45, Gln192, Thr190, Gln189, HIE163, HIE164, Thr26,
Hydrophobic residues	Pro39, Val41, Cys44, Met49, Leu27, Cys145, Leu141, Pro52, Leu50, Ile43, Cys44, Tyr118, Val20, Tyr54,, Phe140, Val42
Negatively charged residues	Asp187, Glu166, Glu47, Asp48
Positively charged residues	Arg40, Arg188

The residues surrounding the binding cavity were shown in Table 1. Among these residues THR26, HIE41, SER46, PHE140, GLY143, CYS145, GLU166 and GLN 189 are important for main protease inhibition.

Materials and Methods

3D-Databases Preparation: Three databases namely Enamine [14], Specs [15], Natural product [15] were refined by eliminating duplicates and adding missing hydrogen atoms.

Protein Preparation: 5R82 was prepared using the Protein Preparation Wizard in Maestro [16]. This involved addition and optimization of hydrogen bonds, atomic clashes removal, addition of formal charges to

the hetero groups and then optimizing at neutral pH. And then the structure was minimized using Optimized Potential for Liquid Simulations-3e (OPLS-3e) force field. The crystal water molecules from the crystal structure near the catalytic site and substrate binding site were retained for the molecular docking studies.

Ligand Preparation: The molecules from the databases were retrieved in structure data file (SDF) format. The molecules were subjected to ligand preparation using Ligand Prep module of Schrodinger suite [17]. The ligand preparation involves definite chirality preservation, generating minimum five low-energy stereoisomers per ligand, using default conditions at pH 7.0 ± 2.0 .

Grid-Generation: The three dimensional receptor interaction, grid was generated at the centroid of the bound ligand in crystal structure of SARS-CoV-2 main protease. The grid box was extended up to 15 Å as the inner box and 26 Å as the outer box covering the entire binding site cavity completely.

Molecular Docking: GLIDE [18] for our docking studies, uses a series of hierarchical filters to search for probable locations of the ligand in the binding-site of a receptor. The grid represents the site where the ligand will be docked and the shape and properties of the receptor provides the ligand a collection of ligand conformations and these screens the ligand to provide ligand conformers to locate promising poses favourable for binding. From the poses selected by initial screening, the ligand is refined in torsional space in the field of the receptor using OPLS3e with a distance-dependent dielectric model. Finally, a small number of poses are minimized within the field of the receptor with full ligand flexibility (post-docking minimization or PDM) [18].

The docking protocol was validated by comparing the ligand interaction of co-crystallized ligand with the reported interaction in PDB ligand interaction data. The compounds were subjected to three-tiered docking strategy in which all the compounds were docked by three stages of the docking protocol that includes High Throughput Virtual Screening (HTVS), Standard Precision (SP) and Extra precision (XP). The first stage of HTVS docking screened the ligands from the databases and all 1000 top scoring compounds were passed on to the second stage of SP docking. 500 screened molecules from SP were then docked using more accurate and computationally intensive XP mode. Based on docking score, glide emodel and glide energy, the docked molecules were analysed [18]. The default parameters were used for the 3 tiered docking strategy as mentioned in Fig 2.

Prime MM-GBSA: Molecular mechanics/generalized born surface area (MM-GBSA) method was used as a post-docking analysis tool (32). Based on the docked complex, it calculates the binding free energy (ΔG_{bind}) of each ligand.

$$\text{Equation: } \Delta G_{\text{bind}} = \Delta E_{\text{intra}} + \Delta G_{\text{solv}} - T\Delta S_{\text{conf}} + E_{\text{vdW}} + E_{\text{elect}} + E_{\text{ptn}}$$

ΔE_{intra} and ΔG_{solv} are the intramolecular strain and the desolvation penalty for each ligand upon binding , - $T\Delta S_{\text{conf}}$ is ligand conformational entropy penalty, multiplied by temperature to convert it into free energy, E_{vdW} and E_{elect} are the protein ligand interaction intermolecular van der Walls and electrostatic interaction energies, whereas E_{ptn} is the protein energy obtained after minimization [18].

Flowchart of the study

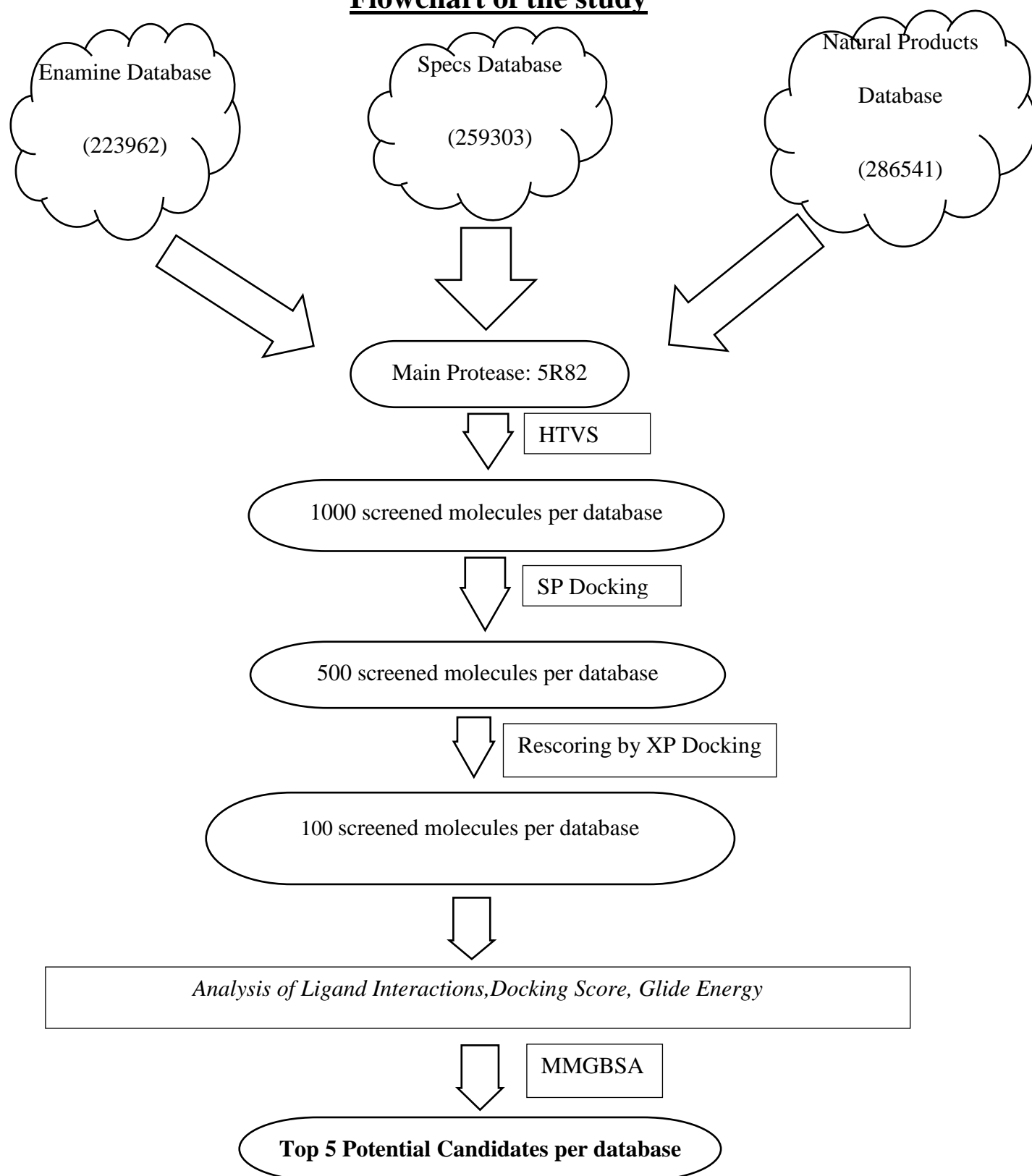


Fig: 2 Representation of the flow of study performed

Results and Discussion:

Main protease crystal structure PDB:5R82 offers the opportunity to employ a structure based drug design strategy in identifying novel covid-19 inhibitors. In the co-crystals structure, RVZ forms a hydrogen bond with GLN189 and hydrophobic interactions with MET49 and MET165 and also a pi-pi stacking interaction with HIS41 as shown in below figure 3. Crystal water mediated interactions of the ligand were also shown. It is indeed the ligand occupied the space of S2, S3 and S4 subpockets of the cavity. The co-crystallized pose was reproduced through molecular docking and the ΔG_{bind} for co-crystallized conformation of RVZ was estimated through MM/GBSA calculation to a value, i.e., -27.50 Kcal/Mol.

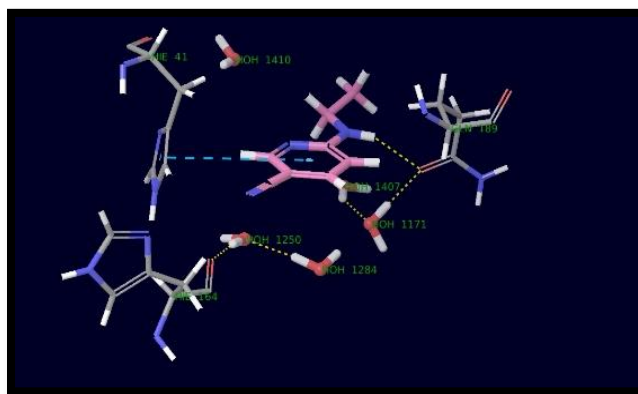


Fig: 3 Co-Crystallized ligand in 5R82 having interaction with protein along with crystal waters in the near vicinity.

It is clear that the crystal waters play a role in the ligand binding. The HOH1171 is having H-bond interaction with GLN189 of protein and docked ligand. Other HOH1410, HOH1250, HOH1284 are residing near the binding site of the ligand and some of them make water mediated interactions with the ligand. These water molecules may have a vital role in ligand binding mechanism.

Molecular docking and data base searching

Approximately 6lakh molecules from three databases were screened by HTVS. We filtered the molecules which show atleast one of the desired hydrogen bond interaction with any of the residues THR26, SER46, GLY143, CYS145 and GLN189 and the pi-pi interaction HIS41 was also a criteria to filter. The top 1000 molecules were later screened through SP docking resulting in 500 molecules from each database. These were further subjected to XP docking (Fig.2). The top 50 molecules from each database were chosen for the analysis. The data of top 20 molecules was listed in supplementary Table S1.

Out of the 50 molecules from each database, the top 5 molecules were represented in the Table II. The 2D structures of the top 5 molecules from each database are shown in supplementary figure.S1. These results include parameters like Docking scores, Glide emodel, Glide energy and Prime MMGBSA ΔG values . The selection of the molecules were based on the fact that these molecules least positional deviation from crystal ligand pose and posses required bonded and non bonded interactions with residues involved in catalytic mechanism engulfing the S1,S2,S3,S4 and S5 sub-pockets of 5R82.

Table 2: Docking results of Top 5 screened database molecules on 5R82

Sr.No	Title	Dock Score	Glide emodel	MMGBSA ΔG Bind*	Glide energy*	Interaction
	Co-crystalized Ligand	-4.555	-34.306	-38.46	-27.50	MET49,MET165,GLN189 (H-bond) HIE41 (pi-pi stack)
Natural products Database						
1	ZINC31167921	-9.776	-69.669	-73.13	-55.876	THR26,THR24,PHE140,SER46(H-bond), HIE41(pi-pi stack)
3	ZINC03870412	-9.157	-63.806	-69.55	-53.35	SER46,THR26(H-bond), HIE41(pi-pi stack)
4	ZINC59732767	-8.959	-59.06	-84.04	-58.276	GLY143(H-bond), GLU166(pi-cation), HIE41(pi-pi stack)
13	ZINC67912528	-7.894	-82.049	-57.45	-60.715	GLN189,GLY143,THR26, THR25(H-bond)
18	ZINC15107337	-7.494	-73.167	-60.89	-51.968	THR26,THR24,HIE164(H-bond), HIE41(pi-pi stack)
Enamine Database						
19	Z1576651068	-8.299	-46.517	-41.05	-40.511	THR26(H-bond) HIE41(pi-pi stack)
20	Z2223094643	-7.991	-55.661	-60.53	-40.8	THR26(H-bond) HIE41(pi-pi stack)
21	Z1129919147	-6.982	-55.268	-51.22	-47.132	THR26(H-bond) HIE41(pi-pi stack)
22	Z2719767438	-6.777	-68.244	-67.22	-47.758	GLY143(H-bond) HIE41(pi-pi stack)
23	Z1538676358	-6.756	-54.336	-70.78	-43.888	THR26(H-bond)
Specs Database						

24	ZINC06025649	-8.266	-58.63	-55.34	-42.687	THR26,ASN142,GLY143 (H-bond) HIE41 (pi-pi stack)
25	ZINC00670776	-7.567	-68.307	-61.51	-54.132	THR26 (H-bond) HIE41 (pi-pi stack)
26	ZINC02753209	-6.98	-67.992	-82.75	-54.016	GLY143 (H-bond) HIE41 (pi-pi stack)
27	ZINC01794727	-6.87	-76.344	-65.5	-58.746	SER46,THR26,GLY143 HIE41 (pi-pi stack)
28	ZINC02097397	-6.84	-72.789	-84.87	-44.762	GLU166,PHE140,ASN142 (H-bond) HIE41 (pi-pi stack)
*indicates kcal/mol						

Molecular Interaction Analysis:

Enamine database: **Z1576651068** showed H-bond interaction with polar residue THR26 and nitrogen atom of pyrazole ring. It also showed H-bond interaction with water molecule via hydroxyl group of the ligand, further leading to interaction of the same water molecule with GLN189. Pi-Pi stacking was observed between HIE41 and benzyl ring of the ligand. **Z2223094543** showed H-bond interaction with polar residue THR26 and oxygen atom of ketone group. Also it showed H-bond interaction with 2 water molecules via hydroxyl group and oxygen atom of the ligand. Pi-Pi stacking was observed between HIE41 and benzyl ring of the ligand. **Z11129919147** showed H-bond interaction with polar residue THR26 and nitrogen atom and oxygen atom. Pi-Pi stacking was observed between HIE41 and benzyl ring of the ligand. **Z2719767438** showed H-bond interaction with SER143 and nitrogen atom of pyrazole ring. Also it showed H-bond interaction with water molecule via hydroxyl group of the ligand. Pi-Pi stacking was observed between HIE41 and benzyl ring of the ligand. **Z1538676358** showed H-bond interaction with polar residue THR25 and oxygen atom of hydroxyl group. Also it showed H-bond interaction with water molecule and nitrogen of the ligand. Pi-Pi stacking was observed between HIE41 and benzyl ring of the ligand. These molecules showed a common interaction pattern where they formed H-bond interactions with THR26, and pi-pi stacking with HIE41. The binding site of the molecules possess hydrophobic as well as hydrophilic region of the protein. The analysis shows that the binding region is surrounded by two negatively charged residues (ASP187, GLU47), one positively charged residue (ARG188) and two water molecules residing at the catalytic site. (Fig 4)

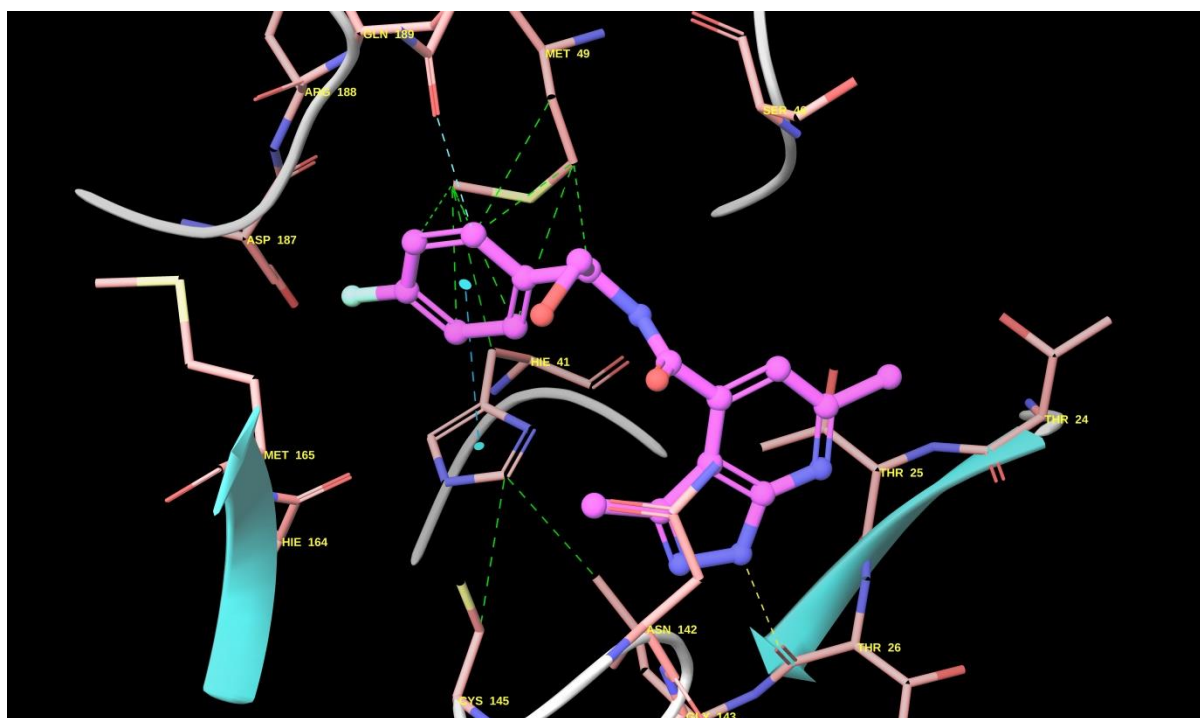


Fig: 4. The 3D interaction diagram of Z1576651068 molecule with top docking score among enamine molecules.

Natural Products Database: **ZINC31167921** showed H-bond interactions with THR24, THR26 and HIE164 and oxygen atom of hydroxyl groups. Also Pi-Pi stacking was observed between HIE41 and benzyl group of the ligand. **ZINC03870412** had H-bond interactions with THR26, GLY143, GLN189, THR25 and hydroxyl group. **ZINC59732767** showed h-bond interactions with THR26, SER46 and hydroxyl group. There is Pi-Pi stacking interaction between HIE41 and benzyl group of the ligand. **ZINC67912528** showed H-bond interactions with PHE140, SER46, THR26, THR24 and hydroxyl group. Also H-bond interaction was seen with the water molecule and oxygen atom of the ligand. There is Pi-Pi stacking between HIE41 and benzyl group ligand. **ZINC15107337** showed H-bond interaction with GLY143 and oxygen of ketone group. There is pi-cation interaction of GLU166 and nitrogen of ligand. Also there is pi-pi stacking between HIE41 and benzyl group of ligand. This group of molecules showed hydrogen bond interactions with PHE140, SER46, THR26, THR24, GLY143, GLN189, THR25 and HIE164. They also formed pi-pi stacking with HIE41 and pi-cation interaction with GLU166. These molecules too entered in the cavity with hydrophobic and hydrophilic residues of protein substrate binding site. (Fig 5)

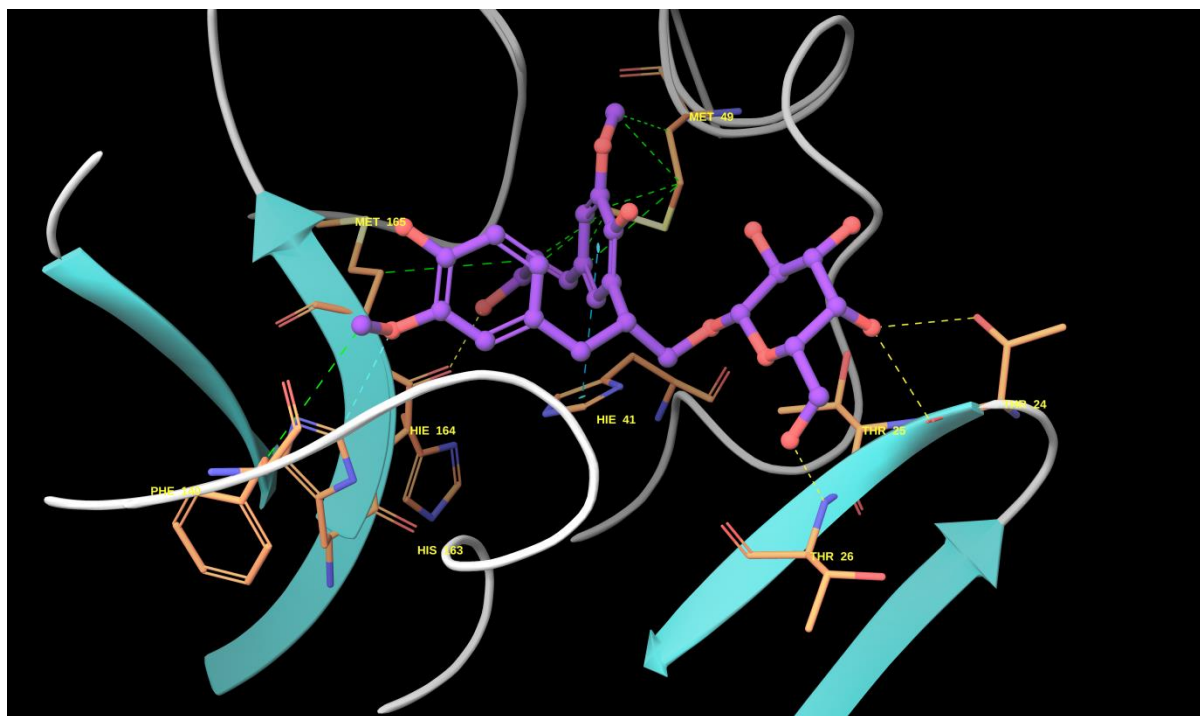


Fig: 5 The 3D interaction diagram of ZINC31167921, molecule with top docking score among natural products database

Specs Database: **ZINC06025649** showed H-bond interactions with THR26 and hydroxyl group, GLY143 and nitrogen, ASN142 and nitrogen of amine group. There is Pi-Pi stacking between HIE41 and benzyl group. **ZINC00670776** showed H-bond interaction with Polar residue THR26 and nitrogen group. There is interaction between water molecule and nitrogen of ligand. Also there is pi-pi stacking between HIE41 and benzyl group of ligand. **ZINC02753209** showed H-bond interactions with GLY143 and nitrile group, also H-bonding between water molecule and nitrogen group of ligand, pi-pi stacking between thiophene rings. **ZINC01794727** showed H-bond interactions with THR26, GLY143 and oxygen atom. There is halogen bond between SER41 and chlorine of ligand. There is pi-pi stacking between HIE41 and benzyl group. **ZINC02097397** showed H-bond interactions with PHE140, ASN142 and nitrogen atom. Also 2 interactions with water molecule are seen. One water molecule has interaction with triazole ring nitrogen and GLU166 pi-pi stacking between HIE41 and benzyl group. These molecules showed Hydrogen bond interactions with GLY143, polar residues like THR26, SER46, ASN142 and hydrophobic residue like PHE140. Also it showed H-bond interaction with water molecules directly. (Fig 6)

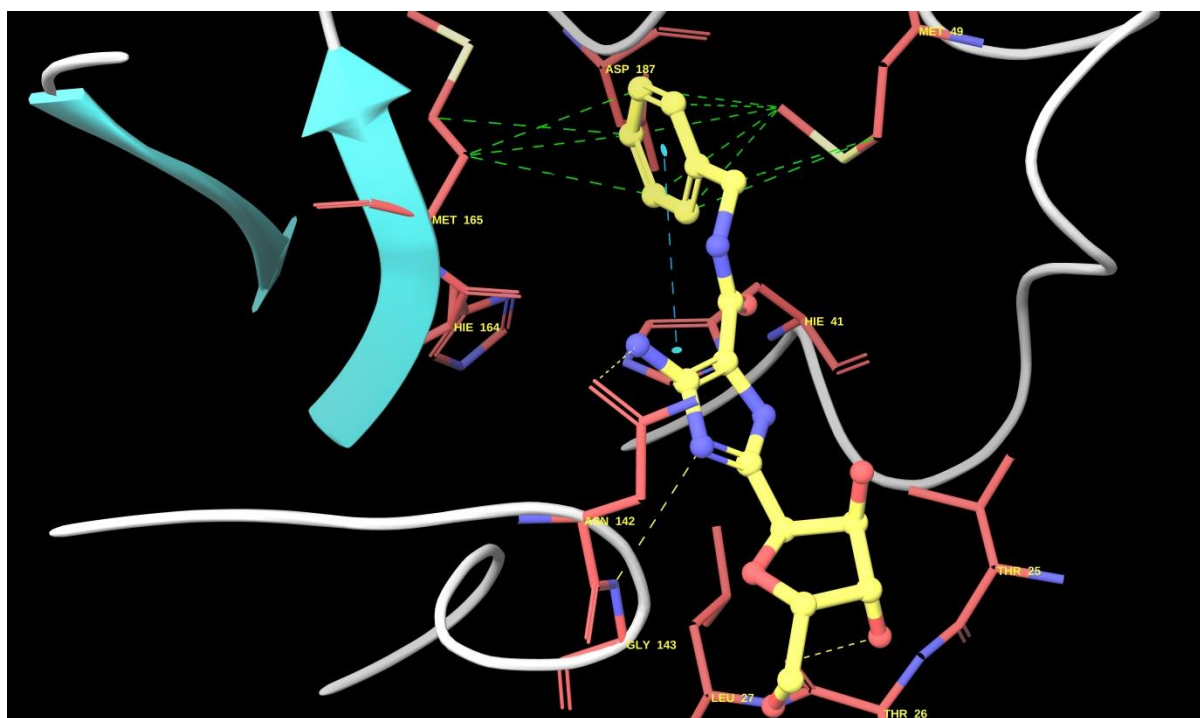


Fig: 6 3D interaction diagram of ZINC06025649, molecule with top docking score among specs database molecules.

Conclusion:

The Covid-19 pandemic calls for a quick response from the scientific realm and with no specific remedy available, computational approach is the key for identification of potential inhibitors that can repurposed for effective therapeutics. The main protease involved in viral replication inside the host cell becomes an important target to inhibit the virus growth. The binding site is considerably optimum with 5 sub-pockets for small molecules to fit in. The functions of the main protease can be inhibited by engaging the residues around the cavity in interactions and presumably stop the viral replication and halt the disease progression. Therefore, identifying molecules which can occupy the whole binding cavity and provide maximum interactions was undertaken through one of the pharmacoinformatic approaches in the study.

An organized approach was followed to perform this study. A three-tiered molecular docking was followed for screening of databases. Our screening studies may pave out a way to design molecules by providing information that can be utilized further *in vitro*, *in vivo* analysis and address the on-going problem of Covid-19 pandemic.

Acknowledgement:

The authors thank the National Institute of Pharmaceutical Education and Research (NIPER) SAS Nagar, Department of Pharmaceuticals, Ministry of Chemicals and Fertilizers, New Delhi, Government of India for providing the facility.

Author Contributions: M.E.S designed the study. All the authors contributed to computations, analysis and preparation of the manuscript.

Competing interests: The authors declare no competing interests

References:

1. Paules, C.I., H.D. Marston, and A.S. Fauci, *Coronavirus infections—more than just the common cold*. *Jama*, 2020. **323**(8): p. 707-708.
2. Wu, F., et al., *A new coronavirus associated with human respiratory disease in China*. *Nature*, 2020. **579**(7798): p. 265-269.
3. WHO, *Coronavirus disease 2019 (COVID-19) Situation Report-93*, in *Situation Report-93*. 2020, WHO: Switzerland.
4. Sawicki, S.G., D.L. Sawicki, and S.G. Siddell, *A contemporary view of coronavirus transcription*. *Journal of virology*, 2007. **81**(1): p. 20-29.
5. Banerjee, A., et al., *Bats and coronaviruses*. *Viruses*, 2019. **11**(1): p. 41.
6. Fung, T.S. and D.X. Liu, *Human Coronavirus: Host-Pathogen Interaction*. *Annual review of microbiology*, 2019. **73**: p. 529-557.
7. Zhu, N., et al., *A novel coronavirus from patients with pneumonia in China, 2019*. *New England Journal of Medicine*, 2020.
8. Ziebuhr, J., E.J. Snijder, and A.E. Gorbalenya, *Virus-encoded proteinases and proteolytic processing in the Nidovirales*. *Journal of General Virology*, 2000. **81**(4): p. 853-879.
9. Anand, K., et al., *Coronavirus main proteinase (3CLpro) structure: basis for design of anti-SARS drugs*. *Science*, 2003. **300**(5626): p. 1763-1767.
10. Zhang, L., et al., *Crystal structure of SARS-CoV-2 main protease provides a basis for design of improved α -ketoamide inhibitors*. *Science*, 2020.
11. Cao, B., et al., *A trial of lopinavir–ritonavir in adults hospitalized with severe Covid-19*. *New England Journal of Medicine*, 2020.
12. Ghosh, A.K., et al., *Structure-based design, synthesis, and biological evaluation of peptidomimetic SARS-CoV 3CLpro inhibitors*. *Bioorganic & medicinal chemistry letters*, 2007. **17**(21): p. 5876-5880.
13. Jin, Z., et al., *Structure of Mpro from COVID-19 virus and discovery of its inhibitors*. *bioRxiv*, 2020.
14. Shivanyuk, A., et al., *Enamine real database: making chemical diversity real*. *Chem. Today*, 2007. **25**(6): p. 58-59.
15. Sterling, T. and J.J. Irwin, *ZINC 15—ligand discovery for everyone*. *Journal of chemical information and modeling*, 2015. **55**(11): p. 2324-2337.
16. Wizard, P.P., *Schrödinger, LLC*. New York, NY, 2010.
17. Release, S., *LigPrep. 2019-2, Schrödinger, LLC, New York, NY*.
18. Release, S., *Glide. 2019-2, Schrödinger, New York, NY*.