



Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.



Rational design of potent anti-COVID-19 main protease drugs: An extensive multi-spectrum in silico approach

Sajjad Ahmad, Yasir Waheed*, Saba Ismail, Muzammil Hasan Najmi, Jawad Khaliq Ansari

Foundation University Medical College, Foundation University Islamabad, Islamabad, Pakistan

ARTICLE INFO

Article history:

Received 30 July 2020

Received in revised form 2 February 2021

Accepted 8 February 2021

Available online 12 February 2021

Keywords:

SARS-CoV-2 M^{Pro}

COVID-19

ASINEX antiviral library

MD simulation

MM-PBSA

WaterSwap

ABSTRACT

The emergence of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) as a novel coronavirus and the etiological agent of global pandemic coronavirus disease (COVID-19) requires quick development of potential therapeutic strategies. Computer aided drug design approaches are highly efficient in identifying promising drug candidates among an available pool of biological active antivirals with safe pharmacokinetics. The main protease (M^{Pro}) enzyme of SARS-CoV-2 is considered key in virus production and its crystal structures are available at excellent resolution. This marks the enzyme as a good starting receptor to conduct an extensive structure-based virtual screening (SBVS) of ASINEX antiviral library for the purpose of uncovering valuable hits against SARS-CoV-2 M^{Pro}. A compound hit (BBB_26580140) was stand out in the screening process, as opposed to the control, as a potential inhibitor of SARS-CoV-2 M^{Pro} based on a combined approach of SBVS, drug likeness and lead likeness annotations, pharmacokinetics, molecular dynamics (MD) simulations, and end point MM-PBSA binding free energy methods. The lead was further used in ligand-based similarity search (LBSS) that found 33 similar compounds from the ChEMBL database. A set of three compounds (SCHEMBL12616233, SCHEMBL18616095, and SCHEMBL20148701), based on their binding affinity for M^{Pro}, was selected and analyzed using extensive MD simulation, hydrogen bond profiling, MM-PBSA, and WaterSwap binding free energy techniques. The compounds conformation with M^{Pro} show good stability after initial within active cavity moves, a rich intermolecular network of chemical interactions, and reliable relative and absolute binding free energies. Findings of the study suggested the use of BBB_26580140 lead and its similar analogs to be explored in vivo which might pave the path for rational drug discovery against SARS-CoV-2 M^{Pro}.

© 2021 Elsevier B.V. All rights reserved.

1. Introduction

Coronavirus disease 19 (COVID-19), declared a pandemic by the World Health Organization (WHO), is a highly infectious disease first reported in late December 2019 as a cluster of pneumonia cases in Wuhan, China, and has since spread globally [1,2]. Genomic sequencing identified a novel coronavirus as the causative agent of this disease. The virus shares 96.2% sequence homology to bat viruses and 79.5% homology to the known SARS-CoV; thus it was named SARS-CoV-2 by the International Committee on Taxonomy of Viruses [3,4]. COVID-19 is affecting more than 219 countries and 2 international conveyances with 103,595,858 confirmed cases that have resulted in 2239,285 fatalities at the time of writing (<https://www.worldometers.info/coronavirus/>). Unfortunately, at this time no well-defined clinically effective antivirals or vaccines are available to treat and prevent COVID-19 infections [5]. Some HIV drugs like Ritonavir and Lopinavir have

been used in combination with α -interferon, but they confer little curative effect and can lead to toxic side effects [6]. A broad-spectrum antiviral, Remdesivir, developed and being explored by Gilead Sciences as a possible COVID-19 treatment needs more data to substantiate its real efficacy [7]. Considering all this, there is an unmet need for development of safe and specific anti-SARS-CoV-2 therapeutics to reduce the severity of this deadly pathogen.

SARS-CoV-2 has an RNA genome with approximate size length of 30,000 nucleotides [8]. It comprises at least 6 open reading frames (ORFs); the first being the overlong ORF (covers 2/3 of the genome) which codes for two polyproteins: polyprotein 1a (pp1a) and polyprotein (pp1ab) [9,10]. Both these proteins are processed into 16 mature non-structural proteins (nsps) by the main protease (M^{Pro}) in the presence of papain-like protease. There are four main nsps: spike (S), envelope (E), membrane (M), and nucleocapsid (N). Considering this prominent and essential role of the M^{Pro} in the life cycle of COVID-19 and because there is no homolog of the M^{Pro} in humans, it has been proposed as an attractive candidate for the discovery of novel drugs [11,12]. M^{Pro} is functional in homodimers, each monomer is 306 residues long and is a three domain cysteine

* Corresponding author at.: Associate Professor & I/C MDL, Foundation University Medical College, Foundation University Islamabad, Pakistan.
E-mail address: yasir.waheed@fui.edu.pk (Y. Waheed).

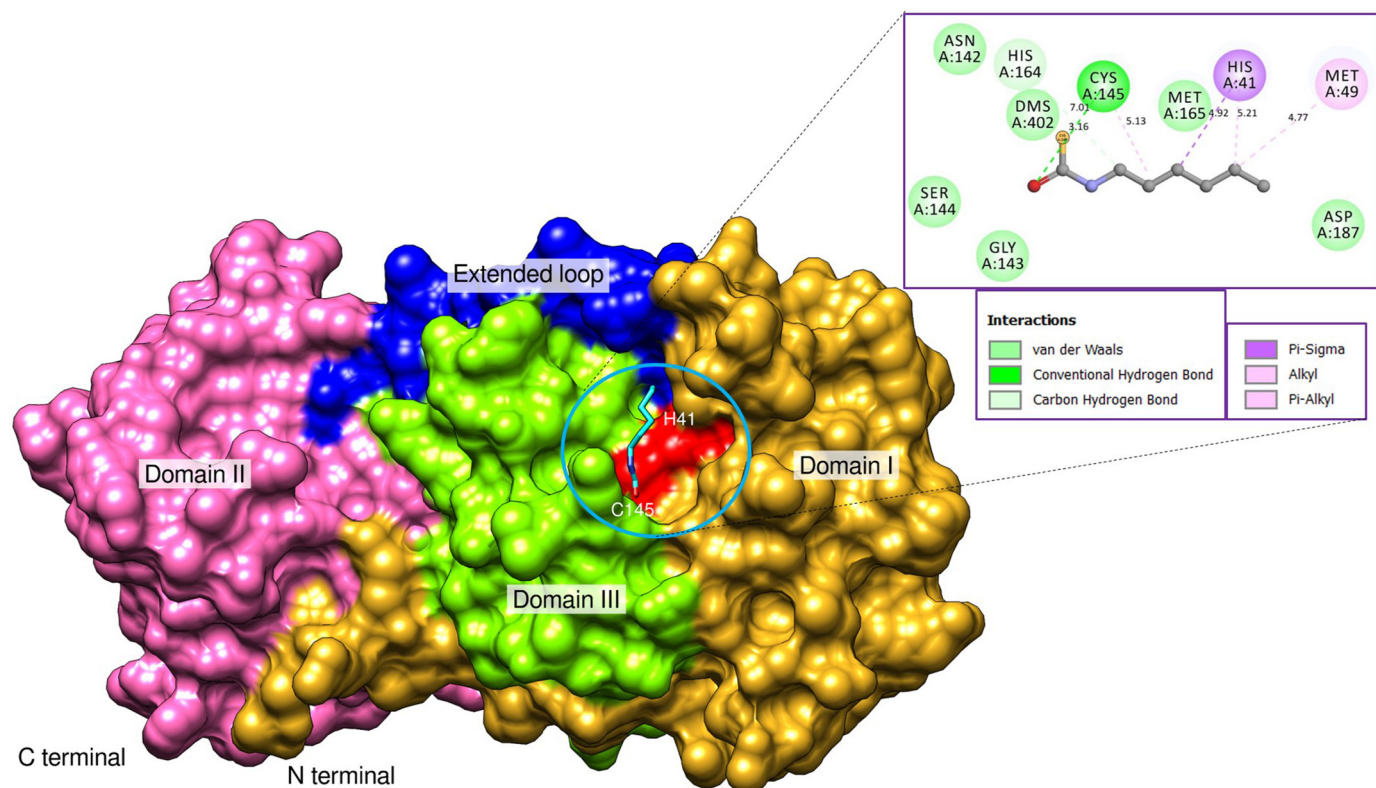


Fig. 1. Surface presentation of M^{Pro} crystal structure. The carmofur compound can be depicted as a stick docked at the active pocket and involved in key interactions with His41 and Cys145. Deep view of the chemical interacting network of the compound is also provided.

protease folded into helices and β -strands [11,13]. The first domain (Domain I) covers residues from 10 to 99 whereas the second (Domain II) is from residue 100–184; both have antiparallel β -barrel structures. The third domain (Domain III), ranging from 201 to 303, is in a large antiparallel globular cluster of five α -helices. Domain III is connected to Domain II by a long loop of 16 residues (residues 185–200). The catalytic activity of the M^{Pro} is mainly due to a dyad (H41 and C145); both of these residues are positioned at the junction of Domain I and II.

Several crystalized monomeric structures of the M^{Pro} have been published recently both in APO (PDB: 6M03) and HOLO (PDB: 7BUY) form [14]. The crystal structure of M^{Pro} co-crystallized with carmofur is shown in Fig. 1. Four active sites of the enzyme have been identified which are highly conserved and include S1', S1, S2, and S4 and have been targeted for the design of several groups of inhibitors. Zhenming et al. detail the inhibition of M^{Pro} by an antineoplastic drug known as carmofur [14]. The reactive carbonyl group of this compound is seen to bind covalently to Cys145, allowing the fatty acid tail to occupy the hydrophobic S2 active site. The compound can inhibit M^{Pro} activity in cells with EC_{50} of 24.30 μ M. Similarly, two lead compounds were identified by Wenhao et al. and co-crystallized with M^{Pro} [12]. The compounds have a covalent interaction with Cys145 and have a good pharmacokinetics profile with low toxicity. As well as this, several *in silico* studies for finding potent leads against M^{Pro} highlight the vital importance of bioinformatics techniques in the identification of potential M^{Pro} inhibitors [15–21]. The current aim is to provide a fast platform to determine potential leads against SARS-CoV-2 M^{Pro} from a pool of antivirals that are active and produce safe pharmacokinetics through a comprehensive computational structure based virtual screening (SBVS), ligand based similarity search (LBSS), drug likeness, lead likeness, molecular dynamics (MD) simulations, binding free energies and entropy techniques.

2. Methodologies

2.1. SARS-CoV-2 M^{Pro} crystal structural retrieval and preparation

The Protein Data Bank (PDB) was accessed to retrieve the crystal structure of SARS-CoV-2 M^{Pro} (PDB ID, 7BUY, and resolution, 1.6 Å) [12]. The structure was subjected to a mini preparatory phase where the attached co-crystallized ligand and water molecules were deleted in UCSF Chimera version 1.15 [22]. The energy of the enzyme was then minimized through 1000 rounds of steepest descent (fast relive

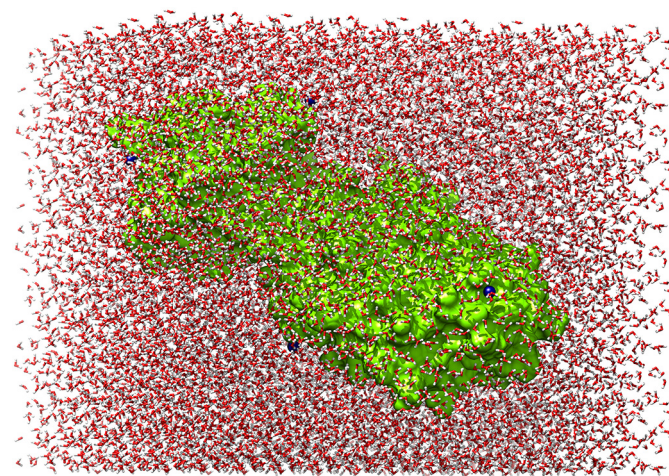


Fig. 2. BBB_26580140- M^{Pro} complex in aqueous solution box. The M^{Pro} is shown by chartreuse surface, water molecules are in sticks whereas Na⁺ ions are presented by blue balls.

of highly unfavorable clashes in the protein structure) and conjugate gradient (slow relive of severe clashes in the protein structure, good at reaching an energy minima) algorithms. The initial step size length of both algorithms are set to default 0.02 Å. All atoms of the enzyme were allowed to move by selecting fixed atoms as none. For assigning parameters to standard residues, AMBER ff14SB was used whereas for non-standard residues AMBER antechamber was employed [23,24]. The overall quality and bad contacts in minimized and un-minimized apo structure of the M^{Pro} were evaluated through the PDBSum Ramachandran plot and the one with better structural features was selected for downward investigations [25,26].

2.2. Retrieval and preparation of Asinex antiviral library

Uncovering chemical entities with improves safety and profound antiviral activity needs high quality compounds as meaningful starting leads. Asinex has developed an antiviral library of macrocycles and

small molecules that could deliver valuable novel leads in target directed virtual screening process. The Asinex antiviral includes 8722 compounds and can be freely accessed at <https://www.asinex.com/antiviral/>. The library was retrieved in .sdf format, subsequently imported to Discovery studio software version 2020 [27] to filtered molecules that meet the criteria of drug likeness [28,29] and lead likeness [30]. The different rules of drug likeness and lead likeness parameters are tabulated in S-Table 1. The compounds were converted to .pdbqt and minimized to achieve to lower energy state in virtual screening software (PyRx software 8.0) [31].

2.3. Site directed virtual screening (SDVS) and ligand based similarity screening (LBSS)

The junction of Domains I and II where catalytic dye is presented was used as the primary site in SDVS [32]. Many of the reported HOLO structures of the M^{Pro} have reported inhibitors bound at this

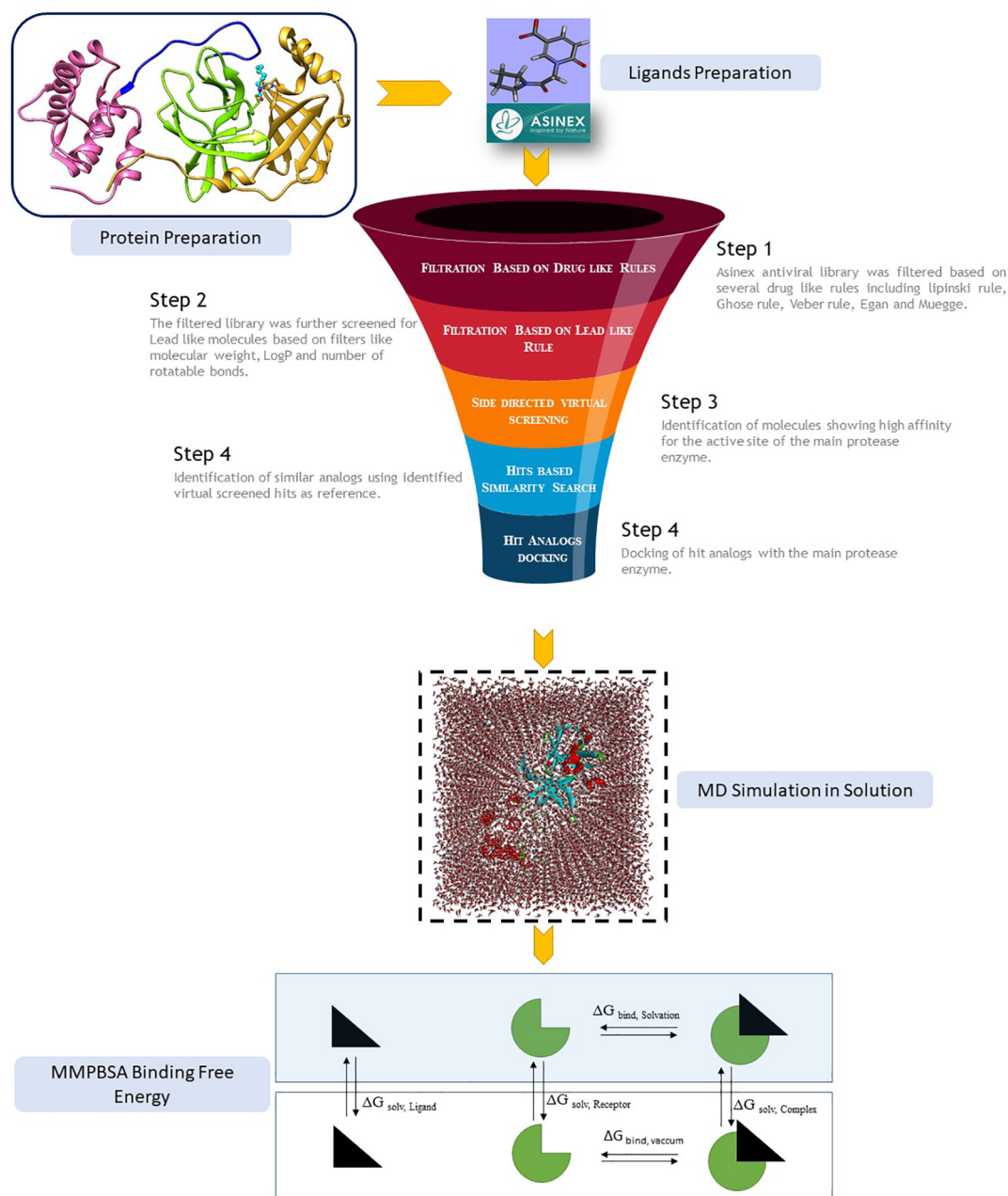


Fig. 3. Methodology flow.

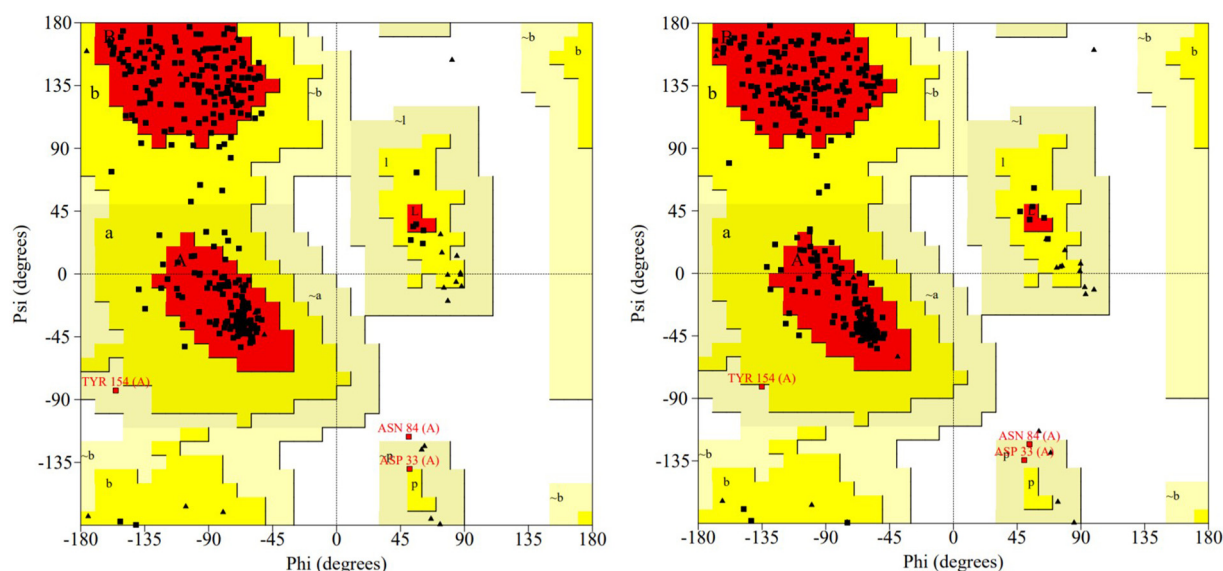


Fig. 4. Ramachandran plot analysis of pre-minimized M^{pro} (left) and minimized M^{pro} (right).

specific site. The prepared compounds library was docked using the coordinates of the Cys145 hotspot residue considered significantly vital for catalytic functionality. The binding site was defined by pointing coordinates of Cys145 sulfur atom (X-axis: -13.610 , Y-axis: 17.354 , Z-axis: 65.955) with dimensions of $15 \text{ \AA}$. The number of iterations set for each compound was 8, and the one with stable binding conformer by acquiring highest negative binding energy (kcal/mol)

was ranked as top. The top complex of the SDVS was selected and subjected to molecular dynamics (MD) simulations [33] and MM-PBSA binding free energy assays [34,35] to affirm docking prediction. The procedure for performing MD simulations and MM-PBSA energy calculations can be seen in Sections 2.5 and 2.7. The screened high affinity binder molecule was then used in ligand-based similarity screening program in ChEMBL database to probe similar analogues

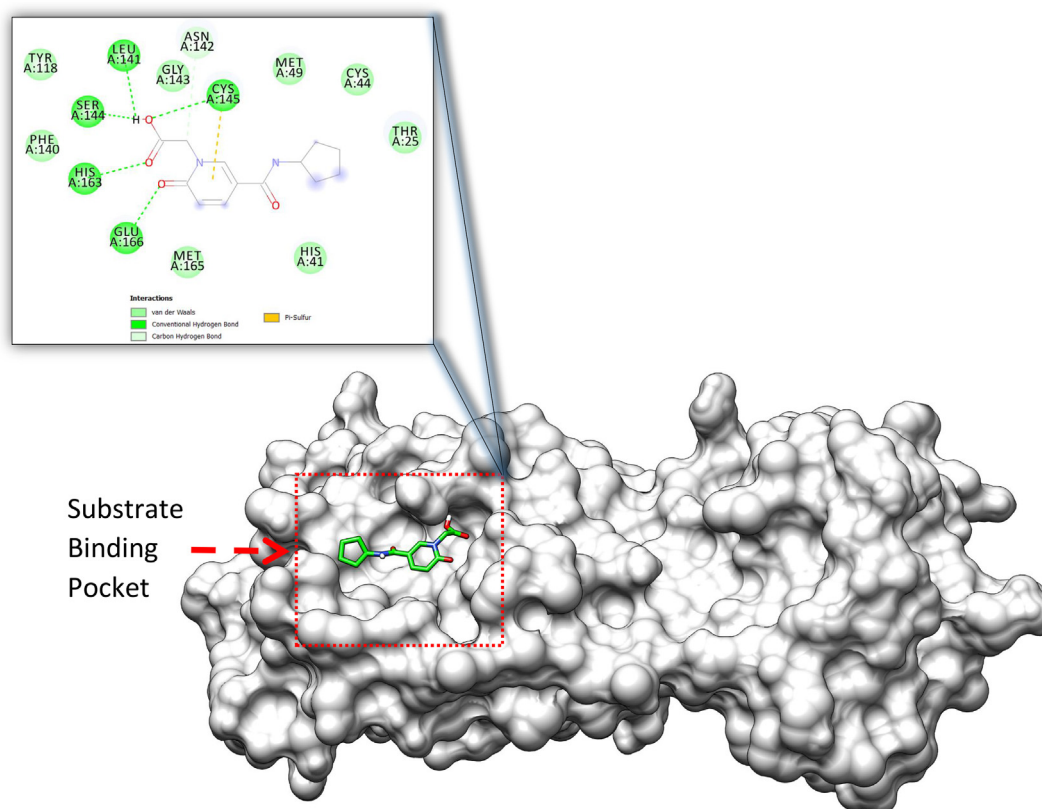


Fig. 5. Binding conformation and chemical interactions network of BBB_26580140 inside M^{pro} binding pocket.

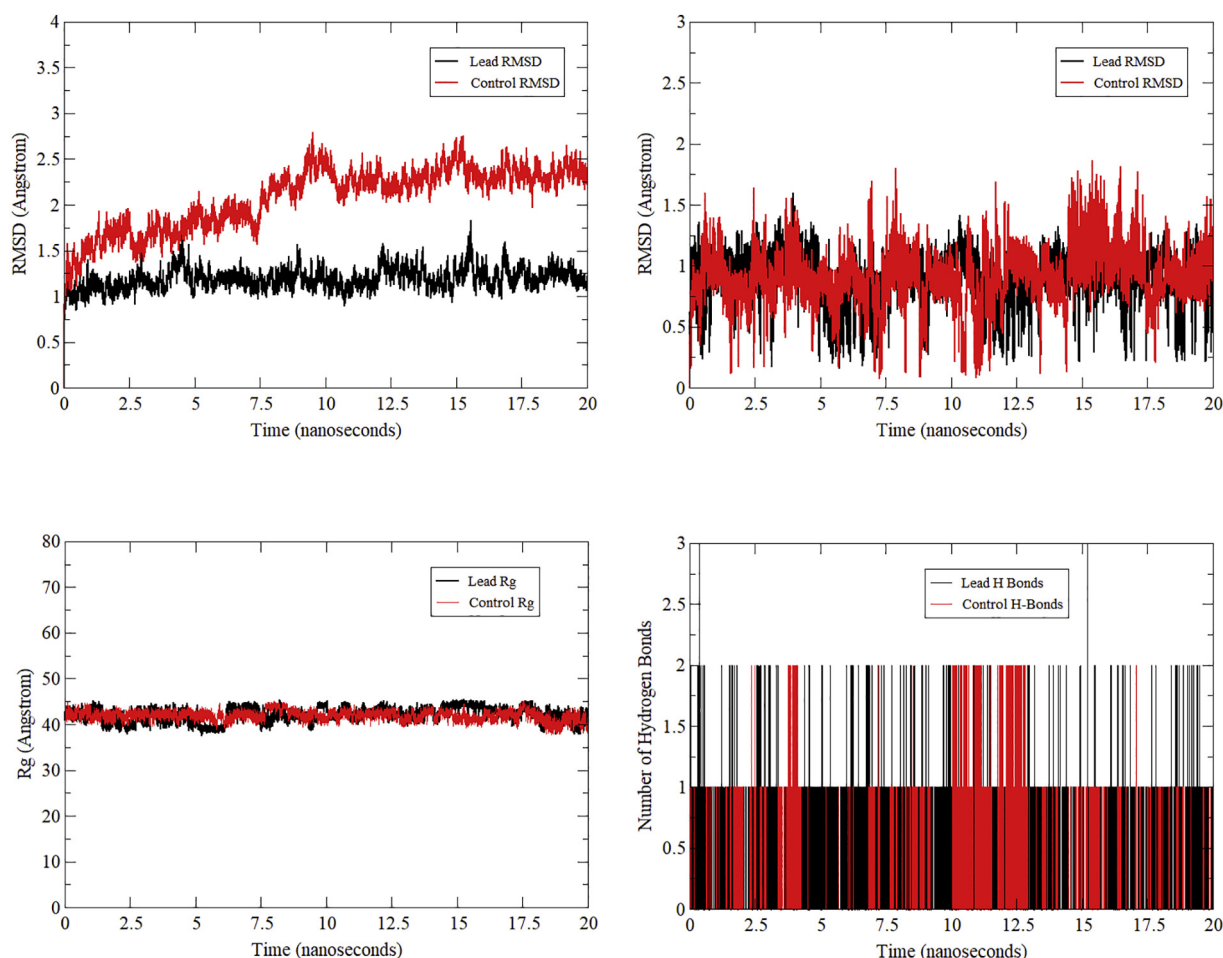


Fig. 6. MD simulation based statistical M^{pro} rmsd (top left), compounds rmsd (top right), M^{pro} Rg (bottom left) and M^{pro} -lead/control H-bonds analysis (bottom right).

of the molecule [36]. The highly similar compounds fulfilling all drug and lead like parameters were selected and processed through the same steps discussed earlier for docking with M^{pro} .

2.4. Pharmacokinetics studies

The different physiochemical properties, pharmacokinetic profile and various other properties of the lead molecule were predicted through online pkCSM [37] and SwissADME [28].

2.5. MD simulations

The prediction made by molecular docking in terms of binding mode of the lead compound and its analogs was evaluated using MD simulation. The MD system chemical interactions profile and stability information were extracted from the trajectories produced during the MD simulation production phase. MD simulations were performed using AMBER 18 simulation package [24]. The MD system parameters library was generated through Antechamber program; the general amber force field (GAFF) [38] was used to create parameters for the compounds whereas the AMBER leap module [39] aided in centering the complex in a 12 Å TIP3P water box. The BBB_26580140- M^{pro} complex in the TIP3P water box is shown in Fig. 2. Charge on the system was neutralized by adding appropriate counter ions. Systems energy was minimized progressively through several rounds: hydrogen atom minimization (for 100 steps), water box minimization (for 500 steps), entire system minimization (500 steps) applying a constraint of 5 kcal/mol $\text{\AA}^2$ on C α , and non-heavy atoms minimization (100 steps)

keeping restraint of 10 kcal/mol $\text{\AA}^2$. Progressive heating of the system from 0 to 300 K was then performed using an NVT ensemble at a time step of 2 femtosecond for a period of 20 ps keeping a restraint of 5 kcal/mol $\text{\AA}^2$ on systems C α . Temperature hold was achieved via langevin dynamics [40] while the SHAKE algorithm [41] was employed to maintain constant hydrogen bond length. Systems were then equilibrated for 1 ns considering time step of 2 femtoseconds. Afterward, constant temperature and pressure (NPT) ensemble [42] was produced for each system by carrying out 1 ns at 300 K and 1 bar pressure. Finally, systems were allowed to equilibrate themselves for 1 ns. Production run was performed 50 ns for the lead compound whereas for lead analogs 100 ns long run was performed. The production ensemble was created using a Berendsen temperature coupling algorithm [43], allowing a time step of 2 fs and non-bound interactions set to 8 Å. Simulation trajectories were performed using the CPPTRA module to examine structural parameters of the systems [44].

2.6. Radial distribution function (RDF) analysis

In biomolecular simulations, the radial distribution function or shortly $g(r)$ is employed to illustrate variations in chemical interaction density as a function of distance from a chosen reference atom [45]. RDF assay was performed using PTRAJ module of AMBER using close receptor-ligand interactions generated in Visual Molecular Dynamics (VMD) [46]. The RDF expression can be presented as:

Table 1
In silico pharmacokinetics of the virtually screened lead molecule.

Property	Model name	Predicted value	Unit
Absorption	Water solubility	−2.738	Numeric (log mol/L)
	Caco2 permeability	−0.012	Numeric (log Papp in 10−6 cm/s)
	Intestinal absorption (human)	59.806	Numeric (% Absorbed)
	Skin Permeability	−2.358	Numeric (log Kp)
	P-glycoprotein substrate	No	Categorical (Yes/No)
Distribution	P-glycoprotein I inhibitor	No	Categorical (Yes/No)
	P-glycoprotein II inhibitor	No	Categorical (Yes/No)
	VDss (human)	−0.546	Numeric (log L/kg)
	Fraction unbound (human)	0.497	Numeric (Fu)
	BBB permeability	−0.522	Numeric (log BB)
Metabolism	CNS permeability	−3.129	Numeric (log PS)
	CYP2D6 substrate	No	Categorical (Yes/No)
	CYP3A4 substrate	No	Categorical (Yes/No)
	CYP1A2 inhibitor	No	Categorical (Yes/No)
	CYP2C19 inhibitor	No	Categorical (Yes/No)
Excretion	CYP2C9 inhibitor	No	Categorical (Yes/No)
	CYP2D6 inhibitor	No	Categorical (Yes/No)
	CYP3A4 inhibitor	No	Categorical (Yes/No)
	Total Clearance	0.533	Numeric (log ml/min/kg)
	Renal OCT2 substrate	No	Categorical (Yes/No)
Toxicity	AMES toxicity	No	Categorical (Yes/No)
	Max. tolerated dose (human)	0.88	Numeric (log mg/kg/day)
	hERG I inhibitor	No	Categorical (Yes/No)
	hERG II inhibitor	No	Categorical (Yes/No)
	Oral Rat Acute Toxicity (LD50)	1.876	Numeric (mol/kg)
	Oral Rat Chronic Toxicity (LOAEL)	1.255	Numeric (log mg/kg_bw/day)
	Hepatotoxicity	Yes	Categorical (Yes/No)
	Skin Sensitization	No	Categorical (Yes/No)
	T. pyriformis toxicity	0.105	Numeric (log ug/L)
	Minnow toxicity	2.575	Numeric (log mM)

$$g(r) = \frac{\langle \rho_{ij}(r) \rangle}{\langle \rho_j \rangle}$$

where, ρ_{ij} represents density frequency at distance (r) ratio to average solvent bulk atom frequency, ρ_j .

2.7. MMPBSA binding free energies estimation

The net binding free energy of complexes containing different high affinity binders was calculated through the AMBER MMPBSA.py method [34]. In total, 100 frames were evenly selected from MD trajectories and each was subjected to solvation free energy and molecular mechanical energy estimation. Parameters set for MMPBSA analysis included internal dielectric constant (1), and external dielectric constant (80). The non-polar solvation energy contribution was estimated using solvent accessible surface area by setting surface tension value to 0.054. The net free energy is estimated for both receptor and ligand molecule individually, add it, and then subtracted from complex net energy [35,47]. This can be presented as,

$$\Delta G_{\text{net binding free energy}} = \Delta G_{\text{binding free energy of complex}} - (\Delta G_{\text{receptor}} + \Delta G_{\text{LG-and}})$$

Each of the above ΔG terms is computed in gas and solvation phase and thus can be classified into,

$$\Delta G_{\text{binding free energy}} = \Delta G_{\text{binding free energy in gas phase}} + \Delta G_{\text{binding free energy in solvation phase}}$$

The gas phase energy is further computed from two components,

$$\Delta G_{\text{binding free energy in gas phase}} = E_{\text{protein-LG-and interaction energy}} - T\Delta S$$

The protein-ligand interactions involve combined energy from electrostatic and van der Waals interactions whereas entropy energy contribution is estimated by $T\Delta S$. The solvation energy term is the output of:

$$\Delta G_{\text{binding free energy in solvation phase}} = \Delta G_{\text{polar solvation energy}} + \Delta G_{\text{non-polar solvation energy}}$$

Estimation of the non-polar energy term is done from the following equation:

$$\Delta G_{\text{non-polar solvation energy}} = \gamma \cdot \text{SASA (solvent accessible surface area)} \cdot \beta$$

In this study, the γ value is set to 0.00542 kcal (mol Å²)^{−1} while β to 0.92 kcal/mol.

2.8. Estimating binding entropy using Normal mode analysis

The entropy contribution to the net binding MM-PBSA energy of the complexes was computed using the AMBER NMODE module [48]. Keeping in mind the extensive need for computational power for this assay, we only considered 10 snapshots from the MD simulation trajectories during the analysis.

2.9. WaterSwap absolute binding free energy calculations

Besides the use of the aforementioned approaches for estimating systems binding free energies, we additionally used the WaterSwap technique from Sire package [49,50]. In contrast to the implicit water model system in MM-PBSA, WaterSwap uses an explicit solvent model. The idea behind WaterSwap is to swap ligand dimensions with equal size and volume of binding pocket explicit water molecules thus considering the free energy contribution of the water molecules present at the protein binding site. A total of 1000 rounds was performed on the last 10 ns MD simulation trajectories. The calculation of absolute binding free energy was done by means of four highly efficient binding free energy methods: free energy perturbation (FEP), Bennett's acceptance ratio (BAR) method, thermodynamic integration (TI), and quadrature-based integration of TI. Good convergence of the predicted values ideally <1 kcal/mol among these statistical methods assessed the good agreement on the high complex stability [51]. The complete flow of steps followed in the current study is presented in Fig. 3.

3. Results and discussion

3.1. Selection of appropriate quality M^{Pro} structure

The first step in the study was the selection of quality M^{Pro} structure for docking simulation studies. This was accomplished by performing energy minimization in UCSF Chimera to get an equilibrium structure. High energy configurations contains steric clashes that potentially can result in a physical perturbation and instability during simulation. However, such minimization may also introduce bad contacts in the structure and affect the protein structure. Therefore, structure evaluation before and after minimization is significant to determine the best energy optimized structure for subsequent docking and simulation assays to put forward the best possible predictions. The energy minimized M^{Pro} achieved a minimal energy state as it has 89.8% residues in the favored region opposed to 87.9% in the pre-minimized M^{Pro} of the Ramachandran plot. Moreover, 9.1% of residues (10.9% for pre-minimized M^{Pro}), 1.1% (0.8% for pre-minimized M^{Pro}), 0% (0.4% for pre-minimized M^{Pro}) residues were plotted and allowed for regions

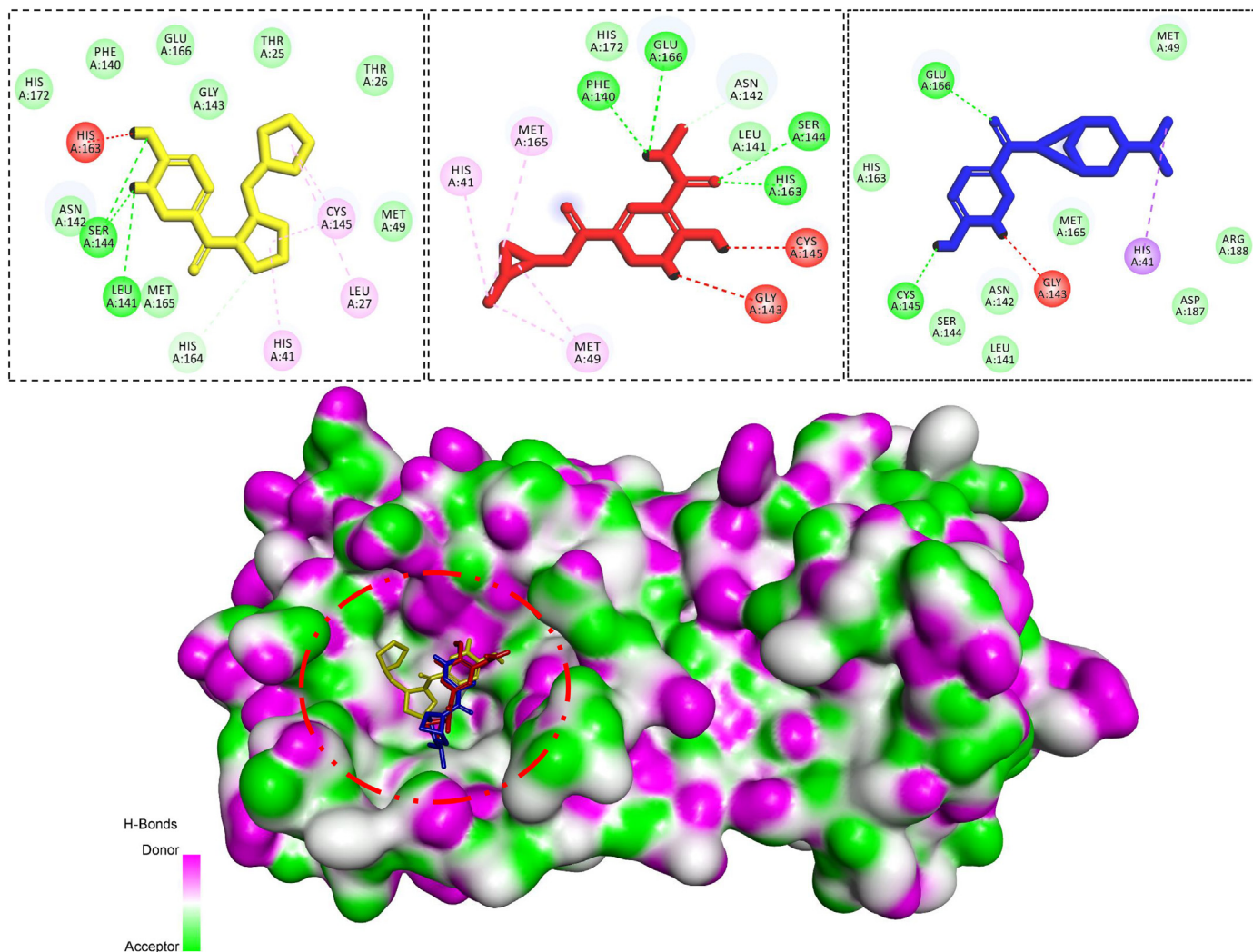


Fig. 7. Docked complex of SCHEMBL12616233 (yellow), SCHEMBL18616095 (red), and SCHEMBL20148701 (blue) at the functional pocket of M^{Pro} (shown in surface). The chemical interactions between the derivatives and functional site residues are also shown. For interpretation of the interactions, the dark green disc represents conventional hydrogen bonds, LIG-ht green discs (van der Waals), LIG-ht aquamarine discs (carbon hydrogen bond), red discs (unfavorable donor), pink discs (alky and pi alkyl bonds), and purple disc (pi sigma).

and disallowed regions, respectively, of the Ramachandran plot. Ramachandran plots of both structures can be seen in Fig. 4.

3.2. Unveiling a lead compound

A lead compound “BBB_26580140” or (1-(carboxymethyl)-5-(cyclopentylcarbamoyl)pyridin-1-ium-2-olate) was unveiled by SBVS, exhibiting good affinity for the functional site of M^{Pro}. The compound binding affinity is -8.1 kcal/mol. The compound selection was based on the fact that it gives an appropriate binding pose and interactions at the binding pocket of M^{Pro} (Fig. 5). The contribution of major interactions was observed from 1-(carboxymethyl)pyridin-1-ium-2-olate moiety of the compound interacting with several critical residues of the S1 subsite of the pocket. The propionic acid moiety of the compound reported to form a network of four hydrogen bonding with Leu141, Leu144, Cys145, and His163 at distance 5.9 Å, 2.8 Å, 4.0 Å, and 5.2 Å, respectively. The central pyridin-1-ium-2-olate moiety positioned closed towards the S1 subsite engaging Glu166 via hydrogen bonding at a distance of 4.12 Å. The N-cyclopentylacetamide moiety is extended in the direction of S' subsite where it is stabilized by establishing several hydrophobic contacts. The conformational stability of the lead molecule with M^{Pro} was considered further through MD simulation assay of 20 ns and the quality was estimated by plotting all C α atoms deviations

along the simulation time. Compared to the control co-crystallized carmofur-M^{Pro} complex (mean rmsd: 2.6 Å), our lead-M^{Pro} can be seen to be structurally stable (mean rmsd: 1.5 Å). Secondly, carmofur-M^{Pro} complex rmsd plot is showing some minor structural variations whereas lead-M^{Pro} rmsd plot is highly stable (Fig. 6-top-left). Further support on our lead-M^{Pro} complex stability was derived from the compound rmsd during the simulation time. Both the control co-crystallized lead and screened lead of this study revealed a significant stable nature (rmsd <1 Å), depicting ligand-reduced movements and complementing on the good agreement favorable binding affinity (Fig. 6-top-right). Detailed insights about protein compactness and conformation equilibrium was achieved by running Rg analysis on MD trajectories of both complexes over the simulation time. Results obtained demonstrated that both complexes are relatively compact (~ 40 Å) (Fig. 6-bottom-left). Hydrogen bonding assay revealed that the stability of both control and lead molecule complexes with M^{Pro} benefits from the regular formation of strong hydrogen bonds throughout simulation time (Fig. 6-bottom-right). The binding strength of our lead towards M^{Pro} was further estimated by MM/PB(GB)SA methods. The net binding energy of the lead-M^{Pro} complex is -43.0937 kcal/mol in MMGBSA that illustrates high stability of the complex compared to the -31.9077 kcal/mol total binding free energy in MMPBSA. To this net energy, the major contributor is the gas phase energy which is -68.0019 kcal/mol in both

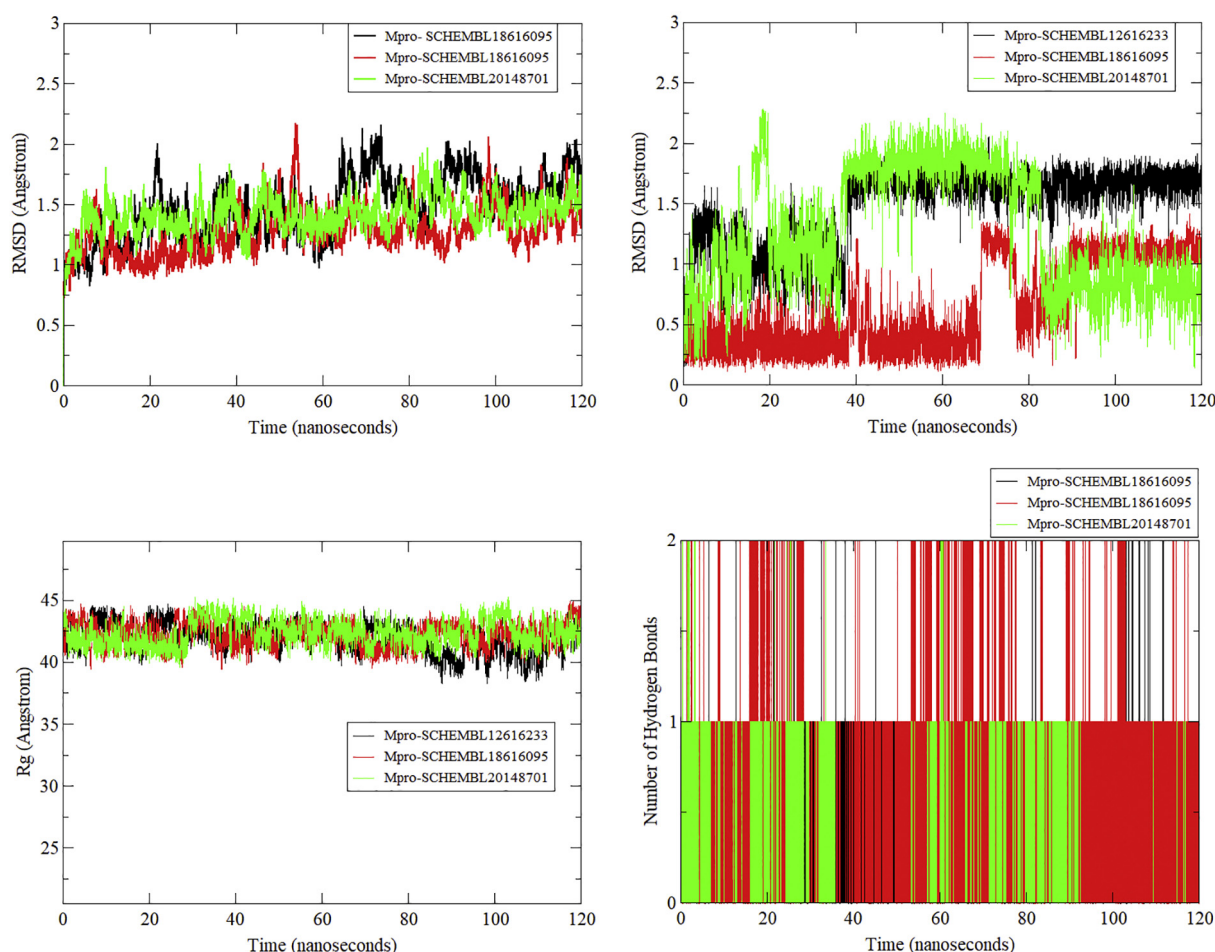


Fig. 8. Time dependent evolution of systems C_{α} rmsd (top left), ligand rmsd (top right), C_{α} Rg (bottom left), and number of hydrogen bond formation (bottom right).

approaches opposed to the highly unfavorable solvation energy (24.9082 kcal/mol in MMGBSA and 36.0942 kcal/mol in MMPBSA). The good gas phase energy is due to 2-fold extra van der Waals energy (−46.3815 kcal/mol) than electrostatic energy (−21.6204 kcal/mol). The polar electrostatic energy (29.0146 kcal/mol in MMGBA and 38.8359 kcal/mol) was found to be the main non-contributor to the solvation energy compared to non-polar energy that contributes favorably (−4.1064 kcal/mol in MMGBSA and −2.7418 kcal/mol in MMPBSA).

3.3. Pharmacokinetics and medicinal chemistry of the lead compound

Drug attrition due to bad pharmacokinetics in the drug discovery process leads to high developmental costs and take extra time [52]. The availability of in silico pharmacokinetics tools in this regard greatly improves the selection of appropriate drug molecules having the safest pharmacokinetics [53]. Therefore, a comprehensive in silico pharmacokinetic profile of the lead molecule was performed to guide future

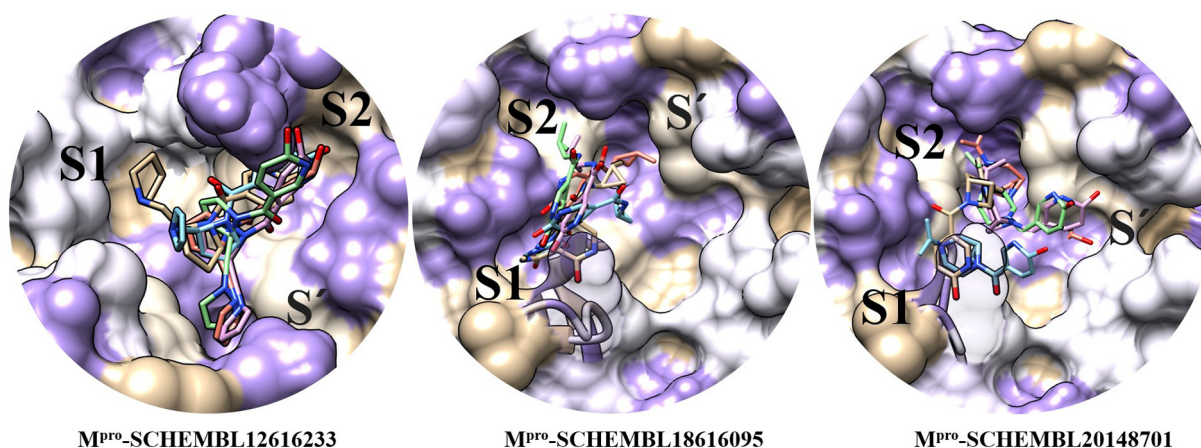
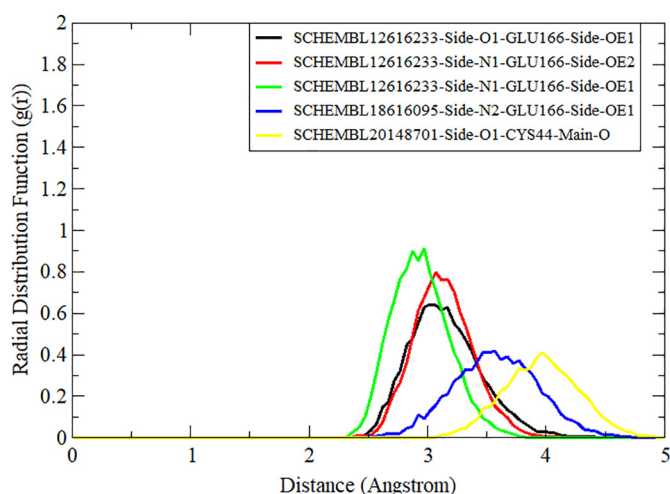


Fig. 9. Close view of compounds adjustment inside binding pocket of M^{Pro}. Snapshots at different time are shown as: 0-ns (tan), 20-ns (sky blue), 40-ns (plum), 80-ns (coral) and 120 (spring green).

Table 2

Time dependent hydrogen bond analysis of the complexes, along with occupancy.

Hydrogen bond analysis								
Mpro-SCHEMBL12616233			Mpro-SCHEMBL18616095			Mpro-SCHEMBL20148701		
Donor	Acceptor	Occupancy	Donor	Acceptor	Occupancy	Donor	Acceptor	Occupancy
LIG-307-Side-O1	GLU166-Side-OE1	3.65%	LIG-307-Side-N2	GLU166-Side-OE1	5.04%	LIG-307-Side-O1	CYS44-Main-O	2.41%
LIG-307-Side-N1	GLU166-Side-OE2	11.26%	LIG-307-Side-N2	GLU166-Side-OE2	2.20%	LIG-307-Side-O1	THR25-Side-OG1	1.49%
LIG-307-Side-N1	GLU166-Side-OE1	12.12%	LIG-307-Side-O2	GLU166-Side-OE2	0.74%	LIG-307-Side-O1	HIE41-Main-O	0.40%
LIG-307-Side-O1	ASN142-Side-OD1	2.73%	LIG-307-Side-O2	ASN142-Side-OD1	1.02%	LIG-307-Side-N2	HIE41-Main-O	0.50%
LIG-307-Side-O1	GLU166-Side-OE2	4.51%	LIG-307-Side-N2	ASN142-Side-OD1	0.12%	LIG-307-Side-O1	THR26-Main-O	2.37%
GLU166-Main-N	LIG-307-Main-O	0.32%	LIG-307-Main-N	ASN142-Side-OD1	0.12%	THR25-Side-OG1	LIG-307-Side-O1	0.02%
LIG-307-Side-N1	LEU141-Main-O	1.29%	LIG-307-Side-O2	GLU166-Side-OE1	0.35%	LIG-307-Side-N2	TYR54-Side-OH	0.06%
LIG-307-Side-N1	SER144-Side-OG	0.19%	GLN189-Side-NE2	LIG-307-Main-O	0.02%	LIG-307-Side-O1	TYR54-Side-OH	0.00%
GLY143-Main-N	LIG-307-Main-O	0.06%	GLN189-Side-NE2	LIG-307-Side-C7	0.01%	CYS145-Main-N	LIG-307-Main-O	0.02%
LIG-307-Side-O1	PHE140-Main-O	0.22%	LIG-307-Side-N1	GLN189-Side-NE2	0.01%	GLY143-Main-N	LIG-307-Main-O	0.00%
LIG-307-Side-O1	SER144-Side-OG	0.01%	LIG-307-Main-N	GLU166-Side-OE2	0.01%	LIG-307-Side-N2	ASN142-Side-OD1	0.32%
LIG-307-Side-N1	ASN142-Side-OD1	0.20%	GLU166-Main-N	LIG-307-Side-C7	0.04%	LIG-307-Side-N2	ASN142-Main-O	0.27%
HIE163-Side-NE2	LIG-307-Side-O1	0.02%	GLU166-Main-N	LIG-307-Side-C8	0.06%	LIG-307-Side-O1	ASN142-Side-OD1	0.12%
HIE163-Side-NE2	LIG-307-Side-C7	0.01%	LIG-307-Side-N2	LEU141-Main-O	0.08%	LIG-307-Side-O1	ASN142-Main-O	0.01%
LIG-307-Side-O1	LEU141-Main-O	0.01%	LIG-307-Side-N2	SER144-Side-OG	0.01%	LIG-307-Side-N2	GLY143-Main-O	0.01%
LIG-307-Side-N1	PHE140-Main-O	0.12%	LIG-307-Side-N1	GLN189-Side-OE1	0.32%	LIG-307-Side-N2	GLN189-Side-OE1	0.32%
LIG-307-Side-N1	GLU166-Main-O	0.01%	LIG-307-Side-N2	GLU166-Side-CD	0.01%	LIG-307-Side-O1	SER46-Main-O	0.73%
LIG-307-Side-O1	ASN142-Side-ND2	0.01%	LIG-307-Side-N1	HIE164-Main-O	0.40%	LIG-307-Side-O1	GLN189-Side-OE1	0.01%
-	-	-	LIG-307-Side-N2	PHE140-Main-O	0.01%	LIG-307-Side-N2	GLN189-Side-NE2	0.00%
-	-	-	LIG-307-Main-N	GLU166-Side-OE1	0.02%	LIG-307-Side-N2	SER46-Side-OG	0.00%
-	-	-	GLU166-Main-N	LIG-307-Side-O1	0.01%	CYS145-Side-SG	LIG-307-Main-O	0.02%
-	-	-	-	-	-	LIG-307-Side-N2	CYS44-Main-O	0.00%

**Fig. 10.** RDF plots for compounds interactions.

chemists in optimizing the structure, keeping pharmacokinetics within the acceptable range. The detailed pharmacokinetics of the lead molecules are provided in Table 1. Drug absorption is a principal focus in medicinal chemistry and was evaluated first in the *in silico* pharmacokinetics studies [54]. This lead molecule is water soluble which is key to oral bioavailability [55]. The molecule is predicted to have good potential to permeate Caco-2 cell lines *in vitro*. The gastrointestinal absorption is also high and the compound is not a substrate for P-glycoprotein transporter [56]. Skin permeability prediction found the compound to be skin permeable making it an extremely good candidate in developing good transdermal delivery products [57]. From the distribution point of view, the molecule has volume of distribution (VD_{ss}) value of -0.546 , indicating the low distribution of the compound in the tissue compared to the plasma [58]. Similarly, the compound fraction unbound (F_u) value is low which could result in its low binding affinity with the serum protein and can improve its diffusion efficiency of the cellular membranes [37]. The capacity of drugs to cross the blood brain barrier (BBB) is critical for evaluation to avoid

toxicity and side effects of a given molecule and to improve its efficiency if its pharmacological activity is within the brain [59]. The compound had poor permeability of the BBB and could not cross the central nervous system (CNS). The lead molecule is a non-inhibitor of detoxifying cytochrome P450 thus allow functional oxidation of xenobiotics and facilitate their excretion. The predicted total clearance of the compound which combines both hepatic clearance and renal clearance was $0.53 \log \text{ ml/min/kg}$. This factor is critical in bioavailability and for calculation of the dosage rate for steady state concentration. From a toxicity perspective, the compound is AMES non-toxic, has a very low LD₅₀ value when administered orally to rats, is predicted to show no skin sensitization and doesn't inhibit hERG I and hERG II thus reducing the chance of developing QT syndrome [60,61]. Medical chemistry prediction categorized the compound as zero alert PAINS (pans assay interference structure) [62]. The compound had good synthetic accessibility and revealed zero alert for lead likeness.

3.4. Lead based similarity search

The study was further extended to identify analogs of the lead molecule to provide an enriched set of molecules that could be used either directly as anti-COVID 19 drugs or might be useful for providing further leads. In total, 33 analogs of the lead molecule were identified from ChEMBLdb as shown in S-Fig. 1. We docked this set of molecules at the same function site described previously. We prioritized three complexes based on their affinity for the M^{Pro} and examined them extensively. These included SCHEMBL12616233 (6-hydroxypyridin-3-yl)(2-(pyrrolidin-1-ylmethyl)pyrrolidin-1-yl)methanone, SCHEMBL18616095 (2-hydroxy-N3-methyl-N5-((1R,2R)-2-methylcyclopropyl)pyridine-3,5-dicarboxamide) and SCHEMBL20148701 ((6-hydroxypyridin-3-yl)(3-isopropyl-3,6-diazabicyclo[3.1.1]heptan-6-yl)methanone) with binding affinity of -9.5 kcal/mol , -9.1 kcal/mol and -9.0 kcal/mol for M^{Pro} functional site, respectively. Much of the docking affinity of the derivatives, like the lead, is the output of interactions dominated by hydrogen bonding with the S1 site of the active site (Fig. 7). Specifically, in the case of SCHEMBL12616233 the pyridin-2-ol favored the S1 residues of the M^{Pro} via balanced hydrogen (with Leu141 and Ser144) and van der Waals bondings. This allows 1-(2-methylpyrrolidin-1-yl)ethanone moiety positioning to the S2 site and

Table 3
Different binding free energies values in kcal/mol for screened analogs predicted by MM-GBSA and MM-PBSA methods. Red to blue color corresponds to non-contributor to the excellent contributor.

Energy Component	MM-GBSA					MM-PBSA					ENPOLAR (MMPBSA non polar solvation energy), ENPOLAR (MMPBSA polar solvation energy), EPB (MMPBSA non-polar solvation energy), EPB (MMPBSA polar solvation energy), EGB (MMGBSA polar solvation energy), ESURF (MMGBSA non-polar solvation energy), DELTA G solv (net solvation energy), DELTA TOTAL (net system energy).				
	Average					Std. Dev.					Std. Err. of Mean				
	SCHEMBL12616233	SCHEMBL18616095	SCHEMBL20148701	SCHEMBL12616233	SCHEMBL18616095	SCHEMBL20148701	SCHEMBL12616233	SCHEMBL18616095	SCHEMBL20148701	SCHEMBL12616233	SCHEMBL18616095	SCHEMBL20148701	SCHEMBL12616233	SCHEMBL18616095	SCHEMBL20148701
VDWAALS	-28.69	-24.87	-31.78	2.98	2.4	2.68	0.29	0.24	0.26	0.29	0.24	0.26	0.29	0.24	0.26
EEL	-101.63	-108.7	-79.46	6.76	8.32	6.58	0.67	0.83	0.65	0.67	0.83	0.65	0.67	0.83	0.65
EGB	110.9	118.62	90.56	6.04	7.54	6.29	0.6	0.75	0.62	0.6	0.75	0.62	0.6	0.75	0.62
ESURF	-2.92	-3.15	-3.06	0.24	0.15	0.18	0.02	0.01	0.01	0.02	0.01	0.01	0.02	0.01	0.01
DELTA G gas	-130.32	-133.57	-111.25	7.53	8.08	7.33	0.75	0.8	0.73	0.75	0.8	0.73	0.75	0.8	0.73
DELTA G solv	107.97	115.46	87.5	5.96	7.49	6.18	0.59	0.74	0.61	0.59	0.74	0.61	0.59	0.74	0.61
DELTA TOTAL	-22.35	-18.1	-23.74	2.73	1.82	2.58	0.27	0.18	0.25	0.27	0.18	0.25	0.27	0.18	0.25
MM-PBSA															
VDWAALS	-28.69	-24.87	-31.78	2.98	2.4	2.68	0.29	0.24	0.26	0.29	0.24	0.26	0.29	0.24	0.26
EEL	-101.63	-108.7	-79.46	6.76	8.32	6.58	0.67	0.83	0.65	0.67	0.83	0.65	0.67	0.83	0.65
EPB	110.26	120.68	94.37	6.22	7.72	6.55	0.62	0.77	0.65	0.62	0.77	0.65	0.62	0.77	0.65
ENPOLAR	-2.48	-2.4	-2.47	0.12	0.11	0.07	0.01	0.01	0	0.01	0.01	0	0.01	0.01	0
DELTA G gas	-130.32	-133.57	-111.25	7.53	8.08	7.33	0.75	0.8	0.73	0.75	0.8	0.73	0.75	0.8	0.73
DELTA G solv	107.78	118.27	91.9	6.18	7.69	6.52	0.61	0.76	0.65	0.61	0.76	0.65	0.61	0.76	0.65
DELTA TOTAL	-22.54	-15.3	-19.35	3.52	2.8	3	0.35	0.28	0.3	0.35	0.28	0.3	0.35	0.28	0.3

leaving downward tilting of the 1-methylpyrrolidine to the S' site. The latter two moieties revealed to interact through a mixture of weak bonding's including van der Waals, carbon hydrogen, alkyl, and pi alkyl bonds. SCHEMBL18616095 conformer is adjusted mainly to S site whereas pushing the tail ((1R,2R)-2-methylcyclopropanamine) to the S2 site of the functional cavity. The bulk of the interactions is the output of strong hydrogen bonding between 2-hydroxy-N-methylnicotinamide of the compound with S1 site residues (Phe140, Ser144, His163, and Glu166). SCHEMBL20148701 occupies the S1 and S2 sites and interact with residues of both sites by hydrophobic and hydrophilic interactions.

3.5. Conformational dynamics evaluation

Conformational stability of each system was assessed by running long MD simulations of 120-ns. System trajectories were statistically analyzed for structural parameters such as C α atoms root mean square deviation (C α rmsd) and C α atoms radius of gyration (C α atoms Rg). C α Rmsd was calculated by superimposing a complete set of complex simulated conformers over initially minimized conformer and plotted versus simulation time. As can be observed in Fig. 8 the systems behave quite similarly and consistent stability can be seen. The mean C α rmsd for M^{pro}-SCHEMBL12616233, M^{pro}-SCHEMBL18616095, and M^{pro}-SCHEMBL20148701 is 1.4 Å, 1.28 Å and 1.57 Å, respectively (Fig. 8-top left). These findings affirm the stability of each system where the lead derivatives sit at the active pocket of M^{pro} owing to enhanced intermolecular affinity. Further, investigation of the ligands rmsd was conducted. This unravels the flexibility of the ligands in the pocket at specific times followed by enhanced conformational stability towards the end of the simulation. (Fig. 8-top right and Fig. 9). For SCHEMBL12616233, compound mean rmsd is 1.49 Å, and flexibility is seen at 40-ns. This flexibility allows compound stretching in the pocket rendering central 1-(2-methylpyrrolidin-1-yl)ethanone positioning away from the S2 site towards the pocket center and easing 1-methylpyrrolidine contact to the S' site. Afterward ligand rmsd deviations corresponded to original docked conformation where pyridin-2-ol adjusting itself very close to the S1 site, central 1-(2-methylpyrrolidin-1-yl)ethanone to the S1 and S2 junction and 1-methylpyrrolidine to the S' site. M^{pro}-SCHEMBL18616095 mean rmsd is 0.66 Å and achieved a more stable conformation compared to the other two ligands despite minor structural adjustment moves at a period of 70-ns to 76-ns. The same adjustments acquired by the first compound were adopted by SCHEMBL18616095 coming close to the S2 site via its tail. SCHEMBL20148701 was reported to have flexibility at two points; at 16-ns where the compound showed minor conformer variations, followed by a second structural variation at time 40 ns-80-ns. Both these flexibilities in compound rmsd was a result of the outcome compound detaching from the S1 site, moving towards the junction between S2 and S' sites. Visual inspection of the simulation trajectories demonstrated such binding mode changes of the compounds could be an approach towards gaining a stable mode as complemented by stable rmsd trend at the end period of simulation. For additional insights on system equilibrium conformation and protein structure compactness, the radius of gyration (Rg) was performed over the simulated trajectories. The mean Rg value for the systems is in the following order: M^{pro}-SCHEMBL12616233 (42.3 Å), M^{pro}-SCHEMBL18616095 (42.21 Å) and M^{pro}-SCHEMBL20148701 (42.61 Å), suggesting good overall protein compactness (Fig. 8-bottom left). These findings also correlated to systems rmsd and the present stability of the protein structural elements during the simulation time. Summing up, MD simulation predicted systems stability making the compound a good candidate be subjected to experimental studies to decipher its real affinity for the M^{pro}.

3.6. Hydrogen bond analysis

Hydrogen bond analysis was done to determine the pattern and duration of hydrogen bonding between the protein and ligand molecules. Hydrogen bonds are the output when a hydrogen atom is shared between the heavy atom donor and acceptor. Such interactions are vital in underpinning intermolecular specificity and are critical to stable protein-ligand complexes. Hydrogen bond formation for M^{pro}-SCHEMBL12616233, M^{pro}-SCHEMBL18616095, and M^{pro}-SCHEMBL20148701 was plotted, considering the cut-off 3.0 Å and cut-off angle of 20 degree. Based on such conditions, all three systems illustrated the consistent formation of hydrogen bonds between the protein and ligands (Fig. 8-bottom right). In total, there were 18, 21, and 22 hydrogen bonds for M^{pro}-SCHEMBL12616233, M^{pro}-SCHEMBL18616095, and M^{pro}-SCHEMBL20148701, respectively. Details of the hydrogen bonding for the complexes along with percent occupancy are tabulated in Table 2.

3.7. RDF analysis

The strong hydrogen bonds between the compounds and the M^{pro} were further utilized in RDF analysis to interpret interactions intensity versus time (Fig. 10). Three interactions from M^{pro}-SCHEMBL12616233 were reported consistently in simulation all involving a Glu166 residue from the active pocket. The pyridin-2-ol ring of this compound favors most of these interactions especially through its oxygen and nitrogen atoms. The rdf plot for the pyridin-2-ol ring of SCHEMBL12616233 with side OE1 of Glu166 represents the highest interaction density at 2.97 Å with g(r) value of 0.91. The second interaction with high interaction density (g(r) value of 0.79) is between compound N1 from pyridin-2-ol ring and Glu166 OE2 atom at distance 3.0 Å. RDF of pyridin-2-ol O1 atoms with the Glu166 OE1 showed less interaction density compared to first two but seems crucial in holding the compound at S1 site. The maximum g(r) of this interaction is 0.64 at distance of 3.04 Å. All the above interactions via RDF quantify the fact of strong affinity of the compound for the S1 site of the M^{pro}. Only one SCHEMBL18616095 chemical association with Glu166 OE1 atom is more dispersed (maximum g(r) value of 0.4 at distance 3.5 Å) which is due to inside active pocket moves. However, still, the compound is demonstrating close affinity for the S1 site. SCHEMBL20148701 RDF showed less interaction density which is attributed to later simulation time. This is because the compound moves from the initial docked site S1 to S1 and S2 junction towards simulation end. The maximum g(r) value of the interaction is 0.40 at a distance of 3.98 Å.

3.8. Estimation of binding free energies

The binding free energy estimation via MM-GBSA and MM-PBSA provides an excellent platform to affirm the affinity of compounds for its receptor in drug discovery protocols [35]. Both techniques can easily be adopted in virtual screening processes and correlate well with experimental activities. All three analogs of the lead showed significantly higher affinity for the M^{pro}, stability at docked site, and enrich pattern of chemical interactions (Table 3). In both protocols, the gas phase energy calculated by molecular mechanics and where no contribution of the solvent is considered found as major driving contributor in complexes stabilization. The gas phase energy for M^{pro}-SCHEMBL12616233 complex, M^{pro}-SCHEMBL18616095 and M^{pro}-SCHEMBL20148701 complex is −130.32 kcal/mol, −133.57 kcal/mol, and −111.25 kcal/mol, respectively. The main contributor to the gas phase energy is electrostatic energy; that is −101.63 kcal/mol for M^{pro}-SCHEMBL12616233, −108.7 kcal/mol for M^{pro}-SCHEMBL18616095, and −79.46 kcal/mol for M^{pro}-SCHEMBL20148701. The van der Waals bond also provides a favorable contribution to the net gas phase energy ranging from −24.87 kcal/mol for M^{pro}-SCHEMBL18616095 to −31.78 kcal/mol for M^{pro}-SCHEMBL20148701 and −28.69 kcal/mol for M^{pro}-SCHEMBL12616233. The net solvation energy for each system is highly non-favorable and the

importance of water molecules in interactions between the receptor and LIG-and seems small. The net solvation energy of MM-GBSA is M^{pro}-SCHEMBL12616233 (107.97 kcal/mol in MM-GBSA and 107.78 kcal/mol in MM-PBSA), M^{pro}-SCHEMBL18616095 (115.46 kcal/mol in MM-GBSA and 118.27 kcal/mol in MM-PBSA), and M^{pro}-SCHEMBL20148701 (87.5 kcal/mol in MM-GBSA and 91.9 kcal/mol in MM-PBSA). The electrostatic energy in solvation phase is the major insignificant contributor and is found to be above 100 kcal/mol for M^{pro}-SCHEMBL12616233 and M^{pro}-SCHEMBL18616095 and above 90 kcal/mol for M^{pro}-SCHEMBL20148701.

3.9. Hotspot M^{pro} residues

The binding interaction pattern of the derivatives was further elucidated by implementing MMPBSA decomposition analysis to clarify residues contributing to inhibitor binding. As all three compounds prefer to bind to the S1 site of the pocket, the majority of the compound stabilizing residues are the same. Another reason for such a similar binding pattern is due to less mobility of the compounds at the docked site. The hotspot residues make regular contacts with the compound during the simulation time and have binding free energy less < −1 kcal/mol. The hotspot residues can be seen in Table 4. For SCHEMBL12616233, 8 residues (Leu27, Met49, Hie41, Cys145, Hie164, Met165, Glu166, Arg188) whereas in case of SCHEMBL18616095 and SCHEMBL20148701 14 (Met49, Hie41, Phe140, Leu141, Ser144, Cys145, Hie163, Hie164, Met165, Glu166, Hie172, Asp187, Arg188, Gln189) and 16 (Leu25, Leu27, Met49, Hie41, Phe140, Leu141, Ser144, Cys145, Hie163, Hie164, Met165, Glu166, Hie172, Asp187, Arg188, Gln189) residues were categorized as hotspot residues. Many of the residues of the compounds are common in all three complexes reflecting their vital role in holding compound at the pocket site and are of considerable interest in future lead optimization. Glu166 in case of M^{pro}-SCHEMBL12616233 complex strongest binding and stabilizing factor for the compound with binding energy of −2.54 kcal/mol. The residues hold the compound via two hydrogen bonding at close distance. Similarly, in the M^{pro}-SCHEMBL20148701 complex Cys145 (−4.1 kcal/mol) shows a strong magnitude of binding affinity for the compound. The M^{pro}-SCHEMBL20148701 interaction is in equilibrium due to the strong affinity exhibition from Cys145 and Hie145 for the compound at strong hydrogen bonds.

3.10. Estimation of binding entropy

Since estimation of binding entropy was skipped for the simulated systems in MM-GBSA and MM-PBSA, a follow up normal mode analysis

Table 4

Hotspot residues of the M^{pro} allowing major interactions and stabilization of the compounds.

Residue	Binding energy in kcal/mol		
	SCHEMBL12616233	SCHEMBL18616095	SCHEMBL20148701
Leu27	−1.36	0	−1.19
Leu25	0	0	−1
Met49	−1	−1.24	−1.79
Hie41	−1.11	−2.56	−1.82
Phe140	0	−2.89	−0.142
Leu141	0	−1.45	−1.01
Ser144	0	−3.5	−2.18
Cys145	−1.47	−4.1	−3.54
Hie163	0	−1.22	−1.1
Hie164	−1	−1	−4.12
Met165	−1.12	−2.45	−2.14
Glu166	−2.54	−2.86	−1
Hie172	0	−3.4	0
Asp187	0	−1.87	−1.23
Arg188	−1.11	−2.12	−1.24
Gln189	0	−2.48	−1.58
Thr190	0	0	0

Table 5

Normal mode binding entropy estimation for the simulated systems. All units are presented in kcal/mol.

	Entropy Term	SCHEMBL12616233	SCHEMBL18616095	SCHEMBL20148701
Complex	Translational	10	9	13
	Rotational	9	8	13
	Vibrational	2594	2585	2678
	Total	2630	2633	2751
Receptor	Translational	12	11	12
	Rotational	11	11	10
	Vibrational	2546	2548	2660
	Total	2569	2570	2682
Ligand	Translational	11	13	14
	Rotational	12	14	13
	Vibrational	34	31	36
	Total	57	58	63
ΔS total	ΔS	4	5	6

Table 6

Absolute binding free energy (in kcal/mol) calculated with WaterSwap.

Complex	BAR	WaterSwap FEP	TI
M ^{Pro} -SCHEMBL12616233	−15.58	−16.78	−15.74
M ^{Pro} -SCHEMBL18616095	−14.51	−15.84	−15.78
M ^{Pro} -SCHEMBL20148701	−18.91	−19.74	−18.76

for entropy calculation was applied to the systems in AMBER package. All three systems clearly can be seen in an entropically unfavorable state which means that ligands are trapped inside the M^{Pro} pocket and have limited mobility. Consequently, the systems have achieved decreased microstate levels and enhanced stability. The entropy energy for each system is tabulated in Table 5.

3.11. WaterSwap absolute binding free energy estimation

To remove limitations of the MM-GBSA and MM-PBSA implicit water model, a more reliable method of WaterSwap that uses an explicit water model was inducted in the study to validate and support the binding free energy conclusion for the systems. WaterSwap uses an algorithm of replica-exchange thermodynamic integration in combination with reaction coordinates that allows swapping of a given ligand to an equivalent explicit water volume present in protein binding pocket. WaterSwap was successfully applied on several snapshots of the last 10-ns simulation trajectories that give an average — 16.02 kcal/mol for M^{Pro}-SCHEMBL12616233, —15.37 kcal/mol for M^{Pro}-SCHEMBL18616095 and —19.13 kcal/mol for M^{Pro}-SCHEMBL20148701. This concludes the higher strength of interactions between the M^{Pro} and compounds. These absolute energies may overestimate the true binding energy as the entropy contribution is ignored during energy estimation. Detailed results of WaterSwap are provided in Table 6.

4. Conclusions

Using rigorous in silico approaches we have successfully determine several antiviral compounds exhibiting high potential against SARS-CoV-2 M^{Pro}. The compounds have improved binding affinity compared to co-crystallized carmofur. Druggability and leadlike rules assessment illustrated the compounds fulfilling all prominent rules and exhibit safe and acceptable pharmacokinetics. MD simulations uncovered compound mobility inside the active pocket that acquired binding mode equilibrium towards end simulation time. The compounds make multiple hydrogen and van der Waals interactions with hot spot residues of the enzyme pocket that contribute to enhancing intermolecular affinity. In particular, all the screened lead molecules interact and form strong

bonding with His41, Cys145, His163, and Glu166. These residues are reported in all SARS-CoV-2 M^{Pro} crystal structures to form bonding with co-crystallized ligands [63]. The binding pocket and catalytic residues of this enzyme are highly conserved among the coronaviruses [21]. So despite of sequence variations in the protein, it is highly unlikely that the virus will evade the screen inhibitory molecules action. As a result of strong interactions, the compounds complexes achieved higher stability by scoring negative binding energies.

Different classes of protease inhibitors are in clinical trials against COVID-19. Hits identified in this study belong to different classes and can accommodate appropriately inside the M^{Pro} pocket. As the compounds can be readily accessed from ASINEX antiviral library and ChEMBL databases and do not require the prior need of synthesis, the compounds can be quickly explored in vivo experimentation for binding affinity and inhibition potential. Concluding, the screened compounds are promising and may provide a good foundation in the hunt of anti-COVID-19 M^{Pro} drug discovery.

Disclosure statement

The authors in this study have no conflict of interest.

Declaration of Competing Interest

The authors declare that they have no conflict of interest.

Acknowledgment

Dr. Yasir Waheed is highly grateful to the Foundation University Islamabad, Islamabad, Pakistan, for granting financial assistance to accomplish this work.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.molliq.2021.115636>.

References

- [1] C. Wang, P.W. Horby, F.G. Hayden, G.F. Gao, A novel coronavirus outbreak of global health concern, *Lancet* 395 (10223) (2020) 470–473, [https://doi.org/10.1016/S0140-6736\(20\)30185-9](https://doi.org/10.1016/S0140-6736(20)30185-9).
- [2] N. Zhu, D. Zhang, W. Wang, X. Li, B. Yang, J. Song, X. Zhao, B. Huang, W. Shi, R. Lu, P. Niu, F. Zhan, X. Ma, D. Wang, W. Xu, G. Wu, G.F. Gao, W. Tan, A novel coronavirus from patients with pneumonia in China, 2019, *N. Engl. J. Med.* 0 (n.d.) null. doi: <https://doi.org/10.1056/NEJMoa2001017>.
- [3] F. Wu, S. Zhao, B. Yu, Y.-M. Chen, W. Wang, Z.-G. Song, Y. Hu, Z.-W. Tao, J.-H. Tian, Y.-Y. Pei, et al., A new coronavirus associated with human respiratory disease in China, *Nature* 579 (2020) 265–269.
- [4] P. Zhou, X.-L. Yang, X.-G. Wang, B. Hu, L. Zhang, W. Zhang, H.-R. Si, Y. Zhu, B. Li, C.-L. Huang, et al., A pneumonia outbreak associated with a new coronavirus of probable bat origin, *Nature* 579 (2020) 270–273.

- [5] H. Li, S.-M. Liu, X.-H. Yu, S.-L. Tang, C.-K. Tang, Coronavirus disease 2019 (COVID-19): current status and future perspective, *Int. J. Antimicrob. Agents* 55 (5) (2020), 105951, <https://doi.org/10.1016/j.ijantimicag.2020.105951>.
- [6] B. Cao, Y. Wang, D. Wen, W. Liu, J. Wang, G. Fan, L. Ruan, B. Song, Y. Cai, M. Wei, et al., A trial of lopinavir–ritonavir in adults hospitalized with severe Covid-19, *N. Engl. J. Med.* 382 (19) (2020) 1787–1799, <https://doi.org/10.1056/NEJMoa2001282>.
- [7] M. Wang, R. Cao, L. Zhang, X. Yang, J. Liu, M. Xu, Z. Shi, Z. Hu, W. Zhong, G. Xiao, Remdesivir and chloroquine effectively inhibit the recently emerged novel coronavirus (2019-nCoV) in vitro, *Cell Res.* 30 (3) (2020) 269–271, <https://doi.org/10.1038/s41422-020-0282-0>.
- [8] Y. Chen, Q. Liu, D. Guo, Emerging coronaviruses: genome structure, replication, and pathogenesis, *J. Med. Virol.* 92 (4) (2020) 418–423, <https://doi.org/10.1002/jmv.25681>.
- [9] Y.-Z. Zhang, E.C. Holmes, A genomic perspective on the origin and emergence of SARS-CoV-2, *Cell* 181 (12) (2020) 223–227, <https://doi.org/10.1016/j.cell.2020.03.035>.
- [10] D. Kim, J.-Y. Lee, J.-S. Yang, J.W. Kim, V.N. Kim, H. Chang, The architecture of SARS-CoV-2 transcriptome, *Cell* 181 (4) (2020) 914–921, <https://doi.org/10.1016/j.cell.2020.04.011>.
- [11] L. Zhang, D. Lin, X. Sun, U. Curth, C. Drosten, L. Sauerhering, S. Becker, K. Rox, R. Hilgenfeld, Crystal structure of SARS-CoV-2 main protease provides a basis for design of improved α -ketoamide inhibitors, *Science* 368 (80) (2020) 409–412.
- [12] W. Dai, B. Zhang, X.-M. Jiang, H. Su, J. Li, Y. Zhao, X. Xie, Z. Jin, J. Peng, F. Liu, et al., Structure-based design of antiviral drug candidates targeting the SARS-CoV-2 main protease, *Science* 368 (80) (2020) 1331–1335.
- [13] M. Gui, W. Song, H. Zhou, J. Xu, S. Chen, Y. Xiang, X. Wang, Cryo-electron microscopy structures of the SARS-CoV spike glycoprotein reveal a prerequisite conformational state for receptor binding, *Cell Res.* 27 (2017) 119–129.
- [14] Z. Jin, Y. Zhao, Y. Sun, B. Zhang, H. Wang, Y. Wu, Y. Zhu, C. Zhu, T. Hu, X. Du, et al., Structural basis for the inhibition of SARS-CoV-2 main protease by antineoplastic drug carmofur, *Nat. Struct. Mol. Biol.* 27 (2020) 529–532.
- [15] Y.W. Chen, C.-P.B. Yiu, K.-Y. Wong, Prediction of the SARS-CoV-2 (2019-nCoV) 3C-like protease (3CL pro) structure: virtual screening reveals velpatasvir, ledipasvir, and other drug repurposing candidates, *F1000Research* 9 (2020).
- [16] D. Gentile, V. Patamia, A. Scala, M.T. Sciortino, A. Piperno, A. Rescifina, Putative inhibitors of SARS-CoV-2 main protease from a library of marine natural products: a virtual screening and molecular modeling study, *Mar. Drugs* 18 (2020) 225.
- [17] A.-T. Ton, F. Gentile, M. Hsing, F. Ban, A. Cherkasov, Rapid identification of potential inhibitors of SARS-CoV-2 main protease by deep docking of 1.3 billion compounds, *Mol. Inform.* 39 (8) (2020), e2000028, <https://doi.org/10.1002/minf.202000028>.
- [18] S. Das, S. Sarmah, S. Lyndem, A. Singha Roy, An investigation into the identification of potential inhibitors of SARS-CoV-2 main protease using molecular docking study, *J. Biomol. Struct. Dyn.* (2020) 1–18.
- [19] Y. Kumar, H. Singh, C.N. Patel, In silico prediction of potential inhibitors for the Main protease of SARS-CoV-2 using molecular docking and dynamics simulation based drug-repurposing, *J. Infect. Public Health* 13 (9) (2020) 1210–1223, <https://doi.org/10.1016/j.jiph.2020.06.016>.
- [20] S.A. Khan, K. Zia, S. Ashraf, R. Uddin, Z. Ul-Haq, Identification of chymotrypsin-like protease inhibitors of SARS-CoV-2 via integrated computational approach, *J. Biomol. Struct. Dyn.* (2020) 1–13.
- [21] R.S. Joshi, S.S. Jagdale, S.B. Bansode, S.S. Shankar, M.B. Tellis, V.K. Pandya, A. Chugh, A.P. Giri, M.J. Kulkarni, Discovery of potential multi-target-directed ligands by targeting host-specific SARS-CoV-2 structurally conserved main protease, *J. Biomol. Struct. Dyn.* (2020) 1–16.
- [22] E.F. Pettersen, T.D. Goddard, C.C. Huang, G.S. Couch, D.M. Greenblatt, E.C. Meng, T.E. Ferrin, UCSF chimera—a visualization system for exploratory research and analysis, *J. Comput. Chem.* 25 (2004) 1605–1612.
- [23] D.A. Case, V. Babin, J.T. Berryman, R.M. Betz, Q. Cai, D.S. Cerutti, T.E. Cheatham III, T.A. Darden, R.E. Duke, H. Gohlke, et al., The FF14SB force field, *Amber* 14 (2014) 29–31.
- [24] D.A. Case, D.S. Cerutti, T.E. Cheatham, T.A. Darden, R.E. Duke, T.J. Giese, H. Gohlke, A.W. Goetz, D. Greene, N. Homeyer, et al., AMBER16 Package, Univ. California, San Fr., 2016.
- [25] R.A. Laskowski, PDBsum: summaries and analyses of PDB structures, *Nucleic Acids Res.* 29 (2001) 221–222.
- [26] R.W.W. Hooft, C. Sander, G. Vriend, Objectively judging the quality of a protein structure from a Ramachandran plot, *Bioinformatics* 13 (1997) 425–430.
- [27] D.S. Biovia, Discovery Studio Visualizer, San Diego, CA, USA, 2017.
- [28] A. Daina, O. Michielin, V. Zoete, SwissADME: a free web tool to evaluate pharmacokinetics, drug-likeness and medicinal chemistry friendliness of small molecules, *Sci. Rep.* 7 (2017) 1–13, <https://doi.org/10.1038/srep42717>.
- [29] C.-Y. Jia, J.-Y. Li, G.-F. Hao, G.-F. Yang, A drug-likeness toolbox facilitates ADMET study in drug discovery, *Drug Discov. Today* 25 (1) (2019) 248–258, <https://doi.org/10.1016/j.drudis.2019.10.014>.
- [30] C.A. Lipinski, Lead- and drug-like compounds: the rule-of-five revolution, *Drug Discov. Today Technol.* 1 (2004) 337–341, <https://doi.org/10.1016/j.ddtec.2004.11.007>.
- [31] S. Dallakyan, A.J. Olson, Small-molecule library screening by docking with PyRx, *Chem. Biol. Springer* 2015, pp. 243–250.
- [32] M.A. Alamri, M. Ul Qamar, M.U. Mirza, R. Bhadane, S.M. Alqahtani, I. Muneer, M. Froeyen, O.M.H. Salo-Ahen, Pharmacoinformatics and molecular dynamics simulation studies reveal potential covalent and FDA-approved inhibitors of SARS-CoV-2 main protease 3CLpro, *J. Biomol. Struct. Dyn.* (2020) 1–13.
- [33] J.D. Durrant, J.A. McCammon, Molecular dynamics simulations and drug discovery, *BMC Biol.* 9 (2011) 71.
- [34] B.R. Miller, T.D. McGee, J.M. Swails, N. Homeyer, H. Gohlke, A.E. Roitberg, MMPBSA.py: an efficient program for end-state free energy calculations, *J. Chem. Theory Comput.* 8 (2012) 3314–3321, <https://doi.org/10.1021/ct300418h>.
- [35] S. Genheden, U. Ryde, The MM/PBSA and MM/GBSA methods to estimate ligand-binding affinities, *Expert Opin. Drug Discovery* 10 (2015) 449–461, <https://doi.org/10.1517/17460441.2015.1032936>.
- [36] A. Gaulton, A. Hersey, A.P. Nowotka Michalando Bento, J. Chambers, D. Mendez, P. Mutowo, F. Atkinson, L.J. Bellis, E. Cibrián-Uhalte, et al., The ChEMBL database in 2017, *Nucleic Acids Res.* 45 (2017) D945–D954.
- [37] D.E.V. Pires, T.L. Blundell, D.B. Ascher, pkCSM: Predicting small-molecule pharmacokinetic and toxicity properties using graph-based signatures, *J. Med. Chem.* 58 (2015) 4066–4072, <https://doi.org/10.1021/acs.jmedchem.5b00104>.
- [38] J. Wang, R.M. Wolf, J.W. Caldwell, P.A. Kollman, D.A. Case, Development and testing of a general amber force field, *J. Comput. Chem.* 25 (2004) 1157–1174.
- [39] C. Schafmeister, W.S. Ross, V. Romanovski, The Leap Module of AMBER, Univ. Calif., 1995.
- [40] J.A. Izaguirre, D.P. Catarella, J.M. Wozniak, R.D. Skeel, Langevin stabilization of molecular dynamics, *J. Chem. Phys.* 114 (2001) 2090–2098.
- [41] V. Krätter, W.F. Van Gunsteren, P.H. Hünenberger, A fast SHAKE algorithm to solve distance constraint equations for small molecules in molecular dynamics simulations, *J. Comput. Chem.* 22 (2001) 501–508.
- [42] S.E. Feller, Y. Zhang, R.W. Pastor, B.R. Brooks, Constant pressure molecular dynamics simulation: the Langevin piston method, *J. Chem. Phys.* 103 (1995) 4613–4621.
- [43] A.S. Lemak, N.K. Balabaev, On the Berendsen thermostat, *Mol. Simul.* 13 (1994) 177–187.
- [44] D.R. Roe, T.E. Cheatham III, PTRAJ and CPPTRAJ: software for processing and analysis of molecular dynamics trajectory data, *J. Chem. Theory Comput.* 9 (2013) 3084–3095.
- [45] J. Donohue, Radial distribution functions of some structures of the polypeptide chain, *Proc. Natl. Acad. Sci.* 40 (1954) 377–381, <https://doi.org/10.1073/pnas.40.6.377>.
- [46] W. Humphrey, A. Dalke, K. Schulten, VMD: visual molecular dynamics, *J. Mol. Graph.* 14 (1996) 33–38.
- [47] T. Hou, J. Wang, Y. Li, W. Wang, Assessing the Performance of the MM_PBSA and MM_GBSA Methods. 1. The Accuracy.pdf, 2011 69–82.
- [48] S. Genheden, O. Kuhn, P. Mikulskis, D. Hoffmann, U. Ryde, The normal-mode entropy in the MM/GBSA method: effect of system truncation, buffer region, and dielectric constant, *J. Chem. Inf. Model.* 52 (2012) 2079–2088.
- [49] C.J. Woods, M. Malaisree, J. Michel, B. Long, S. McIntosh-Smith, A.J. Mulholland, Rapid decomposition and visualisation of protein-ligand binding free energies by residue and by water, *Faraday Discuss.* 169 (2014) 477–499, <https://doi.org/10.1039/c3fd00125c>.
- [50] C.J. Woods, M. Malaisree, S. Hannongbua, A.J. Mulholland, A water-swap reaction coordinate for the calculation of absolute protein-ligand binding free energies, *J. Chem. Phys.* 134 (2011) <https://doi.org/10.1063/1.3519057>.
- [51] B. Nutho, S. Pengthaisong, A. Tankrathok, V.S. Lee, J.R. Ketudat Cairns, T. Rungtongmongkol, S. Hannongbua, Structural basis of specific glucoimidazole and mannoimidazole binding by Os3BGlu7, *Biomolecules* 10 (2020) 907.
- [52] J. Wang, L. Urban, The impact of early ADME profiling on drug discovery and development strategy, *DDW Drug Discov. World.* 5 (2004) 73–86.
- [53] F. Lombardo, P.V. Desai, R. Arimoto, K.E. Desino, H. Fischer, C.E. Keefer, C. Petersson, S. Winiwarter, F. Broccatelli, In Silico absorption, distribution, metabolism, excretion, and pharmacokinetics (ADME-PK): utility and best practices, an industry perspective from the international consortium for innovation through quality in pharmaceutical development: miniperspective, *J. Med. Chem.* 60 (2017) 9097–9113.
- [54] J. Lin, D.C. Sahakian, S.M. De Morais, J.J. Xu, R.J. Polzer, S.M. Winter, The role of absorption, distribution, metabolism, excretion and toxicity in drug discovery, *Curr. Top. Med. Chem.* 3 (2003) 1125–1154.
- [55] D.F. Veber, S.R. Johnson, H.Y. Cheng, B.R. Smith, K.W. Ward, K.D. Kopple, Molecular properties that influence the oral bioavailability of drug candidates, *J. Med. Chem.* 45 (2002) 2615–2623, <https://doi.org/10.1021/jm020017n>.
- [56] Z. Wang, Y. Chen, H. Liang, A. Bender, R.C. Glen, A. Yan, P-glycoprotein substrate models using support vector machines based on a comprehensive data set, *J. Chem. Inf. Model.* 51 (2011) 1447–1456.
- [57] Y. Morimoto, T. Hatanaka, K. Sugibayashi, H. Omiya, Prediction of skin permeability of drugs: comparison of human and hairless rat skin, *J. Pharm. Pharmacol.* 44 (1992) 634–639.
- [58] D.J. Greenblatt, D.R. Abernethy, M. Divoll, Is volume of distribution at steady state a meaningful kinetic variable? *J. Clin. Pharmacol.* 23 (1983) 391–400.
- [59] A.G. De Boer, D.D. Breimer, The blood-brain barrier: clinical implications for drug delivery to the brain, *J. R. Coll. Physicians Lond.* 28 (1994) 502.
- [60] Y.D. Wijeyeratne, E.R. Behr, Long QT syndrome, *Clin. Cardiol.* 51 (2016) 155–173, https://doi.org/10.1007/978-3-319-44203-7_10.
- [61] P.L. Hedley, P. Jørgensen, S. Schlamowitz, R. Wangari, J. Moolman-Smook, P.A. Brink, J.K. Kanters, V.A. Corfield, M. Christiansen, The genetic basis of long QT and short QT syndromes: a mutation update, *Hum. Mutat.* 30 (2009) 1486–1511.
- [62] A. Whitty, Growing PAINS in academic drug discovery, *Future Med. Chem.* 3 (2011) 797–801.
- [63] A. Shitrit, D. Zaidman, O. Kalid, I. Bloch, D. Doron, T. Yarnitzky, I. Buch, I. Segev, E. Ben-Zeev, E. Segev, et al., Conserved interactions required for inhibition of the main protease of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), *Sci. Rep.* 10 (2020) 1–11.