



# Using Chou's 5-steps rule to study pharmacophore-based virtual screening of SARS-CoV-2 Mpro inhibitors

Hemlata Pundir<sup>1</sup> · Tanuja Joshi<sup>2</sup> · Tushar Joshi<sup>3</sup> · Priyanka Sharma<sup>1</sup> · Shalini Mathpal<sup>3</sup> · Subhash Chandra<sup>2,4</sup> · Sushma Tamta<sup>1</sup>

Received: 23 July 2020 / Accepted: 8 October 2020 / Published online: 20 October 2020  
© Springer Nature Switzerland AG 2020

## Abstract

Recently emerged SARS-CoV-2 is the cause of the ongoing outbreak of COVID-19. It is responsible for the deaths of millions of people and has caused global economic and social disruption. The numbers of COVID-19 cases are increasing exponentially across the world. Control of this pandemic disease is challenging because there is no effective drug or vaccine available against this virus and this situation demands an urgent need for the development of anti-SARS-CoV-2 potential medicines. In this regard, the main protease (Mpro) has emerged as an essential drug target as it plays a vital role in virus replication and transcription. In this research, we have identified two novel potent inhibitors of the Mpro (PubChem3408741 and PubChem4167619) from PubChem database by pharmacophore-based high-throughput virtual screening. The molecular docking, toxicity, and pharmacophore analysis indicate that these compounds may act as potential anti-viral candidates. The molecular dynamic simulation along with the binding free energy calculation by MMPBSA showed that these compounds bind to Mpro enzyme with high stability over 50 ns. Our results showed that two compounds: PubChem3408741 and PubChem4167619 had the binding free energy of  $-94.02 \text{ kJ mol}^{-1}$  and  $-122.75 \text{ kJ mol}^{-1}$ , respectively, as compared to reference X77 ( $-76.48 \text{ kJ mol}^{-1}$ ). Based on our work's findings, we propose that these compounds can be considered as lead molecules for targeting Mpro enzyme and they can be potential SARS-CoV-2 inhibitors. These inhibitors could be tested in vitro and explored for effective drug development against COVID-19.

---

Hemlata Pundir and Tanuja Joshi have contributed equally to the work.

---

**Electronic supplementary material** The online version of this article (<https://doi.org/10.1007/s11030-020-10148-5>) contains supplementary material, which is available to authorized users.

---

✉ Subhash Chandra  
scjnu@yahoo.co.in

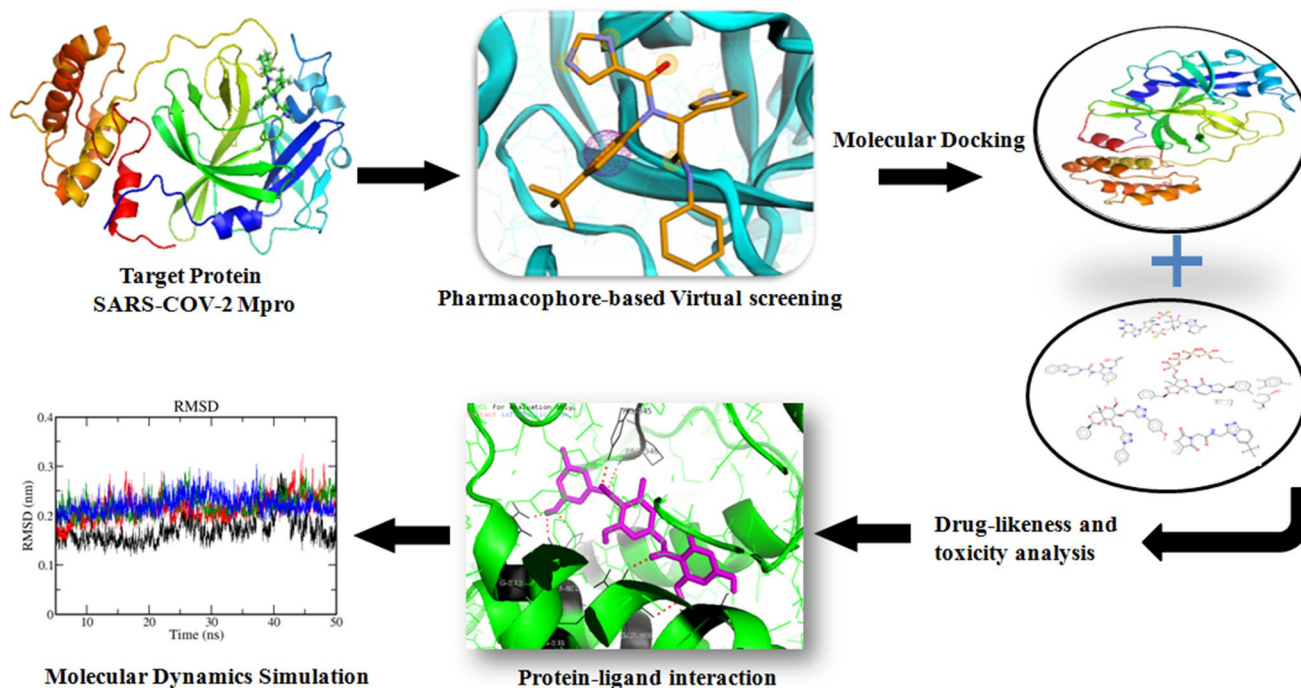
<sup>1</sup> Department of Botany, D.S.B Campus, Kumaun University, Nainital, Uttarakhand 263002, India

<sup>2</sup> Computational Biology and Biotechnology Laboratory, Department of Botany, Soban Singh Jeena University, Almora, Uttarakhand, India

<sup>3</sup> Department of Biotechnology, Kumaun University, Bhimtal Campus, Bhimtal, Uttarakhand 263136, India

<sup>4</sup> Department of Botany, Kumaun University, S. S. J. Campus, Almora, Uttarakhand, India

## Graphic abstract



**Keywords** COVID-19 · Pharmacophore modeling · Molecular docking · Main protease (MPro) · PubChem compounds

## Introduction

In the past two decades, coronaviruses (CoVs) have been associated with significant disease outbreaks in East Asia and the Middle East geographical area. The severe acute respiratory syndrome (SARS) and the Middle East respiratory syndrome (MERS) began to emerge in 2002 and 2012, respectively. The world is currently witnessing a pandemic, the COVID-19 disease, caused by rapidly spreading a new strain of novel coronavirus (SARS-CoV-2), which was identified in late 2019, initially named 2019-nCoV [1]. COVID-19 has been declared a pandemic by the World Health Organization (WHO, March 11 2020). There are two main ways for the spread of COVID-19, first is indirect contact in which the infected droplets of the virus fall, and second is direct physical contact. It has posed a global health threat, causing an ongoing pandemic in all continents except the Antarctica [2]. According to Worldometer's report, about 12,415,582 peoples have been infected with COVID-19, and 557,923 deaths occurred worldwide, till July 10, 2020, and the numbers are abruptly increasing on a daily basis. In India, the numbers of cases have gone on beyond 798,128 and deaths more than 21,654. COVID-19 is not only affecting the healthcare system but also the global economy.

We are not in a situation to effectively treat SARS-CoV-2, because of the absence of specific approved antiviral drugs or vaccines for its treatment. This situation has forced the requirement of such medicines, which can be used against this new pandemic and find a cure for it. Scientists have been trying to develop efficient therapy, but unfortunately, no appropriate vaccines and anti-viral drugs have been made available in the market yet. On March 17, 2020, the USA reported starting a vaccine trial against COVID-19, but stated that it will take more than a year for it to be introduced in markets. Recently, combination therapy of azithromycin and hydroxychloroquine was investigated and its results of the non-randomized clinical trial were published [3, 4]. But it has been found that hydroxychloroquine can cause side effects like drug poisoning in hypersensitive and diabetic patients. Adverse neuropsychiatric condition has been reported in post-treatment of hydroxychloroquine [5]. The administration of high doses of hydroxychloroquine, along with atorvastatin, showed the highest decline of blood glucose in diabetic patients [6]. Therefore, an effective control mechanism is needed to prevent COVID-19 [3].

The knowledge of protein 3D structures or their complexes with the ligand is vitally crucial for rational drug design. To determine 3D structure of a protein, X-ray

crystallography is a powerful tool, but it is time-consuming and expensive. Moreover, not all proteins can be crystallized to resolve 3D structure successfully. Therefore, to acquire the structural information structural of bioinformatics tools can be applied successfully [6]. In this regard, the rapid development in sequential and structural bioinformatics has driven the medicinal chemistry which is undergoing an unprecedented revolution currently [7]. Computational biology has played increasingly important roles in stimulating the development of finding novel drugs and can be used to develop potential novel medicines to combat SARS-CoV-2. Recently, several efforts have been made to design anti-SARS-CoV-2 drugs through computational methods [8, 9]. In March 2020, a group of researchers at UBC screened 1.3 billion small molecules as potential inhibitors against the SARS-CoV-2 Main Protease (Mpro) by deep docking method [10]. According to the Web site (<https://www.guidetopharmacology.org/coronavirus.jsp>), many ligands synthetic in nature have been proposed for the treatment of COVID-19 and are in clinical trials. Several pharmacophore-based studies have also been carried out to identify compounds from PubChem database against different diseases [7, 11]. Consequently, in order to develop novel drugs against COVID-19, we conducted computational screening of compounds from PubChem database.

Nowadays, molecular targets play an important role for specific drug development. Among the validated molecular targets, the COVID-19 Mpro is a potential drug target to discover new drug candidates. Mpro or 3C-like main protease (3CLpro) plays an essential role in the replication and transcription of SARS-CoV-2 which makes it a potential target for searching SARS-CoV-2 inhibitors [12]. 3CLpro directly mediates the maturation of NSP (non-structural protein), which is essential in the life cycle of the virus. Recently, from all over the world, many efforts were made for the investigation of anti-SARS-CoV-2 drugs, which can act as inhibitors against SARS coronavirus Mpro [13–15].

To identify possible inhibitors against SARS-CoV-2, we applied the Pharmacophore-based virtual screening method following Chou's 5-step rule [16], molecular docking, drug-like analysis, and toxicity prediction (Fig. 1). We used the Pubchem database for ligand-based pharmacophore analysis which resulted in 11 compounds out of 93,067,404 PubChem compounds. After that, selected compounds were screened using molecular docking. Further, drug-likeness and toxicity prediction resulted in two-hit compounds, namely PubChem3408741 and PubChem4167619. These compounds were finally subjected to molecular dynamic simulation (MDS) to analyze the Mpro-ligand complex's stability. Our study shows that both compounds form a stable complex with Mpro and are potential anti-viral compounds. These could be further explored in vitro and in vivo for

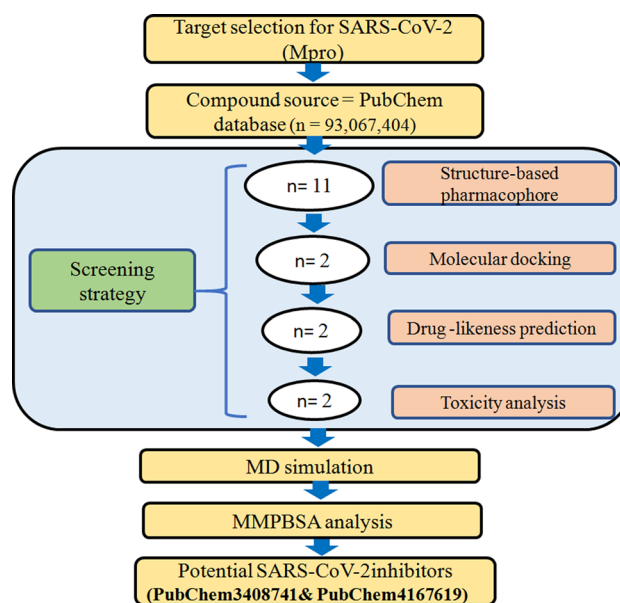


Fig. 1 Overall workflow of various steps of the methodology

effective testing and drug development. The present study's findings will provide other researchers with opportunities to identify the novel drug to combat COVID-19.

## Materials and methods

### Pharmacophore-based screening by Chou's 5-steps rule

To carry out a biological experiment or any other experiment successfully, one can follow “Chou's 5-steps rule” [16]. We adapted the following five steps in our pharmacophore-based screening: (1) selection or construction a valid benchmark dataset to train and test the predictor; (2) representation of the samples with an effective formulation that can truly reflect their intrinsic correlation with the target to be predicted; (3) introduction or development of a powerful algorithm, method, or application of a suitable algorithm to conduct the prediction; (4) properly perform cross-validation tests to evaluate the anticipated prediction accuracy objectively; and (5) use of Pubchem database to screen inhibitors [17, 18]. Chou's 5-steps rules have the following notable merits: (1) crystal clear in logic development, (2) completely transparent in operation, (3) easily to repeat the reported results by other investigators, (4) high potential in stimulating other sequence-analyzing methods, and (5) very convenient to be used by the majority of experimental scientists. Therefore in this study, we used Chou's 5-steps rule to study Pharmacophore-based virtual screening of SARS-CoV-2 Mpro inhibitors.

(1) Selection of train dataset for development of pharmacophore model

To determine the important pharmacophore features which are important for the recognition of ligand by receptor and biological activity, we carried out structure-based pharmacophore modeling of SARS-CoV-2 Mpro (6W63) bound with its inhibitor X77 using Pharmit (<http://pharmit.csb.pitt.edu/>) Web server by loading the docked features.

(2) Preparation of test sample for validation of pharmacophore model

Due to the unavailability of SARS-CoV-2 Mpro dataset, we used SARS-CoV-1 dataset of Mpro inhibitors (ChEMBL ID: ChEMBL3927) containing IC<sub>50</sub> and Ki values of 201 compounds as a test dataset to validate pharmacophore model. ChEMBL3927 dataset contains 201 compounds as inhibitors of the SARS-CoV-1 Mpro. After analyzing the dataset, we obtained 161 compounds (39 active and 122 inactive) considering a threshold activity of 10 μM. After curation by removing the duplicates from them, we collected 117 compounds (31 active and 86 inactive) for preparing the test sample. Recently, NCATS (National Center for Advancing Translational Sciences) has released SARS-CoV-2 3CL protease enzyme assay data for the NCATS Pharmaceutical Collection, containing 10,964 compounds, which are suitable for high-throughput screening. We also analyzed these data and compared them with the predictions from our pharmacophore models. We also collected and curated the data to identify active/ inactive compounds from the assay. At the NCATS Browser, compounds with AC<sub>50</sub> below 12.6 μM ( $-\log AC_{50} = 4.9$ ) are considered “low quality active” [19]. We also applied the same rule for defining active/ inactive compounds. After curation, we found 308 (297 active and 111 inactive) compounds from the assay by removing all duplicates. The active classification was done on the basis of  $-\log AC_{50}$  value ( $-\log AC_{50} \geq 4.9$  active,  $-\log AC_{50} < 4.9$  inactive). Finally, these 308 compounds (297 active and 111 inactive) were tested against our pharmacophore models.

(3) Development of pharmacophore model

By using Pharmit [20], the pharmacophore features of Mpro-X77 complex were visualized and used for developing various pharmacophore models.

(4) Validation of pharmacophore model using experimentally proven inhibitors of Mpro

A test sample prepared from SARS-CoV-1 Mpro dataset (ChEMBL3927) and SARS-CoV-2 3CL protease enzyme assay from NCATS were used for validation of our various

pharmacophore models. The following statistical metrics were used to evaluate the model performance for the classification of test dataset.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

$$\text{Sensitivity} = \frac{TP}{TP + FN}$$

$$\text{Specificity} = \frac{TN}{FP + TN}$$

where TP = true positive, TN = true negative, FP = false positive, FN = false negative.

The pharmacophore model with the highest accuracy was used for screening from PubChem database.

(5) Screening from pharmacophore model

A PubChem library containing 450,708,705 conformers of 93,067,404 compounds was used for screening novel compounds. Finally, compounds were minimized by mRMSD and score computed using AutoDock Vina scoring function. All screened compounds were downloaded as SDF format file from the Pharmit, and then, Open Babel (version 2.3.1) [21] was used to convert SDF files to PDB format files for molecular docking and other analysis. The following filters were used to screen the above library: molecular weight 100–500, hydrogen bond acceptor 1–10, LogP 1–5, and hydrogen bond donor 5.

## Molecular docking

After the pharmacophore-based screening using Chou's 5-steps rule, we performed the molecular docking of all screened compounds with crystal structure of SARS-CoV-2 Mpro. For this, the X-ray crystal structure of SARS-CoV-2 Mpro (PDB ID: 6W63 and Resolution: 2.10 Å) attached with its inhibitor (X77) was obtained from RCSB Protein Data Bank (PDB) (<https://www.rcsb.org>). The X77 (N-(4-tert-butylphenyl)-N-[(1R)-2-(cyclohexylamino)-2-oxo-1-(pyridin-3-yl) ethyl]-1H-imidazole-4-carboxamide) was used as reference compound in this study. For receptor preparation, all the water molecules were removed from the protein molecule with the help of the PyMOL software. After that, hydrogen atoms were added to the receptor molecule by using MGL tool and charge assignment was done using Autodock Tool 1.4 (ADT) [22]. The receptor molecule was further optimized and saved in PDB format for further analysis.

To find potential anti-SARS-CoV-2 compounds, molecular docking of above-screened compounds with Mpro was carried out by AutodockVina using PyRx software (GUI version 0.8) [23]. The grid center for X, Y, and Z coordinates was calculated by using the center of mass function of PyMol. The grid box centered for docking was  $X = -20.11$ ,  $Y = 18.79$ , and  $Z = -27.35$ , and the dimensions of the grid box were set as  $25 \times 25 \times 25$  Å. Initially, for validation of



protocol, molecular docking was performed with the reference molecule X77 at the active site of Mpro to reproduce the same docking conformation similar to co-crystallized X77. After validation of the docking protocol, virtual screening of all ligands was conducted at the active site of Mpro and the docking results were ranked by the binding free energy. Finally, compounds with the lower binding energy (highest affinity) toward Mpro were chosen for further study.

### Drug-likeness and toxicity prediction

The compounds that were finalized by molecular docking were further proceeded to predict their drug-likeness. Lipinski filter [24] was used in DruLiTo software to predict the drug-likeness of compounds. Further, the toxicity profile of the compounds which have a drug-like property and good binding affinity with Mpro receptor were calculated by using OSIRIS property explorer program [25]. The major parameters associated with OSIRIS are toxicity risk properties such as tumorigenicity, mutagenicity, irritation, and reproductive development toxicity.

### Visualization

For visualization of 2D interactions between Mpro-ligand complexes, the LigPlot+ v.1.4.5 software was used. This program was used to identify and depict the interactions of an amino acid of the target protein with ligand by drawing a 2D figure (Ligplot) of hydrogen bonds, hydrophobic bonds, and hydrogen bond lengths for each docked pose.

### Molecular dynamics simulation (MDS)

To determine the structural stability of the native Mpro protein as well as Mpro-ligand complexes under physiological conditions, MDS was carried out. MDS was executed using GROMACS 5.0.7 [26] on a workstation with configuration Ubuntu 16.04 LTS 64-bit, 4 GB RAM, Intel Core i5-6400 CPU processor. The topology of protein and ligands was generated by using ‘pdb2gmx’ script and CGenFF server, respectively, by using CHARMM 36 force field [27]. Then, we assembled protein and ligand topologies to make the protein–ligand complex structure. After that, the solvation of the complex was accomplished by using transferable intermolecular potential water molecules (TIP3Pmodel) with dodecahedral periodic boundary conditions. After solvation, neutralization of the complex was done by adding ions. Further, energy minimization of the complex was done with the steepest descent algorithm by using Verlet cutoff scheme at  $10 \text{ kJ mol}^{-1}$  by taking Particle Mesh Edward (PME) columbic interactions. Equilibration of the system was attained in two phases. First, NVT ensemble was performed at 300 K for 5000 ps, which stabilized the temperature of the system.

While in the second step, NPT ensemble took 1 atm pressure and 5000 ps of steps for NPT simulation. Finally, all systems were subjected to the production MD at a constant temperature of 300 K and a constant pressure of 1 atm with a time step interval of 2 fs, using the Parrinello–Rahman for constant pressure simulation. Molecular dynamics simulations were performed for 50 ns. After successful completion of MDS, the MD trajectories were used to calculate root mean square deviation (RMSD), root mean square fluctuation (RMSF), radius of gyration ( $R_g$ ), hydrogen bonds, solvent accessible surface area (SASA) [28], principal component analysis (PCA) [29], and distance to analyze the stability of Mpro and Mpro-ligand complex.

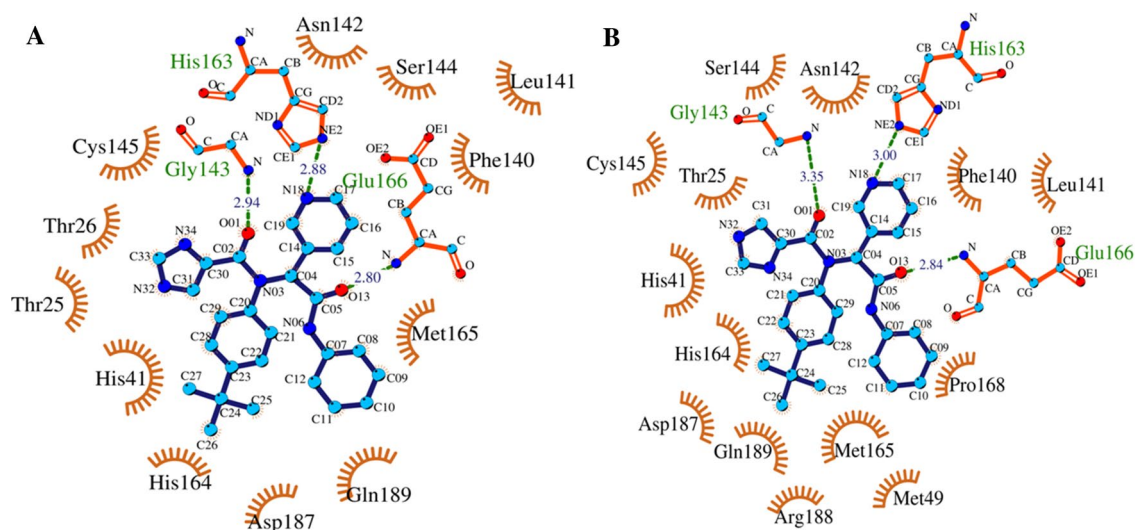
### Binding free energy calculation using MMPBSA

The molecular mechanics Poisson–Boltzmann surface area (MMPBSA) method is widely used to calculate the binding free energy to predict the stability of the complex [30]. Free binding energy calculations consist of free solvation energy (polar and nonpolar solvation energies) and potential energy (electrostatic and van der Waals interactions). Therefore, binding free energy calculations of Mpro-ligand complexes were done by the MMPBSA method. The MD trajectories were processed before doing MMPBSA calculations for the last five ns. Then, average binding energy calculations were done with ‘python’ script provided in g\_mmpbsa.

## Results and discussion

### Pharmacophore-based screening by Chou’s 5-steps rule

X77 binds to the active site of SARS-CoV-2 Mpro with binding energy  $-8.4 \text{ kcal/mol}$  as shown in Fig. 2. X77 showed interaction with the active site residues via three hydrogen bonds with His163, Gly143, and Glu166 and hydrophobic bond with Thr25, Thr26, His41, Phe140, Leu141, Asn142, Ser144, Cys145, His164, Met165, Asp187, and Gln189 of Mpro. According to Fig. 2, O01, and O13 of X77 interact with Gly143 and Glu166, respectively, and N18 interact with His163 of Mpro. Using Pharmit, the Pharmacophore features were developed by loading the crystallographic structure of Mpro-X77 complex (6W63). To develop the pharmacophore models, these pharmacophore features were manually optimized. Finally, the best model was achieved with the following pharmacophore features present in the Mpro-X77 complex—one aromatic ring, one hydrogen bond acceptor, and two hydrophobic groups. The structure of Mpro-X77 pharmacophore is shown in Supplementary Fig. 1. The pharmacophore features, radius, and x, y, z coordinates are shown in Supplementary Table 1.



**Fig. 2** Binding pocket of Mpro showing experimental (a) and docked X77 (b) binds to active site residues. X77 is in blue and active site residues are in orange color representation. Hydrogen bonds that are

formed between protein and ligand are shown by green dotted lines and other residues are hydrophobic bond-forming residues

The pharmacophore model was validated by screening the compounds of ChEMBL3927 and NCATS dataset. On the basis of compounds that matched the pharmacophores, out of 31 active (positive) and 86 inactive (negative) compounds of ChEMBL3927 dataset, our pharmacophore model predicted 26, 5, 62 and 24 as true positive (active compounds that match the pharmacophores), false negative (active compounds that do not match the pharmacophores), true negative (inactive compounds that do not match the pharmacophores), and false positive compounds (inactive compounds that match the pharmacophores), respectively, against SARS-CoV-2 with the accuracy = 75%, sensitivity = 84%, and specificity = 72%. From NCATS dataset, we screened 167, 130, 70, and 41 compounds as true positive, false negative, true negative, and false positive compounds, respectively, from 297 active (positive) and 111 inactive (negative) by using our pharmacophore model with the accuracy = 58%, sensitivity = 56%, and specificity = 63%. Our model had good accuracy ( $\geq 55\%$ ) in both cases; therefore, we conclude that our model can be used for virtual screening. These pharmacophore features of Mpro-X77 were further studied and were used to screen the compounds from PubChem library of Pharmit server. After that, compounds were minimized by mRMSD (minimized RMSD, which is the heavy atom difference between pharmacophore aligned pose and minimized pose) and score computed using AutoDockVina scoring function. Using the pharmacophore model, by applying filters and minimization, 15 compounds (9 from 26 active/true positive and 6 from 24 inactive/false positive, respectively) were screened out from the ChEMBL3927 dataset, and 14 compounds (11 from 167 active/true positive and 3 from 41 inactive/false positive,

respectively) were screened out from the NCATS dataset (Supplementary Table 2 and 3).

Using pharmacophore model, 11 compounds were screened out from the PubChem library (Table 2). These compounds were further screened by docking with Mpro by AutoDockVina using PyRx GUI.

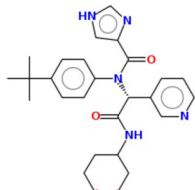
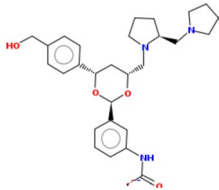
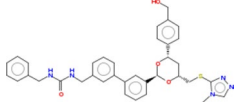
## Molecular docking

Before performing the molecular docking, validation of the protocol was done by re-docking the reference molecule (X77) into the active site of Mpro. The result indicated that the docked X77 was completely superimposed with co-crystallized X77 in PDB and the RMSD of experimental and docked X77 is 0.62. The docked X77 showed interaction with the same amino acid residues by hydrogen and hydrophobic bonds as found in the crystal structure in PDB (Fig. 2). After that, to screen novel compounds against COVID-19, the molecular docking of 11 screened compounds was performed with target protein Mpro using PyRx. The docking results are ranked based on binding energies. Out of 11 compounds, two compounds were finally screened out by docking based on low binding energy/ high binding affinity as shown in Table 1. The binding energy of reference compound X77 was  $-8.3$  kcal/mol, while PubChem-4167619 and PubChem-3408741 had  $-9.3$  kcal mol $^{-1}$  and  $-8.5$  kcal mol $^{-1}$ , respectively. The binding energy of both screened compounds indicated better binding affinity and stability with Mpro as compared to X77. Therefore, these two compounds were preceded for drug-likeness and toxicity prediction.

**Table 1** Summary of pharmacophore-based virtual screening and molecular docking between Mpro and hit compounds

S. no.	Compound	Minimized affinity (Score)	Minimized RMSD (mRMSD)	SMILES format	Docking energy (Kcal mol <sup>-1</sup> )
1	X77			<chem>O=C(N([C@@H](C(=O)NC1CCCC1)[C]1=[CH][CH]=[CH][N]=[CH]\$1)[C]1=[CH][CH]=[C](\$[CH]=[CH]\$1)C(C)(C)C)[C]1=[CH][N]=[CH][N]\$1</chem>	- 8.3
2	PubChem-3408741	- 7.79	1.91	<chem>[C@H]1(O[C@H](C[C@H](O1)c1ccc(cc1)CO)CN1CCC[C@H]1CN1CCCC1)c1cccc(c1)NC(=O)C</chem>	- 8.5
3	PubChem-19600833	- 7.02	1.84	<chem>[C@@H]1(COc2c([C@@H]1O)cc(cc2)O)Cc1cc(c(cc1)OC)OC</chem>	- 7.3
4	PubChem-87961904	- 6.60	1.77	<chem>[C@@H](c1nccc2c1cccc2S(O)(O)O)(C(=O)N)NCCOCc1cc(C)ccc1</chem>	- 7.3
5	PubChem-6402133	- 6.52	1.24	<chem>C(=O)(N1c2cccc2c2c(nc(nn2)SC)O[C@@H]1c1cc(ccc1C)C)CC</chem>	- 7.6
6	PubChem-4167619	- 6.52	1.77	<chem>c1(cccc(c1)c1cc(ccc1)CNC(=O)NCc1cccc1)[C@H]1O[C@@H](C[C@@H](O1)CSc1n(C)cnn1)c1ccc(cc1)CO</chem>	- 9.3
7	PubChem-11175398	- 6.23	1.20	<chem>C1[C@@H](C=C[C@]1(C#N)COCc1cccc1)n1cnc2c1ncnc2N</chem>	- 7.9
8	PubChem-7705524	- 6.17	1.38	<chem>c1(nc2cccc2c(=O)n1C[C@@H]1CCCO1)SCc1n(nnn1)CCCC</chem>	- 7.3
9	PubChem-46322636	- 5.91	1.66	<chem>C(=O)(N1[C@H](Oc2nc(nnc2c2c1cccc2)SCC(C)C)c1ccc(C)cc1)CC</chem>	- 7.5
10	PubChem-45520630	- 5.74	1.21	<chem>c1(c2cnn(c2ncn1)CCNC(=O)c1cncn1)NCCC</chem>	- 6.8
11	PubChem-6406366	- 5.43	1.29	<chem>c1(nc2c(nn1)c1cccc1N([C@H](O2)c1cccc1C)C(=O)C)SCCCC</chem>	- 7.6
12	PubChem-133337777	- 5.28	1.49	<chem>N(C)(c1c(c(C)c(nn1)C)C#N)Cc1csc(n1)[C@@H](C)OC</chem>	- 6.7

**Table 2** Drug-likeness prediction of screened compounds through DruLiTo

S. no.	Compounds name	Structure	Molecular weight	LogP	H-bond acceptor	H-bond donor	Lipinski's rule violation	Drug-likeness alert
1.	X77		458.26	3.39	7	1	0	Accepted
2.	PubChem3408741		456.0	0.75	7	2	0	Accepted
3.	PubChem4167619		600.99	4.17	9	3	1	Accepted

## Drug-likeness and toxicity prediction

The physicochemical properties like molecular weight (MW), logP, hydrogen bond acceptor (HBA), and hydrogen bond donor (HBD) are the important properties of a compound to be considered as a drug candidate according to Lipinski's RO5. DruLiTo was used to predict these physicochemical properties of two compounds obtained after the docking of 11 compounds. These two compounds show better pharmacokinetics and drug-likeness. Both compounds showed better physicochemical properties and satisfied the Lipinski's rule which is shown in Table 2.

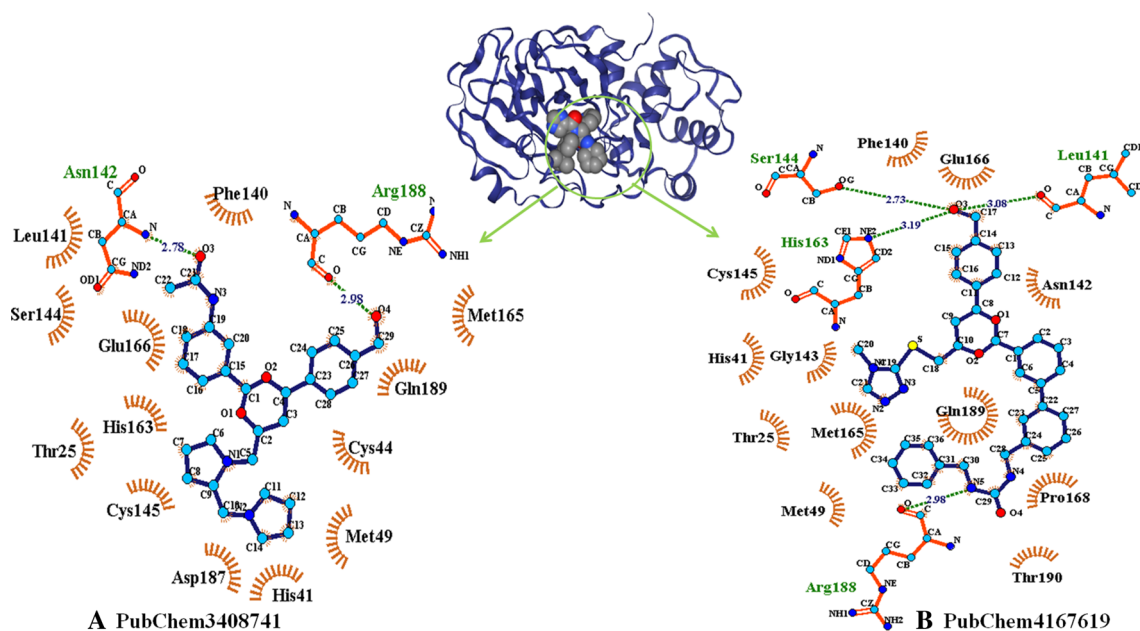
The above two compounds were further analyzed for their tumorigenicity, mutagenicity, irritation, and reproductive toxicity by the OSIRIS tool. The predicted toxicity of both compounds is shown in Table 3. According to Table 3, both compounds were non-toxic (non-mutagenic, non-tumorigenic, non-irritant, and no reproductive effect). Therefore, these two compounds (PubChem3408741 and PubChem4167619) can be further proceeded to study their binding interaction with Mpro and also their dynamic behavior.

## Visualization

LigPlot+ v.1.4.5 was used to visualize the protein–ligand interactions. The docked pose of these two compounds with Mpro is shown in Fig. 3. According to Fig. 3, PubChem3408741 forms two hydrogen bonds with Asn142 and Arg188 which have the bond distance 2.78 Å and 2.98 Å, respectively. It also forms hydrophobic bonds with Phe140, Met165, Gln189, Cys44, Met49, His41, Asp187, Cys145, His163, Thr25, Glu166, Ser144, and Leu141. PubChem4167619 interacts with several residues that make four hydrogen bonds with Ser144, His163, Leu141, and Arg188 having the bond distance 2.73 Å, 3.19 Å, 3.08 Å, and 2.98 Å, respectively. Besides, Asn142, Phe140, Glu166, Asn142, Gln189, Pro168, Thr190, Met49, Met165, Thr25, Gly143, His41, and Cys145 residues were found to participate in hydrophobic interactions in Mpro-PubChem4167619 complex. In both the complexes, His163, Glu166, Asn142, Cys145, His41, Gln189, Ser144, Met165, Thr25, His163, Leu141, and Phe140 are common binding active site residues.

**Table 3** Toxicity profile of hit compounds

S. no.	Compound name	Toxicity risks			
		Mutagenic	Tumorigenic	Irritant	Reproductive effect
1	X77	Non-toxic	Non-toxic	Non-toxic	Non-toxic
2	PubChem3408741	Non-toxic	Non-toxic	Non-toxic	Non-toxic
3	PubChem4167619	Non-toxic	Non-toxic	Non-toxic	Non-toxic



**Fig. 3** 2D interaction showing H-bonds and hydrophobic bonds between hit compounds with Mpro



## Molecular dynamic simulation (MDS)

The Mpro was considered rigid during the docking performance. Therefore, to get a more realistic picture of the Mpro and ligand interactions, the docked complexes were simulated in a solvated dodecahedral box for 50 ns. Four systems (Mpro protein, Mpro-X77, Mpro-PubChem3408741, and Mpro-PubChem4167619 complexes) were subjected to the MD simulation. The structural stability and dynamic behavior were analyzed by the RMSD, RG, RMSF, H-bond, distance, and SASA calculation. The average value of RMSD, RG, RMSF, distance, and SASA is shown in Table 4.

### RMSD

The root mean square deviation (RMSD) was determined to check the structural stability of the Mpro-ligand complex during the 50 ns simulation. The RMSDs profiles were determined for the backbone and heavy atoms of Mpro protein for all the complexes. Figure 4a represents the RMSD plot for native protein Mpro and all Mpro-ligand complexes. The small deviations reflect the more stable nature of protein and protein–ligand complexes. Both the studied complexes along with the reference complex remained stable throughout the simulations. The average RMSD value for protein was  $0.17 \pm 0.03$  nm, whereas in case of protein–ligand complex systems Mpro-X77, Mpro-PubChem3408741, and Mpro-PubChem4167619 were found to be  $0.21 \pm 0.02$  nm,  $0.22 \pm 0.02$  nm, and  $0.22 \pm 0.03$  nm, respectively. Minor fluctuations with few RMSD values were good indicators of the system stability [31]. Both studied complexes show almost

similar RMSD value as reference complex, which confirms the stability of these complexes. Overall, the RMSD analysis suggested that the MD trajectories were stable throughout the 50 ns and were in the acceptable range.

### RMSF

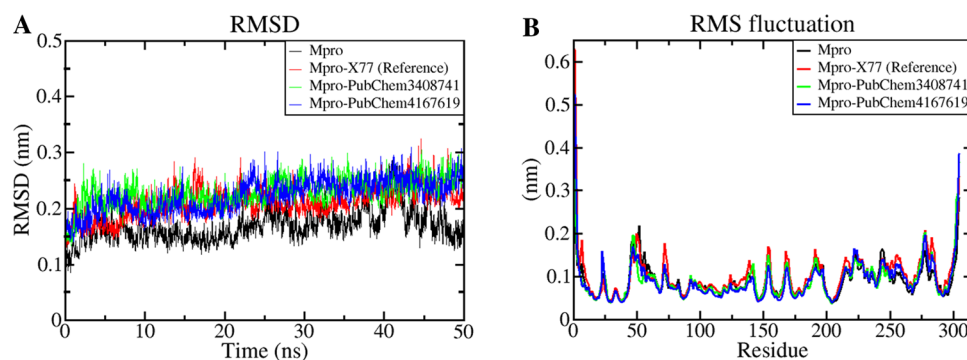
Root mean square fluctuation (RMSF) is used for measuring the motion of each residue around the average position along the trajectory which reveals the flexibility of a specific region of the protein during the MD simulation. The RMSF profile was determined for the Mpro and Mpro-ligand complexes and is depicted in Fig. 4b. The fluctuations in the constituent residues were observed for Mpro and all complexes during the 50 ns, indicating that the binding of these inhibitors has a similar effect on the pattern of residue fluctuations of the Mpro. The most massive fluctuations which were greater than 0.2 nm were observed for residues 50–52. However, these residues are not known to have special functional relevance. Other than this, the fluctuation during all Mpro-ligand interaction was below 0.2 nm which is perfectly acceptable. In conclusion, it indicates that RMSF of all studied complexes is significantly similar to the reference complex, which reflects the good stability of the studied complexes.

To analyze the stability of the binding cavity during the simulation process, the root mean square fluctuation (RMSF) of all active site residues was also calculated. In all complexes, the RMSF for each active site residue is lower than 0.2 nm (Supplementary Table 4), which means that the binding cavity is quite stable during the simulation. This

**Table 4** Average values of RMSD, RMSF, Rg, SASA, and distance for protein–ligand complexes

S. no.	Protein/protein–ligand complex	Average RMSD	Average RMSF	Average RG	Average SASA	Average distance
1.	Mpro	$0.17 \pm 0.03$	$0.09 \pm 0.04$	$1.78 \pm 0.19$	–	–
2.	Mpro-X77 (Reference)	$0.21 \pm 0.02$	$0.10 \pm 0.04$	$1.75 \pm 0.31$	$148.74 \pm 2.61$	$3.67 \pm 0.23$
3.	Mpro-PubChem3408741	$0.22 \pm 0.02$	$0.09 \pm 0.04$	$1.44 \pm 0.23$	$150.59 \pm 3.09$	$3.81 \pm 0.16$
4.	Mpro-PubChem4167619	$0.22 \pm 0.03$	$0.09 \pm 0.05$	$1.79 \pm 0.21$	$148.89 \pm 2.42$	$3.75 \pm 0.19$

**Fig. 4** Plot of RMSD as a function of time (a) and RMSF values of C $\alpha$  atoms (b) for 50 ns trajectories



narrow range of RMSF values of the active site residues of the complexes demonstrates that these compounds are capable of forming stable interactions with the Mpro during MD simulation.

## RG

The radius of gyration ( $R_g$ ) is used to calculate the level of compactness in the structure of protein due to the presence or absence of ligands [32]. A higher  $R_g$  value of the protein–ligand complex reflects the less compactness in that complex. RG is also used to check whether the complexes are stably folded or unfolded during the MD simulation. If the RG shows a relatively consistent value throughout the MD simulation, it can be regarded as a stably folded structure; otherwise, it would be considered an unfolded structure [33]. The RG plot for Mpro and all Mpro–ligand complexes is shown in Fig. 5a. The average RG value for Mpro protein, Mpro–X77, Mpro–PubChem3408741, and Mpro–PubChem4167619 were found to be  $1.78 \pm 0.19$  nm,  $1.75 \pm 0.31$  nm,  $1.44 \pm 0.23$  nm, and  $1.79 \pm 0.21$  nm respectively (Table 4). Through Table 4, it can be observed that all complexes exhibit a relatively similar nature of compactness and show consistent lower  $R_g$  values as compared to the Mpro and reference complex which indicates that these are perfectly superimposed with each other and have good stability.

## SASA

The solvent accessible surface area (SASA) measures the ratio of the complex surface which can be accessed by the water solvent. SASA calculation of the protein–ligand complexes was used for predicting the extent of the conformational changes that occurred during the interaction. Figure 5b shows the plot of SASA value vs. time for all Mpro–ligand complexes. The average SASA value was  $150.59 \pm 3.09$  nm<sup>2</sup> for Mpro–PubChem3408741, and likewise Mpro–PubChem4167619 complex showed the average SASA value  $148.89 \pm 2.42$ , respectively, which were

significantly similar as the reference complex (Mpro–X77), i.e.,  $148.74 \pm 2.61$  nm<sup>2</sup> during the simulation period of 50 ns. So in terms of SASA analysis, we conclude that our studied Mpro–ligand complexes are relatively stable during 50 ns of MD simulation, which indicates no significant changes in the protein structure were caused by these two compounds during simulation.

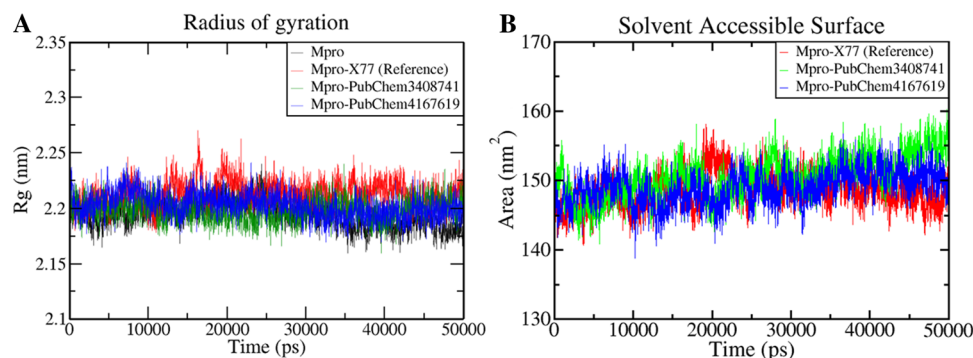
## H-bond analysis

Hydrogen bond plays a very important role in stabilizing the protein–ligand complex. Hydrogen bond is responsible for drug specificity, metabolism, and adsorption in body. Therefore, MD trajectories were also analyzed to examine the number of hydrogen bonds between protein–ligand complexes which were formed during the 50 ns simulation. Figure 6a shows that the number of hydrogen bonds varies from 3 to 5 in Mpro–PubChem3408741 complex. Similarly, it shows variation from 4 to 7 in Mpro–PubChem4167619 complex, while in Mpro–X77, it varies from 2 to 5. We conclude that both the compounds can bind to Mpro effectively and tightly as reference X77.

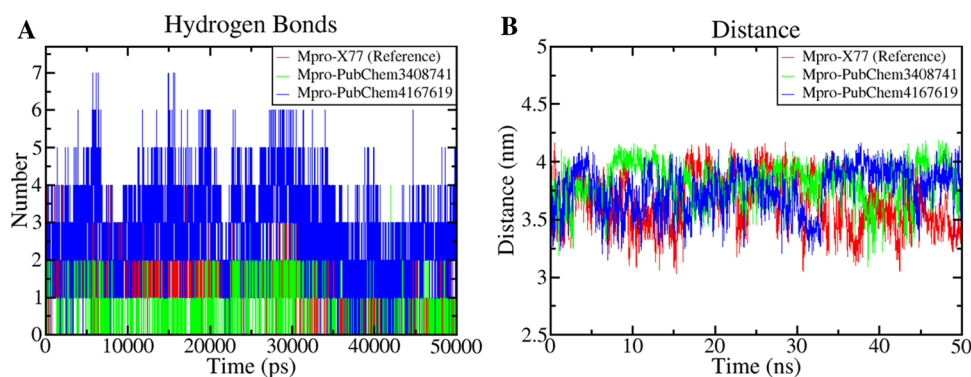
## Distance

The `gmx_mpi pairdist` module of GROMACS was used to calculate the distance between the Mpro and ligand during the simulation. The average distance calculations were done using the 50 ns MD trajectories and are shown in Fig. 6b. The average distance for the Mpro–PubChem3408741 complex was observed to be around 3.81 nm, while for Mpro–PubChem4167619, it was computed to be around 3.75 nm. Likewise, an average distance of 3.67 nm was observed for the reference complex Mpro–X77. Overall, distance analysis suggests that both the compounds have the same distance to the Mpro protein as reference X77.

**Fig. 5**  $R_g$  (a) and SASA (b) as a function of time



**Fig. 6** Number of H-bonds (a) and distance (b) as a function of time



### Principal component analysis (PCA)

To perform PCA, the eigenvectors, eigenvalues, and their projection were calculated using the essential dynamics (ED) method. PCA analysis is used to determine the most significant motions during the ligand binding in different complexes. It is a well-known fact that the first few eigenvectors determine the overall motion of the particular protein. Here, the diagonalization of the matrix was used to calculate the eigenvectors. For this study, we used the first 40 eigenvectors for the analysis of concerted motions for 50 ns MD trajectories. The first ten eigenvector accounts 71.97%, 63.99%, and 65.98% motions in simulation period for Mpro-X77, Mpro-PubChem3408741 and Mpro-PubChem4167619 complexes, respectively (Fig. 7a).

The dynamics of complexes can be explained by 2D projection plot generation in PCA (Fig. 7b).

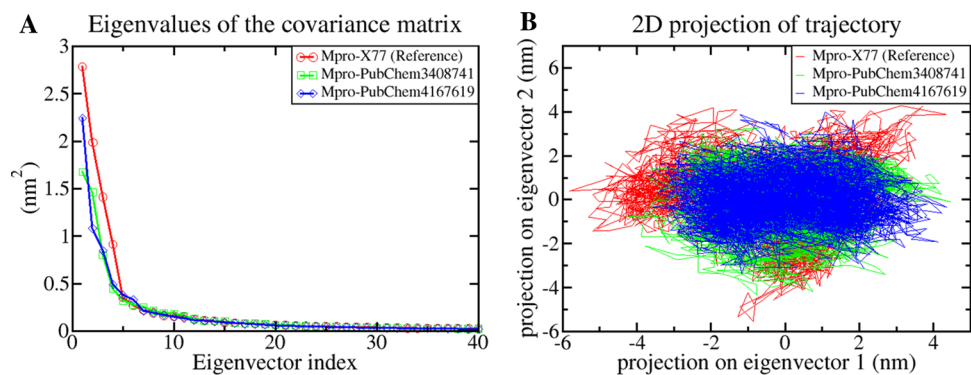
For that, we selected the first two principal components, i.e., PC1 and PC2 for the prediction of significant motions. The complex which occupies less phase space with stable cluster represents a more stable complex. The complex occupying more space with non-stable cluster represents a comparatively less stable complex. From Fig. 7B, it can be observed that Mpro-PubChem3408741 and Mpro-PubChem4167619 complexes showed very stable clusters as they occupy similar phase space as compared to reference X77-Mpro (Red), which also showed a stable cluster.

The Gibbs energy landscape plot for PC1 and PC2 is shown in Supplementary Fig. 2. The plot shows the Gibbs energy value ranges from 0 to 11.4, 0 to 10.5, and 0 to 9.67 for Mpro-X77 (Fig. A), Mpro-PubChem3408741 (Fig. B), and Mpro-PubChem4167619 (Fig. C), respectively. All studied compounds had significantly similar energy as compared to the reference. These observations suggest that Mpro-PubChem3408741 and Mpro-PubChem4167619 complexes can follow energetically more favorable transition from one conformation to another as compared to reference complex and all complexes were thermodynamically stable.

### Binding energy calculation using MMPBSA method and energetic contribution of individual residues

The binding free energy calculation was done using the *g\_mmpbsa* tool for all systems, considering frames from the last five ns of MD trajectories as shown in Table 5. The binding free energy comprises of following energy components: polar solvation energy, SASA nonpolar solvation energy and non-bonded interaction energies (Van der Waal and electrostatic energy). Table 5 shows that Mpro-PubChem3408741 and Mpro-PubChem4167619 complexes possess the lower negative binding energy ( $-94.02 \pm 0.66$  kJ mol<sup>-1</sup> and  $-122.75 \pm 0.70$  kJ mol<sup>-1</sup>, respectively) as compared to the reference complex Mpro-X77 ( $-76.48 \pm 0.53$  kJ mol<sup>-1</sup>). It confirms that both compounds bind efficiently at the active

**Fig. 7** Principal component analysis- The plot of eigenvalues vs. first 40 eigenvectors (a) and first two eigenvectors describing the motion of a protein in phase space for all the complexes (b)



**Table 5** Table showing the van der Waal, electrostatic, polar solvation, SASA, and binding energy for the protein–ligand complexes

S. no.	Name of protein–ligand complex	Van der Waal energy	Electrostatic energy	Polar solvation energy	SASA energy	Total Energy (kJ mol <sup>-1</sup> )
1.	Mpro-X77 (Reference)	-167.10 ± 0.72	-11.93 ± 0.53	122.71 ± 1.04	-20.17 ± 0.06	-76.48 ± 0.53
2.	Mpro-PubChem3408741	-164.43 ± 1.06	-15.85 ± 0.58	107.17 ± 1.05	-20.86 ± 0.10	-94.02 ± 0.66
3.	Mpro-PubChem4167619	-229.60 ± 0.73	-41.46 ± 0.39	166.52 ± 0.56	-18.22 ± 0.41	-122.75 ± 0.70

site of Mpro protein and could be used as lead compounds. Further various energy terms contributing toward binding free energy revealed that in all the studied complexes, the driving component of binding was the van der Waals energy which plays a major contribution in strengthening the binding mode. The polar solvation energy does not show a favorable contribution to the total binding in all the studied complexes. The electrostatic energy and SASA nonpolar solvation energy contribute significantly to the binding free energy.

To analyze the key residues involved in protein–ligand interaction, per residue interaction energy profile was created using the MMPBSA approach for the last five ns of MD trajectories. The per residue decomposition plot of the total binding energy of the Mpro–ligand complexes is shown in Supplementary Fig. 3. For the clear depiction of the results, only active site residues are highlighted in figure. From the plot, it is clear that Thr25, Leu27, Met49, Cys145, Met165, Asp187, and Arg188 are participating amino acids in all complexes. The per residue interaction plot indicates that most of the residues have negative binding energy, while few residues have positive binding energy. The residues that showed negative binding energy play an important role in stabilizing the Mpro–ligand complex. Three active site residues, i.e., Met49, Cys145, and Met165, showed higher binding affinity as compared to other residues. Thus, the results reveal that Met49, Cys145, and Met165 play an important role in Mpro–ligand stabilization.

The comprehensive study shows that both hits have very good binding efficiency against the Mpro. From the overall RMSD, RMSF, Rg, SASA, hydrogen bonds, distance, PCA, and binding free energy analysis results, we conclude that Mpro-PubChem3408741 and Mpro-PubChem4167619 form a very stable complex with Mpro as compared to reference X77 and could be effective drug candidates against COVID-19. However, further studies are necessary to reveal the action of these compounds.

Several therapies for SARS-CoV-2 virus are in a trial to provide a cure for this dreadful viral outbreak. A combination of azithromycin and hydroxychloroquine is being advised to the affected patients in case of emergency. Hydroxychloroquine increases intracellular pH and has been considered as a potent inhibitor of SARS-CoV-2; however, it may cause side effects. Hence, there is an urgent need to

discover potent alternatives. As we were interested in finding potent pharmacophore-based inhibitors against SARS-CoV-2 Mpro, so we discovered the compounds, Mpro-PubChem3408741 and Mpro-PubChem4167619 which show excellent pharmacological properties. These compounds possess excellent properties of drug ability, small molecular weight, and molar refractivity which confirms that these drugs are permeable through particular membranes and can remain constant even in strong or weak solute–solvent, solvent–solvent interactions. Drug-likeness RO5 was followed by these compounds which describes the compound can act as a drug in the biological systems. The toxicity prediction indicates that both the compounds, PubChem3408741 and PubChem4167619, are safe and can be given as a drug to COVID-19 infected patient as predicted by OSIRIS. The MD simulation results indicate that PubChem3408741 and PubChem4167619 bind to the SARS-CoV-2 Mpro efficiently and form a very stable complex. Binding free energy analysis by MMPBSA shows the better binding energy of both compounds to Mpro as compared to X77. It means that these compounds may potent inhibitors of SARS-CoV-2 Mpro and could be used as potential drug candidates against SARS-CoV-2. This study may be helpful to develop effective medications against SARS-CoV-2 in the future.

## Conclusion

COVID-19 is a novel pandemic disease for which treatment is unavailable globally. Therefore, it is necessary to discover effective drugs candidates for treatment of COVID-19 successfully. In this direction, bioinformatics tool-based drug discovery could be a useful approach to combat COVID-19. The present study aimed to identify novel inhibitors against the SARS-CoV-2 Mpro. Therefore, structure-based pharmacophore modeling, virtual screening, molecular docking, and MD simulation were successfully performed to discover novel inhibitors of Mpro. A total of 93,067,404 compounds were screened from PubChem database by the pharmacophore-based virtual screening followed by molecular docking, drug-likeness, and toxicity prediction. Finally, the relative stability of two-hit compounds was validated by 50 ns MD simulation. The MD trajectories analysis showed good structural stability with Mpro during the simulation period.



Hence from this study, two hits, i.e., PubChem3408741 and PubChem4167619, were identified against SARS-CoV-2 Mpro. Thus, the outcome of this study reveals that these two compounds may represent potential treatment options against COVID-19. However, further in vitro and in vivo studies are necessary to investigate pharmacological effects and efficacy of these potential compounds against SARS-CoV-2.

**Acknowledgements** Authors are thankful to Head Department of Botany, Kumaun University, Nainital for providing the facility, space, and resources for this work. The authors also acknowledge to Kumaun University, Nainital for providing high-speed internet facilities. We also extend our acknowledge to Rashtriya Uchchattar Shiksha Abhiyan (RUSA), Ministry of Human Resource Development, Government of India to provide Computational infrastructure for the establishment of Bioinformatics Centre in Kumaun University, S.S.J Campus, Almora.

**Author contributions** Dr. SC contributed to conceptualization, HP and TJ contributed to methodology, TJ and TJ contributed to software, Dr. ST and Dr. SC contributed to supervision, Dr. ST and Dr. SC contributed to validation, PS and HP contributed to visualization, TJ, PS, and SM contributed to formal analysis, HP and TJ contributed to writing—original draft, and Dr. ST and Dr. SC contributed to writing—review and editing.

**Funding** There was no funding source to carry out this research work.

## Compliance with ethical standards

**Conflict of interest** The authors declare that they have no conflict of interest.

## References

- Lee PI, Hsueh PR (2020) Emerging threats from zoonotic coronaviruses—from SARS and MERS to 2019-nCoV. *J Microbiol Immunol Infect* 53(3):365–367. <https://doi.org/10.1016/j.jmii.2020.02.001>
- Kraemer MU, Sinka ME, Duda KA, Mylne AQ, Shearer FM, Barker CM, Moore CG, Carvalho RG, Coelho GE, Van Bortel W, Hendrickx G, Schaffner F, Elyazar IR, Teng HJ, Brady OJ, Messina JP, Pigott DM, Scott TW, Smith DL, Wint GR, Golding N, Hay SI (2015) The global distribution of the arbovirus vectors *Aedes aegypti* and *A. albopictus*. *Elife* 4:e08347. <https://doi.org/10.7554/elife.08347>
- Huang C, Wang Y, Li X, Ren L, Zhao J, Hu Y, Zhang L, Fan G, Xu J, Gu X, Cheng Z, Yu T, Xia J, Wei Y, Wu W, Xie X, Yin W, Li H, Liu M, Xiao Y, Gao H, Guo L, Xie J, Wang G, Jiang R, Gao Z, Jin Q, Wang J, Cao B (2020) Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *Lancet* 395(10223):497–506. <https://doi.org/10.1007/s11684-020-0767-8>
- Gautret P, Lagier JC, Parola P, Hoang VT, Meddeb L, Mailhe M, Doudier B, Courjon J, Giordanengo V, Vieira VE, Dupont HT, Honore S, Colson P, Chabriere E, La Scola B, Rolain JM, Brouqui P, Raoult D (2020) Hydroxychloroquine and azithromycin as a treatment of COVID-19: results of an open-label non-randomized clinical trial. *Int J Antimicrob Agents*. <https://doi.org/10.1016/j.ijantimicag.2020.105949>
- Ali SS, Jones H (2018) An adverse neuropsychiatric reaction following treatment with hydroxychloroquine: a case report. *Rheumatol Adv Pract*. <https://doi.org/10.1093/rap/rky033.014>
- Wondafrash DZ, Desalegn TZ, Yimer EM, Tsige AG, Adamu BA, Zewdie KA (2020) Potential effect of hydroxychloroquine in diabetes mellitus: a systematic review on preclinical and clinical trial studies. *J Diabetes Res* 2020:5214751. <https://doi.org/10.1155/2020/5214751>
- Gonzalez J, Lendebol E, Shen A, Philipp M, Clement CC (2018) In silico-mediated virtual screening and molecular docking platforms for discovery of non  $\beta$ -lactam inhibitors of Y-49  $\beta$ -lactamase from mycobacterium tuberculosis. *MOJ Proteomics Bioinform* 7(1):12–24. <https://doi.org/10.15406/mojpb.2018.07.00207>
- Elmezayen AD, Al-Obaidi A, Sahin AT, Yeleki K (2020) Drug repurposing for coronavirus (COVID-19): in silico screening of known drugs against coronavirus 3CL hydrolase and protease enzymes. *J Biomol Struct Dyn*. <https://doi.org/10.1080/07391102.2020.1758791>
- Muralidharan N, Sakthivel R, Velmurugan D, Gromiha MM (2020) Computational studies of drug repurposing and synergism of lopinavir, oseltamivir and ritonavir binding with SARS-CoV-2 protease against COVID-19. *J Biomol Struct Dyn*. <https://doi.org/10.1080/07391102.2020.1752802>
- Ton AT, Gentile F, Hsing M, Ban F, Cherkasov A (2020) Rapid identification of potential inhibitors of SARS-CoV-2 main protease by deep docking of 1.3 billion compounds. *Mol Inform*. <https://doi.org/10.1002/minf.202000028>
- Pant R, Joshi A, Maiti P, Nand M, Pande V, Chandra S (2020) Identification of potential mycolyltransferase Ag85C inhibitors of mycobacterium tuberculosis H37Rv via virtual high throughput screening and binding free energy studies. *J Mol Graph Model* 98:107584. <https://doi.org/10.1016/j.jmgm.2020.107584>
- Anand K, Ziebuhr J, Wadhwani P, Mesters JR, Hilgenfeld R (2003) Coronavirus main proteinase (3CLpro) structure: basis for design of anti-SARS drugs. *Science* 300(5626):1763–1767. <https://doi.org/10.1126/science.1085658>
- Khan RJ, Jha RK, Amera GM, Jain M, Singh E, Pathak A, Singh RP, Muthukumaran J, Singh AK (2020) Targeting SARS-CoV-2: a systematic drug repurposing approach to identify promising inhibitors against 3C-like proteinase and 2'-O-ribose methyltransferase. *J Biomol Struct Dyn*. <https://doi.org/10.1080/07391102.2020.1753577>
- Das S, Sarmah S, Lyndem S, Singha Roy A (2020) An investigation into the identification of potential inhibitors of SARS-CoV-2 main protease using molecular docking study. *J Biomol Struct Dyn*. <https://doi.org/10.1080/07391102.2020.1763201>
- Enmozhi SK, Raja K, Sebastine I, Joseph J (2020) Andrographolide as a potential inhibitor of SARS-CoV-2 main protease: an in silico approach. *J Biomol Struct Dyn*. <https://doi.org/10.1080/07391102.2020.1760136>
- Chou KC (2011) Some remarks on protein attribute prediction and pseudo amino acid composition. *J Theor Biol* 273(1):236–247. <https://doi.org/10.1016/j.jtbi.2010.12.024>
- Du X, Diao Y, Liu H, Li S (2019) MsDBP: exploring DNA-binding proteins by integrating multiscale sequence information via Chou's five-step rule. *J Proteome Res* 18(8):3119–3132. <https://doi.org/10.1021/acs.jproteome.9b00226>
- Kabir M, Ahmad S, Iqbal M, Hayat M (2020) iNR-2L: a two-level sequence-based predictor developed via Chou's 5-steps rule and general PseAAC for identifying nuclear receptors and their families. *Genomics* 112(1):276–285. <https://doi.org/10.1016/j.ygeno.2019.02.006>
- Alves VM, Bobrowski T, Melo-Filho CC, Korn D, Auerbach S, Schmitt C, Tropsha A (2020) QSAR modeling of SARS-CoV Mpro inhibitors identifies sufugolix, cenicriviroc, proglumetacin,

- and other drugs as candidates for repurposing against SARS-CoV-2. *Mol Inform*. <https://doi.org/10.1002/minf.202000113>
20. Sunseri J, Koes DR (2016) Pharmit: interactive exploration of chemical space. *Nucl Acids Res* 44(W1):W442–W448. <https://doi.org/10.1093/nar/gkw287>
  21. O’Boyle NM, Banck M, James CA, Morley C, Vandermeersch T, Hutchison GR (2011) Open babel: an open chemical toolbox. *J Chem Inform* 3:33. <https://doi.org/10.1186/1758-2946-3-33>
  22. Goodsell DS, Morris GM, Olson AJ (1996) Automated docking of flexible ligands: applications of AutoDock. *J Mol Recognit* 9(1):1–5. [https://doi.org/10.1002/\(sici\)1099-1352\(199611\)9:1%3c1:aid-jmr241%3e3.0.co;2-6](https://doi.org/10.1002/(sici)1099-1352(199611)9:1%3c1:aid-jmr241%3e3.0.co;2-6)
  23. Trott O, Olson AJ (2010) AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J Comput Chem* 31(2):455–461. <https://doi.org/10.1002/jcc.21334>
  24. Lipinski CA, Lombardo F, Dominy BW, Feeney PJ (2001) Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv Drug Deliv Rev* 46(1–3):3–26. [https://doi.org/10.1016/s0169-409x\(00\)00129-0](https://doi.org/10.1016/s0169-409x(00)00129-0)
  25. Halgren TA, Murphy RB, Friesner RA, Beard HS, Frye LL, Pollard WT, Banks JL (2004) Glide: a new approach for rapid, accurate docking and scoring. 2. Enrichment factors in database screening. *J Med Chem* 47(7):1750–1759. <https://doi.org/10.1021/jm030644s>
  26. Pronk S, Pall S, Schulz R, Larsson P, Bjelkmar P, Apostolov R, Shirts MR, Smith JC, Kasson PM, van der Spoel D, Hess B, Lindahl E (2013) GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. *Bioinformatics* 29(7):845–854. <https://doi.org/10.1093/bioinformatics/btt055>
  27. Vanommeslaeghe K, Hatcher E, Acharya C, Kundu S, Zhong S, Shim J, Darian E, Guvench O, Lopes P, Vorobyov I, Mackerell AD Jr (2010) CHARMM general force field: a force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields. *J Comput Chem* 31(4):671–690. <https://doi.org/10.1002/jcc.21367>
  28. Marsh JA, Teichmann SA (2011) Relative solvent accessible surface area predicts protein conformational changes upon binding. *Structure* 19(6):859–867. <https://doi.org/10.1016/j.str.2011.03.010>
  29. Srikumar PS, Rohini K, Rajesh PK (2014) Molecular dynamics simulations and principal component analysis on human laforin mutation W32G and W32G/K87A. *Protein J* 33(3):289–295. <https://doi.org/10.1007/s10930-014-9561-2>
  30. Kumari R, Kumar R, Lynn A (2014) g\_mmpbsa—a GROMACS tool for high-throughput MM-PBSA calculations. *J Chem Inf Model* 54(7):1951–1962. <https://doi.org/10.1021/ci500020m>
  31. Kuzmanic A, Zagrovic B (2010) Determination of ensemble-average pairwise root mean-square deviation from experimental B-factors. *Biophys J* 98(5):861–871. <https://doi.org/10.1016/j.bpj.2009.11.011>
  32. Lobanov M, Bogatyreva NS, Galzitskaia OV (2008) Radius of gyration is indicator of compactness of protein structure. *Mol Biol (Mosk)* 42(4):701–706. <https://doi.org/10.1134/S0026893308040195>
  33. Ghasemi F, Zomorodipour A, Karkhane AA, Khorramizadeh MR (2016) In silico designing of hyper-glycosylated analogs for the human coagulation factor IX. *J Mol Graph Model* 68:39–47. <https://doi.org/10.1016/j.jmgm.2016.05.011>

**Publisher’s Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.