JOURNAL OF BIOMOLECULAR STRUCTURE AND DYNAMICS
https://doi.org/10.1080/07391102.2020.1802341

Taylor & Francis
Taylor & Francis Group

Predictive modeling by deep learning, virtual screening and molecular dynamics study of natural compounds against SARS-CoV-2 main protease

Tanuja Joshi[a], Tushar Joshi[b], Hemlata Pundir[c], Priyanka Sharma[c], Shalini Mathpal[b] and Subhash Chandra[a]

[a]Computational Biology & Biotechnology Laboratory, Department of Botany, Kumaun University, S.S.J Campus, Almora, India; [b]Department of Biotechnology, Kumaun University Uttarakhand, Bhimtal Campus, Bhimtal, India; [c]Department of Botany, Kumaun University, D.S.B Campus, Nainital, India

Communicated by Ramaswamy H. Sarma

ABSTRACT

The whole world is facing a great challenging time due to Coronavirus disease (COVID-19) caused by SARS-CoV-2. Globally, more than 14.6 M people have been diagnosed and more than 595 K deaths are reported. Currently, no effective vaccine or drugs are available to combat COVID-19. Therefore, the whole world is looking for new drug candidates that can treat the COVID-19. In this study, we conducted a virtual screening of natural compounds using a deep-learning method. A deep-learning algorithm was used for the predictive modeling of a CHEMBL3927 dataset of inhibitors of Main protease (Mpro). Several predictive models were developed and evaluated based on R^2, MAE MSE, RMSE, and Loss. The best model with R^2=0.83, MAE = 1.06, MSE = 1.5, RMSE = 1.2, and loss = 1.5 was deployed on the Selleck database containing 1611 natural compounds for virtual screening. The model predicted 500 hits showing the value score between 6.9 and 3.8. The screened compounds were further enriched by molecular docking resulting in 39 compounds based on comparison with the reference (X77). Out of them, only four compounds were found to be drug-like and three were non-toxic. The complexes of compounds and Mpro were finally subjected to Molecular dynamic (MD) simulation for 100 ns. The MMPBSA result showed that two compounds Palmatine and Sauchinone formed very stable complex with Mpro and had free energy of −71.47 kJ mol^{-1} and −71.68 kJ mol^{-1} respectively as compared to X77 (−69.58 kJ mol^{-1}). From this study, we can suggest that the identified natural compounds may be considered for therapeutic development against the SARS-CoV-2.

ARTICLE HISTORY
Received 14 May 2020
Accepted 21 July 2020

KEYWORDS
COVID-19; deep learning; molecular docking; main protease (MPro); natural compounds

1. Introduction

A novel respiratory pathogen, SARS-CoV-2 has recently received worldwide attention, and has been declared a pandemic disease worldwide. The Coronavirus disease (COVID-19), which is caused by SARS-CoV-2 emerged in Wuhan City, Hubei Province, China during the late November 2019, has shown a burgeoning spread and since then as it has been known to infect more than 14.06 M people around the world, resulting in nearly 595 K deaths as of 17 July 2020 (https://www.worldometers.info/coronavirus/?utm_campaign=homeAdvegas1?). SARS-CoV-2 becoming more deadly day by day and symptoms of the disease are also changing continuously. According to WHO, most people infected with the COVID-19 will experience mild to moderate respiratory illness and recover without requiring special treatment. Older people and those with underlying medical problems like cardiovascular disease, diabetes, chronic respiratory disease, and cancer are more likely to develop serious illness. New research by ROBITZSKI on 21 April 2020 suggested that the SARS-CoV-2 virus have already mutated into more than 30 separate strains that make it far harder to fight off infections and facilitates spread (Robitzski, 2020).

Many antiviral drug combinations are being used by doctors to treat the disease but reports are indicating these drugs are not very effective to treat COVID-19. Keeping in mind this problem, many scientists are researching to find novel compounds against SARS-CoV-2, which will give benefit to the near future to fight against coronavirus disease. To fight against SARS-CoV-2 many medicinal plants and their compounds can be used to treat COVID-19 disease (Joshi, Joshi, et al., 2020). For several years, medicinal plants and their compounds have been used as traditional medicines for treating various types of diseases (Lin et al., 2014). Naturally occurring herbal medicine provides a wide variety of natural products, which can serve as an ancillary guide for unlocking many mysteries behind human illnesses (Ganjhu et al., 2015; Mahady, 2001). According to the WHO report, 80% of the population in developing countries relies on conventional plants for health needs (Ganjhu et al., 2015). Therefore to search potential and specific inhibitors of Coronavirus, we carried out the virtual screening of natural compounds against SARS-CoV-2 from Selleck-Natural-Product-Library to identify novel compounds. In this study, we present a computational screening of (1611) natural compounds using deep learning and molecular docking methods. Deep learning is a machine learning method

CONTACT Subhash Chandra scjnu@yahoo.co.in Computational Biology & Biotechnology Laboratory, Department of Botany, Kumaun University, S.S.J Campus, Almora, Uttarakhand 263601, India

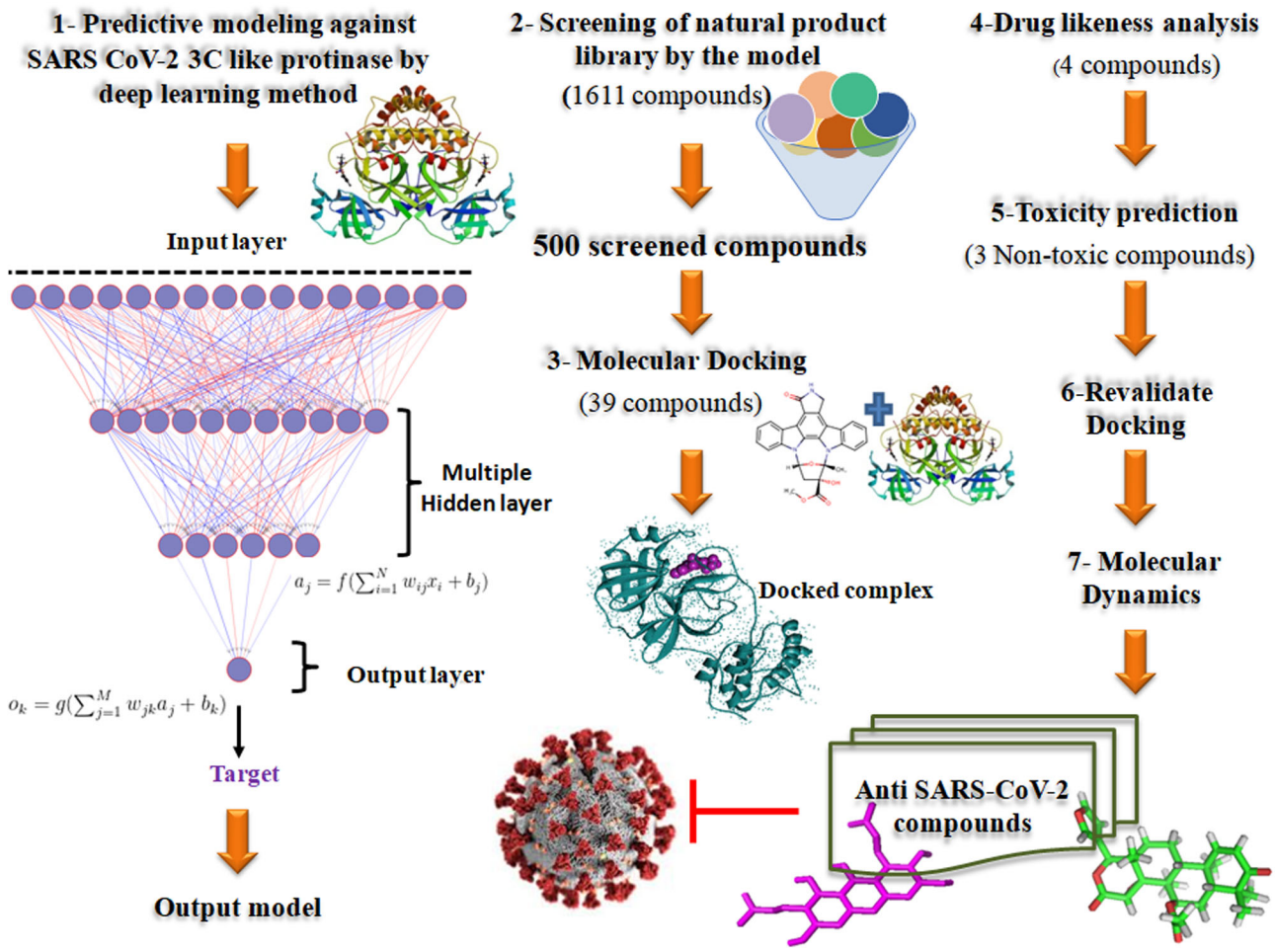


Figure 1. Depiction of the outline of predictive modeling and virtual screening.

that uses advanced algorithms inspired by biological brain functions called artificial neural networks. It is made of multiple processing layers and artificial neurons to simulate function like the human brain (Rusk, 2016). Deep learning has been reported successful in several areas like image processing, self-driving cars, natural language processing, medical diagnosis, and drug discovery, etc (Esteva et al., 2019; He et al., 2019).

To find out new compounds, we have targeted 3C-like proteinase ($3CL^{pro}$) or Main protease (Mpro). Initially, we developed a predictive model by deep learning algorithms using a CHEMBL dataset, CHEMBL3927. This dataset contains 85 compounds and their IC50 values against SARS coronavirus-I 3C-like proteinase. Deep learning models are inspired by information processing and communication patterns similar to biological nervous systems. To make a predictive model, deep learning online server was used (http://deepscreening.xielab.net). CHEMBL3927 dataset was preprocessed by the inbuilt PaDEL tool to develop Pubchem fingerprint features. Pubchem fingerprint features contain 881 features to represent molecular structures. After that, Pubchem fingerprint features were used to construct several regression models. The best regression model with high accuracy and sensitivity was used to predict natural compounds from the Selleck database by virtual screening. The screened compounds by deep learning-based virtual screening were further enriched by the molecular docking process by using AutoDock Vina. Furthermore, all screened hits were subjected to a drug-likeness investigation based on physiochemical properties using the DruLiTo tool. The common screened compounds having drug-like property and high binding affinity with target protein were taken for ADMET analysis. Further, all screened hits that were non-toxic were taken for rescoring using X-Score. Protein-ligand molecular interaction of compounds with remarkable inhibitory characteristics against the target protein was viewed with PyMOL and Discovery studio visualizer to gain structural insight into the binding interaction, including the types of bonding interaction and the amino acids involved in such interactions, compared to the reference compound. Finally, all screened compounds were further preceded to MD simulation to analyze the stability of protein-ligand complexes. In this study, we have identified two anti-SARS-CoV-2 natural compounds using deep learning and structure-based screening approach. The outline of the method is shown in Figure 1. The results of this research work may be helpful in the discovery of novel drug candidates against COVID-19.

2. Material and methods

2.1. Predictive modeling by deep learning

A deep learning algorithm was used to develop the predictive model. To make this model, deep learning online server was used (http://deepscreening.xielab.net) (Liu et al., 2019).

Table 1. Manual optimization of hyperparameters to select the best deep learning model.

S. no.	Model ID	Epoch	Hidden layer	No of neuron	Loss	R2	MSE	RMSE	MAE
1	7GT137487G20GXEW1V6U	30	2	100, 100	17.06	0.82	17.06	4	4
2	1ABB0GA4HLN8K3MHA2IP	80	2	700,200	2	0.84	2	1.44	1.2
3	6362M585L5M11FZ7LKM5	50	2	500,200	2.35	0.83	2.35	1.5	1.32
4	351I50CJ7LE437ZR0UTW	50	2	500,300	1.98	0.83	1.98	1.41	1.2
5	8EP5F65V9088ZE6BB322	30	3	500,200,100	11.36	0.82	11.36	3.37	3.21
6	380ED9D7FYYC9Z08U32Y	50	3	1,000,500,200	4.7	0.86	4.7	2.17	1.92
7	I6V7SFFMS49S8FMOAT20	80	3	1,000,500,200	4.7	0.86	4.7	2.17	1.92
8	042CC5C0GJ6J9162GH2K	80	3	1,000,700,500	17	0.83	17	4	3.86
9	K340SL383UBEJX3347V6	80	3	1500, 1000, 700	3.78	0.85	3.78	1.95	1.72
10	**285O6P887KWHBOE1A054**	**80**	**3**	**1200, 1000, 800**	**1.5**	**0.83**	**1.5**	**1.2**	**1.06**
11	140U19B3FB45MH7K2551	80	3	1300,1000, 700	5.49	0.84	5.49	2.34	2.16
12	Y8M627G1154U90FJO1A5	80	3	1200, 9000,700	21.22	0	21.22	4.61	4.5
13	37V081I70JD08XCQ474X	80	3	1,000,700,300	16.81	0.83	16.81	4	3.81
14	YH2P608A240QV2517S4X	80	3	1,000,700,200	15	0.82	15	3.88	3.52
15	3C0OFWMOF277K24267O9	80	3	1,000,500,100	8.35	0.83	8.35	2.89	2.43

The bold one is the best regression model in terms of R^2, MSE, RMSE, MAE, and Loss.

The CHEMBL dataset (CHEMBL3927) was used, which contained IC50 values for the inhibition activity of the SARS coronavirus 3C-like proteinase. This CHEMBL dataset was preprocessed for molecular vectorization by applying PubChem fingerprint which generates 881 fingerprints using PaDEL software (Yap, 2011). The PubChem fingerprints were used to construct a regression model by applying deep recurrent neural networks (RNN). Several models were generated by manual optimization of hyperparameters like learning rate, epoch, batch size, number of neurons, hidden layers, etc to select the best model (Table 1). All the hidden layers used ReLU activation function (y = max (0, 1)), while the output layer used a sigmoid function.

2.2. Model evaluation and virtual screening

Several deep learning models were built and evaluated based on several statistical matrices. In this study, regression modeling of data set was carried out for developing the model. To evaluate model performance, we used R squared (R^2), Mean squared error (MSE), Root MSE (RMSE), Mean absolute error (MAE), and Loss. The best regression model was deployed on the Selleck-Natural-Product-Library (Library id- L00012) of Selleck database which contains 1611 natural compounds for virtual screening. The model predicted 500 screened hits by virtual screening.

$$\boldsymbol{MSE} = \frac{1}{\boldsymbol{N}}\sum_{\boldsymbol{i}=1}^{\boldsymbol{N}}\left(\boldsymbol{y_i}-\hat{\boldsymbol{y}}\right)^2 \quad \boldsymbol{RMSE} = \sqrt{\boldsymbol{MSE}} = \sqrt{\frac{1}{\boldsymbol{N}}\sum_{\boldsymbol{i}=1}^{\boldsymbol{N}}\left(\boldsymbol{y_i}-\hat{\boldsymbol{y}}\right)^2}$$

$$\boldsymbol{MAE} = \frac{1}{\boldsymbol{N}}\sum_{\boldsymbol{i}=1}^{\boldsymbol{N}}\left|\boldsymbol{y_i}-\hat{\boldsymbol{y}}\right| \quad \boldsymbol{R^2} = 1 - \frac{\sum\left(\boldsymbol{y_i}-\hat{\boldsymbol{y}}\right)^2}{\sum\left(\boldsymbol{y_i}-\hat{\boldsymbol{y}}\right)^2}$$

where,

$\hat{y}$ = Predicted Value of y

y_i = Mean value of y

2.3. Protein receptor and ligand preparation

The 3D structure of COVID-19 Main protease (MPro) co-crystallized with an inhibitor i.e. X77 was retrieved by downloading PDB ID 6W63 from the Protein Data Bank in PDB format (https://www.rcsb.org). All water molecules and existing ligand were removed from the protein molecule using PyMOL software and then hydrogen atoms were added to the receptor molecule by using AutoDockTools (ADT). The 3D structure of reference molecules, X77 which is N-(4-tert-butylphenyl)-N-[(1R)-2-(cyclohexylamino)-2-oxo-1-(pyridin-3-yl) ethyl]-1H-imidazole-4-carboxamide was retrieved from 3D structure of Mpro i.e. 6W63 in SDF format. The natural compounds, as well as X77 were converted from SDF to mol2 chemical format using Open babel. Thereafter, non-polar hydrogens were merged while polar hydrogen was added to protein and ligand, and subsequently saved into dockable pdbqt format for molecular docking analysis.

2.4. Active site confirmation

The CASTp online web server is used in this study for locating, delineating, and measuring geometric and topological properties of protein structure. After predictive modeling, the active site of the protein was checked by Computed Atlas of Surface Topography of proteins (CASTp 3.0) server (http://cast.engr.uic.edu) to confirm that the ligands are linked with protein on the active amino acid. The CASTp 3.0 provides many active amino acid residues which may be vital for protein-ligand interaction. The predicted pocket associated with the residues bound with reference compound was considered for rigid docking. DS Visualizer (BIOvIA, 2015) was also used to visualize the hydrogen and hydrophobic interactions of X77 with Mpro. Therefore, pocket on the target protein provided by CASTp which was associated with the binding of its inhibitor X77 was selected for Molecular docking. Thr25, Thr26, Leu27, His41, Cys44, Met49, Tyr54, Phe140, Leu141, Asn142, Gly143, Ser144, Cys145, His163, His164, Met165, Glu166 Leu167, Pro168, His172, Asp187, Arg188, and Gln189 were taken as the reference amino acids for virtual screening as they were reported to actively participate in the stabilization of its natural substrates (Figure 5(b)).

2.5. Molecular docking

The screened compounds by the best deep-learning model were enriched by molecular docking. Molecular docking simulation was performed with 500 screened compounds

and X77 with target protein by using Autodock Vina (Trott & Olson, 2010). For docking, a three-dimensional grid box was set into X= −23.36, *Y* = 13.84, and Z= −29.63 grid points, and the grid spacing was 67.06 × 35.58 × 31.63 Å for X, Y and Z coordinates respectively. The number of exhaustiveness was set to eight for predicting the accurate result. Throughout the molecular docking process, the ligand molecules were flexible and the receptor was kept as rigid. Finally, the result in the form of binding energy was extracted from the software. The best confirmation with the low binding energy or docking score as compared to X77 were chosen for further analysis.

2.6. Pharmacokinetics and drug-likeness analysis

The compounds that were finalized by Autodock Vina after virtual screening were further proceeded to predict their pharmacokinetics and drug-likeness. Various physicochemical properties, Log P, pharmacokinetics, and drug-likeness were predicted by DruLiTo open-source software. Lipinski (Lipinski et al., 2001), Veber (Veber et al., 2002), Ghose (Ghose et al., 1999), and CMC-like (Ghose et al., 1999) filters were used in DruLiTo to predict the drug-likeness of compounds.

2.7. ADMET analysis

The compounds having drug-like property and good binding affinity with target protein were taken for the extensive ADMET analysis. The ADMET (Absorption, Distribution, Metabolism, Excretion, and Toxicity) properties are very important for approving a drug. ADMET prediction of the compounds was performed using web servers; admetSAR and PreADMET, which are quick, accurate, and easy-to-use prediction servers. The admetSAR server has 95,629 compounds in its dataset that are FDA approved and are used to predict the key features for ADMET. Here, several features including blood-brain barrier (BBB), human intestinal absorption (HIA), caco-2 permeability, P-gp substrate/inhibitor, plasma protein binding, cytochrome p450 (CYP450) substrate/inhibitor, human Ether-a-go-go-Related Gene (hERG) inhibition, AMES toxicity, carcinogenicity, biodegradability, etc. were predicted by these servers.

2.8. Scoring and visualization

The compounds that were drug-like and non-toxic were rescored using X-score (Wang et al., 2002). X-score uses three different empirical scoring functions viz. HPScore (Hydrophobic Pair), HMScore (Hydrophobic Match), and HSScore (Hydrophobic surface). VDW, H-Bond, RT denotes Van der Waals interaction, Hydrogen bonding, rotatable bonds respectively. They can be written as -:

$$\mathbf{HPScore} = C_{0.1} + C_{VDW.1*(}(\mathrm{VDW}) + C_{HB.1*(}(\mathrm{HBond}) + C_{HP*}(\text{Hydrophobic Pair}) + C_{RT.1*}(\mathrm{Rotor})$$

$$\mathbf{HMScore} = C_{0.2} + C_{VDW.2*}(\mathrm{VDW}) + C_{HB.2*}(\mathrm{HBond}) + C_{HM*}(\text{Hydrophobic Match}) + C_{RT.2*}(\mathrm{Rotor})$$

$$\mathbf{HSScore} = C_{0.3} + C_{VDW.3*}(\mathrm{VDW}) + C_{HB.3*}(\mathrm{HBond}) + C_{HS*}(\text{Hydrophobic surface}) + C_{RT.3*}(\mathrm{Rotor})$$

$$\mathbf{X-Score} = (\mathrm{HPScore} + \mathrm{HMScore} + \mathrm{HSScore})/3$$

PyMOL, a molecular viewer (Yuan et al., 2017) was used to visualize the docked pose of hit compounds at the active site of Mpro. Further other interactions along with hydrogen bonds were studied by DS Visualizer.

2.9. MD Simulation

The MD Simulation study of Mpro and screened ligand-protein complexes was performed using GROMACS 5.0 (Pronk et al., 2013) package of Molecular Dynamics as described in a publication (Joshi, Sharma, et al., 2020). MD Simulation was performed on a work station with configuration Ubuntu 16.04 LTS 64-bit, 8GB RAM, Intel®CoreTM i7-9900K CPU, and 6 GB GPU. Four systems were created, one for predicting the stability of the Mpro-X77 complex and others for Mpro-ligand complexes and subjected for 100 ns MD Simulation studies. The topology file for ligand and protein was generated by using CGenFF server and pdb2gmx respectively by applying CHARMM 36 force field (Vanommeslaeghe et al., 2009). After that, a water solvated system was built by using the water model of TIP3P with dodecahedral periodic boundary conditions having box vectors of equal length 9.81 nm. The solutes were centered in the simulation box with a minimum distance to the box edge of 10 Ð (1.0 nm). After defining the box, all the systems were solvated using the TIP3P water model in a dodecahedral box. Each solvated system was neutralized by the addition of 4 Na + ions. Energy minimization was done at 10 KJ/mol with steepest descent Algorithm by using Verlet cut off-scheme taking Particle Mesh Edward (PME) columbic interactions and the total nsteps taken by all systems during energy minimization cycle were 50,000. After that, position restraints were applied in the equilibration step. Then, NVT equilibration was done in 300 K and 5000 ps of steps and NPT equilibration taking Parrinello-Rahman (pressure coupling), 1 bar reference pressure, and 5000 ps of steps. At last, the production MD of the protein and protein-ligand complex was run for 100 ns. All the MD Simulations were performed with a time step interval of 2 fs. After successful completion of MD Simulation for 100 ns, the Root mean square deviation (RMSD), Root mean square fluctuation (RMSF), and Radius of Gyration (Rg) were calculated using g_rms, g_rmsf, g_gyrate, tools of GROMACS 5.0.7.

2.10. Post-MD simulation

After successful completion of MD Simulation for 100 ns, we computed the numbers of H-bonds, solvent accessible surface area (SASA), and the average distance between protein and ligand to quantify the strength of the interaction between protein-ligand complexes. The numbers of H-bonds, SASA, and average distance were calculated by g_hbond, g_sasa, g_dist tool of GROMACS 5.0.7. Further, Principal component analysis (PCA) was carried out by g_covar, g_anaeig, and g_sham tools. The MD trajectories were analyzed by visual molecular

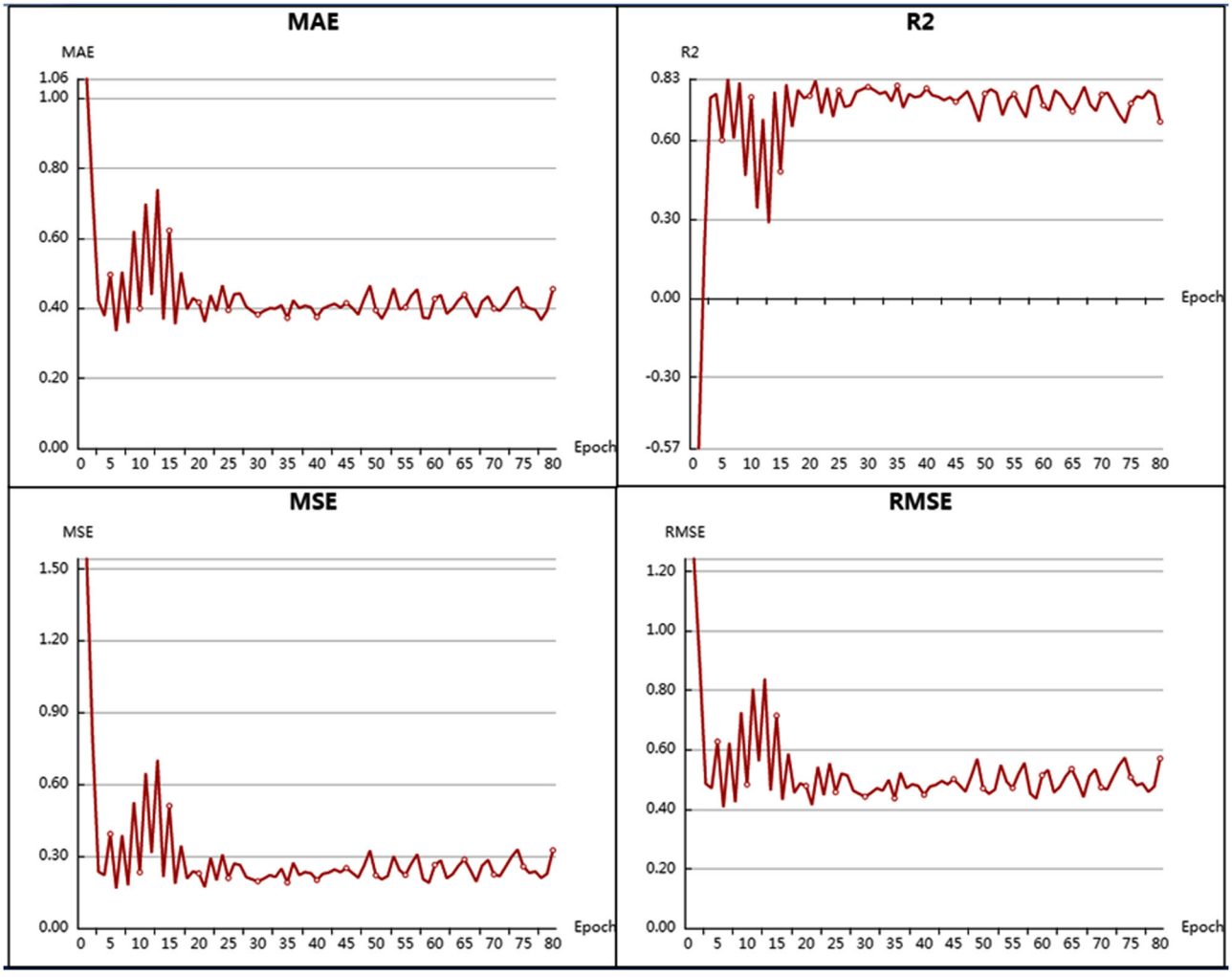


Figure 2. Performance of the best deep learning regression model for Mpro enzyme of SARS-CoV.

dynamics (VMD) software. Finally, the xmgrace tool was used for generating and visualizing the plots.

2.11. Binding free energy calculation using MM-PBSA

MMPBSA (Molecular Mechanics Poisson-Boltzmann surface area) method is widely used to calculate the binding free energy to predict the stability of the protein-ligand complex after MD Simulation (Kumari et al., 2014). Binding free energy calculations consist of free solvation energy (polar and non-polar solvation energies) and potential energy (electrostatic and Vander Waals interactions). Here, binding free energy calculations of protein-ligand complexes were done by the MMPBSA method. The MD trajectories were processed before doing MM-PBSA calculations for the last 10 ns. Then average binding energy calculations were done with 'python' script provided in g_mmpbsa.

3. Results and discussion

3.1. Predictive modeling and virtual screening

The interrelations between IC50 values and molecular fingerprints were modeled by a deep learning algorithm to build a predictive model. To develop the best model, several hyperparameters were manually optimized and statistical parameters were analyzed. Finally, the best model was achieved with learning rate 0.01, Epochs 80, batch size 16, hidden layers 3, number of neuron 1200, 1000, and 800, activation function ReLU, Drop out 0, and output function sigmoid. The best model showed the acceptable range of statistical parameters like R^2, MSE, RMSE, MAE, and loss and yielded good performance with R^2 value (0.83), MSE (1.5), RMSE (1.2) MAE (1.06) and loss (1.5) (Figure 2).

After that, the deep learning model was subjected to carry out virtual screening on the Selleck-Natural-Product-Library of Selleck database which contains the library of 1611 natural compounds. Virtual screening resulted in 500 hits showing a range of value scores from 6.9 to 3.8 as shown in Figure 3. All the 500 screened natural compounds were retrieved in a single file in SDF format from the server. The single SDF file of 500 compounds was split into individual SDF files by using the"ChemmineR" package of R (version 3.4.3).

3.2. Active site confirmation

Pockets for probable active sites on the target protein (Mpro) were identified by using CASTp online tool. The

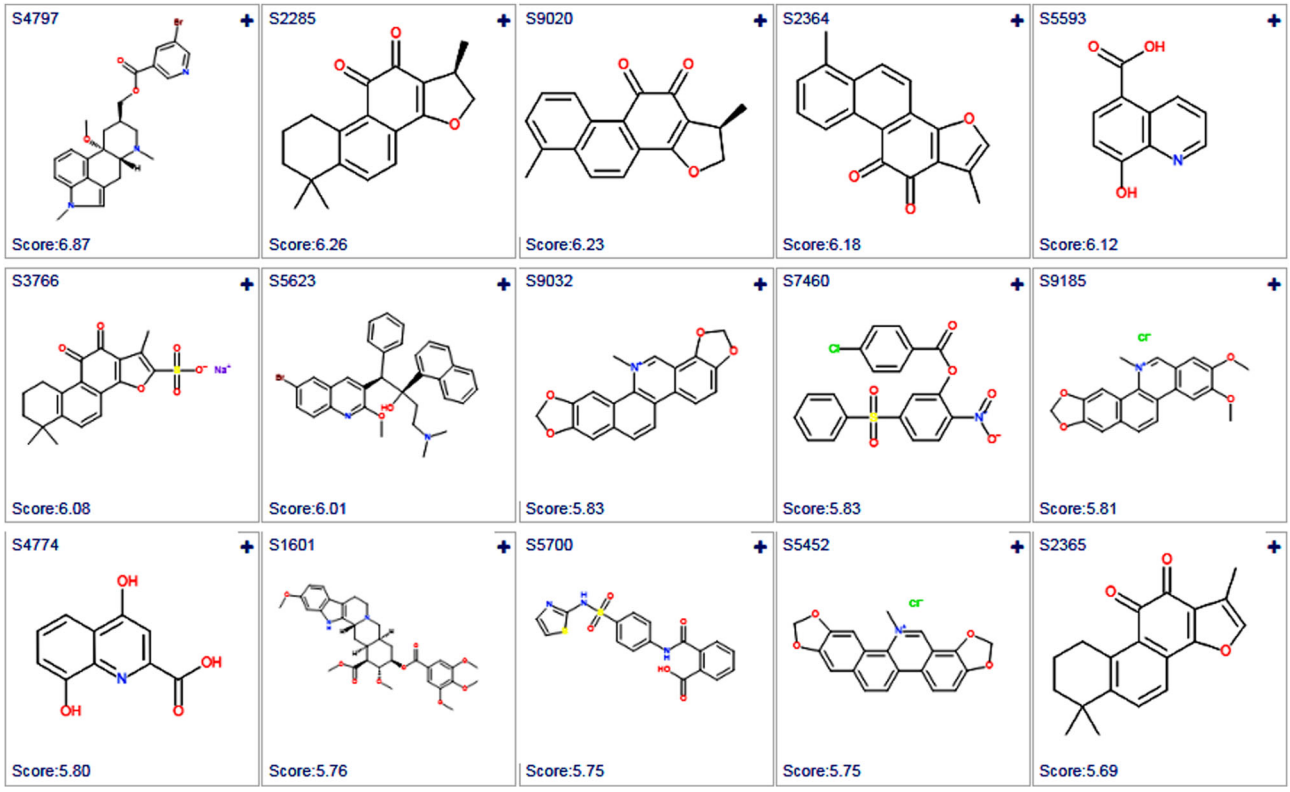


Figure 3. Screening results of Mpro model against Selleck natural product library.

pocket selected for virtual screening has the following amino acid residues;-

Thr25, Thr26, Leu27, His41, Cys44, Thr45, Ser46, Met49, Pro52, Tyr54, Phe140, Leu141, Asn142, Gly143, Ser144, Cys145, His163, His164, Met165, Glu166, Leu167, Pro168, His172, Asp187, Arg188, Gln189, Thr190, and Gln192 (Figure 4(b)). This pocket was selected because it contains all the amino acid residues (Thr25, Thr26, Leu27, His41, Cys44, Met49, Tyr54, Phe140, Leu141, Asn142, Gly143, Ser144, Cys145, His163, His164, Met165, Glu166 Leu167, Pro168, His172, Asp187, Arg188, and Gln189) that are associated with the binding of X77 which is an inhibitor of Mpro (Figure 5(b)). According to CASTp results, the active site area covered by Mpro enzymes was 304.26 and the volume was 296.68 (Figure 4(a)).

3.3. Molecular docking

Before screening the ligands using molecular docking, the docking protocol was validated by re-docking the X77 into its binding pocket within the Mpro crystal structure to obtain the correct coordinates and docked pose. The result showed that the docked X77 was completely superimposed with crystallized X77 (Figure 5(a)). Thus, this protocol was considered good enough for reproducing the docking results similar to the X-ray crystal structure and can be applied for further docking experiments.

All compounds ($n = 500$), screened by the deep learning model were docked in the active site of Mpro by using AutoDock Vina for predicting the best possible binding pose of ligands and lower binding energy. From molecular docking, a total of 39 compounds were selected which showed binding energy ranging from −11.8 to −8.2 kcal mol^{-1}. The binding energy of X77 was −8.2 kcal mol^{-1} (Table 2). All 39 compounds showed lower binding energy in comparison with X77. Among the screened compounds, 7-Epitaxol (S9265) showed the lowest binding energy i.e. −11.8 kcal mol^{-1} followed by Rifapentine (S1760) which had the binding energy of −10.5 kcal mol^{-1} and Indacaterol (S5654) that showed the highest binding energy i.e. −8.2 kcal mol^{-1} which was comparable to the reference molecule. The results show that screened compounds may have the same mechanism of action as the reference molecules. Then, all these 39 compounds and the X77 were further used for Drug-likeness prediction.

3.4. Pharmacokinetics and drug-likeness analysis

Various physicochemical properties like molecular formula, molecular weight (Mw), rotatable bonds, hydrogen bond acceptor (HBA), hydrogen bond donor (HBD), topological polar surface area (TPSA), etc. are important for the compound to be considered for a drug candidate. DruLiTo software was used to predict the physicochemical properties of 39 compounds that were obtained after molecular docking. DruLiTo calculates more than 23 physicochemical properties which are important for evaluating the drug-likeness of a molecule. Here, the drug-likeness was measured under the different filters of drug-likeness i.e. Lipinski, Veber, Ghose, and CMC-like filters. Among the 39 compounds, 18 compounds follow Lipinski RO5 and only four compounds i.e. Sanguinarine, Palmatine, Sauchinone, and

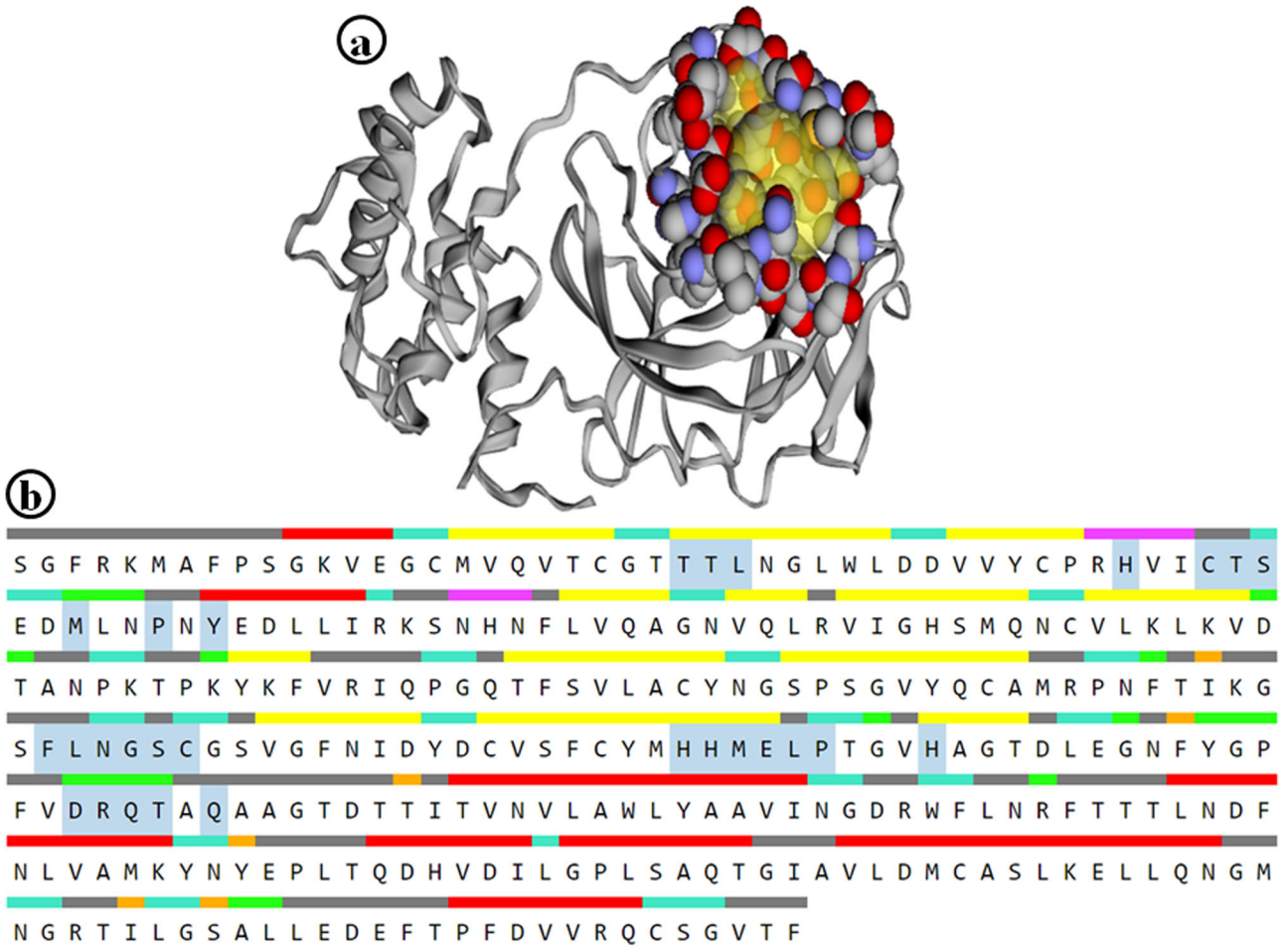


Figure 4. Active binding site of target protein-(a) Active site area (b) Active amino acid residue (Highlighted).

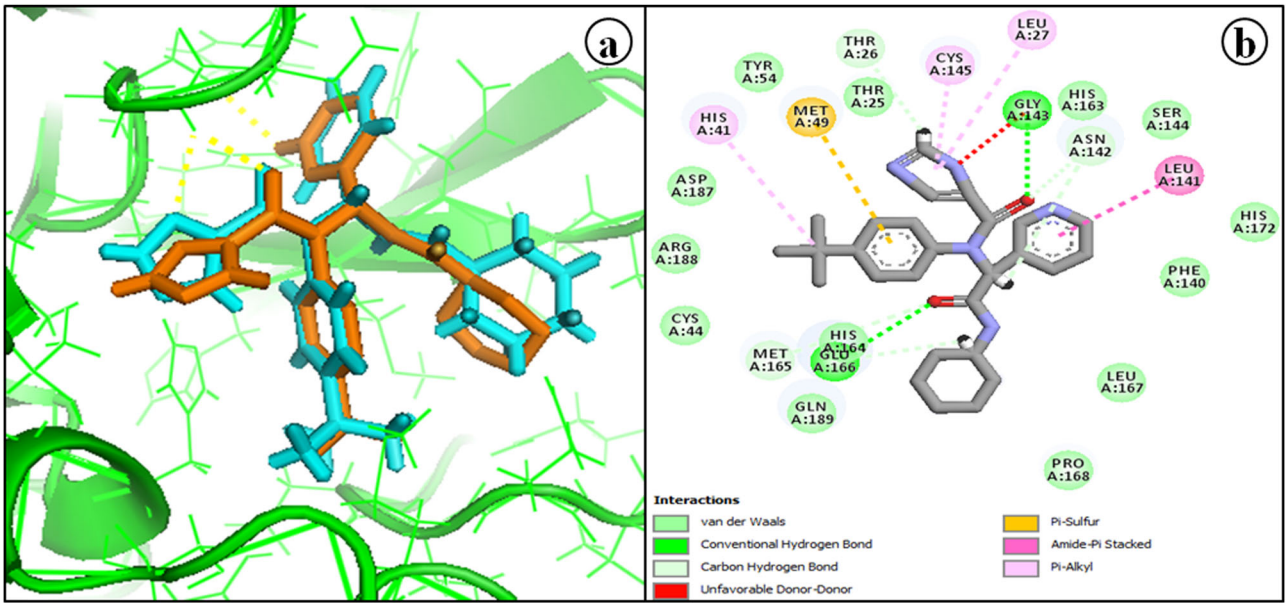


Figure 5. The superimposition of the docked X77 with its X-ray crystal structure, Orange color indicates Docked X77 and Blue color indicates Experimental X77 (a). 2D interaction of X77 with Mpro crystal structure as identified from Protein data bank (b).

Tabersonine show better pharmacokinetics and successfully passed all filters (Table 3). The compounds which show better pharmacokinetics and satisfied the fundamental drug-likeness rules are accepted for showing drug-like nature. After that, these four compounds were subjected to ADMET prediction. So from out of 1611 natural compounds, we have selected 4 compounds for ADMET analysis.

Table 2. Summary of Molecular docking results between Mro and screened hits.

S. No.	Name of Hit Compound	Compound ID	Molecular formula	Deep learning score	Binding energy (kcal mol^{-1})
1	Reference (X77)	145998279	$C_{27}H_{33}N_5O_2$	–	−8.2
2	Nicergoline	S4797	$C_{24}H_{26}BrN_3O_3$	6.78	−8.6
3	Rifapentine	S1760	$C_{47}H_{64}N_4O_{12}$	5.80	−10.5
4	Rifampin	S1764	$C_{43}H_{58}N_4O_{12}$	5.72	−10.3
5	Reserpine	S1601	$C_{33}H_{40}N_2O_9$	6.20	−8.7
6	Sanguinarine	S9032	$C_{20}H_{14}NO_4^+$	6.05	−8.3
7	BTB06584	S7460	$C_{19}H_{12}ClNO_6S$	5.51	−8.8
8	Bedaquiline	S5623	$C_{32}H_{31}BrN_2O_2$	6.16	−9.3
9	Tanshinone IIA sulfonate (sodium)	S3766	$C_{19}H_{17}NaO_6S$	5.68	−8.9
10	Ecabet sodium	S4853	$C_{20}H_{27}NaO_5S$	5.31	−8.8
11	Cephalomannine	S2408	$C_{45}H_{53}NO_{14}$	4.64	−9.8
12	7-Epitaxol	S9265	$C_{47}H_{51}NO_{14}$	4.63	−11.8
13	Nafcillin Sodium	S4042	$C_{21}H_{21}N_2NaO_5S$	4.44	−9
14	Sennoside B	S4018	$C_{42}H_{38}O_{20}$	4.50	−9.1
15	Sennoside A	S4033	$C_{42}H_{38}O_{20}$	4.50	−9.1
16	Lithospermic acid	S9259	$C_{27}H_{22}O_{12}$	4.51	−8.9
17	Salvianolic acid B	S4735	$C_{36}H_{30}O_{16}$	4.51	−10
18	Coptisine chloride	S5249	$C_{19}H_{14}ClNO_4$	5.22	−8.5
19	10-Deacetylbaccatin-III	S2409	$C_{29}H_{36}O_{10}$	4.40	−9.9
20	Protopine	S3883	$C_{20}H_{19}NO_5$	4.85	−8.9
21	Dicloxacillin Sodium	S4111	$C_{19}H_{16}Cl_2N_3NaO_5S$	4.69	−9
22	Berberrubine	S9237	$C_{19}H_{16}ClNO_4$	5.12	−8.4
23	Cefpiramide sodium	S5353	$C_{25}H_{23}N_8NaO_7S_2$	5.14	−9.5
24	Doxycycline	S5159	$C_{22}H_{24}N_2O_8$	4.58	−8.9
25	Dibenzoyl Thiamine	S5474	$C_{26}H_{26}N_4O_4S$	4.44	−8.9
26	Palmatine	S3769	$C_{21}H_{22}NO_4^+$	4.98	−8.7
27	Itraconazole	S2476	$C_{35}H_{38}Cl_2N_8O_4$	4.17	−8.3
28	Indacaterol	S5654	$C_{24}H_{28}N_2O_3$	4.61	−8.2
29	Rotenone (Barbasco)	S2348	$C_{23}H_{22}O_6$	4.38	−9.2
30	Sauchinone	S9406	$C_{20}H_{20}O_6$	4.72	−8.9
31	Tenacissoside I	S9030	$C_{44}H_{62}O_{14}$	4.18	−8.7
32	Tabersonine	S9427	$C_{21}H_{24}N_2O_2$	4.12	−8.6
33	Anamorelin	S4980	$C_{31}H_{42}N_6O_3$	4.89	−9.7
34	Chelidonine	S9154	$C_{20}H_{19}NO_5$	4.33	−8.4
35	Terconazole	S5033	$C_{26}H_{31}Cl_2N_5O_3$	4.27	−8.5
36	Corynoline	S9085	$C_{21}H_{21}NO_5$	4.24	−8.9
37	Buparvaquone	S4971	$C_{21}H_{26}O_3$	4.80	−8.4
38	Folic acid	S4605	$C_{19}H_{19}N_7O_6$	4.62	−8.5
39	Calcium folinate	S5136	$C_{20}H_{21}CaN_7O_7$	4.45	−8.7
40	NAD+	S2518	$C_{21}H_{27}N_7O_{14}P_2$	4.39	−9.1

3.5. ADMET analysis

ADMET analysis is one of the major factors for drug testing and design. The screened four compounds were further used for ADMET prediction using PreADMET and admetSAR database. Out of the selected four compounds, three (Palmatine, Sauchinone, and Tabersonine) compounds have acceptable ADMET properties and are non-toxic while the remaining one (Sanguinarine) is toxic (Table 4).

The BBB (Blood-Brain Barrier) permeability describes the efficiency of drugs in terms of crossing the BBB and its action on the central nervous system. The admetSAR gave the + ve and − ve sign for the compounds which can pass and fail the BBB, respectively. In our study, we observed that all four compounds were able to cross this barrier. The computational BBB value corresponds to its entry into the central nervous system. The acceptable range of BBB values for an ideal drug candidate ranges between −3.0 and 1.2 (Nisha et al., 2016). All the compounds have the BBB value under this range. Absorption of the drug in the gastrointestinal tract is a crucial factor for oral drug delivery and it is described by HIA. All four compounds showed absorption in the human intestine. The permeability of the drug molecule from the large intestine can be assessed by CaCo-2 (colorectal carcinoma) permeability. In our study, all four compounds can pass from the Caco-2 cell line while X77 failed in these criteria. Log S refers to the solubility of the drug molecule that ideally ranges between −6.5 and 0.5. All the compounds are showing Log S values under this range. P-glycoprotein (P-gp) receptor present on the cell surface is involved in the efflux of xenobiotics. admetSAR predicts two classes, as either predicted hit is substrate/non-substrate of P-gp or predicted hit is an inhibitor/non-inhibitor of P-gp. The P-gp substrate indicates that this molecule can be effluxed by these P-gp proteins while P-gp non-substrate indicated that this compound cannot be effluxed by P-gp proteins. Likewise, P-gp inhibitor and non-inhibitor compounds can inhibit and non-inhibit the P-gp proteins, respectively. Out of the four studied compounds, two compounds acted as non-substrates while the other two compounds and X77 were substrates of P-gp. All the studied four compounds act as a non-inhibitor of P-gp (Table 4). The distribution factor is renal organic cationic transporter inhibition/non-inhibition. Out of the four compounds, two compounds have shown inhibition while the other two compounds and X77 showed non-inhibition of renal organic cationic transporter. The major enzyme involved in the metabolism of xenobiotics inside the cell is Cytochrome P450 (CYP450). admetSAR server can predict substrate and inhibitor of Cyp450 enzymes. The data on the metabolism profile of hit compounds are also compiled in

Table 3. The parameters showing different physiochemical properties of hits to satisfy drug-likeness rules.

Name of Compound	Mw	LogP	HBA	HBD	TPSA	AMR	No. of Rotatable Bond	No. of Atom	No. of Rigid Bond	No. of Aromatic Ring	Lipinski Rule	Ghose Filter	CMC Like Rule	Veber Filter	PAINS filter	Drug-likeness alert
X77 (Reference)	458.26	3.397	7	1	74.13	129.53	9	66	28	3	+	+	–	+	+	Accepted
Sanguinarine	332.09	3.002	4	0	39.93	97.3	0	39	30	4	+	+	+	+	+	Accepted
Palmatine	352.15	2.546	4	0	39.93	106	4	48	25	3	+	+	+	+	+	Accepted
Sauchinone	356.13	2.978	6	0	63.22	93.64	0	46	31	1	+	+	+	+	+	Accepted
Tabersonine	336.18	2.755	4	1	41.57	103.46	3	49	26	1	+	+	+	+	+	Accepted

Table 4. The renal clearance of the compounds is predicted by an excretion parameter, MDCK (Madin Darby Canine Kidney). All hit compounds showed better results in terms of MDCK than reference. admetSAR also predicts toxicity and carcinogenicity of the predicted hits. In our study, three compounds and X77 were found to be non-toxic while only one compound, Sanguinarine was toxic. We found that the reference and hit compounds were non-carcinogenic. The Human ether-a-go-go-related gene (HERG) encodes a membrane channel protein (potassium ion channel) and its inhibition can lead to QT syndrome. In our study, we found that all the compounds have shown non-inhibition toward HERG i.e. they showed no risk of QT syndrome. Lethal dose50 (LD50) refers to the dose a compound required to kill 50% of the population of an organism. The LD50 was predicted *in silico* in a rat simulation model. Low LD50 values denote the high efficacy of the compound. In this study, all hit compounds showed LD50 values between 2 to 3 $mol{\cdot}kg^{-1}$ similar to the reference. From the ADMET profile, we selected Palmatine, Sauchinone, and Tabersonine for rescoring which showed acceptable ADMET properties and fulfill all the enlisted criteria.

3.6. Scoring and visualization of the docked complex

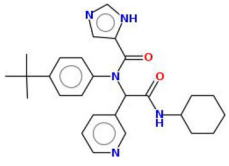
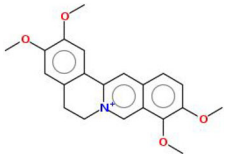
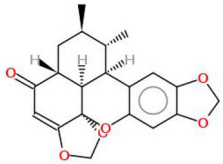
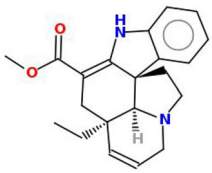
From the virtual screening, molecular docking, drug-likeness, and ADMET prediction, we found Palmatine, Sauchinone, and Tabersonine as potential inhibitors of SARS-CoV-2 Mpro as they are drug-like, non-toxic, approved ADMET and all other criteria. The re-scoring of the hits by X-Score was performed for predicting the accurate binding affinity. The virtual screening score from deep learning, docking score (binding energy) from Auto Dock Vina, and X-Score of these three compounds are given in Table 5. The reference molecule X77 showed a binding affinity of $-8.2\,Kcal.mol^{-1}$ from Autodock Vina and $-9.67\,Kcal.mol^{-1}$ from X-Score. X-score results show that the predicted three compounds have a good binding affinity towards Mpro. Among the screened hits, Palmatine showed the deep learning score 4.98 and binding affinity of $-8.7\,Kcal.mol^{-1}$ and $-8.12\,Kcal.mol^{-1}$ from Autodock Vina and X-Score respectively, other compound Sauchinone showed the deep learning score 4.72 and binding affinity of $-8.9\,Kcal.mol^{-1}$ from Autodock Vina and $-8.74\,Kcal.mol^{-1}$ from X-Score while Tabersonine showed the deep learning score 4.12 and binding affinity of $-8.6\,Kcal.mol^{-1}$ from Autodock Vina and $-8.27\,Kcal.mol^{-1}$ from X-Score.

PyMOL was used to visualize the 3D interactions of the protein-ligand complex. The docked poses of reference and these three compounds with Mpro are shown in Figure 6. According to Figure 6, Palmatine, forms one hydrogen bond having distance 2.7 Å with Gln189 (Figure 6(b)), another compound Sauchinone form two hydrogen bonds of 2.4 Å and 2.9 Å distance with the Met49 and His41 respectively (Figure 6(c)) while Tabersonine forms a hydrogen bonds of 2.2 Å distance with Cys44. The reference molecule, X77 found to interact with Gly143, Ser 144, and Glu166 of Mpro with 2.6 Å, 2.9 Å, and 1.9 Å distance respectively through hydrogen

Table 4. The ADMET profile of screened compounds obtained from PreADMET and admetSAR server.

	Hit Compounds				
Parameters	X77 (Reference)	Sanguinarine	Palmatine	Sauchinone	Tabersonine
Absorption					
BBB probability	−/0.5768	+/0.9651	+/0.9287	+/0.8851	+/0.9573
HIA probability	+/0.9273	+/0.7267	+/0.8017	+/0.9959	+/0.9941
Caco-2 permeability probability	−/0.6563	+/0.7712	+/0.8444	+/0.6721	+/0.5687
Caco-2 permeability	0.7730	1.2338	1.3889	1.5354	0.9175
Aques solubility/ logS	−3.6185	−3.2332	−3.0227	−4.1626	−3.0809
Distribution					
P-glycoprotein Substrate	Substrate/0.5920	Non-substrate/0.7655	Non-substrate/0.5248	substrate/0.6015	substrate/0.8960
P-glycoprotein Inhibitor	inhibitor/0.7331	Non-inhibitor/0.9594	Non-inhibitor/0.6853	Non-inhibitor/0.7964	Non-inhibitor/0.6589
Renal Organic Cation Transporter	Non-inhibitor/0.7989	Non-inhibitor/0.6641	inhibitor/0.5390	Non-inhibitor/0.7869	inhibitor/0.6649
Metabolism					
CYP-2C9 substrate/inhibitor	Non-substrate/ Non-inhibitor	Non-substrate/ Non-inhibitor	Non-substrate/Non-inhibitor	Non-substrate/inhibitor	Non-substrate/ Non-inhibitor
CYP-2D6 substrate/inhibitor	Non-substrate/ Non-inhibitor	Non-substrate/ inhibitor	substrate/inhibitor	Non-substrate/inhibitor	Non-substrate/inhibitor
CYP-3A4 substrate/inhibitor	substrate/inhibitor	Non-substrate/ inhibitor	substrate/Non-inhibitor	substrate/inhibitor	substrate/Non-inhibitor
CYP-1A2 inhibitor	Non-inhibitor	inhibitor	Non-inhibitor	inhibitor	Non-inhibitor
CYP-2C19 inhibitor	inhibitor	inhibitor	Non-inhibitor	inhibitor	Non-inhibitor
CYP inhibitory promiscuity	High	High	Low	High	Low
Excretion					
MDCK	0.05	34.41	1.20	64.94	72.45
Toxicity					
AMES Toxicity	Non-AMES-toxic	**AMES-toxic**	Non-AMES-toxic	Non-AMES-toxic	Non-AMES-toxic
Carcinogens	Non-Carcinogens	Non-Carcinogens	Non-Carcinogens	Non-Carcinogens	Non-Carcinogens
Human Ether-a-go-go-Related Gene Inhibition	Non-inhibitor	Non-inhibitor	Non-inhibitor	Non-inhibitor	Non-inhibitor
Biodegradation	Not ready biodegradable	Ready biodegradable	Not ready biodegradable	Not ready biodegradable	Not ready biodegradable
Acute Oral Toxicity	III/ 0.6450	III/0.7774	III/0.7551	III/0.5024	III/0.5367
Rat LD50	2.4787	2.3612	2.6332	2.8354	2.9389

Table 5. Details of the three selected compounds and the reference compound. Structure, Deep learning score obtained after the virtual screening, docking energies obtained by molecular docking, and X-Score analysis are provided in the table.

				Binding energy with Mpro					
					X-Score				
S. No.	Name of Hit Compound	Structure	Deep learning score	AutoDock Vina (kcal mol^{-1})	HP SCORE (−log(Kd))	HM SCORE (−log(Kd))	HS SCORE (−log(Kd))	AVERAGE_ SCORE (−log(Kd))	BINDING_ ENERGY (kcal mol^{-1})
1	Reference (X77)		–	−8.2	6.7	7.55	7.02	7.09	−9.67
2	Palmatine		4.98	−8.7	5.73	6.35	5.79	5.96	−8.12
3	Sauchinone		4.72	−8.9	6.03	7.2	5.99	6.41	−8.74
4	Tabersonine		4.12	−8.6	6.03	6.17	5.99	6.07	−8.27

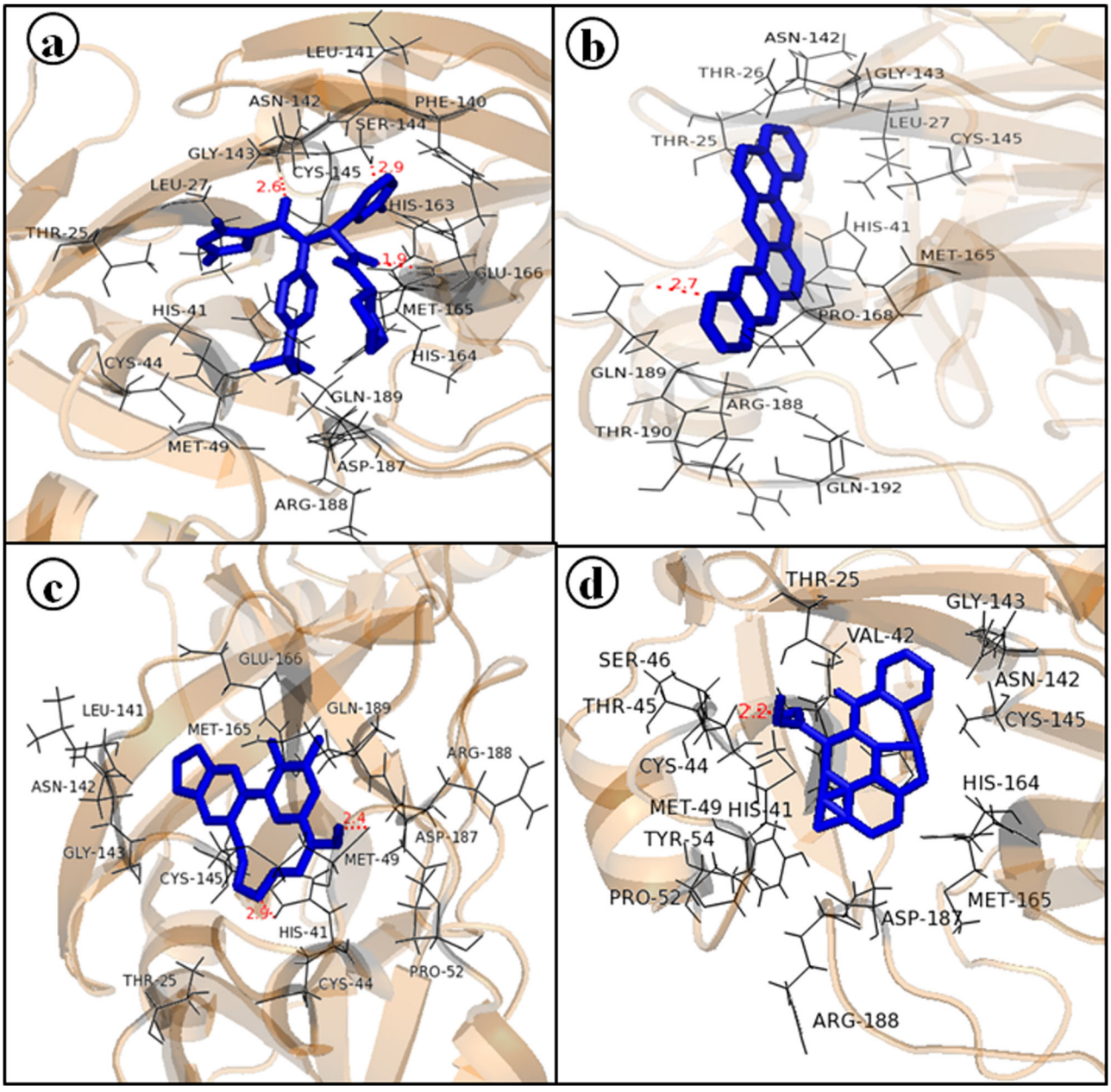


Figure 6. Docked poses of the reference (a) and top hit compounds, Palmatine (b), Sauchinone (c), and Tabersonine (d) with Mpro. The target protein is in brown color cartoon representation. Active site residues are in black colored lines. Hydrogen bonds that are formed in between protein and compound are shown as red dotted lines.

bond (Figure 6(a)). According to protein-ligand interaction, Palmatine, Sauchinone, and Tabersonine bind with the active site residues of Mpro protein, therefore these hit compounds may inhibit the Mpro of SARS-CoV-2.

Further, to get insights into the binding mechanism of the screened compound in the active sites of the Mpro, we performed 2D interactions analysis of the docked complexes by Discovery studio (DS) visualizer software as shown in Figure 7. Reference molecule, X77 is interacting with several residues via significant interactions, including hydrogen and hydrophobic interactions. It formed three conventional-hydrogen bonds with Glu166, His163, and Gly143; four Carbon-hydrogen bond with Met165, Glu166, Asn142, and Leu141; and other interactions are Vander Waals interaction with Thr25, Leu27, His164, His41, Asp187, Cys44, Arg188, Cys145, Ser144, Phe140, and Gln189; and Pi-sulphur bond with Met 49 of Mpro indicating a strong binding with Mpro (Figure 7(a)). Palmatine formed hydrogen bonds as well as other interactions with active site residues. It formed a total of five hydrogen bonds with active site residues; one conventional hydrogen bond with Gln189 and four carbon-hydrogen bonds with Thr190, Gln189, His41, and Thr26. The Mpro-Palmatine complex was also stabilized by Vander Waals interaction with Asn142, Gly143, Thr25, Leu27, Arg188, Gln192, Pro168; Pi-Alkyl Hydrophobic interaction with Cys145 and Met165; Pi-sigma interaction with Gln189; and Pi-Pi T-shaped interaction with His41 (Figure 7(b)). Sauchinone formed interaction with active site residues. It formed two

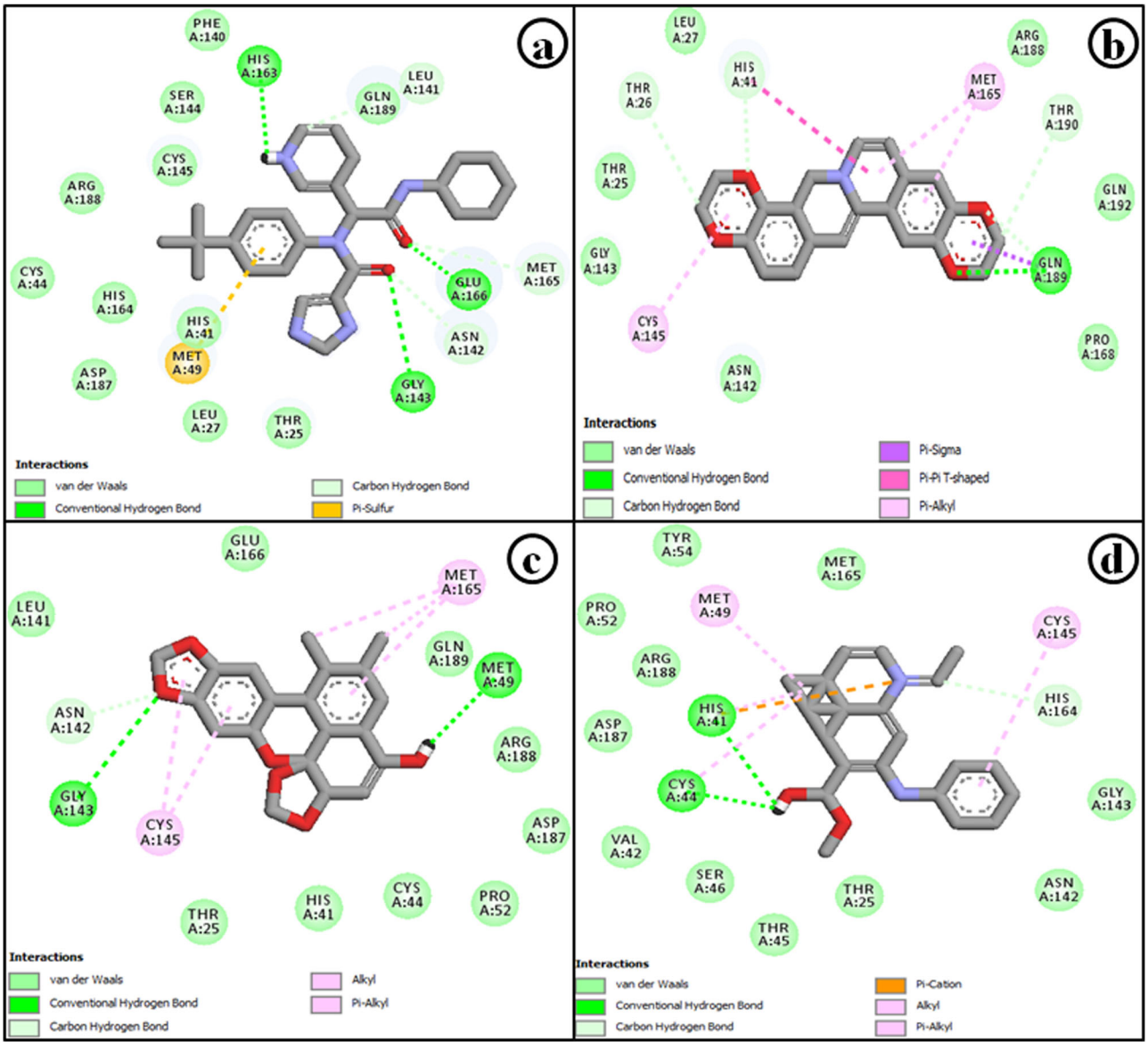


Figure 7. 2D molecular interaction of the reference (a) and top hit compounds, Palmatine (b), Sauchinone (c), and Tabersonine (d) with target protein Mpro.

conventional hydrogen bonds with Met49 and Gly143 and one carbon-hydrogen bond with Asn142. Moreover, Mpro-Sauchinone complex also showed Vander Waals interaction with Arg188, Asp187, Pro52, Cys44, His41, Thr25, Leu141, Glu166, and Gln189; Alkyl Hydrophobic interaction with Cys145 and Met165; Pi-Alkyl Hydrophobic interaction with Cys145 and Met165 (Figure 7(c)). Another compound, Tabersonine also formed interaction with active site residues. It formed two conventional hydrogen bonds with residues Cys44 and His41 and a carbon-hydrogen bond with His164. Besides, Mpro-Tabersonine complex was also stabilized by other interactions like Vander Waals interaction with Thr45, Thr25, Gly143, Ser46, Val42, Asp187 Arg188, Tyr54, Pro52, Met165, and Asn142; Alkyl Hydrophobic interaction with Cys44, His41, Met49; Pi-Alkyl Hydrophobic interaction with Cys145; and Pi-sigma interaction with His41 (Figure 7(d)).

From the molecular interaction analysis of docked complexes, we observed that all the hit compounds show H-bond interaction and other interactions with Mpro were bound to the same binding cavity having the active site residues similar to reference molecule which suggested the crucial role of these interactions to hold the ligand at the active site of the target protein. Active side residues viz. Thr25, Leu27, His41, Cys44, His164, Asp187, Arg188, Cys145, Met49, and Met165 were the common interacting residues between Mpro and hit compounds. Ultimately, from these results, we finalize these three compounds viz. Palmatine, Sauchinone, and Tabersonine for further analysis using MD Simulation to study their stability and dynamic properties.

3.7. MD Simulation

The MD Simulation can be used to explain the dynamics like structural details, conformational behavior, and stability of the target-ligand complexes, etc. The stability analysis of the native Mpro, reference complex (Mpro-X77), and the protein-ligand complexes (Mpro-Palmatine, Mpro-Sauchinone, and

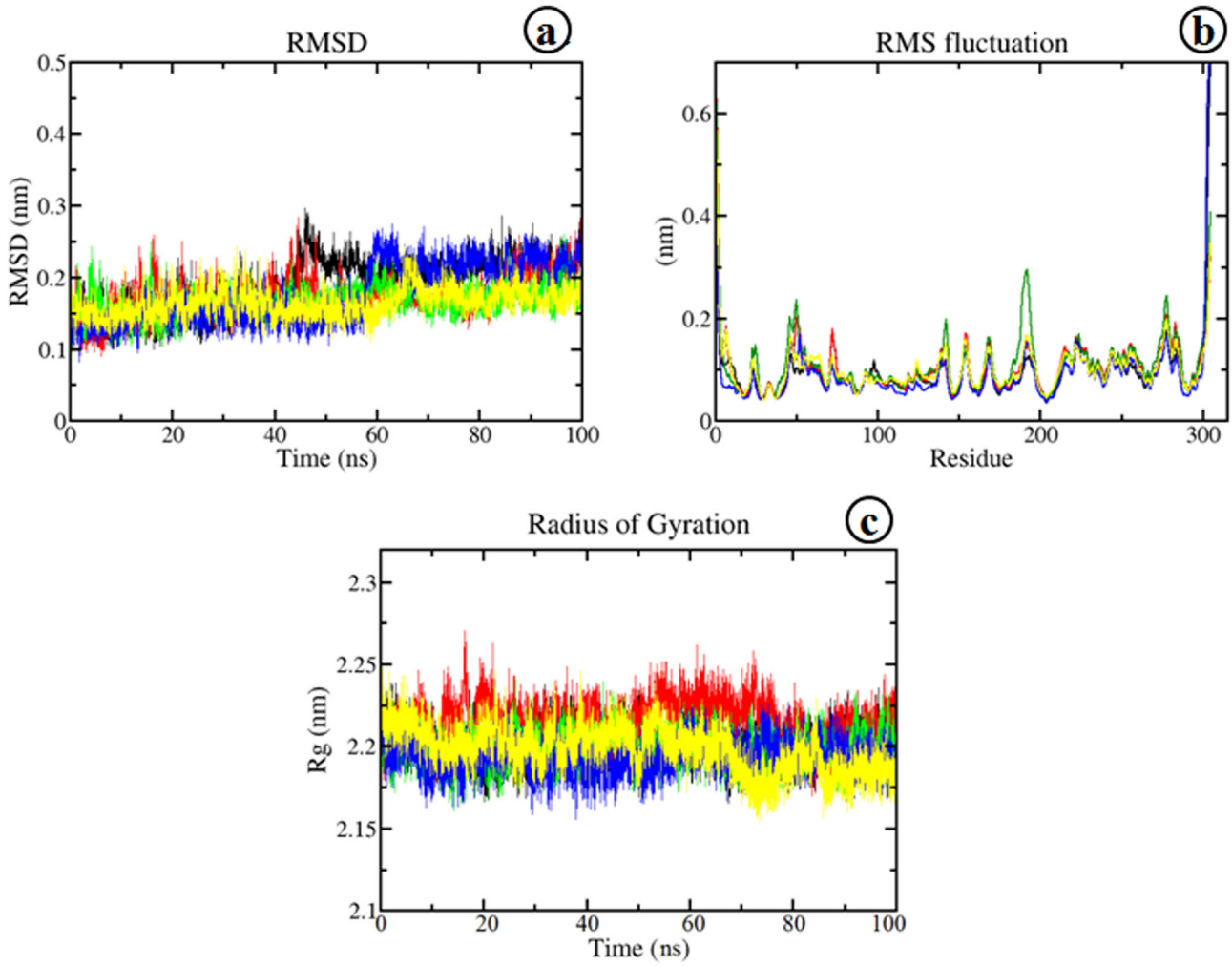


Figure 8. MD simulation studies. a. RMSD, b. RMSF, and c. Radius of gyration as a function of time. In all systems, the color code indicates- Mpro protein (black), Mpro-X77 (red), Mpro-Palmatine (green), Mpro-Sauchinone (blue), and Mpro-Tabersonine (yellow).

Mpro-Tabersonine) was performed by surrounding them into a dodecahedral box of size 669.34 nm3 at a computationally maintained temperature of 300 K. TIP3P water model was used to solvate the system. Water molecules present in the box in the case of Palmatine were 20622, in the case of Sauchinone they were 20628, in the case of Tabersonine they were 20613 and 20626 water molecules in X77. Four Na + ions were used to neutralize the system. After that, each complex was subjected to the process of energy minimization for 50,000 steps of the steepest descent, followed by equilibration and finally subjected to 100 ns MD Simulation. The structural changes and dynamic behavior in complexes were analyzed by the various computational analyses like RMSD, RG, and RMSF calculation.

3.7.1. Root mean square deviation (RMSD)

To determine the conformational and structural stability of Mpro and Mpro-ligand complexes, we monitored the differences between the backbone atoms of native protein from initial conformation to its final position through RMSD analysis. The deviations produced in protein during its simulation describe the stability of that protein's conformation. Smaller deviations in protein reflect its more stable nature. RMSD score for the C-alpha backbone was calculated for 100 ns MD Simulation to evaluate the stability of all the complexes. Figure 8(a) shows the plot of RMSD (nm) vs. time (ns) for native protein Mpro, reference complex Mpro-X77, and three Mpro-ligand complexes (Mpro-Palmatine, Mpro-Sauchinone, and Mpro-Tabersonine). From this figure, we can see that the RMSD trajectory of protein and all the complexes attained the equilibration and produced stable trajectories. The average value of RMSD for protein and all the complexes is shown in Table 6. The average RMSD values for protein were 0.18 ± 0.03 nm while for complexes; Mpro-X77, Mpro-Palmatine, Mpro-Sauchinone, and Mpro-Tabersonine were found to be 0.17 ± 0.02 nm, 0.16 ± 0.02 nm, 0.17 ± 0.04 nm, 0.16 ± 0.01 nm respectively. The reference complex Mpro-X77and Mpro-Sauchinone showed the same RMSD value, while Mpro-Palmatine and Mpro-Tabersonine showed the least value which confirmed the stability of the complexes.

3.7.2. Root mean square fluctuation (RMSF)

To investigate the conformational fluctuations of Mpro and the residues involved in the Mpro-ligand interactions, we calculated the average fluctuation of each amino acid of Mpro

Table 6. The average values of RMSD, RMSF, Rg, SASA, and Distance between protein-ligand in Mpro-ligand complexes.

		MD simulation			Post-MD simulation	
S. No	Protein/ Protein-ligand complex	Average RMSD	Average RMSF	Average RG	Average SASA	Average Distance
1	Mpro	0.18 ± 0.03	0.09 ± 0.05	1.92 ± 0.13	–	–
2	Mpro-X77 (Reference)	0.17 ± 0.02	0.13 ± 0.07	1.88 ± 0.26	149.29 ± 2.77	3.51 ± 0.21
3	Mpro-Palmatine	0.16 ± 0.02	0.11 ± 0.05	1.82 ± 0.18	148.04 ± 2.33	3.58 ± 0.24
4	Mpro-Sauchinone	0.17 ± 0.04	0.08 ± 0.07	1.74 ± 0.24	148.77 ± 2.52	3.73 ± 0.17
5	Mpro-Tabersonine	0.16 ± 0.01	0.09 ± 0.37	1.49 ± 0.25	152.43 ± 3.01	3.58 ± 0.29

Table 7. Residues of the Active site and their RMSF values (Angstrom).

Residues	Mpro	Reference Mpro-X77	Mpro-Palmatine	Mpro-Sauchinone	Mpro-Tabersonine
Thr25	0.07	0.06	0.14	0.06	0.09
Leu27	0.05	0.04	0.07	0.04	0.06
His41	0.05	0.04	0.08	0.04	0.01
Cys44	0.07	0.07	0.11	0.06	0.08
Met49	0.09	0.15	0.20	0.11	0.11
Cys145	0.05	0.05	0.07	0.04	0.06
His164	0.05	0.06	0.06	0.06	0.08
Met165	0.06	0.07	0.08	0.06	0.09
Asp187	0.07	0.09	0.14	0.07	0.08
Arg188	0.07	0.10	0.16	0.08	0.09

protein by RMSF analysis. The RMSF value describes how the binding of the ligand can change the confirmation of the protein during the complex. In proteins, the rigid structures containing regions like helix and sheets showed low RMSF value, while loose structures containing region of protein like sheets and turns showed higher RMSF value. The RMSF plot of protein and all protein-ligand complexes is shown in Figure 8(b). RMSF plot shows that the secondary conformations of Mpro remain stable during the MD Simulation of 100 ns. The average RMSF values for Mpro protein, Mpro-X77, Mpro-Palmatine, Mpro-Sauchinone, and Mpro-Tabersonine complexes were recorded as 0.09 ± 0.05 nm, 0.13 ± 0.07 nm, 0.11 ± 0.05 nm, 0.08 ± 0.07 nm, and 0.09 ± 0.37 nm respectively (Table 6).

All the complexes showed similar or less average RMSF value as compared to the Mpro and Mpro-X77 complex, which suggests that they did not cause much fluctuation in protein after binding. The RMSF results represented that all predicted complexes were stable, and hence, these predicted compounds had the potential to inhibit the catalytic activity of Mpro. Using protein-ligand interaction analysis, we found that Thr25, Leu27, His41, Cys44, His164, Asp187, Arg188, Cys145, Met49, and Met165 active site residues were involved in maintaining the catalytic activity of the Mpro-ligand complex. From Table 6, it becomes clear that in all the studied complexes, the RMS fluctuation value decreased due to the ligand binding. The results suggested that due to the binding of the ligand, the RMS fluctuation of the active site residues decreased, resulting in the change in conformation of Mpro and thereby, inhibiting the activity of Mpro-X77 complex. The RMS fluctuation values of catalytically important residues are shown in Table 7. In all studied complexes, the RMSF for each active site residue is lower than 0.2 nm, which means that the binding cavity is quite stable during the MD Simulation. This narrow range of RMSF values of the active site residues of the studied complexes demonstrated that these compounds were capable of forming stable interactions with the Mpro during the MD Simulation.

3.7.3. Radius of gyration (RG)

The radius of gyration (*Rg*) is an effective parameter to understand the level of compaction in the structure of the protein in the absence and presence of ligands. The time evolution plot of Rg for Mpro protein and all Mpro-ligand complexes is shown in Figure 8(c). The average Rg value for Mpro protein, Mpro-X77, Mpro-Palmatine, Mpro-Sauchinone, and Mpro-Tabersonine complexes were found to be 1.92 ± 0.13 nm, 1.88 ± 0.26 nm, 1.82 ± 0.18 nm, 1.74 ± 0.24 nm, and 1.49 ± 0.25 nm, respectively (Table 6). The Mpro-Tabersonine and Mpro-Sauchinone complex showed much less Rg value as compared with the Mpro protein and other complexes, suggesting that they form more compact and stable complex as compared to other systems, although other hits also showed relatively good Rg value similar to reference complex. If the Rg values remain relatively consistent throughout the MD Simulation, it can be regarded as a stably folded structure; otherwise, it would be considered unfolded. From Table 6, it is visible all the systems exhibited relatively similar and consistent values of Rg as reference (X77) which indicates that these are perfectively superimposed with each other and they exhibit similar compactness and stability similar to X77. These results show that all complexes achieved relatively stable folded conformation during the 100 ns trajectory of MD Simulation at the constant temperature of 300 K and the constant pressure of 1 atm. Overall, it can be concluded that the complexation of protein with hit compounds increases the compactness/rigidity of the Mpro structure, leading to increased overall stability.

3.8. Post-MD simulation

3.8.1. Hydrogen bonds

The receptor-ligand complexes are stabilized by different kinds of interactions like hydrogen bonds, hydrophobic bonds, electrostatic, and other interactions but out of them, the hydrogen bonds are very specific interactions that play a crucial role in the stabilization of protein-ligand complex.

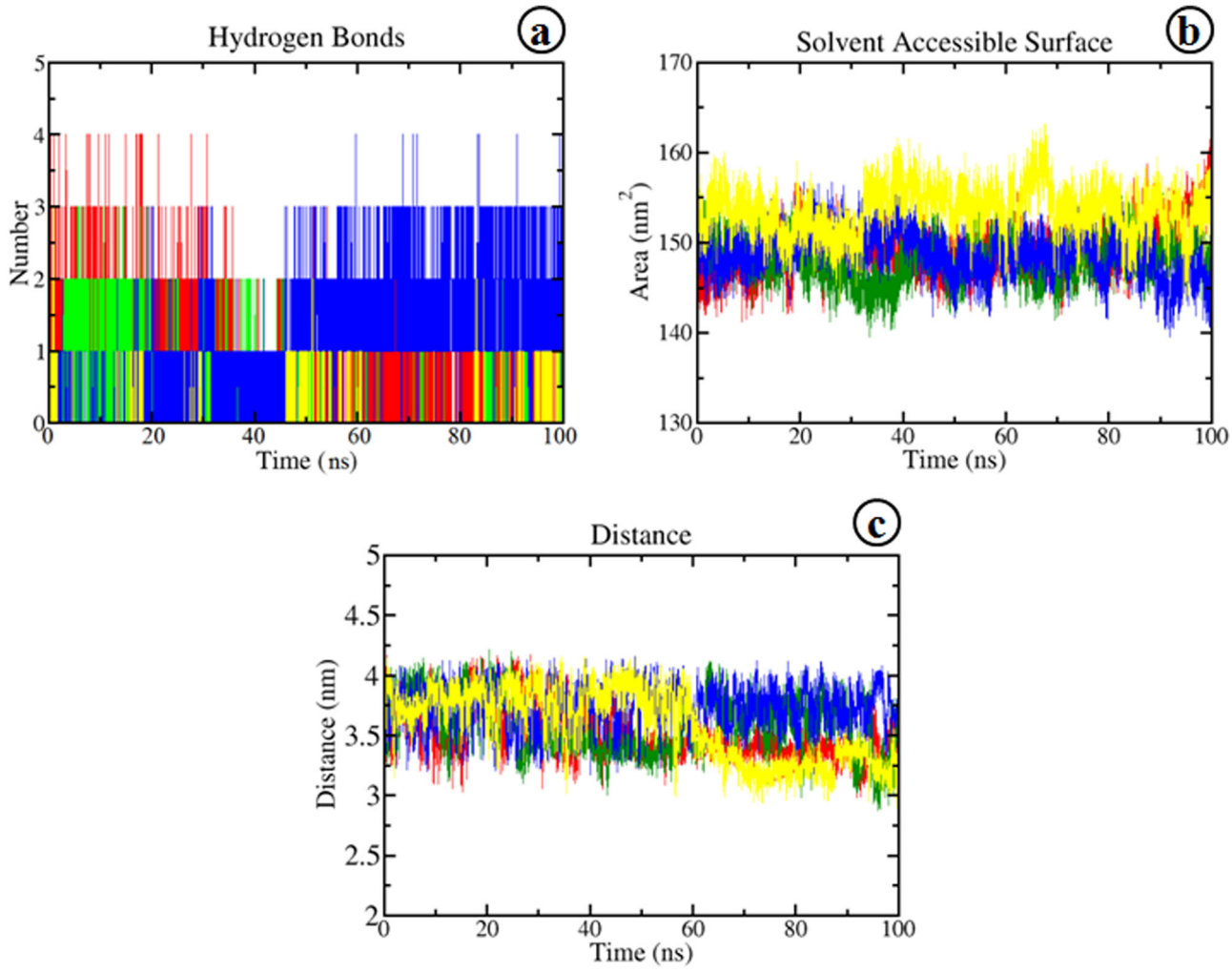


Figure 9. Post-MD simulation studies. a. Number of H-bonds, b. SASA, and c. Distance as a function of time. In all panels, the colors indicate- Mpro-X77 (red), Mpro-Palmatine (green), Mpro-Sauchinone (blue), and Mpro-Tabersonine (yellow).

Along with the structural stability of protein, hydrogen bonds also play a significant role in ligand binding at the active site of the receptor as well as strongly influence drug specificity, metabolization, and adsorption in drug design. Furthermore, the bonding patterns were assessed by observing the fluctuations of the hydrogen bonds in all the complexes. Figure 9(a) shows the maximum number of hydrogen bonds versus time for all complexes during 100 ns MD Simulation. The result shows the appearance of a maximum of four H-bond interactions between reference compound X77 and Mpro during the MD Simulation period of 100 ns. Maximum three hydrogen bonds were observed in Mpro-Palmatine and Mpro-Tabersonine complexes while Mpro-Palmatine showed a maximum of four H-bonds with Mpro as X77. These observed bonding parameters indicated that all compounds were bound to the Mpro as effectively and tightly as X77.

3.8.2. Solvent accessible surface area (SASA)

Solvent accessible surface area (SASA) analysis enables us to measure the proportion of the protein surface which can be accessible by the water solvent and analyze interactions between complex and solvent during the MD Simulation. Figure 9(b) shows the plot of SASA value vs. time for all the protein-ligand complexes (Mpro-X77, Mpro-Palmatine, Mpro-Sauchinone, and Mpro-Tabersonine). The average SASA of 148.04 nm^2 was calculated for Mpro-Palmatine, 148.77 nm^2 for Mpro-Sauchinone, while for Mpro-Tabersonine it was found to be 152.43 nm^2. Likewise, reference complex Mpro-X77 showed the average value of SASA to be around 149.29 nm^2 (Table 6). All complexes showed average SASA values approximately similar to the reference indicating their stability similar to Mpro-X77.

3.8.3. Distance

The gmx_mpi pairdist module of GROMACS was used to calculate the distance between the structure of Mpro and ligands during the MD Simulation. Figure 9(c) shows the plot of distance value vs. time for all the protein-ligand complexes (Mpro-X77, Mpro-Palmatine, Mpro-Sauchinone, and Mpro-Tabersonine). The average distance for the reference complex Mpro-X77 was observed to be around 3.51 nm. Likewise, an average distance of 3.58 nm was observed for the Mpro-Palmatine complex while for the complexes Mpro-Sauchinone and Mpro-Tabersonine it was computed to be around 3.73 nm and 3.58 nm respectively (Table 6). From the

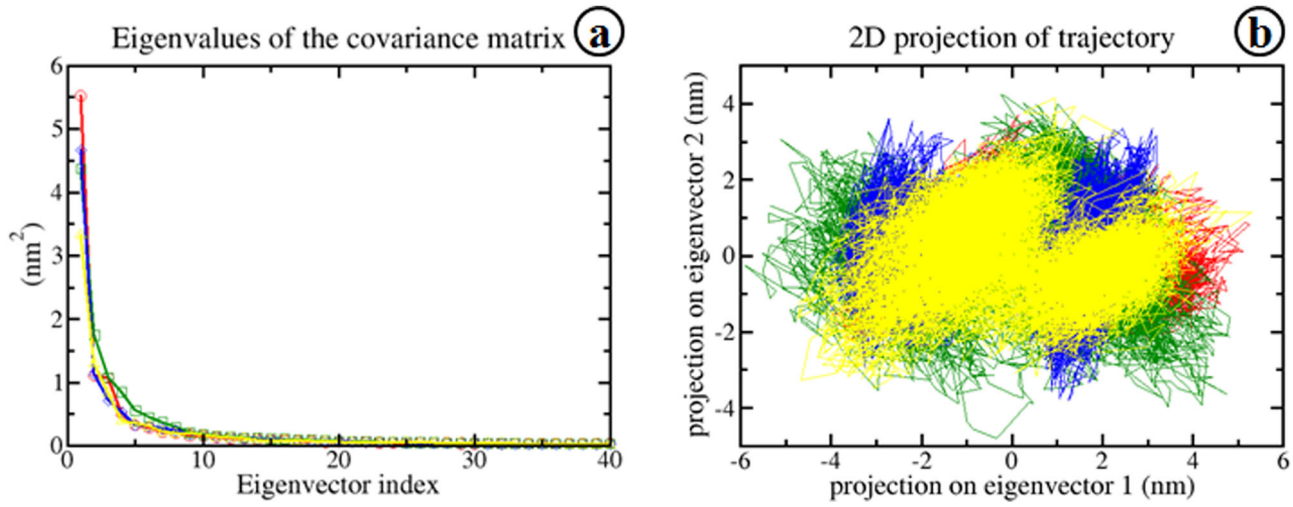


Figure 10. Principle Component Analysis. a. A Plot of eigenvalue *vs* eigenvector index. The first 40 eigenvectors were considered for PCA analysis and b. PCA scatter plot along with the first two principal components, PC1 and PC2 showing all-atom fluctuations. In both panels, the color code indicates- Mpro-X77 (red), Mpro-Palmatine (green), Mpro-Sauchinone (blue), and Mpro-Tabersonine (yellow).

result, it is obvious that all hits show a similar value of average distance as X77 which indicates that during MD Simulation these ligands have the same distance toward Mpro.

3.8.4. Principal component analysis (PCA)

To investigate the significant concerted motions during ligand binding, the PCA analysis was carried out. It is well known that the overall motion of the protein is determined by only the first few eigenvectors (Yang et al., 2014). In this study, the diagonalization of the matrix is used to calculate the eigenvectors. For this study, we selected the first 40 eigenvectors for the calculation of concerted motions. Figure 10(a) represents the eigenvalues that are obtained from the diagonalization of the covariance matrix of atomic fluctuations in decreasing order versus the corresponding eigenvector for Mpro-X77, Mpro-Palmatine, Mpro-Sauchinone, and Mpro-Tabersonine. It was observed that out of the 40 eigenvectors, the first ten eigenvectors accounted for 74.59%, 73.35%, 71.67%, and 67.81% of total motions for Mpro-X77, Mpro-Palmatine, Mpro-Sauchinone, and Mpro-Tabersonine (Figure 10(a)). All the studied complexes showed very fewer motions as compared with the reference compound. So from the PCA, we concluded that all compounds showed fewer motions and form a stable complex with Mpro. From PCA, we conclude that ligand binding leads to the change in protein conformation as well as in dynamics.

The 2D projection plot generation in PCA is another way to achieve the dynamics of complexes. Figure 10(b) shows the 2D projection of the trajectories in the phase space for the first two principal components, PC1 and PC2 for Mpro-X77, Mpro-Palmatine, Mpro-Sauchinone and Mpro-Tabersonine complexes. The complex which occupied less phase space showed and stable cluster represent a more stable complex while the complex that occupied more space and showed a non-stable cluster represented a less stable complex. From the figure, it can be concluded that the Mpro-Sauchinone (Blue) and Mpro-Tabersonine (Yellow) complexes were highly stable as they occupied less space in the phase space, and the cluster was well defined as compared to Mpro-X77 (Red), and Mpro-Palmatine (Green) complexes. All results indicate that Tabersonine and Sauchinone formed a more stable complex with Mpro. The 2D PCA result was also in agreement with the above PCA and other MD Simulation results.

The Gibb's energy plot for PC1 and PC2 was also calculated and is shown in Figure 11(a–d). The plot shows Gibbs energy value ranging from 0 to 12.5, 0 to 11.9, 0 to 13, and 0 to 11.9 kJ·mol^{-1} for Mpro-X77, Mpro-Palmatine, Mpro-Sauchinone, and Mpro-Tabersonine, respectively. All the studied complexes showed significantly similar or low energy as compared with the Mpro-X77, which suggestes that these complexes follow the energetically more favorable transition from one conformation to another.

3.9. Binding energy calculation and energetic contribution of individual residues

The Binding free energy calculation was carried out using the g_mmpbsa tool for all systems, considering the last 10 ns of MD trajectories as shown in Table 8. The binding free energies were composed of their energy components: polar solvation energy, SASA non-polar solvation energy and non-bonded interaction energies (Van der Waal and electrostatic energy) get insights into their contributions. Through Table 8, it can be observed that Mpro-Sauchinone and Mpro-Palmatine complexes possess the least negative binding energy i.e. −71.68 +/− 9.23 kJ mol^{-1} and −71.47 +/− 9.750 kJ mol^{-1} respectively, whereas, Mpro-Tabersonine complex displayed positive free energy (59.21 +/− 41.96 kJ mol^{-1}). Two hit compounds (Sauchinone and Palmatine) showed significantly better binding energy as compared to the reference complex Mpro-X77 (−69.58 +/− 29.43 kJ mol^{-1}). It confirms that both compounds can bind efficiently at the binding site of Mpro and could be used as lead compounds against COVID-19. Further various energy terms

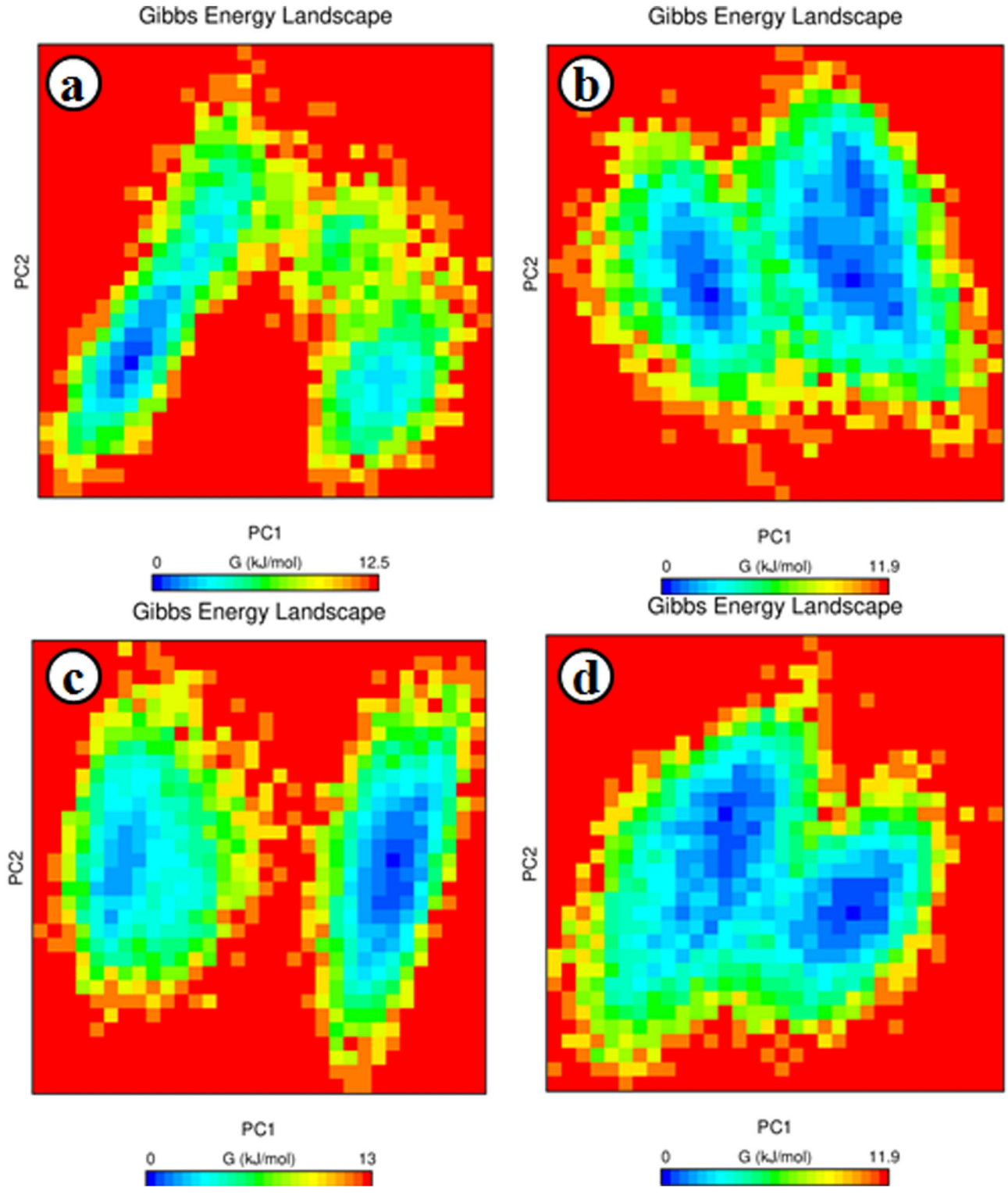


Figure 11. The Gibbs free energy landscape for Mpro-X77 (a), Mpro-Palmatine (b), Mpro-Sauchinone (c), and Mpro-Tabersonine (d).

Table 8. A Table showing the Van der Waal, electrostatic, polar salvation, SASA, and total 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 binding Energy (kJ mol^{-1})
1	Mpro-X77 (Reference)	−118.54 +/−10.21	−5.42 +/− 6.19	70.21 +/− 29.85	−15.83 +/− 1.23	−69.58 +/− 29.43
2	Mpro-Palmatine	−117.65 +/−12.25	−8.49 +/− 6.01	68.50 +/− 12.25	−13.82 +/− 1.15	−71.47 +/− 9.750
3	Mpro-Sauchinone	−120.16 +/− 9.55	−27.01 +/− 5.93	89.11 +/− 10.28	−13.61 +/− 0.81	−71.68 +/− 9.23
4	Mpro-Tabersonine	0.000 +/− 0.00	0.05 +/− 0.06	59.03 +/− 42.02	0.13 +/− 1.91	59.21 +/− 41.96

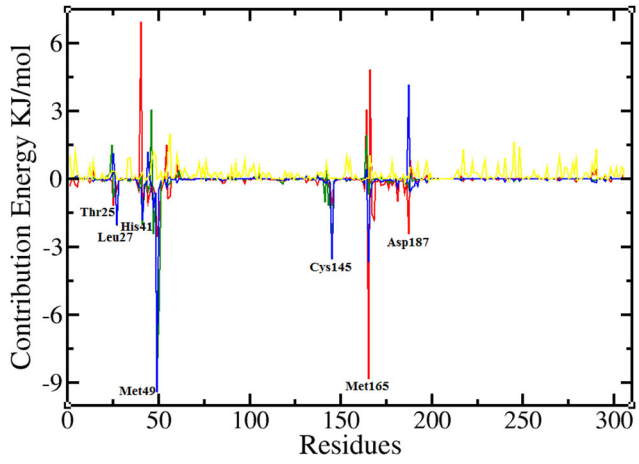


Figure 12. The contributions of individual amino acid residues of Mpro to the total binding energies of Mpro-ligand complexes. In all systems, the color code indicates- Mpro-X77 (red), Mpro-Palmatine (green), Mpro-Sauchinone (blue), and Mpro-Tabersonine (yellow). Negative values indicate a stabilization effect on Mpro-ligand interactions.

contributing towards binding free energy revealed that in all the studied complexes, the driving component of binding was the van der Waals energy which played a major contribution in strengthening the binding mode. The polar solvation energy did not showed a favorable contribution to the total binding in all the studied complexes. The electrostatic Energy and SASA non-polar solvation energy contribute similarly to the binding free energy.

To analyze the key residues involved in protein-ligand interaction, per residue interaction energy profile was created using the MM-PBSA approach for the last 10 ns of MD trajectories. The per-residue decomposition plot of the total binding energy of the Mpro-ligand complexes is shown in Figure 12. For the clear depiction of results, only active site residues are highlighted in Figure. From the plot, it was revealed that Thr25, Leu27, His41, Met49, Cys145, Met165, and Asp187 were the actively participating amino acids in all complexes. The per residue interaction plot indicated that most of the residues showed negative binding energy, while few residues showed 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. The results revealed that Met49, Cys145, and Met165 play an important role in Mpro-ligand stabilization which is in agreement in a previous study (Joshi, Sharma, et al., 2020).

From the overall MD Simulation (including RMSD, RMSF, and Rg analysis) and Post-MD Simulation (including hydrogen bonds, SASA, distance, and PCA analysis) and binding free energy analysis results, we conclude that two predicted hits Palmatine and Sauchinone are very stable complexes that show excellent binding affinities as compared to the reference compound.

Drug discovery from natural sources is a basic and novel idea in a current situation where the whole world is finding a solution to treat COVID-19. Many studies have proven that natural compounds are very effective to find potential drug candidates against different viral diseases. A recent study suggests that some natural compounds may be helpful for the treatment of COVID-19 infection (Joshi, Sharma, et al., 2020; Kumar et al., 2020). *In vitro* tests have reported that natural compound lycorine isolated from Lycoris radiata extract is candidates for the development of new anti-SARS-CoV drugs in the treatment of SARS (Li et al., 2005). Recently, an *in silico* study conducted by Wahedi et. al 2020 showed that Stilbene-based Natural Compounds particularly resveratrol, has shown *in vitro* antiviral activity against SARS-CoV-2 through disruption of its spike protein (Wahedi et al., 2020). Several recent studies have used Mpro as a molecular target to find the anti-SARS-CoV-2 compounds using *in silico* techniques (Joshi, Sharma, et al., 2020; Mittal et al., 2020). Drug-repurposing studies have also screened many compounds against COVID-19 (Elmezayen et al., 2020).

Deep learning methods are an important tool which can be used to develop a predictive model of an experimentally validated dataset of compounds and used for prediction or virtual screening of unknown dataset. Therefore, the current study was undertaken to find some natural compounds that can be used against the SARS-CoV-2 virus using various computational techniques like deep-learning-based virtual screening, molecular docking, drug-likeness analysis, ADMET, X-Score and MD Simulation. From the overall structures-based drug discovery methods, we found two anti-SARS-CoV-2 natural compounds; Palmatine and Sauchinone.

These natural compounds are also used to treat some other disease as Palmatine is an isoquinoline alkaloid that has sedative, antidepressant, antioxidative, anti-ulcerative, antacid, anticancer, and anti-metastatic activities (Long et al., 2019) and Sauchinone is an active lignan isolated from the roots of Saururus chinensis, possesses diverse pharmacological properties, such as hepatoprotective, anti-inflammatory and anti-tumor effects, etc (He et al., 2018). Based on our results, we suggest that Palmatine and Sauchinone show better scores by deep learning model and better binding energy against Mpro receptor and hence they may be considered for evaluation *in vitro* experiment against SARS-CoV-2.

4. Conclusion

The inhibition of Mpro enzyme represents a promising strategy for anti-SARS-CoV-2 drug discovery. In this study, the best deep learning model predicted many potential natural inhibitors against Mpro. After that various computational methods were performed for the identification of natural molecules as potential inhibitors of SARS-CoV-2 Mpro. From the overall study, we conclude that two compounds Palmatine and Sauchinone form a very stable complex with Mpro that show excellent binding affinities higher as compared to the reference complex. These observations suggest that these natural compounds can inhibit the activity of Mpro enzyme of SARS-CoV-2 and may be explored for the innovation and development of suitable drug candidates against COVID-19.

Acknowledgements

The authors are thankful to the Department of Botany, Kumaun University, S.S.J Campus, Almora for providing the facility, space, and resources for this work. The authors also acknowledge 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.

Disclosure statement

The authors declare that there is no conflict of interest in this paper.

ORCID

Subhash Chandra http://orcid.org/0000-0002-8978-5427

References

BIOvIA, D. S. (2015). *Discovery studio modeling environment, Release 4*. Dassault Systemes.

Elmezayen, A. D., Al-Obaidi, A., Sahin, A. T., & Yelekci, K. (2020). Drug repurposing for coronavirus (COVID-19): In silico screening of known drugs against coronavirus 3CL hydrolase and protease enzymes. *Journal* of Biomolecular Structure and Dynamics, 1–13.

Esteva, A., Robicquet, A., Ramsundar, B., Kuleshov, V., DePristo, M., Chou, K., Cui, C., Corrado, G., Thrun, S., & Dean, J. (2019). A guide to deep learning in healthcare. *Nature Medicine*, *25*(1), 24–29. https://doi.org/10.1038/s41591-018-0316-z

Ganjhu, R. K., Mudgal, P. P., Maity, H., Dowarha, D., Devadiga, S., Nag, S., & Arunkumar, G. (2015). Herbal plants and plant preparations as remedial approach for viral diseases. *Virusdisease*, *26*(4), 225–236. https://doi.org/10.1007/s13337-015-0276-6

Ghose, A. K., Viswanadhan, V. N., & Wendoloski, J. J. (1999). A knowledge-based approach in designing combinatorial or medicinal chemistry libraries for drug discovery. 1. A qualitative and quantitative characterization of known drug databases. *Journal of Combinatorial Chemistry*, *1*(1), 55–68. https://doi.org/10.1021/cc9800071

He, J., Baxter, S. L., Xu, J., Xu, J., Zhou, X., & Zhang, K. (2019). The practical implementation of artificial intelligence technologies in medicine. *Nature Medicine*, *25*(1), 30–36. https://doi.org/10.1038/s41591-018-0307-0

He, Z., Dong, W., Li, Q., Qin, C., & Li, Y. (2018). Sauchinone prevents TGF-beta-induced EMT and metastasis in gastric cancer cells. *Biomedicine & Pharmacotherapy = Biomedecine & Pharmacotherapie*, *101*, 355–361. https://doi.org/10.1016/j.biopha.2018.02.121

Joshi, T., Joshi, T., Sharma, P., Mathpal, S., Pundir, H., Bhatt, V., & Chandra, S. (2020). In silico screening of natural compounds against COVID-19 by targeting Mpro and ACE2 using molecular docking. *European Review for Medical and Pharmacological Sciences*, *24*(8), 4529–4536. https://doi.org/10.26355/eurrev_202004_21036

Joshi, T., Sharma, P., Joshi, T., Pundir, H., Mathpal, S., & Chandra, S. (2020). Structure-based screening of novel lichen compounds against SARS Coronavirus main protease (Mpro) as potentials inhibitors of COVID-19. *Molecular Diversity*, 1–13.

Kumar, A., Choudhir, G., Shukla, S. K., Sharma, M., Tyagi, P., & Bhushan, A. (2020). Identification of phytochemical inhibitors against main protease of COVID-19 using molecular modeling approaches. *Journal* of Biomolecular Structure and Dynamics, 1–11.

Kumari, R., Kumar, R., Open Source Drug Discovery Consortium, & Lynn, A. (2014). g_mmpbsa-a GROMACS tool for high-throughput MM-PBSA calculations. *Journal of Chemical Information and Modeling*, *54*(7), 1951–1962. https://doi.org/10.1021/ci500020m

Li, S.-Y., Chen, C., Zhang, H.-Q., Guo, H.-Y., Wang, H., Wang, L., Zhang, X., Hua, S.-N., Yu, J., Xiao, P.-G., Li, R.-S., & Tan, X. (2005). Identification of natural compounds with antiviral activities against SARS-associated coronavirus. *Antiviral Research*, *67*(1), 18–23. https://doi.org/10.1016/j.antiviral.2005.02.007

Lin, L. T., Hsu, W. C., & Lin, C. C. (2014). Antiviral natural products and herbal medicines. *Journal of Traditional and Complementary Medicine*, *4*(1), 24–35. https://doi.org/10.4103/2225-4110.124335

Lipinski, C. A., Lombardo, F., Dominy, B. W., & Feeney, P. J. (2001). Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Advanced Drug Delivery Reviews*, *46*(1–3), 3–26. https://doi.org/10.1016/j.addr.2012.09.019

Liu, Z., Du, J., Fang, J., Yin, Y., Xu, G., & Xie, L. (2019). DeepScreening: A deep learning-based screening web server for accelerating drug discovery. *Database*, *2019*, 1–11. https://doi.org/10.1093/database/baz104

Long, J., Song, J., Zhong, L., Liao, Y., Liu, L., & Li, X. (2019). Palmatine: A review of its pharmacology, toxicity and pharmacokinetics. *Biochimie*, *162*, 176–184. https://doi.org/10.1016/j.biochi.2019.04.008

Mahady, G. B. (2001). Global harmonization of herbal health claims. *The Journal of Nutrition*, *131*(3s), 1120S–1123S. https://doi.org/10.1093/jn/131.3.1120S

Mittal, L., Kumari, A., Srivastava, M., Singh, M., & Asthana, S. (2020). Identification of potential molecules against COVID-19 main protease through structure-guided virtual screening approach. *Journal* of Biomolecular Structure and Dynamics, 1–19.

Nisha, C. M., Kumar, A., Vimal, A., Bai, B. M., Pal, D., & Kumar, A. (2016). Docking and ADMET prediction of few GSK-3 inhibitors divulges 6-bromoindirubin-3-oxime as a potential inhibitor. *Journal of Molecular Graphics & Modelling*, *65*, 100–107. https://doi.org/10.1016/j.jmgm.2016.03.001

Pronk, S., Páll, S., Schulz, R., Larsson, P., Bjelkmar, P., Apostolov, R., Shirts, M. R., Smith, J. C., Kasson, P. M., van der Spoel, D., Hess, B., & Lindahl, E. (2013). GROMACS 4.5: A high-throughput and highly parallel open source molecular simulation toolkit. *Bioinformatics (Oxford, England)*, *29*(7), 845–854. https://doi.org/10.1093/bioinformatics/btt055

Robitzski, D. (2020). Scientists: The coronavirus has already mutated into 30+ strains. Neoscope.

Rusk, N. (2016). Deep learning. *Nature Methods*, *13*(1), 35–35. https://doi.org/10.1038/nmeth.3707

Trott, O., & Olson, A. J. (2010). AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *Journal of Computational Chemistry*, *31*(2), 455–461. https://doi.org/10.1002/jcc.21334

Vanommeslaeghe, K., Hatcher, E., Acharya, C., Kundu, S., Zhong, S., & Shim, J. (2009). CHARMM general force field: A force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields. *Journal* of Computational Chemistry., *31*(4), 671–690.

Veber, D. F., Johnson, S. R., Cheng, H. Y., Smith, B. R., Ward, K. W., & Kopple, K. D. (2002). Molecular properties that influence the oral bioavailability of drug candidates. *Journal of Medicinal Chemistry*, *45*(12), 2615–2623. https://doi.org/10.1021/jm020017n

Wahedi, H. M., Ahmad, S., & Abbasi, S. W. (2020). Stilbene-based natural compounds as promising drug candidates against COVID-19. *Journal* of Biomolecular Structure and Dynamics., 1–16.

Wang, R., Lai, L., & Wang, S. (2002). Further development and validation of empirical scoring functions for structure-based binding affinity prediction. *Journal of Computer-Aided Molecular Design*, *16*(1), 11–26. https://doi.org/10.1023/a:1016357811882

Yang, L.-Q., Sang, P., Tao, Y., Fu, Y.-X., Zhang, K.-Q., Xie, Y.-H., & Liu, S.-Q. (2014). Protein dynamics and motions in relation to their functions: Several case studies and the underlying mechanisms. *Journal of Biomolecular Structure & Dynamics*, *32*(3), 372–393. https://doi.org/10.1080/07391102.2013.770372

Yap, C. W. (2011). PaDEL-descriptor: An open source software to calculate molecular descriptors and fingerprints. *Journal of Computational Chemistry*, *32*(7), 1466–1474. https://doi.org/10.1002/jcc.21707

Yuan, S., Chan, H. C. S., & Hu, Z. (2017). Using PyMOL as a platform for computational drug design. *Wiley Interdisciplinary Reviews: Computational Molecular Science*, *7*(2), e1298.