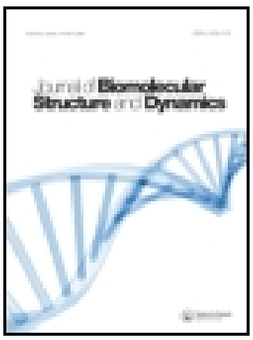


Taylor & Francis
Taylor & Francis Group

Journal of Biomolecular Structure and Dynamics

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/tbsd20

Virtual screening and molecular simulation study of natural products database for lead identification of novel coronavirus main protease inhibitors

Nancy Tripathi , Bharat Goel , Nivedita Bhardwaj , Bharat Sahu , Hemant Kumar & Shreyans K. Jain

To cite this article: Nancy Tripathi , Bharat Goel , Nivedita Bhardwaj , Bharat Sahu , Hemant Kumar & Shreyans K. Jain (2020): Virtual screening and molecular simulation study of natural products database for lead identification of novel coronavirus main protease inhibitors, Journal of Biomolecular Structure and Dynamics, DOI: 10.1080/07391102.2020.1848630

To link to this article: https://doi.org/10.1080/07391102.2020.1848630

JOURNAL OF BIOMOLECULAR STRUCTURE AND DYNAMICS
https://doi.org/10.1080/07391102.2020.1848630

Virtual screening and molecular simulation study of natural products database for lead identification of novel coronavirus main protease inhibitors

Nancy Tripathi[a], Bharat Goel[a], Nivedita Bhardwaj[a], Bharat Sahu[a], Hemant Kumar[b] and Shreyans K. Jain[a]

[a]Department of Pharmaceutical Engineering and Technology, Indian Institute of Technology (Banaras Hindu University), Varanasi, Uttar Pradesh, India; [b]Molecular Biology Unit, Institute of Medical Science, Banaras Hindu University, Varanasi, Uttar Pradesh, India

Communicated by Ramaswamy H. Sarma

ABSTRACT

3CL like protease (3CLpro or Mpro) is one of the main proteases of 2019-nCoV. The 3CLpro is a non-structural protein of SARS-CoV and has an essential role in viral replication and transcription, thus, could be a potential target for anti-SARS drug development. The present study employed ligand- and structure-based approaches to identify the potent inhibitors of 2019-nCoV protease. The e-pharmacophore developed from 3CLpro-1 yielded virtual hits, that were subjected through drug likeliness and PAINS filters to remove interfering compounds. Further comprehensive docking studies, free energy calculations and ADMET studies resulted in two virtual leads- MolPort-000-410-348 and MolPort-002-530-156. The compounds MolPort-000-410-348 and MolPort-002-530-156 displayed good docking score of −12.09 and −13.38 Kcal/mol and free binding energy of −63.34 ± 2.03 and −61.52 ± 2.24 Kcal/mol, respectively. The compounds also exhibited satisfactory predicted ADMET profile and were subjected to molecular dynamic (MD) studies. The MD simulation produced stable complexes of these ligands with 3CLpro protein and ligand RMSD in acceptable limits.

Abbreviations: ADME: absorption distribution metabolism and excretion; AMBER: assisted model building with energy refinement; BBB: blood–brain barrier; HIA: human intestinal absorption; HTVS: high throughput virtual screening; LGA: Lamarckian genetic algorithm; MD: molecular dynamics; MM-GBSA: molecular mechanics, the generalised Born surface area study; PAINS: pan assay interference compounds; RMSD: root mean square deviation; RMSF: root mean square fluctuation; SP: standard precision; XP: extra precision

ARTICLE HISTORY
Received 1 May 2020
Accepted 5 November 2020

KEYWORDS
COVID-19; natural products; virtual screening; molecular docking; protease inhibitors

1. Introduction

Coronaviruses (CoVs), positive sense, single-stranded RNA viruses belonging to family Coronaviridae, are highly diverse in nature, causing various respiratory, enteric, hepatic and neurological diseases in humans as well as animals. A novel betacoronviruses (βCoVs), known as severe acute respiratory syndrome coronavirus (SARS-CoV) is existing since more than a decade and is responsible for causing severe mortality and morbidity associated epidemic as no effective treatment is available yet (Zumla et al., 2016). The SARS-CoV first emerged in the Foshan, Guangdong, Southern China, during 2002–2003 (Zhao et al., 2003). A novel infectious disease (COVID-19), was detected in Wuhan city of Hubei province in China, caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) in December 2019. The COVID-19 reached epidemic proportions in China and declared as a pandemic by WHO due to rapid spreading, and as of 21 July 2020, 14,859,811 confirmed cases, and 613,637 deaths have been reported around the globe (World Health Organization, 2020). The 2019-nCoV possesses a high degree of sequence similarity with original SARS-CoV with their spike proteins possessing identical 3D structure in the receptor-binding domain (Zhang et al., 2020). The principal route of transmission is via respiratory droplets and can also be transmitted by contact. The clinical manifestations of COVID-19 infection at onset of infections include fever, cough, myalgia, malaise; dyspnoea as the infection progresses; and pneumonia (Huang et al., 2020). In lack of proper treatment and drug against COVID-19, the strategy for its management includes supportive care to prevent complications like acute respiratory distress syndrome, acute cardiac injury, RNAemia and secondary nosocomial infections (Cheng et al., 2007; Huang et al., 2020). In absence of FDA approved drug against 2019-nCoV, some countries are practising the use of combination of antiviral drugs and also several clinical trials are in progress to evaluate the activity of antiviral agents against the same.

Upon the viral entry, the positive genomic RNA gets directly attached to the host ribosome for the translation of its major polyproteins that are processed by proteolysis into new components for packaging new virions (Morse et al.,

CONTACT Shreyans K. Jain sjain.phe@itbhu.ac.in Department of Pharmaceutical Engineering and Technology, Indian Institute of Technology, Varanasi, Uttar Pradesh 221005, India

Supplemental data for this article can be accessed online at https://doi.org/10.1080/07391102.2020.1848630

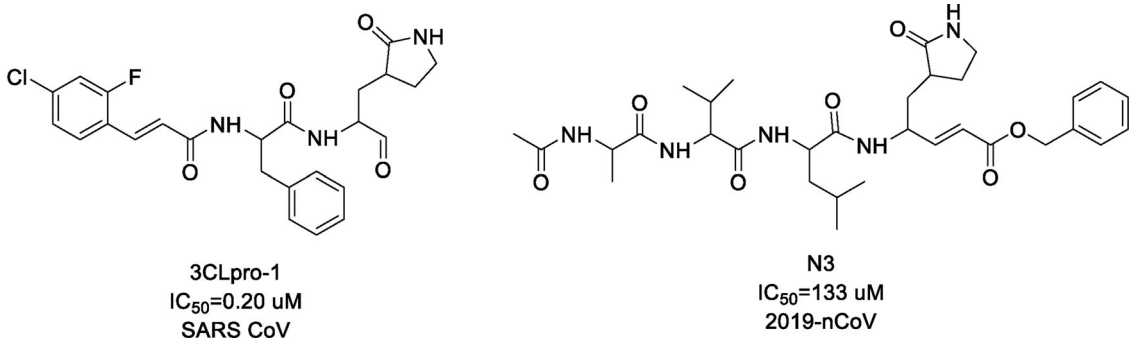


Figure 1. Chemical structures of main coronavirus protease inhibitors, 3CLpro-1 and N3.

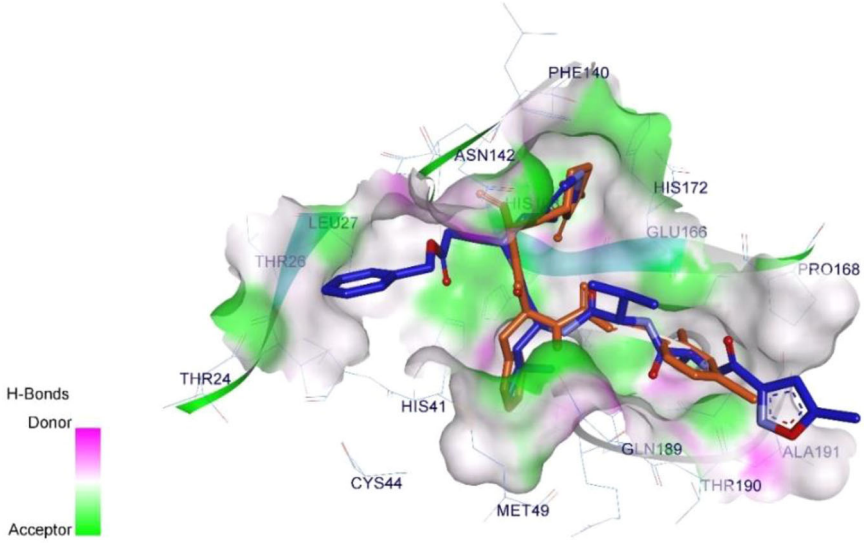


Figure 2. Binding site of co-crystallised ligand N3 (Blue color) and docked pose of 3CLpro-1 (Orange color).

2020).Two proteases of coronavirus that participate in this process are the coronavirus main protease, 3CL like protease (3CLpro, also called as Mpro) and the papain-like protease, PLpro. Both proteases are cysteine proteases, and therefore, covalent inhibitors with high potencies could potentially be developed for targeting 2019-nCoV.

The last two decades have seen much of focus on drugging SARS-CoV via the development of small-molecules, peptide and peptidomimetic inhibitors of 3CLpro. The 3CLpro could be propitious target to screen small molecule libraries as it inhibits the cleavage of viral polyproteins and stop the spread of infection. The functional importance of 3CLpro makes it a potential therapeutic target for the drug development against 2019-nCoV. The 3CLpro share high sequence similarity (96%) and structural similarity between the SARS-CoV and 2019-nCoV (Joshi et al., 2020) and its structure could be used for an initial *in silico* high throughput screening. Thus, it is possible that an inhibitor developed for the SARS-CoV may also work for the 2019-nCoV 3CLpro.

Kumar et al. (2017) identified a potent SARS-CoV's 3CLpro inhibitor 3CLpro-1 having IC_{50} of 0.6 μM. This inhibitor is the most potent peptide-based inhibitor (Cheng et al., 2015). Recently, Jin et al. (2020) identified a mechanism-based inhibitor of 2019-nCoV 3CLpro, N3 by computer-aided-drug design and subsequently determined the crystal structure of 2019-nCoV 3CLpro in complex with the compound N3 (PDB ID.6LU7). The IC_{50} of this compound was found to be 133 μM (Figure 1).

In the absence of co-crystallised structure of 3CLpro-1 with 2019-nCoV main protease, we docked the 3CLpro-1 inhibitor with the protein structure of 2019-nCoV 3CLpro (PDB ID. 6LU7). Upon docking it was observed that 3CLpro-1 occupied the same binding site as of the co-crystalized inhibitor, N3 (Figure 2). The 3CLpro is a homodimeric protein and is highly conserved among the Coronaviridae members (Muralidharan et al., 2020). The 3D structure of 3CLpro contain three domains, i.e. domain I from 8 to 101 amino acid residues, domain II from 102 to 184 amino acid residues and domain III from 201 to 306 amino acid residues. The active site is located at the cleft of domains I and II, which comprises two catalytic residues, i.e. His41 and Cys145 (Das et al., 2020).

The identification of 2019-nCoV main protease inhibitor can be extremely fruitful considering its vital role in the viral life-cycle. The objective of the study is to identify inhibitors from the natural sources that can inhibit main protease of 2019-nCoV through *in silico* and molecular modeling approach.

Table 1. Acceptable limits of physicochemical properties for ligands of natural origin.

Physicochemical properties	Acceptable limits
cLogP	<5
Molecular weight	<600
Number of HBD	<5
Number of HBA	<10
TPSA	<100 Å^2
Number of rotatable bonds	<10

2. Material and methods

All the computational studies were performed on Dell Power edge with Intel® Xeon® Silver 4114 processor running on Red Hat Linux Enterprise version 7.3 operating system.

2.1. Generation of e-pharmacophore and pharmacophore-based virtual screening

The pharmacophoric features of 3CLpro-1 showing interactions with main protease were used to develop e-pharmacophore. The features were mapped and pharmacophore hypothesis was developed using RDkit. The pharmacophore-based virtual screening was performed on Pharmit server (http://pharmit.csb.pitt.edu). The obtained pharmacophore hypothesis was used to screen MolPort natural compound database at https://www.molport.com/shop/index containing 1,18,315 molecules. The RMSD cut-off value was set to 0.5 Å between queried features and hit compounds.

2.2. Drug-likeliness and PAINS filtering

The obtained virtual hits were filtered for drug likeliness using a workflow designed on KNIME® analytics platform (ver. 3.7.1) (Supporting Information Figure S1). The molecular properties of the ligands were calculated using descriptor calculator node provided by RDkit. The acceptance criteria of molecular properties are summarised in Table 1. All the ligands having two or less than two violations were selected for further studies.

PAINS are the compounds which are expected to provide false result in high-throughput screening. They interact with numerous biological targets in nonspecific manner, rather than single specific desired target. The PAINS filtering was performed using the PAINS remover web server (www.cbligand.org/PAINS/) (Baell & Holloway, 2010).

2.3. Ligand preparation for HTVS and docking studies

The structural optimization was performed using RD kit. The steps included removal of nonpolar hydrogen, stripping of salt, removal of fragments and defining aromatic criteria. The energy minimization of ligands was performed using Open Babel 3.0.0 (O'Boyle et al., 2011). The General Amber Force Field was used with 4000 steps of steepest descent algorithm. The minimized ligands were converted to Autodock compatible format, i.e. PDBQT using MGL Tools 1.5.7 rc.

2.4. Protein preparation for HTVS and docking studies

The crystal structure of 2019-nCoV main protease was obtained from protein data bank (PDB) (www.rcsb.org) with PDB ID 6LU7. The crystal structure was evaluated for several parameters viz. experimental resolution, *R*-value free and *R*-value work. The protein was checked for missing residues using UCSF Chimera 1.14 rc (Pettersen et al., 2004).

2.4.1. Protein-energy minimization

The correct protonation state of residues was assigned at pH 7.4 using the PDB2PQR server. It employs PROPKA to assign the pK_a values to the ionizable groups of the protein. The energy minimization of the protein was performed on Chimera 1.14 using amber minimization plugin. The protein minimization was performed using 1000 steps of conjugate gradient and 4000 steps of steepest descent using step size of 0.02 Å. The incomplete side chains were replaced using Durnbrack 2010 rotamer library (Bhuyan & Gao, 2011). The backbone refinement of protein was done using amber ff14SB force field. The minimized protein was evaluated by VERIFY3D and Ramachandran plot accessible via SAVES v5.0 server (https://servicesn.mbi.ucla.edu/SAVES/).

The minimized protein was further processed in MGL tools 1.5.7.rc to merge nonpolar hydrogens, add Gastieger charge and assign atom types. The protein was then converted into PDBQT format.

2.5. High throughput virtual screening and molecular docking study

2.5.1. Grid generation and validation

The active site of the protein was identified using CASTp 3.0 web-server (http://sts.bioe.uic.edu/castp/) (Tian et al., 2018). Autodock was used to perform high throughput virtual screening and molecular docking study. Autodock employs autogrid 4.0 to calculate grid maps of interaction energies with various type of atoms viz. A, C, HD, NA, N, OA, S, Br, Cl and I present in the ligand. The grid dimensions used in the study was set to $68 \times 76 \times 64$, *xyz* points for the proteins with grid spacing of 0.375 Å. The grid centre was placed at coordinates (*x*, *y* and *z*) −14.331, 15.48 and 68.064, respectively.

The grid validation was performed by redocking the co-crystallized ligand N3 and calculating the RMSD between experimentally determined co-crystallized ligand structure and docked pose using Discovery studio 2019.

2.5.2. Docking

The docking study was performed in three steps viz. HTVS, standard precision (SP) and extra precision (XP) via Autodock 4.2 using the Lamarckian genetic algorithm. The binding free energy of the ligand–receptor complex was used for scoring various conformations. The scoring function is based on a semi-empirical free energy force field to evaluate conformations obtained by molecular docking (Singh et al., 2020). The ligand 3CLpro-1 was used as an internal standard. The

ligands having higher cluster percentage and less binding energy were filtered and subjected to further docking study. The parameters of different docking studies are summarised in Supporting Information Table S1. The compounds with binding energy less than −9.41 Kcal/mol and cluster size percentage more than 40% were selected to perform XP docking and were subjected to the next steps of *in silico* screening study.

2.6. *In silico* ADMET property prediction

The selected compounds obtained after docking studies were further analyzed for their ADME and toxicity using preADMET (https://preadmet.bmdrc.kr/) and swiss ADME (http://www.swissadme.ch/) web server (Daina et al., 2017). The compounds were predicted for Blood–brain barrier (BBB) permeability, aqueous solubility, human intestinal absorption (HIA), human ether-a-go-go related gene (hERG) inhibition, rodent carcinogenicity, CYP2D6 inhibition and CYP3A4 inhibition.

2.7. Binding free energy calculation

The best five ligands from the docking study were selected to perform molecular mechanics, the generalized Born surface area study (MM-GBSA). In order to calculate binding free energy, we ran a 5 ns molecular dynamics (MD) simulation. MD study was performed using the *pmemd* module of Amber 20 (Ganeshpurkar et al., 2020). The parameter and coordinates files of complex were prepared using *tleap* by employing amber ff14SB force field. The refined structures were solvated using TIP3P water molecules in a cubic periodic box with 12 Å edge length. Sodium ions were added to neutralize the overall charge of the system. A non-bonded cut-off was set to 10 Å and Particle Mesh Ewald (PME) method was applied to treat long range electrostatic interactions. The system was then refined via energy minimization using steepest descent and conjugate gradient algorithms. The NVT ensemble was used to heat the system to 310 K with Langevin thermostat, and collision frequency was set to 5 ps^{-1}. After the system attained the desired temperature, the system was run for 100 ps MD run at NPT ensemble at a pressure of 1 bar using Berendsen barostat. After attaining the equilibrium state 5 ns production MD was performed.

The MMPBSA.py package of Amber tools 20 was used to analyse the 50 snapshots from 5 ns of MD simulation by GB method with salt concentration of 0.1 nM, inner dielectric of 1 and exterior dielectric of 80. The result of MM-GBSA calculation was then compared with 3CLpro-1 protein complex. The ligands showing better free binding energy were selected, and a 150 ns MD simulation study was performed.

2.8. MD simulation and MM-GBSA study

The two ligand–protein complexes having better binding free energy than 3CLpro-1 were selected for detailed MD simulation study for 150 ns. The procedure for MD simulation study is discussed in subsection 2.7. Only the production MD duration was changed to 150 ns.

The trajectories obtained from the MD simulations were analysed for RMSD, root-mean-square fluctuation (RMSF) and H-bond analyses using *cpptraj* tool of amber 20. The MM-GBSA study was performed for the last 120 ns of the simulation run at an interval of 15 ns using the procedure discussed in subsection 2.7.

3. Result and discussion

3.1. Generation of e-pharmacophore and pharmacophore-based virtual screening

3CLpro-1 is a coronavirus main protease inhibitor with proven efficacy (Cheng et al., 2015). The pharmacophoric features viz. hydrogen bond acceptor (HA), Aromatic ring (R), hydrogen bond donor (HD) and Hydrophobic groups (HY) were mapped from the 3CLpro-1 and two e-pharmacophores were developed (Figure 3). The identified features were used to screen MolPort's natural product library using Pharmit server (Sunseri & Koes, 2016). The screening resulted in identification of 341 ligands having RMSD less than 0.5 Å with queried features.

3.2. Drug-likeliness and PAINS filter

The selected compounds were filtered for drug likeliness. The acceptable limits were based on the analysis of physicochemical properties for drug of natural origin (Camp et al., 2015). This resulted in 256 drug-like compounds. The distribution data of physicochemical properties of screened hits has been provided in Supporting Information (Figure S4). The PAINS filtering resulted in removal of 29 PAINS compound. The chemical space of hits is represented in Figure 3.

3.3. Ligand preparation for docking studies

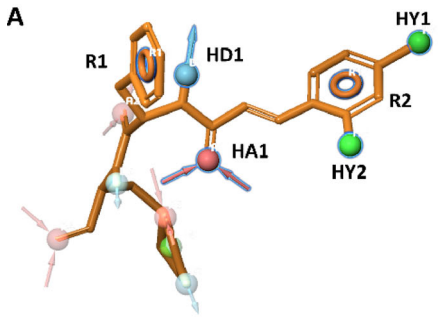
Energy minimization of ligands optimizes the bond length and bond angle before docking study. GAFF employs a simple harmonic form for angles and bonds. It covers most of the organic chemical space.

3.4. Selection of protein structure

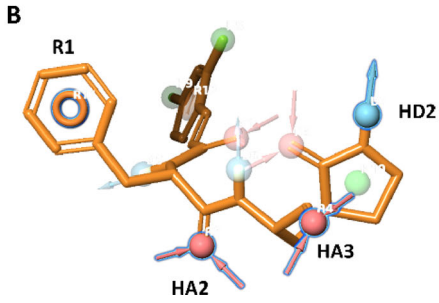
The X-ray derived crystal structure of nCoV-2019 main protease in complex with an inhibitor N3 was obtained from PDB (PDB ID-6LU7). It had experimental resolution 2.16 Å, *R*-value of 0.235. It consists of one chain of 306 residues. The structure examination on chimera revealed no missing residues and loops.

3.5. Protein structure energy minimization

Energy minimization and optimization is an important step in docking and simulation studies. Verify 3D score of minimized protein was found to be 93.76%, ERRAT quality factor



R1	0					
HD1	3.78	0				
HA1	5.60	3.14	0			
HY1	10.77	8.46	8.69	0		
R2	8.18	5.70	6.51	3.09	0	
HY2	8.03	6.29	5.58	4.93	2.66	0
	R1	HD1	HA1	HY1	R2	HY2



R1	0			
HA2	5.91	0		
HA3	8.78	5.08	0	
HD2	9.14	5.88	4.48	0
	R1	HA2	HA3	HD2

Figure 3. Pharmacophore mapping and distance matrix of pharmacophore models. A. Model 1. B. Model 2.

Table 2. Protein structure quality evaluation.

Evaluation parameter	Score
Ramachandran plot:	100 %
a) Residues in allowed region	0 %
b) Residues in disallowed region	
Verify 3D	93.76 %
ERRAT	95.48
Molprobity score	1.46
Bad bonds	0
Bad angles	0

was found to be 95.48. The minimized protein was also evaluated on the MolProbity web server, and the results are summarised in Table 2. The Ramachandran plot analysis showed no residues in the disallowed region (Figure 4).

3.6. HTVS and docking studies

CASTp web-server was used to determine the active site of protein. The grid for HTVS and docking studies were prepared using residues viz. Ser144, Gly143, Cys145, His163, Glu166, Leu167, Pro168, Ala191, Thr190, Met49 and His41 as a reference to ascertain the dimensions of grid box (Supporting Information Figure S2). The grid validation resulted in RMSD of 1.021 between docked ligand and co-crystalized ligand, which indicates that the selected grid displayed less deviation and can be used for docking studies.

HTVS is a method of choice to identify potential drug candidates from a large database. Autodock 4.2 was used to perform HTVS and docking studies. HTVS was performed using the Lamarckian-Genetic algorithm (LGA) followed by clustering of the obtained pose with RMSD tolerance of 2.0 Å. The scoring function employed by the Autodock is:

$$V = W_{vdw} + \sum_{i,j}\left\{\frac{A_{ij}}{r_{ij}^{12}} + \frac{B_{ij}}{r_{ij}^{6}}\right\} + W_{hbond}\sum_{i,j} E(t)\left\{\frac{A_{ij}}{r_{ij}^{12}} + \frac{B_{ij}}{r_{ij}^{6}}\right\} + W_{elec}\left(\frac{q_{i\cdot}q_{j}}{e(r_{ij})r_{ij}}\right) + W_{sol}\sum_{ij}(SiVj - SjVi)\, e^{(-rij^2/2\sigma^2)}$$

On the basis of scoring function, the ligands with better binding energy and cluster size were selected for standard and extended precision docking.

A total of 37 ligands were obtained after HTVS, which were narrowed down to 10 ligands using standard precision docking. The binding energy and ligand efficiency of all the 10 ligands after XP docking is summarized in Table 3.

MolPort-000-410-348 showed H-bond interactions with His163 and Glu166. The indole part of the molecule showed π-alkyl interaction with Met165. The pyrimidine part of the molecule showed π-alkyl interaction with Leu141.

Molport-002-530-156 showed H-bond interactions with Glu166, Phe140, His163 and Thr190. The phenyl ring of indoline moiety showed π-cation interaction with His41 (Figure 5).

3.7. ADME and toxicity prediction

The screened compounds were evaluated for their ADME and toxicity property. BBB permeability is an important parameter to predict the amount of drug entering in the brain. The drugs which tend to enter into the brain might show serious CNS side-effects. Only one compound, i.e. MolPort-002-701-723 showed high BBB permeability, rest all the compounds showed low to medium BBB permeation. Aqueous

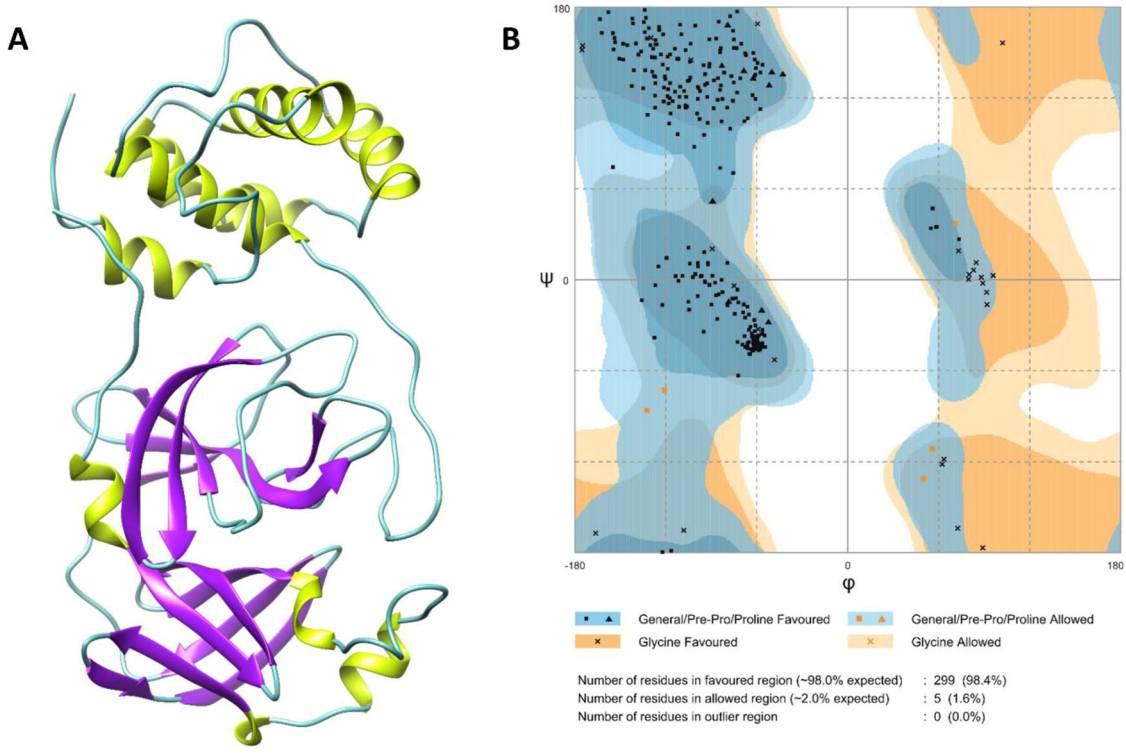


Figure 4. Structure of protein PDB ID 6LU7. (A) Three-dimensional structure of minimized protein-helices are represented in yellow, and sheets are represented in violet. (B) Ramachandran plot of minimized protein.

Table 3. Summary of extended precision (XP) docking of ligands on nCoV-2019 main protease (PDB ID. 6LU7).

MolPort database ID	Interaction with nCoV-2019 main protease (PDB ID-6LU7)	Binding energy	Ligand efficiency
MolPort-039-338-091	His41 (H-bond), Ser144 (H-bond), Glu166(H-bond), Cys145 (π-sulphur), His41 (π-π T-shaped)	−14.56	−0.231
MolPort-002-530-156	Glu166 (H-bond), Thr190 (H-bond), Glu166 (H-bond), Phe140 (H-bond), His41 (π-cation), Leu27 (π-alkyl)	−13.38	−0.212
MolPort-027-852-450	Ser144 (H-bond), Asn142 (H-bond), His41 (π-alkyl), Cys145 (π-alkyl)	−12.47	−0.29
MolPort-000-410-348	His163 (H-bond), Glu166 (H-bond), His172 (π-alkyl) Leu141 (π-sigma)	−12.09	−0.224
MolPort-002-701-723	Ser144 (H-bond), Glu166 (H-bond), Gln189 (H-bond), Cys145 (π-alkyl)	−12.04	−0.28
MolPort-000-717-985	Glu166 (H-bond), Gln192 (H-bond), Met165 (π-sulphur), Cys44 (H-bond)	−11.2	−0.311
MolPort-001-749-200	Pro168 (π-alkyl), Cys145 (H-bond), Gln189 (H-bond), Cys145 (H-bond)	−10.7	−0.267
MolPort-002-512-655	Glu166 (H-bond), Cys145 (π-alkyl), His14 (π-sigma)	−10.6	−0.286
MolPort-002-553-694	His163 (H-bond), Met49 (π-alkyl), Met165 (π-alkyl)	−10.55	−0.301
MolPort-002-533-520	Glu166 (H-bond), Met49 (π-alkyl), Met165 (π-alkyl)	−10.14	−0.290

solubility is a major concern in drug discovery as low solubility is related to poor potency. The compound MolPort-039-338-091 showed poor aqueous solubility of 9.58 mg/mL. HIA is one of the most important ADME properties. It is also a key step for the drug transporting to their targets (Yan et al., 2008). The compound MolPort-039-338-091 showed no HIA. The blockade of the cardiac ion channel encoded by hERG can lead to cardiac dysfunctions. It is a major concern in drug discovery and development. Numerous drug discovery programs have been hampered by this issue (Danker & Möller, 2014). In the present study, MolPort-039-338-091 showed a high risk of hERG inhibition. The filtered compounds were also tested for rodent carcinogenicity. No ligands showed rodent toxicity. The compounds were also predicted for CYP2D6 and CYP3A4 inhibition. These are the families of cytochrome P450 enzymes responsible for the metabolism of drugs. All the ligands were noninhibitor of CYP2D6. MolPort-002-701-723 and MolPort-039-338-091 showed CYP3A4 inhibition (Table 4).

3.8. Binding free energy calculation

MM-GBSA is a popular method for binding free energy prediction and is considered to be more accurate than most of the scoring functions of molecular docking (Wang et al., 2019). In MM-GBSA the free energy for binding of a ligand (L) to the receptor (R) to form a complex (RL),

$$\Delta G_{\text{binding}} = \Delta G_{\text{RL}} - [\Delta G_P + \Delta G_L]$$

$$\Delta G_{\text{binding, aq.}} = \Delta G_{\text{binding, vac}} + \Delta G_{\text{binding, solv}}$$

$$\Delta G = \Delta E_{\text{mm}} + \Delta G_{\text{solvation}} - T\Delta S$$

$$\Delta E_{\text{mm}} = \Delta E_{\text{bonding}} + \Delta E_{\text{vdW}} + \Delta E_{\text{ele}}$$

$$\Delta G_{\text{solvation}} = \Delta G_{\text{GB}} + \Delta G_{\text{SA}}$$

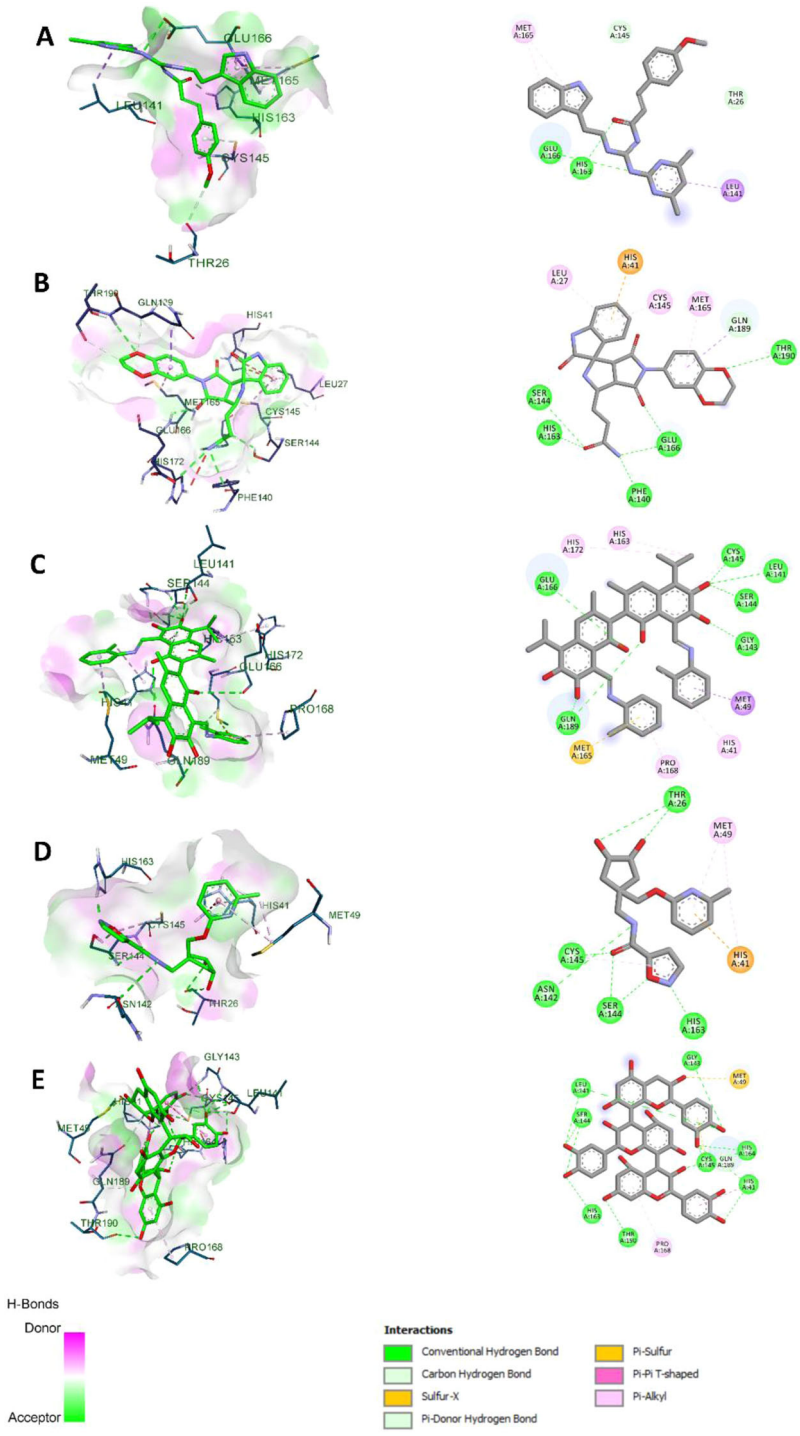


Figure 5. 2D and 3D interaction diagrams of best hits obtained via HTVS. (A) MolPort-000-410-348. (B) MolPort-002-701-723. (C) MolPort-002-530-156. (D) MolPort-027-852-450. (E) MolPort-039-338-091.

Where, $\Delta E_{bonding}$, ΔE_{vdW} and ΔE_{ele} are energy from bonded, van der Waals and electrostatic interactions, respectively, of molecular mechanics. ΔG_{GB} and ΔG_{SA} are the contributions to the solvation free energies. ΔG_{GB} was calculated from the Generalised Born equation and ΔG_{SA} was calculated from the solvent-accessible surface area (SASA). The entropy correction factor was neglected due to the heavy computational cost.

From this study, it was observed that two ligands, i.e. MolPort-002-530-156 and MolPort-000-410-348, showed better binding free energy than 3CLPro-1. Both these ligands show better electrostatic and van der Waals interaction energy. Thus, it can be said that the stable complex is formed. The result has been summarised in Figure 6.

3.9. MD simulation and MM-GBSA assay

MD simulation is a computational method for studying the physical movement of atoms and molecules in a biophysical system. In this study, MD simulation was done to establish

Table 4. ADME and toxicity prediction of selected compounds.

MolPort ID	BBB (C_{brain}/C_{plasma})	Aqueous solubility (mg/mL)	Human intestinal absorption (%)	hERG inhibition	Rodent carcinogenicity	CYP2D6 inhibition	CYP 3A4 inhibition
MolPort-002-530-156	0.03	76.38	89.52	No risk	Noncarcinogen	Noninhibitor	Weak inhibitor
MolPort-002-701-723	7.95	8547.79	93.41	No risk	Noncarcinogen	Noninhibitor	Inhibitor
MolPort-000-410-348	1.33	89.69	92.07	Medium risk	Noncarcinogen	Noninhibitor	Weak inhibitor
MolPort-027-852-450	0.1	83.80	79.26	Medium risk	Noncarcinogen	Noninhibitor	Weak inhibitor
MolPort-039-338-091	0.03	9.58	0	High risk	Noncarcinogen	Noninhibitor	Inhibitor

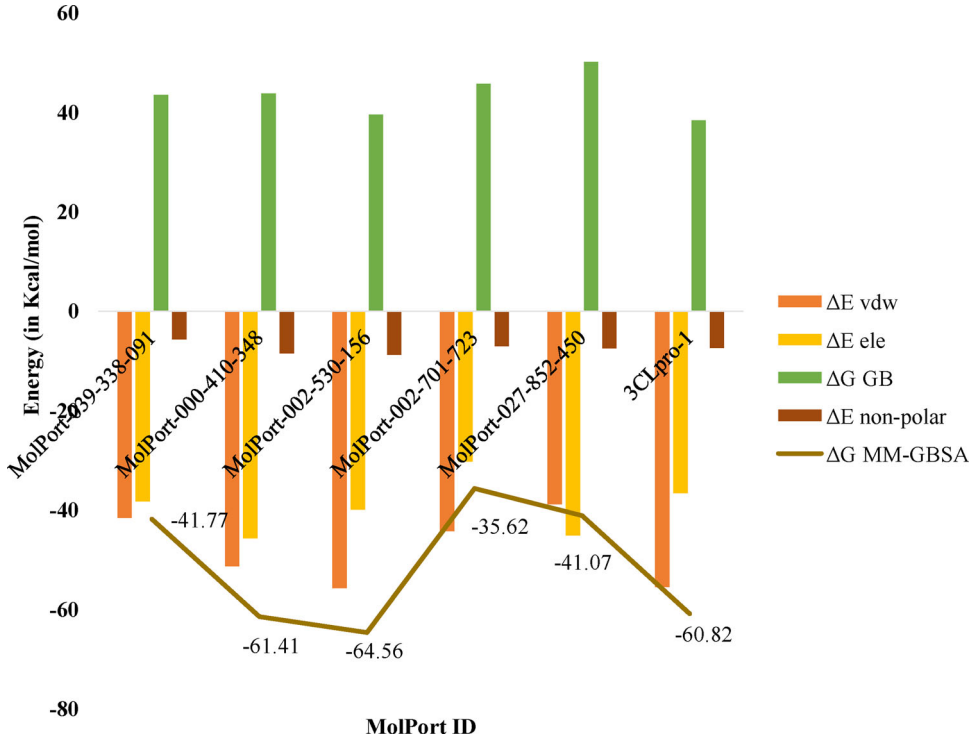


Figure 6. Energy contribution of P-L complex in binding free energy calculation.

the stability of protein–ligand complex of MolPort-002-530-156 and MolPort-000-410-348 complexed with protein. The trajectories obtained from the following MD run were analysed for protein and ligand RMSD, protein root-mean square fluctuation (RMSF), H-bond interactions and MM-GBSA study.

3.9.1. RMSD analyses

The actual movement and structural changes of a protein in biological environment can be visualised using MD simulation. The RMSD indicates the equilibration and stability of a complex. It is evident from the Figures 7 and 8 that protein attains a stable RMSD during the simulation run. In case of MolPort-000-410-348 and MolPort-002-530-156 protein complex, the maximum protein backbone RMSD was found to be 1.45 Å and 1.84 Å, respectively, which is considered perfectly fine for small and globular protein. The average protein backbone RMSD was found to be 1.31 Å and 1.16 Å for of MolPort-000-410-348 and MolPort-002-530-156 protein complex, respectively. In the same plot, it was observed that the ligands RMSD were not significantly larger than the protein RMSD which indicates that ligand does not diffuses away from the initial binding site. The average ligand RMSD was found to be 1.38 Å and 1.46 Å for of MolPort-000-410-348 and MolPort-002-530-156 protein complex, respectively.

3.9.2. RMSF analyses

The RMSF is useful for characterising local changes along the protein chain. The peaks in the plot represent the residues which fluctuates the most during simulation. The RMSF analysis was calculated for C-α atom of protein residues. The average protein RMSF was found to be 0.51 Å and 0.54 Å for MolPort-000-410-348 and MolPort-002-530-156, respectively. The residues participating in interactions with ligand, i.e. Hid41, Gly143, Ser144, Glu166, Arg188, Gln189 and Thr190 remained highly stable throughout the simulation.

3.9.3. H-Bond interaction analyses

Protein interactions with the ligand were monitored throughout the simulation. The H-bond interaction analyses are

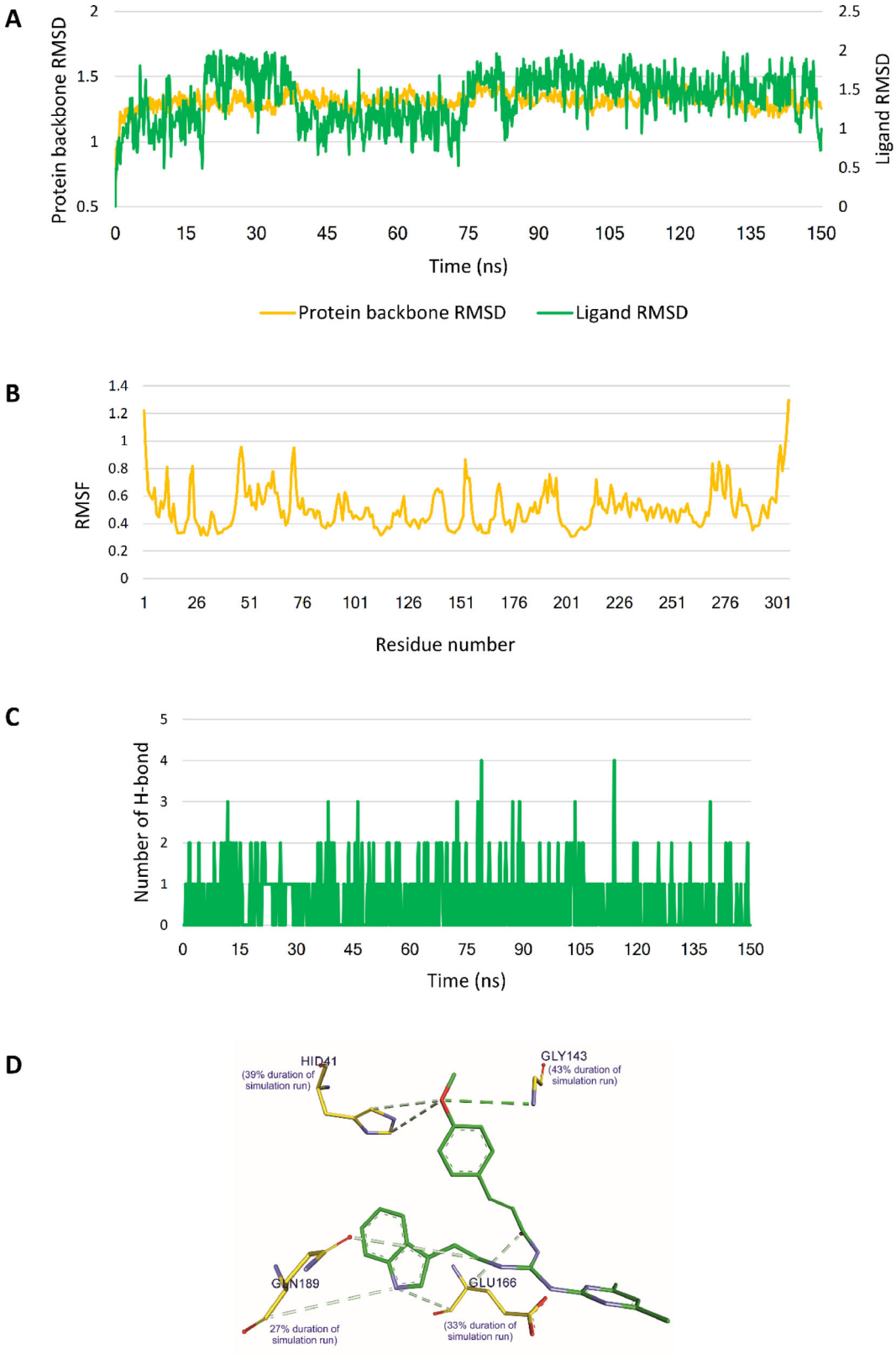


Figure 7. MD simulation of MolPort-000-410-348 complexed with protein. (A) RMSD plot of protein and MolPort-000-410-348. (B) RMSF plot of protein. (C) Number of H-bond interaction between MolPort-000-410-348 and protein during MD run. (D) Interactions of MolPort-000-410-348 with protein during MD run.

important to understand the stability of predicted protein–ligand complex. H-bonding plays a significant role in accommodating the ligand inside the binding site. MolPort-000-410-348 showed H-bond interaction with one of the catalytic dyad Hid41, Gly143, Gln189 and Glu166 (Sk et al., 2020). MolPort-002-530-156 showed H-bond interaction with, Arg188, Ser144, Asn142 and Thr190. The similar interactions were also observed in docking studies as well.

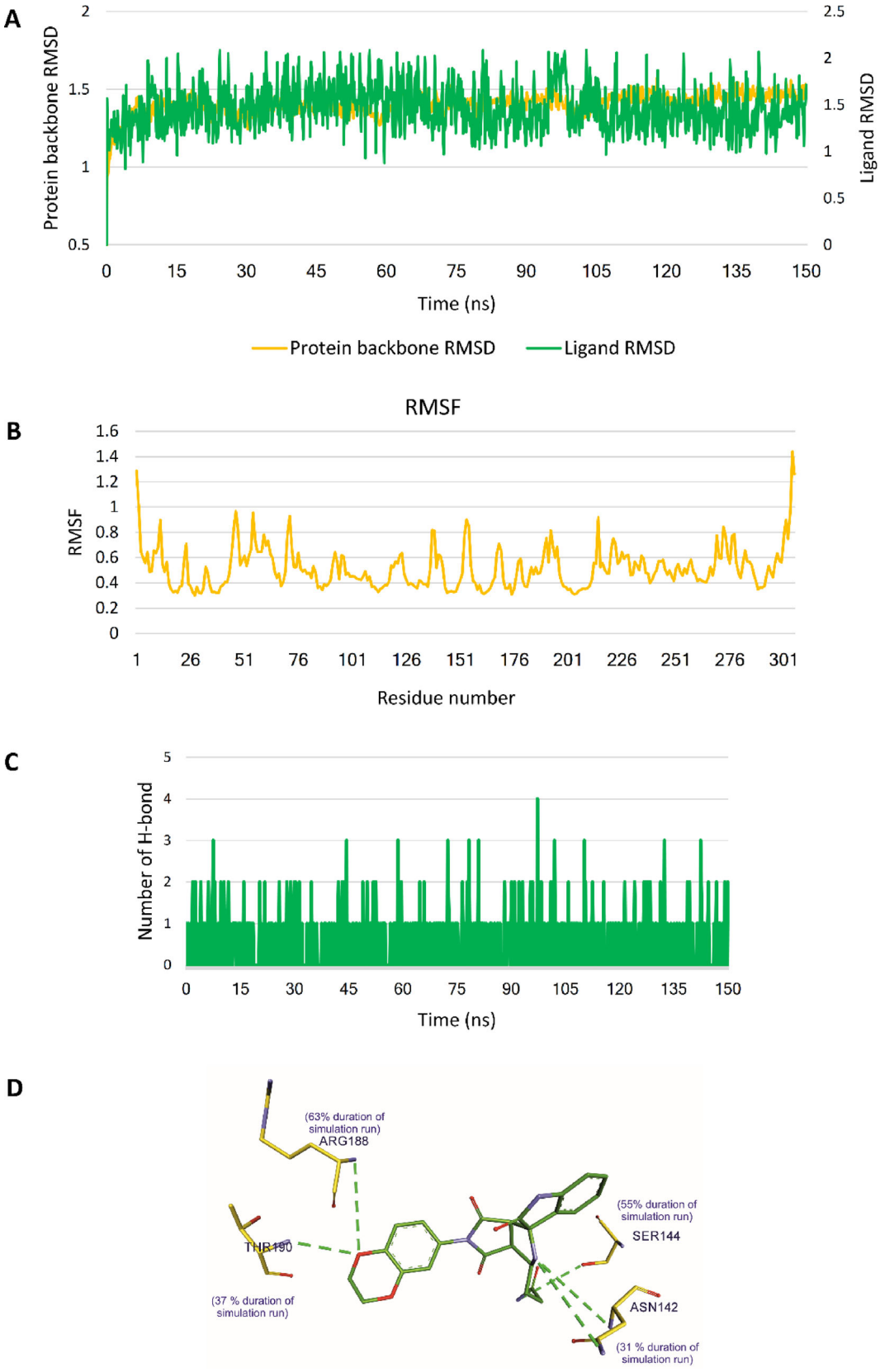


Figure 8. MD simulation of MolPort-002-530-156 complexed with protein. (A) RMSD plot of protein and MolPort-002-530-156. (B) RMSF plot of protein. (C) Number of H-bond interaction between MolPort-002-530–156 and protein during MD run. (D) Interactions of MolPort-002-530-156 with protein during MD run.

3.9.4. MM-GBSA assay

The meaning of different energy terms has been explained in subsection 3.8. The MM-GBSA study is significant to validate the findings of docking and MD study. The detailed contribution of binding free energies in MM-GBSA is listed in Table 5. The binding-free energy of ligands was found to be

Table 5. Energy contributions of protein–ligand complexes in MM-GBSA study.

Ligand	ΔE_{vdw}	ΔE_{ele}	ΔE_{GB}	$\Delta E_{nonpolar}$	$\Delta G_{MM\text{-}GBSA}$
MolPort-000-410-348	−74.11 ± 2.46	−9.22 ± 2.25	26.37 ± 2.69	−6.38 ± 0.24	−63.34 ± 2.03
MolPort-002-530-156	−66.91 ± 2.71	−6.59 ± 1.21	17.93 ± 1.78	−5.95 ± 0.26	−61.52 ± 2.24

Data represented in mean ± standard deviation.

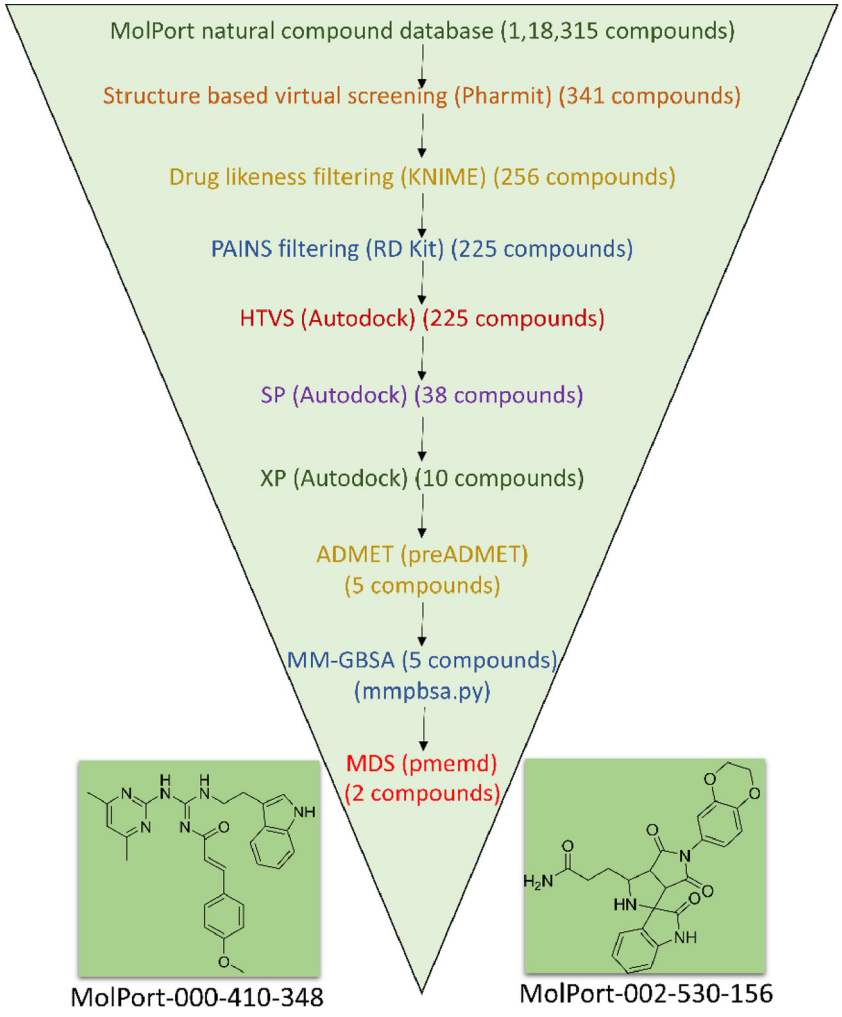


Figure 9. Summary of virtual screening workflow.

−63.34 ± 2.03 and −61.52 ± 2.24 kcal mol^{-1} for MolPort-000-410-348 and MolPort-002-530-156, respectively. These low energies are an indication of high stability of the complexes.

4. Conclusion

In the current scenario, there is an emergent requirement of discovering small molecules having potential to cure the 2019-nCoV disease. In such a situation of crisis, this article focused on computational biology and molecular modeling approaches for aiding the drug discovery process against 2019-nCoV. The pharmacophore model for 2019-nCoV was developed from 3CLpro-1 and utilized for a pharmacophore-based virtual screening of Molport natural database of 1,18,315 compounds. Three hundred forty-one compounds were identified from the virtual screening that was subsequently subjected to drug-likeliness and PAINS filters to remove interfering compounds. The crystal structure of 2019-nCoV main protease was obtained from PDB (PDB ID. 6LU7), and the protein had some bad contacts and incorrect side chain(s) positions that were removed by energy minimization before docking studies to obtain a relatively favorable protein state. Through HTVS, SP and XP docking, 225, 38 and finally, 10 ligands, respectively, were obtained. The *in silico* ADME and toxicity parameters were estimated for all the compounds to determine the molecules having clinical potential. Binding free energy calculations were performed on five molecules, and two of them displayed better binding free energy than 3CLpro-1. The MD simulation study of the two selected molecules, i.e. MolPort-000-410-348 [(E)-N-((E)-((2-(1H-indol-3-yl)ethyl)amino)((4,6-dimethylpyrimidin-2-yl)amino)methylene)-3-(4-methoxyphenyl) acrylamide)] and MolPort-002-530-156 [3-(5′-(2,3-dihydrobenzo[b][1,4]dioxin-6-yl)-2,4′,6′-trioxo-3′,3a′,4′,5′,6′,6a′-hexahydro-2′H-spiro[indoline-3,1′-pyrrolo[3,4-c]pyrrol]-3′-yl)propanamide] displayed protein–ligand complex stability in simulated fluid conditions.

Thus, the compounds, i.e. MolPort-000-410-348 and MolPort-002-530-156, are anticipated as the virtual leads of the study and are expected to display predicted behavior in experimental studies (Figure 9).

Acknowledgements

The authors would like to thank MolPort for providing access to their natural compound database. NT, BG, NB and BS are grateful to IIT (BHU), Varanasi for providing teaching assistantship. The authors would also like to extend their gratitude toward Professor David A. Case, Department of Chemistry & Chemical Biology, Rutgers University, New Jersey, USA, for granting a license for Amber 20.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

The authors acknowledge financial support of Seed Research Grant from the IIT (BHU).

ORCID

Nancy Tripathi http://orcid.org/0000-0003-4193-6748
Bharat Goel http://orcid.org/0000-0002-6297-0067
Nivedita Bhardwaj http://orcid.org/0000-0002-7857-9441
Bharat Sahu http://orcid.org/0000-0002-1077-8153
Hemant Kumar http://orcid.org/0000-0001-7698-2301
Shreyans K. Jain http://orcid.org/0000-0001-6160-8755

References

Baell, J. B., & Holloway, G. A. (2010). New substructure filters for removal of pan assay interference compounds (PAINS) from screening libraries and for their exclusion in bioassays. *Journal of Medicinal Chemistry*, *53*(7), 2719–2740. https://doi.org/10.1021/jm901137j

Bhuyan, M. S. I., & Gao, X. (2011). A protein-dependent side-chain rotamer library. *BMC Bioinformatics*, *12*(S14), S10–S10. https://doi.org/10.1186/1471-2105-12-S14-S10

Camp, D., Garavelas, A., & Campitelli, M. (2015). Analysis of physicochemical properties for drugs of natural origin. *Journal of Natural Products*, *78*(6), 1370–1382. https://doi.org/10.1021/acs.jnatprod.5b00255

Cheng, K.-W., Cheng, S.-C., Chen, W.-Y., Lin, M.-H., Chuang, S.-J., Cheng, I. H., Sun, C.-Y., & Chou, C.-Y. (2015). Thiopurine analogs and mycophenolic acid synergistically inhibit the papain-like protease of Middle East respiratory syndrome coronavirus. *Antiviral Research*, *115*(1159), 9–16.10.1016/j.antiviral.2014.12.011 https://doi.org/10.1016/j.antiviral.2014.12.011

Cheng, V. C. C., Lau, S. K. P., Woo, P. C. Y., & Yuen, K. Y. (2007). Severe acute respiratory syndrome coronavirus as an agent of emerging and reemerging infection. *Clinical Microbiology Reviews*, *20*(4), 660–694.https://doi.org/10.1128/CMR.00023-07

Daina, A., Michielin, O., & Zoete, V. (2017). SwissADME: A free web tool to evaluate pharmacokinetics, drug-likeness and medicinal chemistry friendliness of small molecules. *Scientific Reports*, *7*(1), 42717. https://doi.org/10.1038/srep42717

Danker, T., & Möller, C. (2014). Early identification of hERG liability in drug discovery programs by automated patch clamp. *Frontiers in Pharmacology*, *5*, 203–203. https://doi.org/10.3389/fphar.2014.00203

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. *Journal of Biomolecular Structure and Dynamics*. https://doi.org/10.1080/07391102.2020.1763201

Ganeshpurkar, A., Singh, R., Gore, P. G., Kumar, D., Gutti, G., Kumar, A., & Singh, S. K. (2020). Structure-based screening and molecular dynamics simulation studies for the identification of potential acetylcholinesterase inhibitors. *Molecular Simulation*, *46*(3), 169–185. https://doi.org/10.1080/08927022.2019.1682572

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., … Cao, B. (2020). Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *The Lancet*, *395*(10223), 497–506.10.1016/S0140-6736(20)30183-5

Jin, Z., Du, X., Xu, Y., Deng, Y., Liu, M., Zhao, Y., Zhang, B., Li, X., Zhang, L., Peng, C., Duan, Y., Yu, J., Wang, L., Yang, K., Liu, F., Jiang, R., Yang, X., You, T., Liu, X., … Yang, H. (2020). Structure of Mpro from COVID-19 virus and discovery of its inhibitors. *Nature*, *582*(7811), 289–293. https://doi.org/10.1038/s41586-020-2223-y

Joshi, R. S., Jagdale, S. S., Bansode, S. B., Shankar, S. S., Tellis, M. B., Pandya, V. K., Chugh, A., Giri, A. P., & Kulkarni, M. J. (2020). Discovery of potential multi-target-directed ligands by targeting host-specific SARS-CoV-2 structurally conserved main protease. *Journal of Biomolecular Structure and Dynamics*. https://doi.org/10.1080/07391102.2020.1760137

Kumar, V., Shin, J. S., Shie, J.-J., Ku, K. B., Kim, C., Go, Y. Y., Huang, K.-F., Kim, M., & Liang, P.-H. (2017). Identification and evaluation of potent Middle East respiratory syndrome coronavirus (MERS-CoV) 3CLPro inhibitors. *Antiviral Research*, *141*, 101–106.10.1016/j.antiviral.2017.02.007 https://doi.org/10.1016/j.antiviral.2017.02.007

Morse, J. S., Lalonde, T., Xu, S., & Liu, W. R. (2020). Learning from the past: Possible urgent prevention and treatment options for severe acute respiratory infections caused by 2019-nCoV. *Chembiochem: A European Journal of Chemical Biology*, *21*(5), 730–738. https://doi.org/10.1002/cbic.202000047

Muralidharan, N., Sakthivel, R., Velmurugan, D., & Gromiha, M. M. (2020). Computational studies of drug repurposing and synergism of lopinavir, oseltamivir and ritonavir binding with SARS-CoV-2 protease against COVID-19. *Journal of Biomolecular Structure and Dynamics*. https://doi.org/10.1080/07391102.2020.1752802

O'Boyle, N. M., Banck, M., James, C. A., Morley, C., Vandermeersch, T., & Hutchison, G. R. (2011). Open Babel: An open chemical toolbox. *Journal of Cheminformatics*, *3*(1), 33. https://doi.org/10.1186/1758-2946-3-33

Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., & Ferrin, T. E. (2004). UCSF Chimera-a visualization system for exploratory research and analysis. *Journal of Computational Chemistry*, *25*(13), 1605–1612. https://doi.org/10.1002/jcc.20084

Singh, R., Ganeshpurkar, A., Kumar, D., Kumar, D., Kumar, A., & Singh, S. K. (2020). Identifying potential GluN2B subunit containing N-Methyl-D-aspartate receptor inhibitors: An integrative in silico and molecular modeling approach. *Journal of Biomolecular Structure & Dynamics*, *38*(9), 2533–2545. https://doi.org/10.1080/07391102.2019.1635530

Sk, M. F., Roy, R., Jonniya, N. A., Poddar, S., & Kar, P. (2020). Elucidating biophysical basis of binding of inhibitors to SARS-CoV-2 main protease by using molecular dynamics simulations and free energy calculations. *Journal of Biomolecular Structure and Dynamics*. https://doi.org/10.1080/07391102.2020.1768149

Sunseri, J., & Koes, D. R. (2016). Pharmit: Interactive exploration of chemical space. *Nucleic Acids Research*, *44*(W1), W442–W448. https://doi.org/10.1093/nar/gkw287

Tian, W., Chen, C., Lei, X., Zhao, J., & Liang, J. (2018). CASTp 3.0: Computed atlas of surface topography of proteins. *Nucleic Acids Research*, *46*(W1), W363–W367. https://doi.org/10.1093/nar/gky473

Wang, E., Sun, H., Wang, J., Wang, Z., Liu, H., Zhang, J. Z. H., & Hou, T. (2019). End-point binding free energy calculation with MM/PBSA and MM/GBSA: Strategies and applications in drug design. *Chemical Reviews*, *119*(16), 9478–9508. https://doi.org/10.1021/acs.chemrev.9b00055

World Health Organization. (2020). *COVID-19 situation reports*. Retrieved April 28, 2020, from https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports/

Yan, A., Wang, Z., & Cai, Z. (2008). Prediction of human intestinal absorption by GA feature selection and support vector machine regression. *International Journal of Molecular Sciences*, *9*(10), 1961–1976. https://doi.org/10.3390/ijms9101961

Zhang, H., Penninger, J. M., Li, Y., Zhong, N., & Slutsky, A. S. (2020). Angiotensin-converting enzyme 2 (ACE2) as a SARS-CoV-2 receptor: Molecular mechanisms and potential therapeutic target. *Intensive Care Medicine*, *46*(4), 586–590. https://doi.org/10.1007/s00134-020-05985-9

Zhao, Z., Zhang, F., Xu, M., Huang, K., Zhong, W., Cai, W., Yin, Z., Huang, S., Deng, Z., Wei, M., Xiong, J., & Hawkey, P. M. (2003). Description and clinical treatment of an early outbreak of severe acute respiratory syndrome (SARS) in Guangzhou, PR China. *Journal of Medical Microbiology*, *52*(Pt 8), 715–720.10.1099/jmm.0.05320-0 https://doi.org/10.1099/jmm.0.05320-0

Zumla, A., Chan, J. F. W., Azhar, E. I., Hui, D. S. C., & Yuen, K.-Y. (2016). Coronaviruses - Drug discovery and therapeutic options. *Nature Reviews. Drug Discovery*, *15*(5), 327–347. https://doi.org/10.1038/nrd.2015.37