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

Informatics in Medicine Unlocked 23 (2021) 100516

Contents lists available at ScienceDirect

Informatics in Medicine Unlocked

journal homepage: http://www.elsevier.com/locate/imu

In silico exploration of novel protease inhibitors against coronavirus 2019 (COVID-19)

Elham Aghaee [a], Marzieh Ghodrati [b], Jahan B. Ghasemi [a,*]

[a] Drug Design in Silico Lab, Chemistry Faculty, School of Sciences, University of Tehran, Tehran, Iran
[b] Department of Neurology, Faculty of Medicine, Qom University of Medical Sciences, Qom, Iran

ARTICLE INFO

Keywords:
Virtual screening
COVID-19
Main protease
Molecular docking
Molecular dynamics simulation
ADME studies

ABSTRACT

The spread of SARS-CoV-2 has affected human health globally. Hence, it is necessary to rapidly find the drug-candidates that can be used to treat the infection. Since the main protease (M^{pro}) is the key protein in the virus's life cycle, M^{pro} is served as one of the critical targets of antiviral treatment. We employed virtual screening tools to search for new inhibitors to accelerate the drug discovery process. The hit compounds were subsequently docked into the active site of SARS-CoV-2 main protease and ranked by their binding energy. Furthermore, in-silico ADME studies were performed to probe for adoption with the standard ranges. Finally, molecular dynamics simulations were applied to study the protein-drug complex's fluctuation over time in an aqueous medium. This study indicates that the interaction energy of the top ten retrieved compounds with COVID-19 main protease is much higher than the interaction energy of some currently in use protease drugs such as ML188, nelfinavir, lopinavir, ritonavir, and α-ketoamide. Among the discovered compounds, Pubchem44326934 showed druglike properties and was further analyzed by MD and MM/PBSA approaches. Besides, the constant binding free energy over MD trajectories suggests a probable drug possessing antiviral properties. MD simulations demonstrate that GLU166 and GLN189 are the most important residues of $M^{pro,}$ which interact with inhibitors.

1. Introduction

The new coronavirus disease (COVID-19) is caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). It is highly contagious and has forced many local and government officials to step in to slow down the rate of infection since the virus was first detected in the December of 2019 in Wuhan, China [1]. The pandemic caused by this virus gave rise to devastating health threats and has created a new front-line for discovering effective drugs and new vaccines.

Coronaviruses (CoVs) are a single-stranded positive-sense RNA genome coated with a membrane envelope. This membrane is covered with spike glycoprotein, giving a crown shape to the virus. The length of RNA genomes is 26–32 kilobases containing 6 to 12 open reading frames (ORFs) [2,3]. The first two-thirds of the whole genome of coronavirus encodes two polyproteins that are divided into 15 or 16 nonstructural proteins denoted as nsp1 to nsp16. The rest of the ORFs holds the genetic codes of four essential structural proteins: envelop (E), membrane (M), nucleocapsid (N), and spike (S) proteins. These proteins play significant roles, from the virus's survival to viral development and viral attack capabilities [4–7].

For the virus to enter the cell and cause infection, the S protein has to interact with the host cellular receptors - specifically, the angiotensin-converting enzyme 2 (ACE2) receptors – and have the host translational machinery hijacked [8–11]. When done, the host cell proteins get exposed to the virus. The ORF lab translates a polyprotein in which this protein breaks into sixteen nsp proteins, namely nsp1 – nsp16 [12,13]. Among 16 nsp proteins, nsp5 or Main protease (M^{pro}) is a key coronavirus enzyme for viral replication and transcription and makes it one of the most popular drug targets [14–17]. M^{pro} inhibitors bind to M^{pro} and prevent this viral replication and block the proteolytic cleavage of protein precursors essential for the infection's inception.

Research-based on homology modeling, molecular docking, and binding free energy calculations explored by Xu et al. identified that nelfinavir has a potential inhibitory against COVID-19 main protease [18]. Lopinavir and ritonavir are other protease inhibitors currently being used for treating HIV patients. Since the sequence of SARS-CoV-2 main protease is similar to other CoV proteases such as HIV, China's national health commission has proposed using these drugs for COVID-19, despite there is no official approval of these drugs to treat COVID-19. Kumar and colleagues screened the FDA-approved antiviral

* Corresponding author.
E-mail address: jahan.ghasemi@ut.ac.ir (J.B. Ghasemi).

https://doi.org/10.1016/j.imu.2021.100516
Received 15 November 2020; Received in revised form 6 January 2021; Accepted 8 January 2021
Available online 12 January 2021

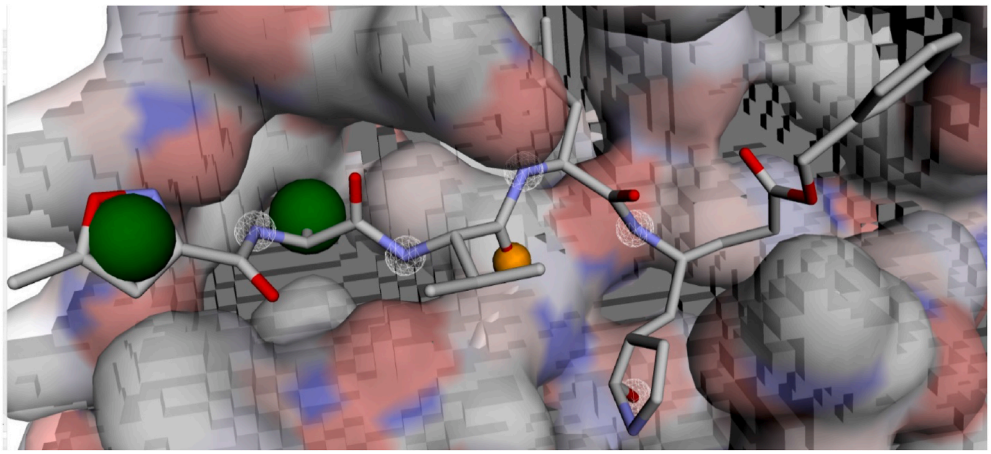


Fig. 1. Pharmacophore model based on N3 inhibitor in the pocket of COVID-19 M^{pro} generated with Pharmit server. Hydrogen-bond donor features are color-coded with white, hydrogen-bond acceptors are colored with yellow, and hydrophobic features are colored in green. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

drugs through molecular docking and performed molecular dynamics simulation of the top three selected drugs (lopinavir-ritonavir, tipranavir, and raltegravir) with the main protease of SARS-CoV [19]. They declared that these drugs show the best interaction with M^{pro} and result in a stable complex. Alpha-ketoamides are other inhibitors of the coronavirus main protease. Zhang et al. designed and synthesized α-ketoamides as broad-spectrum inhibitors of the main proteases of coronaviruses. They modified their former best inhibitor to increase the half-life of inhibitor in plasma, increase the solubility, and decrease the inhibitor's binding to plasma proteins. Then they reported the X-ray structure of SARS-CoV-2 M^{pro} in complex with α-ketoamide [20]. By using computer-aided drug design, Jin and colleagues identified the N3 inhibitor and then determined the crystal structure of COVID-19 M^{pro} in complex with this inhibitor. Subsequently, they performed structure-based virtual screening of over 10,000 compounds (approved drugs, drug candidates in clinical trials, and other pharmacologically active compounds as inhibitors of M^{pro}). They also evaluated the in vitro inhibitory of the top six selected compounds in cell-based assays. They subsequently claimed that these compounds could inhibit the COVID-19 M^{pro} with IC_{50} values ranging from 0.67 to 21.4 μM [14].

This work aims to introduce novel Mpro inhibitors by using pharmacophore modeling, virtual screening, molecular docking, and molecular dynamics simulation. A pharmacophore model is a combination of electronic and steric features essential for interacting with a particular receptor to activate or block its biological activity. Virtual screening is a method of discovering new active molecules and is widely used in drug discovery. In this method, large databases of small molecules are screened to find structures with a high affinity toward the target receptor. Molecular docking is a computational method of identifying the essential interactions between a drug and its receptor. Molecular dynamics (MD) is a computer simulation method for investigating atoms and molecules' physical movements. It can be used for studying the structure of a biomolecular system. For a while, the atoms and molecules are free to move and interact, so a dynamic system is observed [21–23].

For performing a quick search of large compound databases, the Pharmit web service (http://pharmit.csb.pitt.edu) was utilized to screen the databases using pharmacophores, molecular shape, and energy minimization. The investigated compounds were transferred into Discovery Studio 4.1 and docked into the pocket of M^{pro} using CDOCKER (CHRMm-based DOCKER) algorithm and ranked by their CDOCKER energy. However, in the docking methods, a rigid protein is considered, and water molecules were deleted. These deficiencies reduce the accuracy of the prediction of docking methods. Molecular dynamics (MD) simulations resolve protein flexibility and solvent problems [24,25]. But the MD computational calculations are very time-consuming, so these simulations cannot screen large databases, and just a few ligand-protein systems can be studied. Consequently, at first, a fast docking method is performed to search large databases, and later on, more accurate but expensive MD simulations are applied for just a few molecules [26–28]. Eventually, in silico ADME studies were carried out on deriving molecules to investigate the pharmacokinetic parameters of the discovered compounds.

2. Materials and methods

2.1. Computational system

The computational studies were carried out on an Intel Xeon (R) CPU E5-2650 v2 @ 2.60 GHz × 32 core, 32 GB RAM system. The GPU unit used for molecular modeling and dynamic simulations was Nvidia GeForce GTX (4 GB), running on Linux Ubuntu 16.04 LTS.

2.2. Pharmacophore and virtual screening

Pharmit website (http://pharmit.csb.pitt.edu) was employed to construct the pharmacophore model using the crystal structure of COVID-19 main protease in complex with the N3 inhibitor (pdb: 6LU7) (http://www.pdb.org). A pharmacophore defines the necessary features of interaction. The coordinates of features were determined by calculating the average of all atoms' coordinates in their SMARTS chemical expression. The Pharmit, searches the selected databases using the Pharmer search method. This search method is similar but different from geometric hashing and the generalized Hough methods (two object recognition methods). An extensive explanation of Pharmer algorithm was given in the article written by David Ryan Koes and Carlos J. Camacho [29]. In this study, 7 features were used to construct the model: four donors, one acceptor, and two hydrophobic features. The shape constraints were also applied to ensure no heavy atom center within the receptor (exclusive shape) and at least one heavy atom center of the hits places within the inclusive shape (ligand). The exclusive shape eliminates molecules following the pharmacophore but has considerable steric conflicts with the receptor [30]. In order to validate the pharmacophore model, a set of decoys were built based on N3 and active inhibitors obtained from previous researches [14,31] using the DUD.E web tool (http://dude.docking.org/) [32]. The developed model with an Enrichment Factor of 7.6 was used to search the PubChem and

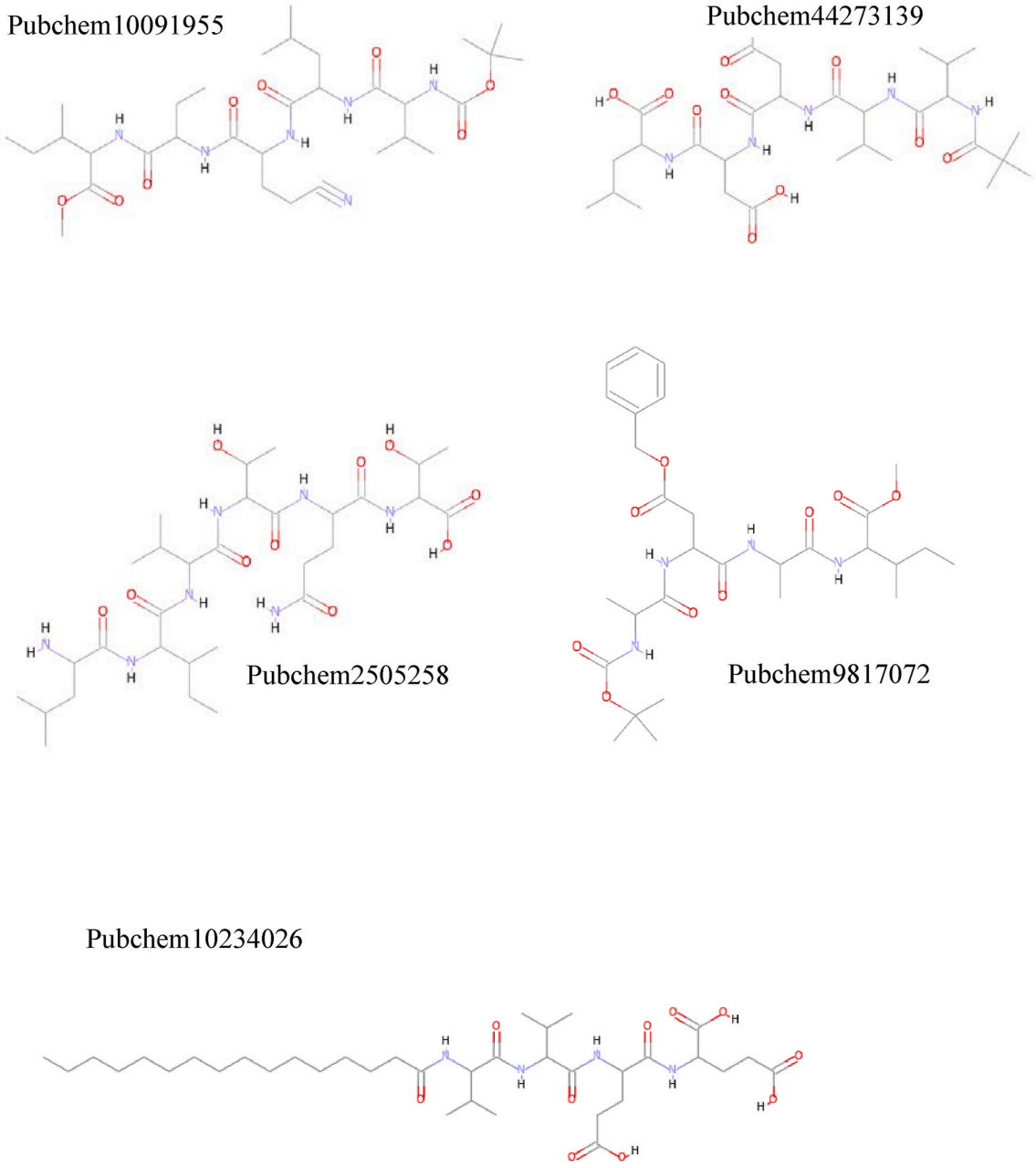


Fig. 2. Structure of retrieved molecules from pharmacophore-based virtual screening.

antiviral drug database (CAS COVID-19 antiviral candidate compound database) [33].

2.3. Molecular docking

Discovery Studio 4.1 (Accelrys Inc, San Diego, CA, USA) was applied to dock the resulting screening structures. The compounds were typed with CHARMm forcefield and minimized with Smart Minimizer. The structure of M^{pro} was prepared using protein preparation protocol. CDOCKER (CHRMm-based DOCKER) and a molecular dynamics (MD) simulated-annealing based algorithm were used to dock inhibitors into the receptors [34]. CDOCKER is a docking method that uses rigid receptor and flexible ligands. The active site was defined around the bounded ligand (N3), and the radius of the site sphere was set to 14 Å. Due to the limitation of CDOCKER as it considers a rigid receptor, a more accurate docking algorithm was utilized as another scoring function. GOLD is a genetic algorithm for docking flexible ligands into a protein's binding site, which accounts for the flexibility of the target protein's side chains [35].

2.4. Molecular dynamics simulation (MD) and MM/PBSA (Binding free energy)

For evaluation of the trustworthiness of the docking outcomes and the study of the changes of the protein-drug complex over time in an aqueous medium, molecular dynamics simulations (MD) have proceeded with the GROMACS 5.1.2 simulation package [36]. GROMOS96 53a6 force field [37] was applied to simulate the protein-ligand system. The PRODRG 2.5 server was utilized to generate the molecular topology files [38]. The ligand-protein complex structure was solvated with

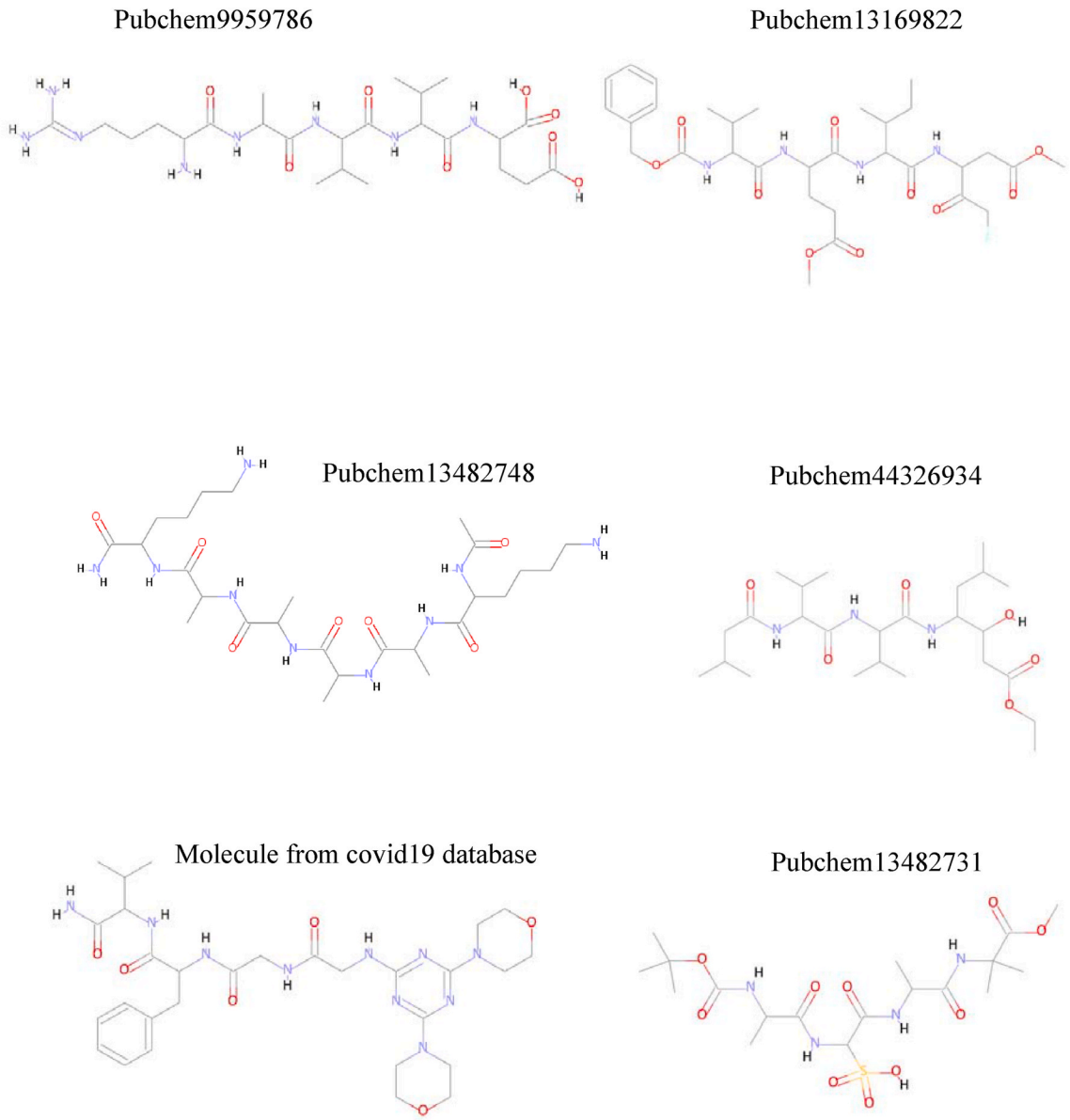


Fig. 2. (*continued*).

simple point charge (SPC) water molecules in a dodecahedral box [39]. The ligand-protein complex was centralized in the box with a minimum distance of 1 nm between the complex and the edge of the box. Four Na^+ ions were added to neutralize the charge of the system. To minimize the system, the steepest descent integrator for 50,000 steps, up to a tolerance of 10 kJ/mol without any constraints, was performed. Afterwards, the system was equilibrated with a short (200 ps) position restrained equilibration simulation at 300 K followed by a 200 ps position restrained simulation at the pressure of 1 bar. Finally, the constant temperature and pressure (NPT) MD simulations with periodic boundary conditions were performed for a period of 100 ns with a time step of 2 fs. The LINCS algorithm (Linear Constraint Solver) constrained the length of covalent bonds [40]. The particle-mesh Ewald (PME) summation technique was utilized to compute long-ranged electrostatic interactions [41]. The van der Waal's and coulomb cut-offs were set to 1.2. For keeping the temperature constant, the Berendsen thermostat was applied. Initial velocities were generated using the Maxwell distribution at 300 K.

The Molecular Mechanics Poisson-Boltzmann Surface Area (MM/PBSA) is an effective and reliable free energy simulation method. The g_mmpbsa application was proceeded to calculate the free energy between M^{pro} and the discovered drug. g_mmpbsa is a console application which is executed from terminal/console by command options similar to other GROMACS module [42,43]. This application was downloaded from https://rashmikumari.github.io/g_mmpbsa/website.

2.5. ADME studies

The four crucial subjects in pharmacokinetics are absorption, distribution, metabolism, and excretion (ADME). The compound

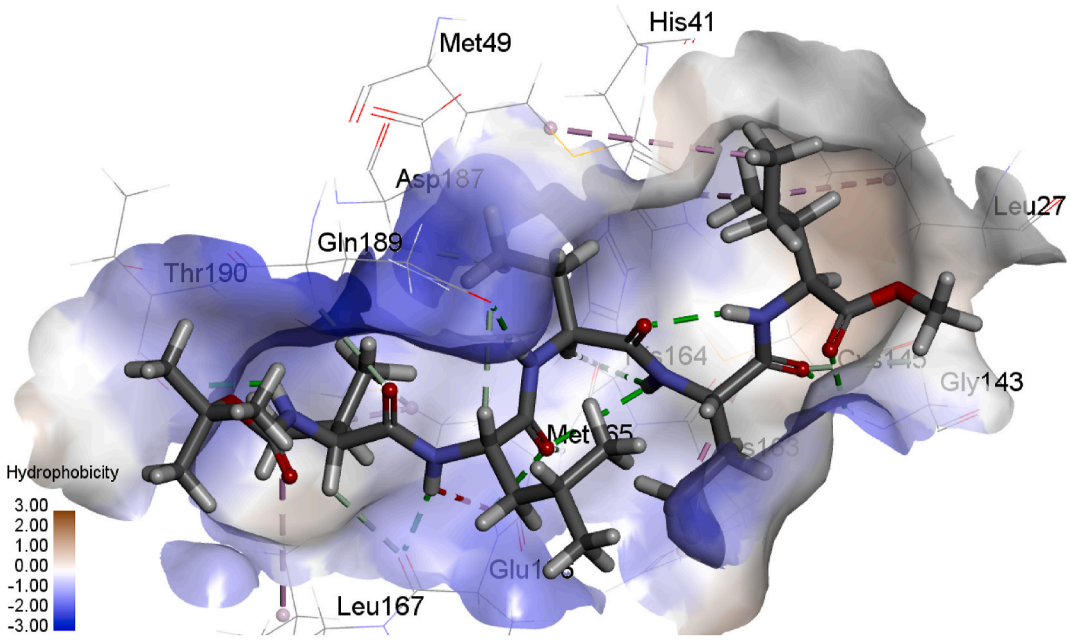


Fig. 3. The best docking pose of the most active compound (Pubchem100919551) in the pocket of COVID-19 virus M^{pro} obtained from CDOCKER. The surface of the binding site is colored by the hydrophobicity of the receptor residues, from blue for hydrophilic to brown for hydrophobic. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

possessing the best binding energy with the receptor might not be the best drug. A good drug should entirely and quickly absorb from the gastrointestinal tract, explicitly distributed to its target, metabolized in a manner that does not immediately stop its activity, and finally got out without resulting any harm [44]. A relationship between physiological parameters and chemical structures exists; thus, chemical descriptors can be used to calculate pharmacokinetic properties.

The blood-brain barrier (BBB) was used to predict the blood-brain penetration after oral consumption. The cytochrome P450 2D6 (CYP2D6) inhibition was measured using 2D chemical structure as input. CYP2D6 conducts the metabolism of a wide range of compounds in the liver. So, inhibition of CYP2D6 activity by a drug creates the problem of a drug interaction. Thus, it is essential to measure the CYP2D6 inhibition as a part of the drug development process. The hepatotoxicity model was performed to predict the potential toxicity of compounds.

The plasma protein binding model (PPB) was conducted to predict whether a molecule is probable to bound tightly (≥90% bound) to the plasma carrier proteins. A drug's efficiency can be affected by binding to the plasma protein because the attached fraction is temporarily covered from metabolism and only the free fraction exhibits pharmacological effects [45]. Molecular weight is a significant descriptor since the compounds with a molecular weight of more than 500 Da have too many rotatable bonds and functional groups able to form hydrogen bonds. The skin-permeability coefficient (log Kp) predicts the ability of compounds to permeate this barrier. The octanol-water partition coefficient (log Po/w) exhibits the drug's lipophilicity. Higher hydrophobicity results in an increase in the metabolism of compounds and low absorption. On the other hand, a hydrophobic drug is more likely to bind to undesirable hydrophobic macromolecules. Thus, the drug's hydrophobicity is an important descriptor.

ADME parameters, physicochemical descriptors, pharmacokinetic properties, and druglike nature of molecules were computed by using the SwissADME website (http://www.swissadme.ch/) and ADMET protocol of Discovery Studio software.

3. Results and discussion

3.1. Pharmacophore and virtual screening

To obtain a pharmacophore model and in silico screening for discovering lead inhibitors of COVID-19 virus M^{pro}, the crystal structure of COVID-19 M^{pro}-N3 inhibitor was used. By submitting the pdb code of M^{pro} to the Pharmit website, a set of interacting pharmacophore features will be automatically generated from the M^{pro}-N3 complex. Pharmit identifies 22 pharmacophore features. Searching with this default query yields no hits. To find the best features and to validate the pharmacophore model, a set of active and decoys were built and submitted to Pharmit. Based on the interaction of N3 with the active site of M^{pro} obtain from docking, 7 essential features plus shape constraints were applied to build the model: four donors, one acceptor, and two hydrophobic features. Fig. 1 shows the pharmacophore model based on N3 inhibitor. This model was the best model with an Enrichment Factor of 7.6 and was used to screen PubChem and antiviral drug database (CAS COVID-19 antiviral candidate compound database). Twenty-five hits with Max score of 0 and Min RMSD value of 3 was obtained. These hits were imported to Discovery Studio 4.1 to investigate the best inhibitors of COVID-19 M^{pro}.

3.2. Molecular docking

Docking studies were performed to predict binding conformations of

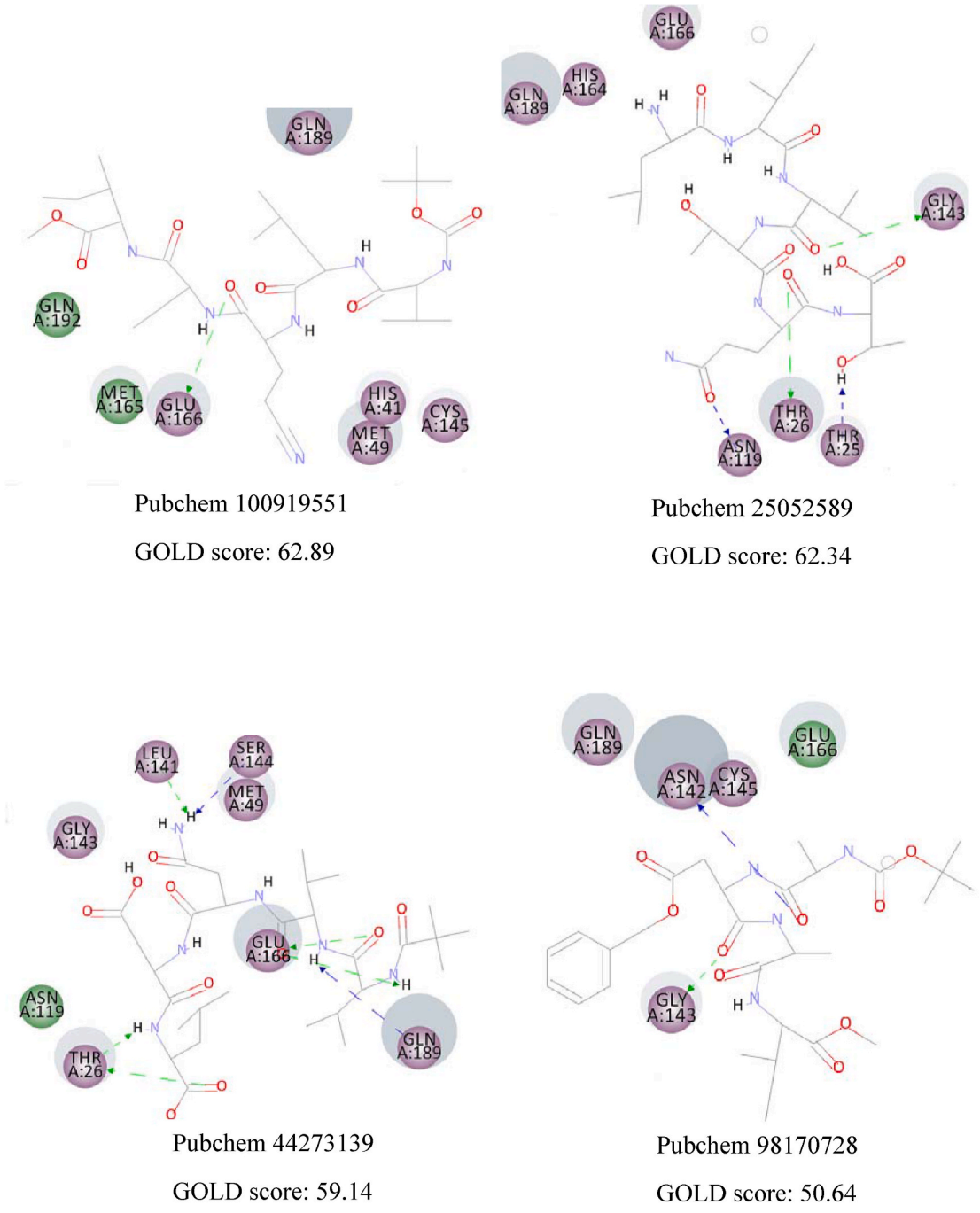
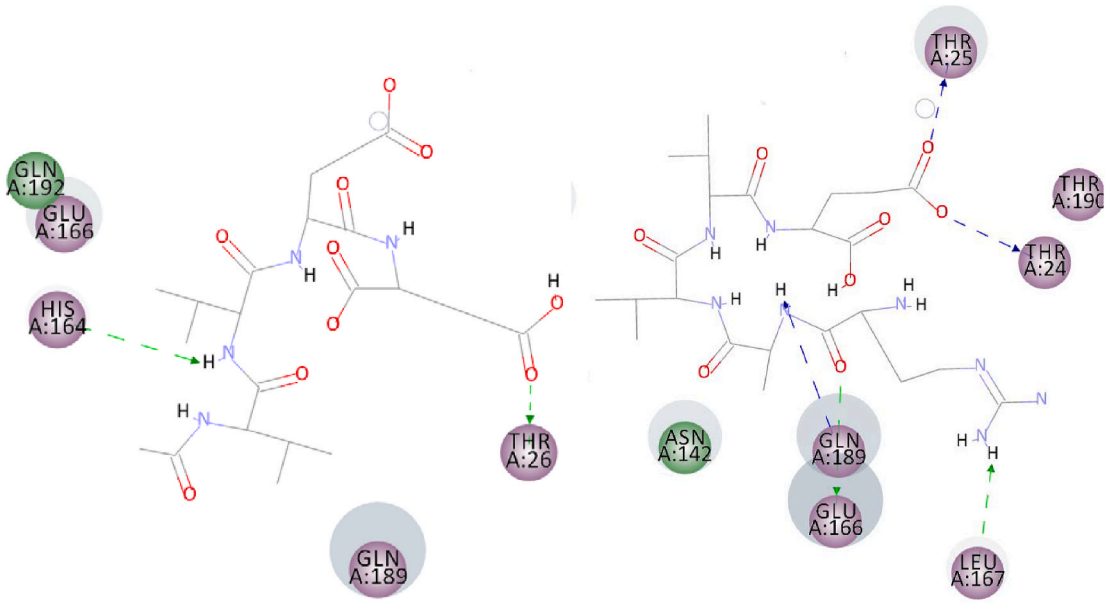


Fig. 4. The 2D interactions of the selected compounds obtained from molecular docking using GOLD algorithm. Residues with electrostatic and van der Waals interactions are colored in violet and green, respectively. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

all hits and investigate the best inhibitors of COVID-19 M^{pro}. CDOCKER algorithm was used to dock the hits into the COVID-19 M^{pro} active site. Root-mean-square distance (RMSD) value was calculated between coc-rystal (N3 inhibitor) and the re-docked inhibitor, which was 1.50 Å. This value of 1.50 Å demonstrates that CDOCKER is a reliable method to reproduce the experimental bonded conformation of the ligand-receptor complex. The retrieved compounds from virtual screening were docked into the active site of M^{pro} and ranked by their CDOCKER energy. CDOCKER score is the negative value of CDOCKER_ENERGY that a higher value is related to a higher affinity of binding. Thus, the energy can be applied like a score. This score contains internal ligand and receptor-ligand interaction energies. The top-ranked compounds are listed in Table 1. Fig. 2 shows the structures of these compounds. The best docking conformation of the most active molecule

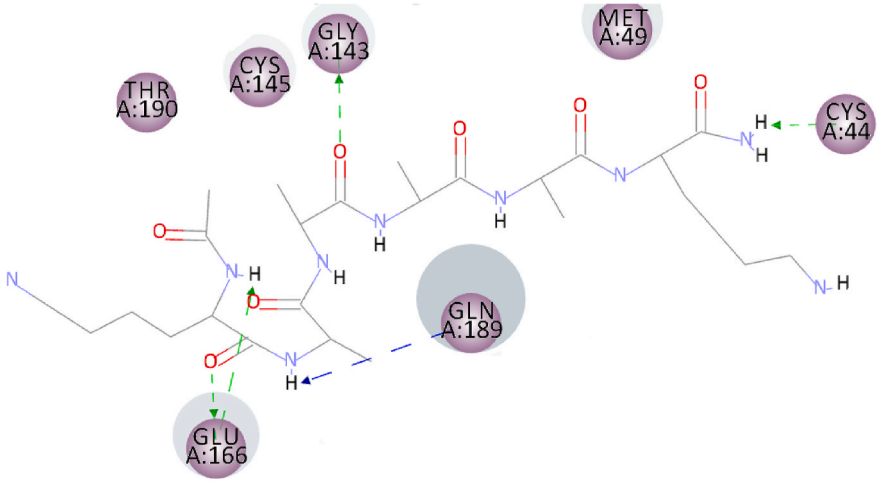


Pubchem 102340264

GOLD score: 58.35

Pubchem 9959786

GOLD score: 55.42



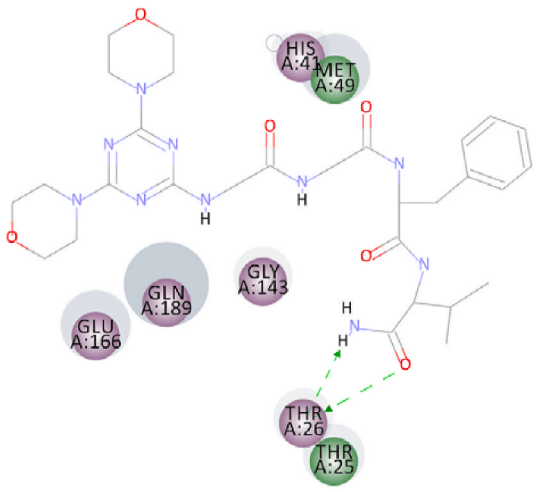
Pubchem 134827489

GOLD score: 56.09

Fig. 4. (*continued*).

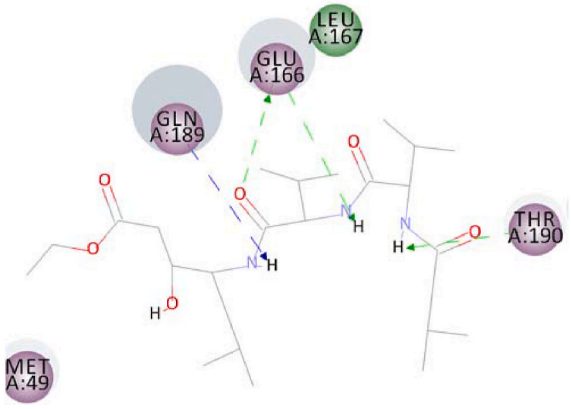
(Pubchem100919551) is shown in Fig. 3. The binding site's surface is colored by the receptor residues' hydrophobicity, from blue for hydrophilic to brown for hydrophobic. As can be seen, the binding site of M^{pro} is a hydrophilic pocket. So, electrostatic interactions are more critical than van der Waals interactions. Carbonyl groups of the best-docked pose of Pubchem100919551 represent hydrogen bonds with Cys145, Gly143, His164, Glu166, Asp 187, and Gln189, while the N–H moieties of this ligand show hydrogen bonding with residues Thr190, Glu166, Gln189, and His164. Some hydrophobic interactions were observed between the methyl moieties of ligand and the hydrophobic parts of His41, Leu27, Met49, His163, Cys145, Met165, and Leu167.

CDOCKER algorithm docks flexible ligands in a rigid receptor. Thus, GOLD algorithm was utilized as another scoring function to account for the protein's side chains' flexibility. The results are similar, and the same interactions can be observed in both methods. Fig. 4 displays the 2D interactions of the selected compounds obtained from molecular



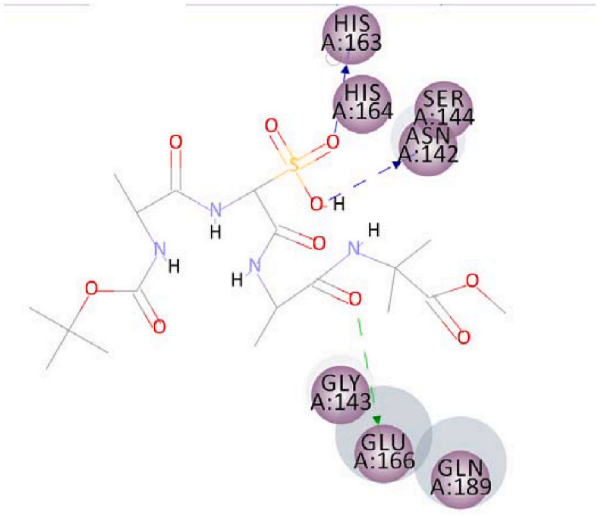
Molecule from Covid19 database

GOLD score: 58.69



Pubchem 44326934

GOLD score: 59.46



Pubchem 134827316

GOLD score: 50.80

Fig. 4. (*continued*).

docking using GOLD algorithm. As can be seen, Glu166 and Gln189 have an essential contribution in forming hydrogen bonds with most compounds. Residues with electrostatic and van der Waals interactions are colored in violet and green, respectively. It is quite evident that the majority of interactions are electrostatic and colored in violet.

Furthermore, docking of some protease inhibitor drugs such as ML188, nelfinavir, lopinavir, ritonavir, and α-ketoamide was conducted for comparison purposes. The 2D interactions of these protease inhibitors are presented in Fig. 5. It is clearly showed that the majority of interactions are electrostatic and colored in violet. Remdesivir interacts with the viral RNA-dependent RNA polymerase, but some articles stated that remdesivir is a protease inhibitor [46,47], so we decided to dock remdesivir into the main protease. The resulting CDOCKER energy and GOLD score value of remdesivir is very low, demonstrates that remdesivir is not a protease inhibitor. It is interesting to see that our discovered compounds show much more binding energy with COVID-19 M^{pro} (CDOCKER energy of the discovered molecule range from 73.99 to 100.09 kcal/mol, while the CDOCKER energy of the drugs currently in use range from 21.04 to 61.21 kcal/mol). These findings suggest the newly discovered compounds as more potent inhibitors (Table 2).

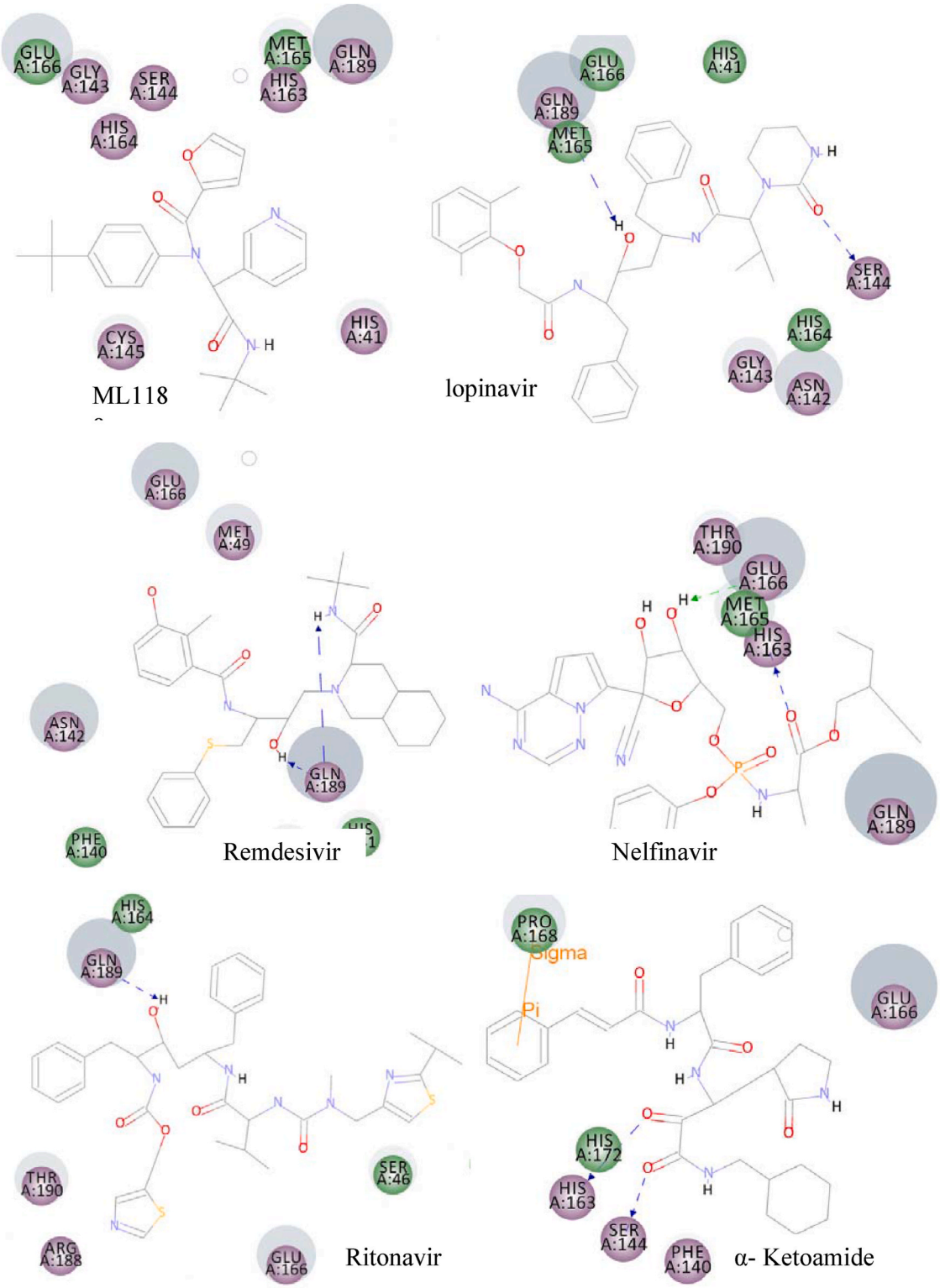


Fig. 5. The 2D interactions of the protease inhibitors of COVID19 currently in use. Residues with electrostatic and van der Waals interactions are colored in violet and green, respectively. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

Table 1
CDOCKER energy, Gold score, ADME prediction of molecules obtained from virtual screening using.

compounds	Molecular weight (g/mol)[a]	-Cdocker Energy (kcal/mol)	GOLD score	Log $p_{o/w}$[a]	Log S[a]	drug likeness based on solubility level[b]	GI level[a&b]	BBB permeant[a&b]	Log K_p (skin permeation) cm/s[a]	CYP2D6 inhibitor[b]	Hepatotoxic[b]	PPB[b] (plasma protein binding)	drug likeness based on Lipinski rule[a]
Pubchem 100919551	652.82	100.09	62.89	4.24	−4.72 Moderately soluble	Yes, good	low	No	−7.41	−8.79526 false	2.69562 true	false	No; 2 violations: MW > 500, N or O > 10
PubChem 25052589	943.14	96.60	60.34	1.84	−3.88 soluble	Yes, good	low	No	−11.60	−11.5093 false	−2.86896 true	false	No; 3 violations: MW > 500, N or O > 10 NH or OH > 5
Pubchem 44273139	642.74	96.57	59.14	2.74	−2.74 Soluble	Yes, good	low	No	−9.66	−11.9028 false	−0.0144951 true	false	No; 3 violations: MW > 500, N or O > 10 NH or OH > 5
Pubchem 98170728	592.68	94.94	50.64	5.19	−3.91 soluble	Yes, good	low	No	−7.96	−6.9075 false	−7.77278 false	false	No; 2 violations: MW > 500, N or O > 10
Pubchem 102340264	712.91	94.07	58.35	4.49	−6.43 Poorly soluble	Yes, good	low	No	−5.67	−6.13563 false	−11.9212 false	false	No; 3 violations: MW > 500, N or O > 10 NH or OH > 5
Pubchem 9959786	572.65	93.01	55.42	1.20	0.45 Highly soluble	Yes, good	low	No	−12.48	−10.6017 false	−1.6592 true	false	No; 3 violations: MW > 500, N or O > 10 NH or OH > 5
Pubchem 131698223	652.71	92.16	62.67	3.24	−4.01 Moderately soluble	Yes, good	low	No	−8.31	−3.88847 false	−9.86718 false	false	No; 2 violations: MW > 500, N or O > 10
Pubchem 134827489	599.72	86.49	56.09	1.40	−0.06 Highly soluble	Yes, optimal	low	No	−11.97	−5.7308 false	−1.1849 true	false	No; 3 violations: MW > 500, N or O > 10 NH or OH > 5
Molecule from covid19 database	626.71	86.15	58.69	3.06	−3.30 soluble	Yes, good	low	No	−9.56	−12.9677 false	−8.22286 false	false	No; 2 violations: MW > 500, N or O > 10
Pubchem 44326934	485.66	74.01	59.46	4.18	−3.83 soluble	Yes, good	high	No	−6.74	−6.32064 false	−6.96698 false	false	Yes; 0 violation
Pubchem 134827316	496.53	73.99	50.80	1.64	−1.42 Very soluble	Yes, good	low	No	−9.83	−11.7972 false	−0.12871 true	false	Yes; 1 violation: N or O > 10

[a] SwissADME website.
[b] ADMET protocol of Discovery Studio software.

Table 2
Prediction of ADME properties of some potential protease inhibitors of covid19 using.

compounds	Molecular weight (g/mol)[a]	-Cdocker Energy (kcal/mol)	Log $p_{o/w}$[a]	Log S[a]	drug likeness based on solubility level[b]	GI level[a&b]	BBB permeant[a&b]	Log K_p (skin permeation) cm/s[a]	CYP2D6 inhibitor[b]	Hepatotoxic[b]	PPB[b] < (plasma protein binding)	drug likeness based on Lipinski rule[a]
ML188	433.54	32.31	4.13	−5.46 Moderately soluble	Yes, low	high	no	−5.42	−10.425 false	−5.04757 false	false	Yes; 0 violation
Lopinavir	628.80	50.95	4.22	−6.64 Poorly soluble	Yes, good	high	no	−5.93	−3.68473 false	2.1839 true	true	Yes; 1 violation: MW > 500
Remdesvir	602.58	21.04	3.40	−4.12 Moderately soluble	Yes, low	low	no	−8.62	−9.08578 false	−0.643255 true	false	No; 2 violations: MW > 500, N or O > 10
Nelfinavir	567.78	23.94	4.24	−6.36 Poorly soluble	Yes, low	low	no	−5.74	−44.2708 false	−3.35042 true	true	Yes; 1 violation: MW > 500
Ritonavir	720.94	61.21	4.38	−6.99 Poorly soluble	Yes, low	low	no	−6.40	24.7098 true	28.7074 true	false	No; 2 violations: MW > 500, N or O > 10
α-Ketoamide (11r)	566.65	50.28	2.51	−4.81 Moderately soluble	Yes, good	low	no	−7.28	−10.5278 false	−8.63411 false	false	Yes; 1 violation: MW > 500

[a] SwissADME website.
[b] ADMET protocol of Discovery Studio software.

3.3. ADME studies

For the investigation of the pharmacokinetic parameters of retrieved compounds, log$P_{o/w}$ for octanol/water, aqueous solubility (log S), human oral absorption in the gastrointestinal tract (GI), blood-brain barrier (BBB), skin-permeability coefficient (log K*p*), CYP2D6 inhibitor, Hepatotoxicity, and plasma protein binding (PPB) were computed (Table 1). An extreme hydrophobic drug reveals low solubility in the gut and solvate in fat globules [48]. Based on solubility level, all compounds have druglikness properties. As shown in Table 1, just Pubchem44326934 can be absorbed from the gastrointestinal tract (GI). All compounds are not able to cross the blood-brain barrier (BBB) and can not inhibit the cytochrome P450 2D6 (CYP2D6) activity (inhibition of CYP2D6 by a drug creates the problem of drug interaction).

The resulting hepatotoxicity model predicts the potential toxicity of some of the compounds (Table 1). Moreover, a drug's efficiency can be affected by binding to the plasma protein as only the free fraction exhibits pharmacological effects. The plasma protein binding model (PPB) predicts that non of the molecules can bound tightly to the plasma carrier proteins. In addition, most orally administered drugs are relatively small molecules with molecular weight less than 500Da. As can be observed, only two compounds have a molecular mass of less than 500 (g/mol).

According to the resulting ADME parameters, only Pubchem44326934 has high gastrointestinal absorption (GI), molecular mass less than 500Da, no toxicity, and has druglike properties (based on Lipinski's rule of 5). However, the discovered compounds can be modified to show more druglike properties.

3.4. Molecular dynamics simulation

The binding stability of the M^{pro}-drug complex was analyzed through molecular dynamics simulation. The structure of the M^{pro}-drug complex was obtained from molecular docking and was simulated for 100 ns using the GROMACS package. To investigate the stability of the M^{pro}-drug complex under the simulation conditions, root mean square deviation (RMSD) of the trajectories from their primary structures was computed. The RMSD value of the backbone of protein ranged from 0.03 to 0.42 nm, while the ligand RMSD ranged from 0.04 to 0.31 nm, as shown in Fig. 6. The RMSD of the ligand got to about 0.2 nm after 700 ps, and was maintained at this amount through the entire MD simulation. This indicates that the structure of M^{pro}-drug complex is stable all through the simulation procedure. However, the structure of M^{pro} represents much more significant fluctuations and a more considerable RMSD value of about 0.21 nm, while the drug (Pubchem44326934) reveals smaller conformational changes. These results demonstrate the high affinity of the proposed drug toward the active site of COVID-19 M^{pro}, forming a stable complex with low flexibility.

Hydrogen bonds between the proposed drug (Pubchem44326934) and M^{pro} are illustrated in Fig. 7. There are Lots of H-bonds within the entire MD simulations, but only some of them are stable. These unstable H-bonds present due to the motion of the atoms of ligand and protein and may have a small contribution to the drug's affinity toward M^{pro}. It can be seen from Fig. 6 that there are three stable H-bonds during most of the simulation time. These H-bonds play an important role in forming the ligand-protein complex. They exist between the carbonyl group of Pubchem44326934 and the NH group of Gln189 as well as two hydrogen bonds based on Glu166 (1. The hydrogen atom of the amino group of Glu166 and the carbonyl group of ligand, 2. The carbonyl group of Glu166 and the hydrogen atom of the amino group of the ligand). These stable H-bonds are also present in the first pose of docking. Consequently, the protein-ligand contacts demonstrate the highest H-bond and water bridge interaction with Glu166; the second-highest contacts were obtained with Gln189. Afterwards, Asn119 and Gly143 can form semi-stable hydrogen bonds.

The most relevant hydrophobic interaction can be observed between

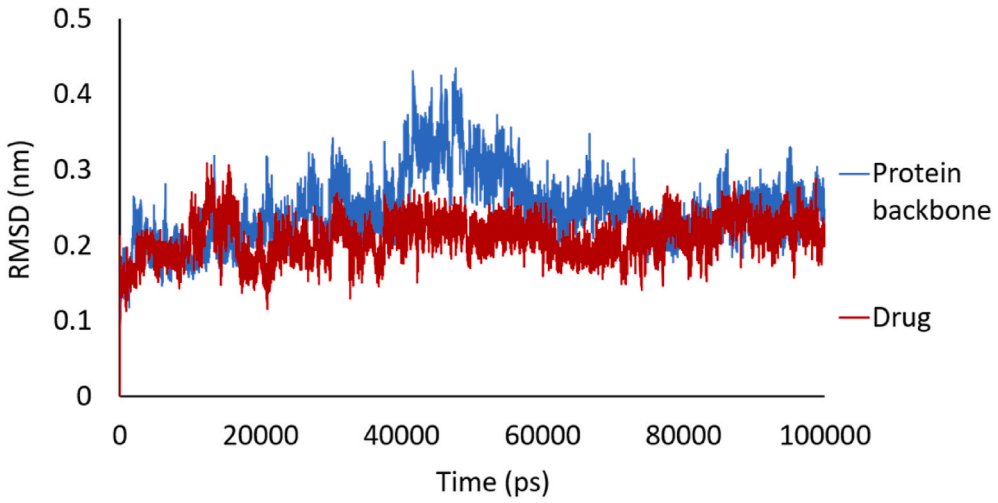


Fig. 6. The root mean square deviation (RMSD) of M^{pro} (blue) and Pubchem44326934 (red) versus MD simulation time (picoseconds). (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

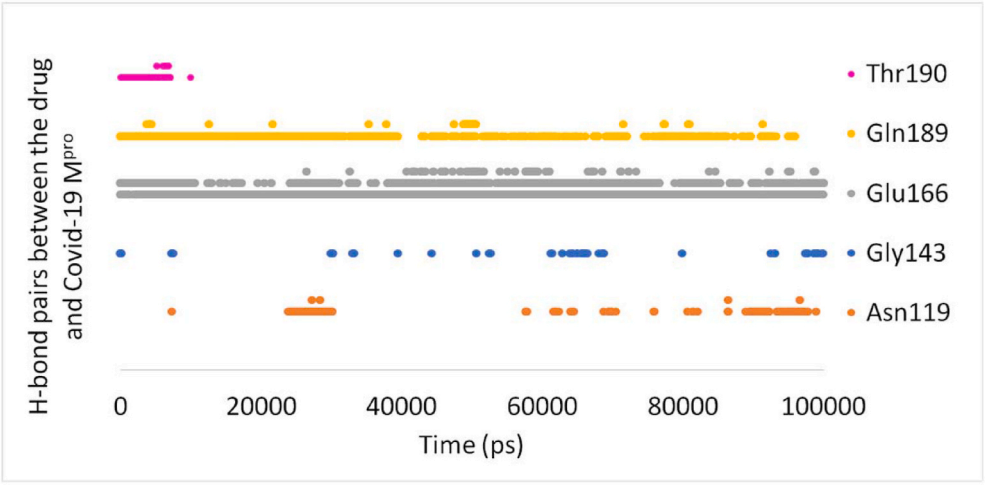


Fig. 7. Hydrogen bonds between the ligand (Pubchem44326934) and COVID-19 M^{pro} throughout the MD simulations. A colored mark displays H-bonds according to protein residue.

the hydrophobic residues of the binding site and the selected compounds. The minimum distance between these hydrophobic residues and Pubchem44326934 was studied by MD simulations. No notable changes were observed in the distances between Leu27, Met165, Leu167, Gln189, and the ligand during the simulation, demonstrating that these hydrophobic interactions are important and form a stable complex with only small conformational changes (Fig. 8). However, compared with Leu27, Met165, and Leu167, Gln189 has a less distance of about 0.1 nm from Pubchem44326934.

The minimum distance between some hydrophobic residues such as Asp187, Val186, and Gln192 increase and also fluctuate more widely during the last 4 ns of the simulation, indicating the instability of these hydrophobic interactions at the end of the simulation.

Water molecules are usually removed before docking. However, they are the main portion of the simulation that can affect ligand-protein binding and bridge between the ligand and protein. The final snapshot was used to study the important effect of water molecules. Fig. 9 shows five water-mediated H-bonds. Residues with electrostatic and van der Waals interactions are colored in violet and green, respectively. Two hydrogen bonds can be observed between Glu166 and ligand, while a π-sigma interaction presents between Pubchem44326934 and His41.

3.5. *Binding free energy calculations*

The MM/PBSA approach was applied to calculate the free energy between M^{pro} and the discovered drug. Fig. 10 represents the vacuum molecular mechanics energy during the simulation. As can be seen, the trend lines of MM/PBSA for MD trajectories show low fluctuations, indicate its high binding affinity toward M^{pro}. The average of M^{pro}-drug (Pubchem44326934) electrostatic energy is −90 kJ/mol, while the van der Waals energy between M^{pro} and Pubchem44326934 is −200 kJ/mol. The average total energy of M^{pro}-drug is −290 kJ/mol, confirming that Pubchem44326934 has enough affinity for being considered as M^{pro} inhibitor. The binding energy obtained through molecular docking of

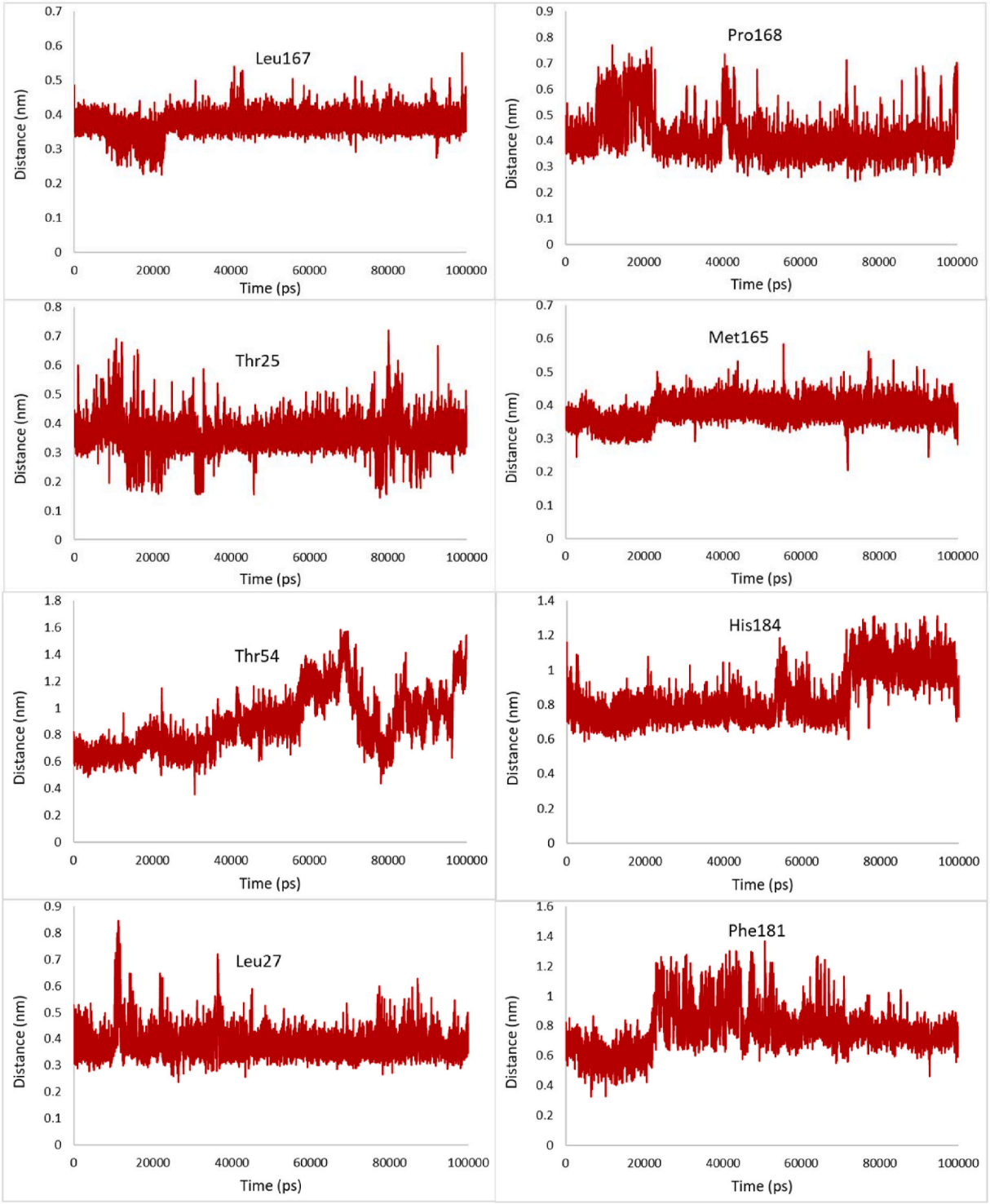


Fig. 8. Changes in distances of the most relevant hydrophobic interactions between hydrophobic residues of the binding site and Pubchem44326934 during the 100ns MD simulation.

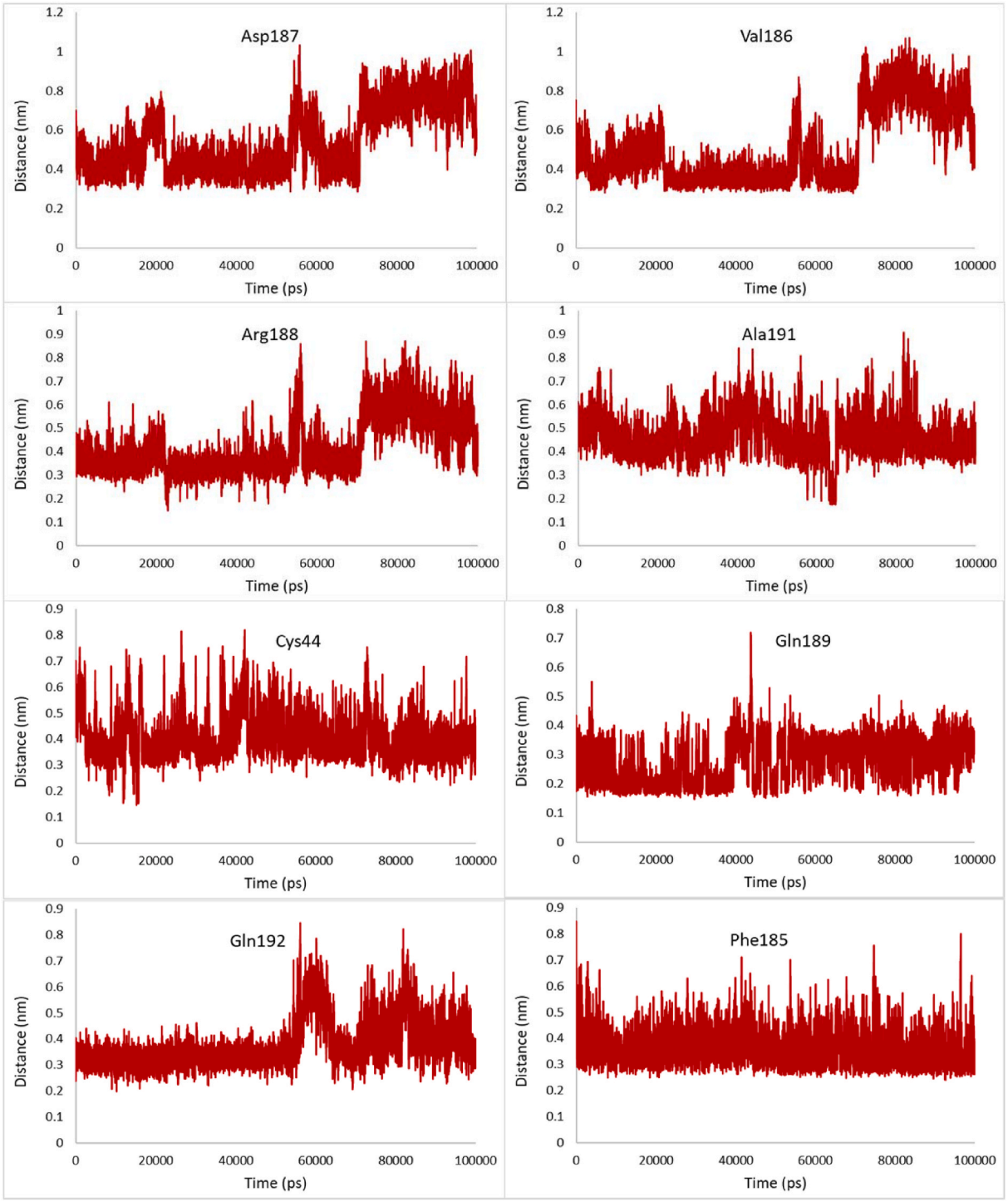


Fig. 8. (*continued*).

the discovered molecule was compared to the MM/PBSA. It can be observed from Table 1 that the CDOCKER energy of Pubchem44326934 is −74.01 kcal/mol (−309.66 kJ/mol), which is close to the MM/PBSA total energy. Therefore, the reliability of the docking results was confirmed by MD and MM/PBSA approaches. The high negative binding free energy demonstrates this complex's stable configuration, ensuring that Pubchem44326934 has enough affinity for being considered as COVID-19 M^{pro} inhibitor drug.

4. Conclusions

In this work, more potent coronavirus protease inhibitors were identified. Based on the crystal structure of COVID-19 M^{pro} with N3 inhibitor, pharmacophore-based virtual screening was conducted. The resulting compounds were docked into the protease pocket and ranked by their interaction energy. In silico ADME studies were performed on docking outcomes to see whether the compounds are suitable to be considered as a drug or not. Pubchem44326934 showed the druglike properties and was further analyzed by MD and MM/PBSA approaches, confirmed that it could be regarded as a new potent COVID-19 M^{pro}

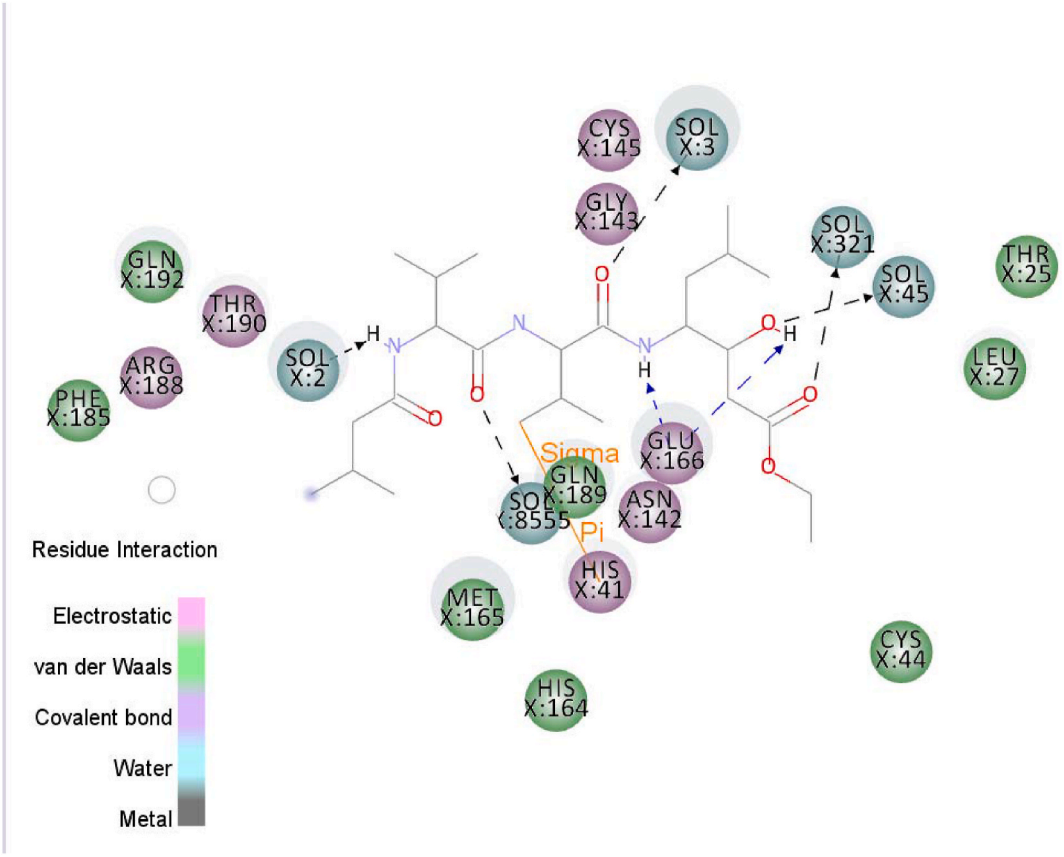


Fig. 9. The 2D interactions of Pubchem44326934 obtained from MD simulation. Residues with electrostatic and van der Waals interactions are colored in violet and green, respectively. Water molecules are shown in blue. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

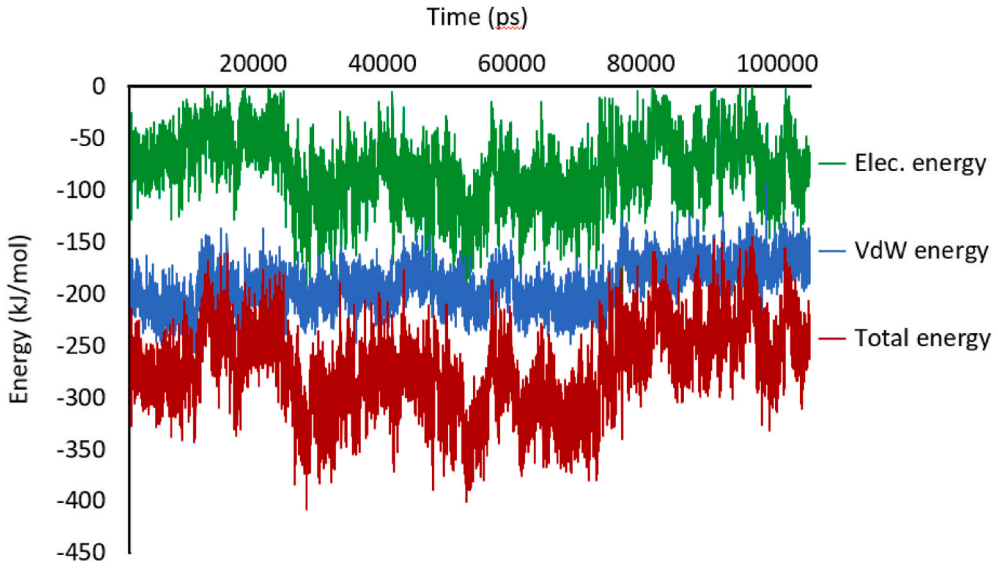


Fig. 10. The vacuum molecular mechanics energy MM/PBSA during the simulation.

inhibitor drug. Future structural-functional studies are recommended to improve the ADME properties of the screened compounds. However, the discovered compounds should be tested, and their in vitro potential should be determined for further validation of the results.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

We thank Dr. Somayeh Pirhadi for her expertise and assistance.

References

[1] Mittal L, Kumari A, Srivastava M, Singh M, Asthana S. Identification of potential molecules against COVID-19 main protease through structure-guided virtual screening approach. ChemRxiv. Preprint. http://doi.org/10.26434/chemrxiv.12086565.v2.
[2] Xu C, Ke Z, Liu C, Wang Z, Liu D, Zhang L, et al. Systemic in silico screening in drug discovery for Coronavirus Disease (COVID-19) with an online interactive web server, J Chem Inf Model. http://doi.org/10.1021/acs.jcim.0c00821.
[3] Luk HKH, Li X, Fung J, Lau SKP, Woo PCY. Molecular epidemiology, evolution and phylogeny of SARS coronavirus. Infect Genet Evol 2019;71:21–30. https://doi.org/10.1016/j.meegid.2019.03.001.
[4] Ziebuhr J. Molecular biology of severe acute respiratory syndrome coronavirus. Curr Opin Microbiol 2004;7:412–9. https://doi.org/10.1016/j.mib.2004.06.007.
[5] Weiss SR, Leibowitz JL. Coronavirus pathogenesis. Adv Virus Res 2011;81:85–164. https://doi.org/10.1016/B978-0-12-385885-6.00009-2.
[6] Brian DA, Baric RS. Coronavirus genome structure and replication. Curr Top Microbiol Immunol 2005;287:1–30. https://doi.org/10.1007/3-540-26765-4_1.
[7] Narayanan K, Huang C, Makino S. SARS coronavirus accessory proteins. Virus Res 2008;133:113–21. https://doi.org/10.1016/j.virusres.2007.10.009.
[8] Yan R, Zhang Y, Li Y, Xia L, Guo Y, Zhou Q. Structural basis for the recognition of the SARS-CoV-2 by full-length human ACE2. Science 2020;367:1444–8. https://doi.org/10.1126/science.abb2762.
[9] Walls AC, Park YJ, Tortorici MA, Wall A, McGuire AT, Veesler D. Structure, function, and antigenicity of the SARS-CoV-2 spike glycoprotein. Cell 2020;181:281–92. https://doi.org/10.1016/j.cell.2020.02.058.
[10] Hoffmann M, Kleine-Weber H, Schroeder S, Kruger N, Herrler T, et al. SARS-CoV-2 cell entry depends on ACE2 and TMPRSS2 and is blocked by a clinically proven protease inhibitor. Cell 2020;181:271–80. https://doi.org/10.1016/j.cell.2020.02.052.
[11] Dimitrov DS. The secret life of ACE2 as a receptor for the SARS virus. Cell 2003;115:652–3. https://doi.org/10.1016/s0092-8674(03)00976-0.
[12] Subissi L, Imbert I, Ferron F, Collet A, Coutard B, Decroly E, et al. SARS-CoV ORF1b-encoded nonstructural proteins 12-16: replicative enzymes as antiviral targets. Antivir Res 2014;101:122–30. https://doi.org/10.1016/j.antiviral.2013.11.006.
[13] Graham RL, Sparks JS, Eckerle LD, Sims AC, Denison MR. SARS coronavirus replicase proteins in pathogenesis. Virus Res 2008;133:88–100. https://doi.org/10.1016/j.virusres.2007.02.017.
[14] Jin Z, Du X, Xu Y, Deng Y, Liu M, Zhao Y, et al. *Struc*ture of Mpro f^{rom} COVID-19 virus and discovery of its inhibitors. Nature 2020;582:289–93. https://doi.org/10.1038/s41586-020-2223-y.
[15] Yin J, Niu C, Cherney MM, Zhang J, Huitema C, et al. A mechanistic view of enzyme inhibition and peptide hydrolysis in the active site of the SARS-CoV 3C-like peptidase. J Mol Biol 2007;371:1060–74. https://doi.org/10.1016/j.jmb.2007.06.001.
[16] Chou KC, Wei DQ, Zhong WZ. Binding mechanism of coronavirus main proteinase with ligands and its implication to drug design against SARS. Biochem Biophys Res Commun 2003;308:148–51. https://doi.org/10.1016/s0006-291x(03)01342-1.
[17] Muramatsu T, Kim YT, Nishii W, Terada T, Shirouzu M, Yokoyama S. Autoprocessing mechanism of severe acute respiratory syndrome coronavirus 3C-like protease (SARS-CoV 3CLpro) from its polyproteins. FEBS J 2013;280:2002–13. https://doi.org/10.1111/febs.12222.
[18] Xu Z, Peng C, Shi Y, Zhu Z, Mu K, Wang X, et al. Nelfinavir was predicted to be a potential inhibitor of 2019-nCov main protease by an integrative approach combining homology modelling, molecular docking and binding free energy calculation. BioRxiv 2020. https://doi.org/10.1101/2020.01.27.921627v1.
[19] Kumar Y, Singh H, Patel CN. In silico prediction of potential inhibitors for the main protease of SARS-CoV-2 using molecular docking and dynamics simulation based drug-repurposing. J Infect Public Health 2020;13:1210–23. https://doi.org/10.1016/j.jiph.2020.06.016.
[20] Zhang L, Lin D, Sun X, Curth U, Drosten C, Sauerhering L, et al. Crystal structure of SARS-CoV-2 main protease provides a basis for design of improved α-ketoamide inhibitors. Science 2020;368:409–12. https://doi.org/10.1126/science.abb3405.
[21] Pirhadi S, Ghasemi JB. Pharmacophore identification, molecular docking, virtual screening, and in silico ADME studies of non-nucleoside reverse transcriptase inhibitors. Mol Inf 2012;31:856–66. https://doi.org/10.1002/minf.201200018.
[22] Prinz H, Ridder AK, Vogel K, Böhm KJ, Ivanov I, Ghasemi JB, et al. N-heterocyclic (4-phenylpiperazin-1-yl)methanones derived from phenoxazine and phenothiazine as highly potent inhibitors of tubulin polymerization. J Med Chem 2017;60:749–66. https://doi.org/10.1021/acs.jmedchem.6b01591.
[23] Aghaee E, Ghasemi JB, Manouchehri F, Balalaie S. Combined docking, molecular dynamics simulations and spectroscopic studies for the rational design of a dipeptide ligand for affinity chromatography separation of human serum albumin. J Mol Model 2014;20:2446–59. https://doi.org/10.1007/s00894-014-2446-7.
[24] Alonso H, Bliznyuk AA, Gready JE. Combining docking and molecular dynamic simulations in drug design. Med Res Rev 2006;26:531–68. https://doi.org/10.1002/med.20067.
[25] Ghasemi JB, Hooshmand S. 3D-QSAR, docking and molecular dynamics for factor Xa inhibitors as anticoagulant agents. Mol Simulat 2013;39:453–71. https://doi.org/10.1080/08927022.2012.741235.
[26] Manetti F, Locatelli GA, Maga G, Schenone S, Modugno M, Forli S, et al. A combination of docking/dynamics simulations and pharmacophoric modeling to discover new dual c-Src/Abl kinase inhibitors. J Med Chem 2006;49:3278–86. https://doi.org/10.1021/jm060236z.
[27] Li L, Wei DQ, Wang JF, Chou KC. Computational studies of the binding mechanism of calmodulin with chrysin. Biochem Biophys Res Commun 2007;358:1102–7. https://doi.org/10.1016/j.bbrc.2007.05.053.
[28] Kesner TF, Elcock AH. Computational sampling of a cryptic drug binding site in a protein receptor: explicit solvent molecular dynamics and inhibitor docking to p38 map kinase. J Mol Biol 2006;359:202–14. https://doi.org/10.1016/j.jmb.2006.03.021.
[29] Koes DR, Camacho CJ. Pharmer: efficient and exact pharmacophore search. J Chem Inf Model 2011;51:1307–14. https://doi.org/10.1021/ci200097m.
[30] Sunseri J, Koes DR. Pharmit: interactive exploration of chemical space. Nucleic Acids Res 2016;44. https://doi.org/10.1093/nar/gkw287. W442-W8.
[31] Zhang L, Lin D, Kusov Y, Nian Y, Ma Q, Wang J, et al. α-Ketoamides as broad-spectrum inhibitors of coronavirus and enterovirus replication: structure-based design, synthesis, and activity assessment. J Med Chem 2020;63:4562–78. https://doi.org/10.1021/acs.jmedchem.9b01828.
[32] Mysinger MM, Carchia M, Irwin JJ, Shoichet BK. Directory of useful decoys, enhanced (DUD-E): better ligands and decoys for better benchmarking. J Med Chem 2012;55:6582–94. https://doi.org/10.1021/jm300687e.
[33] COVID-19 antiviral candidates compounds database. available at, https://www.cas.org/covid-19-antiviral compoundsdataset?utm_source=hootsuite&utm_medium=facebook&utm_term=&utm_content=dbcae960-b771-4732-9e4a-40215709e07a&utm_campaign=COVID-19.
[34] Wu G, Robertson DH, Brooks CL, Vieth M. Detailed analysis of grid-based molecular docking: a case study of CDOCKER - a CHARMm-based MD docking algorithm. J Comput Chem 2003;24:1549–62. https://doi.org/10.1002/jcc.10306.
[35] Politi A, Durdagi S, Moutevelis-Minakakis P, Kokotos G, Mavromoustakos T. Development of accurate binding affinity predictions of novel rennin inhibitors through molecular docking studies. J Mol Graph Model 2010;29:425–35. https://doi.org/10.1016/j.jmgm.2010.08.003.
[36] Spoel DVD, Lindahl E, Hess B, Groenhof G, Mark AE, Berendsen HJC. GROMACS: fast, flexible, and free. J Comput Chem 2005;26:1701–18. https://doi.org/10.1002/jcc.20291.
[37] Oostenbrink C, Villa A, Mark AE, Gunsteren WFV. A biomolecular force field based on the free enthalpy of hydration and solvation: the GROMOS forcefield parameter sets 53A5 and 53A6. J Comput Chem 2004;25:1656–76. https://doi.org/10.1002/jcc.20090.
[38] Schüttelkopf AW, Aalten DMFV. PRODRG - a tool for high-throughput crystallography of protein-ligand complexes. Acta Crystallogr D Biol Crystallogr 2004;60:1355–63. https://doi.org/10.1107/S0907444904011679.
[39] Berendsen HIC, Postma JPM, Gunsteren WV, Hermans J. Interaction models for water in relation to protein hydration. In: Pullman B, editor. Intermolecular forces. The jerusalem symposia on quantum chemistry and biochemistry, vol. 14. Dordrecht: Springer; 1981. p. 331–42. https://doi.org/10.1007/978-94-015-7658-1_21.
[40] Hess B, Bekker H, Berendsen HJC, Fraaije JGEM. LINCS: a linear constraint solver for molecular simulations. J Comput Chem 1997;18:1463–72. https://doi.org/10.1002/(SICI)1096-987X(199709)18:12<1463::AID-JCC4>3.0.CO;2-H.
[41] Darden T, York D, Pedersen L. Particle mesh ewald-an N.log(N) method for Ewald sums in large systems. J Chem Phys 1993;98:10089–92. https://doi.org/10.1063/1.464397.
[42] Kumari R, Kumar R, Lynn A. g_mmpbsa - a GROMACS tool for high-throughput MM-PBSA calculations. J Chem Inf Model 2014;54:1951–62. https://doi.org/10.1021/ci500020m.
[43] Baker NA, Sept D, Joseph S, Holst MJ, McCammon JA. Electrostatics of nanosystems: application to microtubules and the ribosome. Proc Natl Acad Sci U S A 2001;98:10037–41. https://doi.org/10.1073/pnas.181342398.
[44] Hodgson J. ADMET-turning chemicals into drugs. Nat Biotechnol 2001;19:722–6. https://doi.org/10.1038/90761.
[45] ADMET descriptors in discovery Studio. https://www.3dsbiovia.com/products/datasheets/ds_admet_descriptors.pdf.
[46] Naik VR, Munikumar M, Ramakrishna U, Srujana M, Goudar G, Naresh P, et al. Remdesivir (GS-5734) as a therapeutic option of 2019-nCOV main protease - *in silico* approach. J Biomol Struct Dyn 2020:1–14. https://doi.org/10.1080/07391102.2020.1781694.
[47] Mothay D, Ramesh KV. Binding site analysis of potential protease inhibitors of COVID-19 using AutoDock. Virus (Tokyo) 2020;31:194–9. https://doi.org/10.1007/s13337-020-00585-z.
[48] Palm K, Stenberg P, Luthman K, Artursson P. Polar molecular surface properties predict the intestinal absorption of drugs in humans. Pharm Res (N Y) 1997;14:568–71. https://doi.org/10.1023/A:1012188625088.