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.

Computers in Biology and Medicine 157 (2023) 106785

Contents lists available at ScienceDirect

Computers in Biology and Medicine

journal homepage: www.elsevier.com/locate/compbiomed

Identification of medicinal plant-based phytochemicals as a potential inhibitor for SARS-CoV-2 main protease (M^{pro}) using molecular docking and deep learning methods

Alomgir Hossain [a,*], Md Ekhtiar Rahman [a], Md Siddiqur Rahman [a], Khondokar Nasirujjaman [a], Mohammad Nurul Matin [a], Md Omar Faruqe [b], Muhammad Fazle Rabbee [c,**]

[a] Department of Genetic Engineering and Biotechnology, University of Rajshahi, Rajshahi, 6205, Bangladesh
[b] Department of Computer Science and Engineering, University of Rajshahi, Rajshahi, 6205, Bangladesh
[c] Department of Biotechnology, Yeungnam University, Gyeongsan, Gyeongsangbuk-do, 38541, Republic of Korea

ARTICLE INFO

Keywords:
Deep learning
Catechin gallate
Main protease
Molecular docking
SARS-CoV-2

ABSTRACT

Highly transmissive and rapidly evolving Coronavirus disease-2019 (COVID-19), a viral disease caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), triggered a global pandemic, which is one of the most researched viruses in the academia. Effective drugs to treat people with COVID-19 have yet to be developed to reduce mortality and transmission. Studies on the SARS-CoV-2 virus identified that its main protease (M^{pro}) might be a potential therapeutic target for drug development, as this enzyme plays a key role in viral replication. In search of potential inhibitors of M^{pro}, we developed a phytochemical library consisting of 2431 phytochemicals from 104 Korean medicinal plants that exhibited medicinal and antioxidant properties. The library was screened by molecular docking, followed by revalidation by re-screening with a deep learning method. Recurrent Neural Networks (RNN) computing system was used to develop an inhibitory predictive model using SARS coronavirus M^{pro} dataset. It was deployed to screen the top 12 compounds based on their docked binding affinity that ranged from −8.0 to −8.9 kcal/mol. The top two lead compounds, Catechin gallate and Quercetin 3-O-malonylglucoside, were selected depending on inhibitory potency against M^{pro}. Interactions with the target protein active sites, including His41, Met49, Cys145, Met165, and Thr190 were also examined. Molecular dynamics simulation was performed to analyze root mean square deviation (RMSD), root mean square fluctuation (RMSF), radius of gyration (RG), solvent accessible surface area (SASA), and number of hydrogen bonds. Results confirmed the inflexible nature of the docked complexes. Absorption, distribution, metabolism, excretion, and toxicity (ADMET), as well as bioactivity prediction confirmed the pharmaceutical activities of the lead compound. Findings of this research might help scientists to optimize compatible drugs for the treatment of COVID-19 patients.

1. Introduction

The COVID-19 pandemic, the extremely contagious viral infection induced by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), has been posing serious threats to the world in a multitude of ways. As of February 2023, there have been 677.74 million confirmed cases of COVID-19 along with 6.78 million deaths, reported by the World Health Organization (WHO). Although significant advancements in clinical research have improved our knowledge of SARS-CoV-2 and COVID-19 management, there is growing worried about how to prevent the dissemination of this virus and its mutations. SARS-CoV-2 is still causing havoc around the globe and the advent of mutant variations is primarily to blame for the second or third wave of breakouts of this viral illness that are currently affecting more than 230 countries [1].

Abbreviations: ADMET, Absorption, distribution, metabolism, excretion and toxicity; COVID-19, Coronavirus disease-2019; SARS-CoV-2, Severe acute respiratory syndrome coronavirus 2; M^{pro}, Main protease; RNN, Recurrent Neural Networks; RMSD, Root mean square deviation; RMSF, Root mean square fluctuation; RG, Radius of gyration; SASA, Solvent accessible surface area; MD, Molecular Dynamics.

* Corresponding author.
** Corresponding author.
E-mail addresses: s1910861106@ru.ac.bd (A. Hossain), rabbi.biotech@gmail.com (M.F. Rabbee).

https://doi.org/10.1016/j.compbiomed.2023.106785
Received 19 November 2022; Received in revised form 15 February 2023; Accepted 10 March 2023
Available online 11 March 2023

SARS-CoV-2 is a new beta coronavirus that belongs to the identical subgenus (SARS-CoV, 2002) and the Middle East Respiratory Syndrome Coronavirus (MERS-CoV, 2013), and their previous mortality rates were up to 10% and 35%, respectively [2]. Throughout this pandemic, many SARS-CoV-2 variants have been identified among which merely a few are recognized as variants of concern by the WHO due to their effect on the world's public health such as Alpha (B.1.1.7) in the United Kingdom, Beta (B.1.351) in South Africa, Gamma (P.1) in Brazil, Delta (B.1.617.2) in India and Omicron (B.1.1.529) in South Africa [1].

SARS-CoV-2 is a single-stranded positive-sense RNA virus containing ~30,000 nucleotides and 11 protein-coding genes, with 12 expressed proteins [3–5]. Among these few encoded proteins, is the M^{pro}, a cysteine protease that mediates the maturation cleavage of polyproteins at the time of viral replication [6,7]. The M^{pro} is a homodimer that consists of three domains (domain I, II, and III), however, is situated in the region adjacent to domains I and II and the protomers. They play a significant role in the formation of the substrate-binding site and bind to each other through their N-terminal residues 1 through 7, which are positioned between domains II and III [8–12]. The M^{pro} offers a promising target for the development of broad-spectrum anti-coronaviral therapeutic agents due to its highly conserved three-dimensional structure [13].

All coronaviruses share the M^{pro} protein [14], which is essential for the maturation of both the M^{pro} itself and other significant polyproteins [7]. Additionally, it plays a crucial role in how viruses enter host cells; when this enzyme is inhibited, the viral entrance and subsequent infection are halted [15]. Moreover, it also plays a crucial role in the proteolytic release of enzymes necessary for viral replication, such as nsp 13, which acted as both an RNA helicase and an NTPase [16,17]. These crucial roles of the viral protease enzyme, M^{pro} make them an intriguing therapeutic target for reducing illnesses associated with coronaviruses [5,16]. M^{pro} is highly conserved among various coronaviruses and due to its relatively conserved nature, structurally tailored broad-spectrum medications substantially block the M^{pro} of coronavirus [18,19]. Similar to how SARS-CoV-2 M^{pro} has garnered considerable interest in creating medications to combat the existing COVID-19 pandemic.

Structure-based drug design strategies became widespread to explicate antiviral activity through active constituents present in traditional medicinal plants. Several natural products have been identified as adjuvant therapy for COVID-19 treatment, direct antiviral therapy [20], as well as immune enhancer [21]. Studies both *in silico* and *in vitro/in vivo* showed that numerous types of inhibitors effectively bound to and inhibited SARS-CoV-2 M^{pro} [22]. Several studies demonstrated that some potential inhibitors such as carmofour, ebselen, cinanserin, famotidine, nelfinavir, lopinavir, and ritonavir bind with SARS-CoV-2 M^{pro} and inhibit viral replication, thereby reducing viral activity through the inhibition of M^{pro} [23–26]. These inhibitors are unlikely to have any adverse side effects or toxicity because humans do not have any proteases with a similar cleavage specificity [27–29]. Using molecular dynamics simulations, the binding characteristics of the Delta and Omicron (BA.1) variants were investigated at the interface of the spike protein receptor binding domain and human angiotensin-converting enzyme-2 ectodomain and compared the molecular modeling systems with others [30].

In addition, a number of commonly used medications, such as valrubicin (antitumor), colistin (antibiotic), bepotastine (prescribed for rhinitis), icatibant (used to treat hereditary angioedema), perphenazine (antipsychotic), and caspofungin (antifungal), also bind to the M^{pro} even though they are more tolerant to mutation than lopinavir/ritonavir, suggesting potential drug candidates [26]. In another investigation, the FDA-approved antiviral medications tipranavir, lopinavir-ritonavir, and raltegravir were found to bind to the SARS-CoV-2 M^{pro} active site with robust, persistent, and versatile affinity [31]. Therefore, the replication process of the virus can be inhibited by searching for a potential inhibitor against SARS-CoV-2 M^{pro}.

Drug development activities are achieved efficiently in the burgeoning field of computational modeling. For successful binding to the SARS-CoV-2 M^{pro} active site, researchers have examined new and well-known antiviral substances such as vaniprevir, sovaprevir, boceprevir, glecaprevir, simeprevir, danoprevir, paritaprevir, and grazoprevir those firmly adhere to SARS-CoV-2 M^{pro} [32]. Medicinal plants are renowned as complementary approaches in which antiviral natural products have been established as safe and effective drugs [33]. Furthermore, among other things, the best medicine can be predicted with the help of computational drug design tools. Even drug development depends on safety measures regarding toxicity in the human body through a series of clinical trials, however, computer-based toxicity detection is also facilitated as it can be done before development. Therefore, a number of studies focused on the *in-silico* design of effective medications targeting SARS-CoV-2 M^{pro} is continuously rising. However, given the potential drawbacks of passing clinical studies, the therapeutic efficacy of these desired medications is debatable. In this study, computer-aided drug design is used for its more time and cost-efficient way of developing novel target drug candidates for inhibiting SARS-CoV-2 M^{pro} activity, in turn, the severity and mortality due to SARS-CoV-2 will decrease.

2. Materials and methods

2.1. Retrieval and protein preparation

The three-dimensional crystal structure of SARS-CoV-2 M^{pro} with N3 inhibitor (PDB ID: 6LU7) was retrieved in pdb format from the Protein Data Bank. Then the protein structure was prepared by removing all heteroatoms and water molecules using PyMol and Discovery Studio client. The clean protein structure was prepared by optimizing important factors like missing hydrogen bonds, side-chain geometry, and improper bond order were energy minimized using the GROMOS 43B1 force field [34] in Swiss-PDB Viewer.

2.2. Preparation of ligand

Initially, 104 Korean medicinal plants and their 2431 phytochemicals, identified by GC-MS, were listed (Supplementary Table 01) through a rigorous literature review. Then the compounds were downloaded in 3D SDF format from the PubChem Database. Energy minimization of the ligand was utilized using the mmff94 force field and the steepest descent optimization algorithm steps 2000 in PyRx software before performing molecular docking analysis.

2.3. Molecular docking study

Docking is a technique used to predict the binding interaction of a drug with its target molecule. The AutoDock Vina virtual screening tool was used to perform molecular docking analysis to determine the specific residues in the target protein that interact with the particular ligands [35]. For docking analysis, AutoDock vina was used to convert pdb into a pdbqt format to input protein and ligands. Re-docking is the process of removing the ligand molecule from the receptor model. Therefore, to verify the docking studies, 6LU7 was re-docked with its crystallographic inhibitor N3 and compared the differences between the re-docking pose and conformation of the ligand in the crystal structure. This protein structure has several cavity sites and inhibitor N3 bound to at site 1 (X: 11.90, Y: 20.33, Z: 58.94) to the main protease. Consequently, site 1 was selected as an inhibition site. In this study, docking was used to understand the binding of ligands to the active sites of the M^{pro} protein. To dock the compounds against 6LU7, the center of the grid box was kept at X: 26.294, Y: 12.6009, Z: 58.9473, and dimensions were kept at X: 51.3565, Y: 66.9335, and Z: 59.6050 Å. The binding affinity of the ligand was by kcal/mol as a unit for negative binding score [36]. The compounds' docking results were calculated and ranked based on their highest negative values, representing the highest binding

affinities [35,37]. The interactions of drug molecules with the M^{pro} active site residues were confirmed by examining the docking and re-docking results at each docking position.

2.4. Re-screening using deep learning

The compounds with above −8.0 kcal/mol binding affinity were selected for re-evaluation with deep learning method using Recurrent Neural Networks (RNN) to predict models using the CHEMBL3927 dataset consisting of inhibitors for M^{pro} of SARS-CoV-1. The DeepScreeing server was used for this process and the server shows the average and median area under the ROC curve value of 0.86, and 0.89 under several constructed deep-learning models indicating good performance in screening [38]. PubChem fingerprint was used by PaDEL to generate 881 molecular fingerprints [39]. The hyperparameters for this developed model were manually set like learning rate (0.0001), number of neurons (2000,700,200), and hidden layer (3). The activation function and sigmoid function were used in hidden and output layers, respectively. The predicted pIC50 score was calculated using this equation.

$$pIC_{50} = -\log_{10}IC_{50}\ (M) = 9 - \log_{10}IC_{50}\ (nM)$$

This model was employed to re-screen the top docked 12 compounds and presented in Table 1.

2.5. ADME/T and bioactivity analysis

The ADME/T were predicted using SwissADME, pKCSM, and ProTox-II server using SMILES (simplified molecular-input line-entry system) specification as an input syntax of the compounds. Pharmacokinetic and pharmacodynamic properties related to the metabolism, absorption, distribution, excretion, and toxicity of the compounds were evaluated. Molinspiration server was used to evaluate the biological activities of the compounds. Here, bioactivity score >0 is regarded as a biologically active compound, −5.0 to <0 as a moderately active compound, and <-5.0 as a biologically inactive compound.

2.6. Molecular dynamics simulation study

Molecular Dynamics (MD) simulation helps to determine and validate the structural flexibility and entropic effects of protein-ligand complexes in an artificial environment. Prior to performing the production simulations, energy minimization and equilibration simulations were conducted to remove bad contacts between the solute and solvent water molecules [40]. The solute and lipid chains were subjected to a harmonic restraint in the first step of energy minimization, and the second step was to allow all atoms to move freely [41]. The YASARA dynamics software was used in a molecular dynamics simulation to validate the predictions from the docking study [42]. In this simulation study, the AMBER14 force field [43] was utilized, which is widely used in macromolecular systems. The M^{pro} protein was used as a control system to compare the structural integrity against the complexes. The particle mesh Ewald method [44] was used to calculate long-range electrostatic interactions, and 8 Å cutoff radius was used to analyze short-range van der Waals and Coulomb interactions [45]. The total environmental condition of the system was set to a temperature of 298 K, pH 7.4, and 0.9% NaCl. The steepest descent algorithm was used to minimize the system [35]. A Monte Carlo barostat and a Langevin thermostat were used to maintain the pressure and temperature, respectively [46]. The MD simulation trajectories were saved after every 100 ps using a time step of 1.25 fs [47]. To ensure the simulation convergence, multiple parallel trajectories were conducted for each complex [40]. The MD simulation study was conducted for 100 ns to analyze the root mean square deviation (RMSD), the root mean square fluctuation (RMSF), the radius of gyration (RG), the solvent accessible surface area (SASA), and the number of hydrogen bonds [45,48–50]. The MD plots were visualized by Matplotlib.

3. Results and discussion

Drug discovery against viruses facilitated the successful identification of potential inhibitors by rational screening and understanding the bio-physio-chemical basis of inhibitor binding mechanism through molecular dynamics simulations as well as a series of conformational and toxicity analyses [51]. Molecular docking has been widely used to predict the bioactive compounds from plants to develop drugs against target proteins that cause infection by viruses. M^{pro} is the important therapeutic target of a virus as it cleaves the polyprotein of the virus that is crucial in viral transcription and replication that maintains the viral life cycle [23,52,53]. Therefore, this study aimed at the molecular docking and MD simulation approaches using Korean medicinal plants to predict their inhibitory role against virus infection. The target protein of this study has well-defined active sites that help in designing novel drugs against coronaviruses [54]. The binding propensity of the compounds against M^{pro} was calculated and ranked based on the AutoDock Vina docking score. Lesser binding affinity was considered to be a good docking score for screening the compound library. The lowest affinity from several binding poses of each compound was selected because it indicates the lowest energy release for bond formation [55].

At first, the native inhibitor N3 of COVID-19 M^{pro} has been docked in order to legitimatize the virtual molecular docking accuracy and then re-docked. The re-docking of inhibitor N3 exhibited a binding affinity of −8.1 kcal/mol against the inhibition site of the target protein (Table 1). Further, validated the re-docking results of their conformational superimposition of ligand binding modes in the target protein (Fig. 1). The RMSD value between the re-docked and co-crystal native ligand: pose of inhibitor N3 was 0.037 Å which was within 2 Å indicating effective docking protocol. This conformational superimposition of ligands RMSD value represented good reproduction of the correct pose mostly with identical inhibition sites and residues. The inhibitors of the molecules in the active pocket of M^{pro} of the protein and their interaction type are shown in Fig. 2.

After docking of the phytochemicals of Korean medicinal plants, their lower binding affinity than inhibitor N3 was analyzed. Based on the lowest affinity, the top 12 compounds, i.e. Catechin gallate, Cynaroside, Cosmosiin, Isoquercitrin, Rutin, Hyperoside, Isochlorogenic acid b, Quercetin 3-O-malonylglucoside, Cacticin, Narcissoside, Guaijaverin, Luteolin-7-O-Rutinoside were selected which showed negative binding affinity (kcal/mol) of −8.5, −8.2, 8.0, −8.6, −8.9, −8.6, −8.2, −8.2, −8.4, −8.7, −8.5, −8.5, respectively (Table 1).

Other molecular properties and 2D structure of the predicted compounds are presented in Table 1. Further, catechin gallate and quercetin 3-O-malonylglucoside had been indented as top potential candidates those had binding affinities of −8.5 and −8.2, kcal/mol (Fig. 3). Moreover, their other physio-chemical properties, solubility, drug-likeness, etc. have been demonstrated in Table 2, and their non-bonded interactions within the active and catalytic sites of the M^{pro} have been illustrated in Fig. 4. Re-docking performed the N3 inhibitor against 6LU7, which was used as a positive reference molecule. The top two lead compounds exhibited notable docking conformations with a binding affinity of more than −8.1 kcal/mol (binding affinity of N3) in the active site of a target protein. The protein-ligand interaction profile and binding affinity between the top two compounds and the inhibitor N3 in the binding site of M^{pro}, and the comparative docking score are presented in Table 1. Several natural compounds like theaflavin, quercetin, withanone, dihydrowithaferin, and caffeic acid phenethyl ester have been reported as potential inhibitors of SARS-CoV-2 [51,56,57].

The Deep learning method uses various artificial neural network frameworks on the dataset to predict molecular properties and bioactivities from large libraries unitizing different chemical, topological, and biological data patterns of compounds [58–60]. It was performed to

Table 1
Binding affinity, Predicted IC_{50} score and non-bond interactions between SARS-CoV-2 main protease and the top twelve compounds.

Compound name and CID	2D Structure	Predicted pIC_{50} score	Binding affinity (kcal/mol)	Residue in contact	Interaction type	Bond distance in Å
Catechin gallate 6419835	fx1	4.7227	−8.5	THR190	Hydrogen Bond	2.19304
				ARG188	Hydrogen Bond	2.97179
				HIS163	Hydrogen Bond	1.68962
				SER144	Hydrogen Bond	2.21275
				PRO168	Carbon Hydrogen Bond	2.67908
				CYS145	Pi-Sulfur	5.40505
				MET165	Pi-Sulfur	5.40764
				HIS41	Pi-Pi Stacked	4.73238
				MET49	Pi-Alkyl	5.26499
Quercetin 3-O-malonylglucoside 5282159	fx2	4.7172	−8.2	ASN238	Hydrogen Bond	1.88031
				ASP197	Hydrogen Bond	2.75917
				LYS137	Hydrogen Bond	2.49689
				THR199	Hydrogen Bond	2.72559
				TYR239	Hydrogen Bond	2.15711
				LEU272	Carbon Hydrogen Bond	2.6859
				TYR237	Pi-Pi T-shaped	5.05381
Rutin 5280805	fx3	4.6303	−8.9	GLY143	Hydrogen Bond	2.25632
				GLU166	Hydrogen Bond	2.31189
				PHE140	Hydrogen Bond	2.90128
				ARG188	Hydrogen Bond	2.12475
				THR26	Carbon Hydrogen Bond	1.84333
				GLN189	Carbon Hydrogen Bond	2.90367
				ASN142	Carbon Hydrogen Bond	2.74658
				CYS145	Pi-Sulfur	4.9642
				MET165	Pi-Sulfur	5.16804
Isoquercitin 10813969	fx4	4.6179	−8.6	ARG188	Hydrogen Bond	2.40792
				THR190	Hydrogen Bond	2.12519
				HIS163	Hydrogen Bond	2.05993
				GLN192	Hydrogen Bond	2.42437
				GLU166	Carbon Hydrogen Bond	2.22159
				PRO168	Carbon Hydrogen Bond	2.88806
				CYS145	Pi-Sulfur	5.81078
				MET165	Pi-Sulfur	5.57255
				HIS41	Pi-Pi Stacked	4.9543
				MET49	Pi-Alkyl	4.94603
Hyperoside 5281643	fx5	4.6179	−8.6	SER144	Hydrogen Bond	2.84073
				LEU141	Hydrogen Bond	2.56989
				ARG188	Carbon Hydrogen Bond	2.83337
				CYS145	Pi-Sulfur	5.06909
				HIS41	Pi-Pi T-shaped	4.59608
				MET49	Pi-Alkyl	5.04361
Quercetin-3-D-Xyloside 5320863	fx6	4.6179	−8.5	LYS137	Hydrogen Bond	2.73634
				ASN238	Hydrogen Bond	2.30343
				GLU290	Hydrogen Bond	2.36161
				THR199	Carbon Hydrogen Bond	2.24907
				ASP289	Pi-Anion	3.65208
				LEU286	Pi-Alkyl	4.87541
Cacticin 5318644	fx7	4.6171	−8.4	THR190	Hydrogen Bond	2.73319
				ARG188	Hydrogen Bond	2.42079
				MET165	Hydrogen Bond	2.30981
				GLN189	Carbon Hydrogen Bond	2.62132
				GLU166	Carbon Hydrogen Bond	2.4619
				THR190	Pi-Donor Hydrogen Bond	2.7937
				MET49	Alkyl	4.15532
				HIS41	Pi-Alkyl	3.72376
Narcissoside 5481663	fx8	4.595	−8.7	GLY143	Hydrogen Bond	2.25307
				GLU166	Hydrogen Bond	2.35394
				LEU141	Hydrogen Bond	2.24284
				THR190	Hydrogen Bond	2.27931
				THR26	Hydrogen Bond	2.31897
				HIS41	Hydrogen Bond	2.82795
				GLN189	Carbon Hydrogen Bond	2.76494
				PHE140	Carbon Hydrogen Bond	2.5103
				THR26	Carbon Hydrogen Bond	3.09029
				CYS145	Pi-Sulfur	4.84568
				MET165	Pi-Sulfur	5.1467
3,4-Dicaffeoylquinic acid5281780	fx9	4.5823	−8.2	LEU271	Hydrogen Bond	2.30224
				ALA285	Hydrogen Bond	2.78731
				LEU272	Hydrogen Bond	2.31922
				ASN238	Hydrogen Bond	2.75446
				THR199	Hydrogen Bond	2.11357
				ASP197	Hydrogen Bond	2.91729

(*continued on next page*)

Table 1 (*continued*)

Compound name and CID	2D Structure	Predicted pIC_{50} score	Binding affinity (kcal/mol)	Residue in contact	Interaction type	Bond distance in Å
				ARG131	Hydrogen Bond	2.35626
				LYS137	Hydrogen Bond	2.12948
				MET276	Pi-Alkyl	5.48501
				LEU287	Pi-Alkyl	4.78701
luteolin-7-O-rutinoside 14032966	fx10	4.4184	−8.5	THR199	Hydrogen Bond	2.18848
				ASN133	Hydrogen Bond	2.8903
				ASP197	Hydrogen Bond	2.38578
				TYR237	Carbon Hydrogen Bond	2.50399
				ASP197	Carbon Hydrogen Bond	2.45529
				LEU272	Pi-Alkyl	5.11035
				LEU287	Pi-Alkyl	5.38559
Cynaroside 5280637	fx11	4.4176	−8.2	MET165	Hydrogen Bond	2.7413
				ARG188	Hydrogen Bond	2.68482
				GLU166	Hydrogen Bond	2.39233
				HIS163	Hydrogen Bond	3.08162
				THR26	Hydrogen Bond	2.57113
				SER144	Carbon Hydrogen Bond	2.4719
				CYS145	Pi-Sulfur	4.70184
Cosmosiin 5280704	fx12	4.3704	−8.0	THR199	Hydrogen Bond	2.42785
				LEU287	Hydrogen Bond	2.14488
				THR199	Hydrogen Bond	2.65619
				LEU286	Carbon Hydrogen Bond	3.09014
				TYR237	Pi-Pi T-shaped	4.72839
Inhibitor N3 146025593	fx13	4.690	−8.1	GLN189	Carbon Hydrogen Bond	2.48559
				GLU166	Hydrogen Bond	1.93628
				GLY143	Hydrogen Bond	1.86317
				HIS172	Carbon Hydrogen Bond	2.39426
				HIS41	Pi-Alkyl Bond	4.3148
				MET165	Carbon Hydrogen Bond	2.72823
				MET49	Alkyl	4.66427
				ALA191	Pi-Alkyl	4.53889
				PRO168	Pi-Alkyl	4.84869
				LEU167	Alkyl	5.46094
				THR190	Hydrogen Bond	1.79854
				HIS164	Hydrogen Bond	2.95502
				PHE140	Hydrogen Bond	2.4722
				ASN142	Carbon Hydrogen Bond	2.8316

CID; PubChem CID, pIC_{50} score in nM.

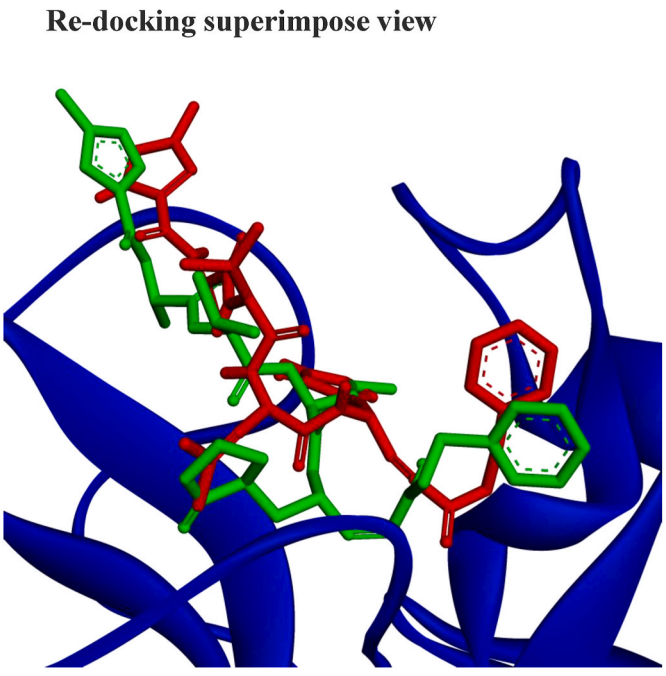

Fig. 1. Three-dimensional view of the best re-docking pose and the conformational superposition of ligands (inhibitor N3) and the RMSD value of 0.037 Å.

screen selected compounds based on inhibitory activities against the target protein. Compounds with inhibitory properties against M^{pro} of SARS-CoV-1 were used for building many regression models based on IC_{50}. Evaluation matrices, i.e., coefficient of determination (R^2), mean square error (MSE), root mean square error (RMSE), and mean absolute error (MAE) were measured to evaluate the performance and quality of the selected model. These parameters represent the fitness of the regression model and the errors between predicted and actual data [61,62]. The values of R^2, MSE, RMSE, and MAE were 0.81, 0.26, 0.51, and 0.41, respectively (Fig. 5). High R^2 value showed that the model had higher fitness and a good performance against the test data. Small MSE, RMSE, and MAE values indicated the higher accuracy of the model [63]. Selected compounds were screened in the regression model against the biologically active inhibitors for M^{pro} of SARS-CoV-1. Both catechin gallate and quercetin 3-O-malonylglucoside showed pIC_{50} values of 4.723 and 4.717, respectively, demonstrating that these two compounds were biologically active and able to inhibit the protein.

It is reported that M^{pro} of SARS-CoV-2 has three domains consisting of residues 8–101, 102–184, and 201–303 [51,64]. Among these residues, Cys145-His41 acts as a catalytic dyad between domains I and II, and is responsible for protease activity and substrate binding [65–68]. In domain I, residue Phe140, His172, Asn142, His163, Glu166, and Leu141 are involved in substrate binding. Residue Asp187, Met165, Tyr54, Met49, and His41 in domain II and residue Gln192, Met165, Gln189, Phe185, and Leu167 in domain III are also important for substrate binding [69,70]. In the activity of the protein, the large number of hydrogen bonds between ligands and residues signifies a strong binding affinity [49]. Post-dock analysis showed Catechin gallate had a better binding affinity with the protein because of having more electrostatic

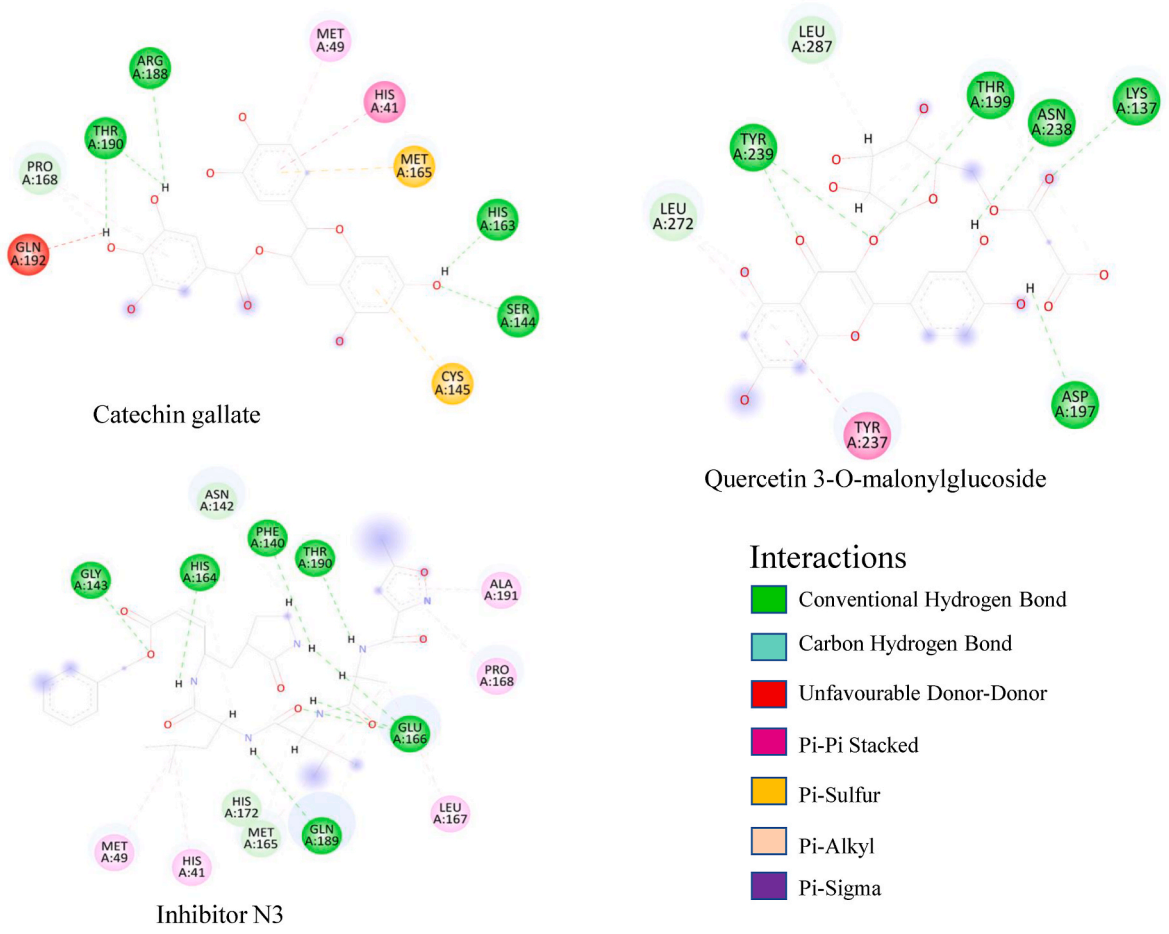

Fig. 2. Molecular docking interaction and two-dimensional (2D) binding mode of selected ligands in SARS-CoV-2 main protease active site (PDB ID: 6LU7).

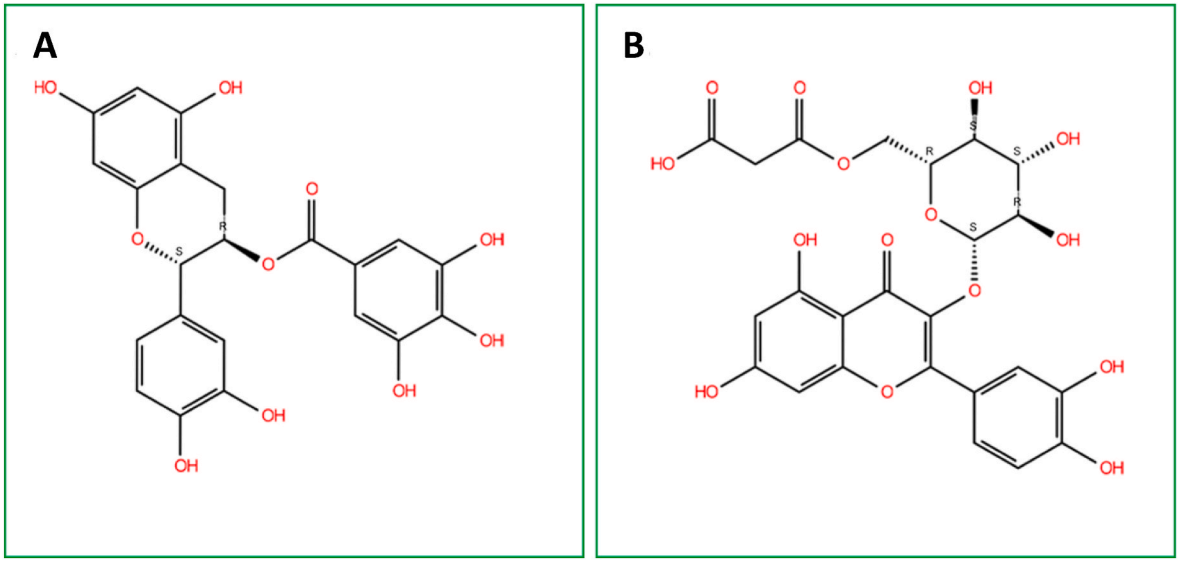

Fig. 3. 2D chemical structure of (A) Catechin gallate and (B) Quercetin 3-O-malonylglucoside. The structures were drawn using the 2D sketcher (Beta) software.

force than the Quercetin 3-O-malonylglucoside. Catechin gallate had no interaction with domain III of the protein. It formed four conventional hydrogen bonds with residue Thr190, Arg188, His163, and Ser144 residue creating strong affinity on the domain I and II. Two non-covenant pi-sulfur bonds were formed with Cys145 and Met165 residue. Pi-Pi Stacked bond with residue His41 interacting with both aromatic rings and pi-alkyl bond with residue Met49 was also observed, that created a pi-electron cloud over the alkyl group. Carbon-hydrogen bonds with Pro168 residue were also formed, which gave this compound higher interaction affinity with the protein (Fig. 4). The residues Gln189 and Leu455 have already been found to be potent drug binding sites [51, 71]. Catechin gallate complex interactions also match with another

Table 2
Pharmacological profiles of the top two potential candidates derived from the SwissADME and pKCSM webservers.

Properties		Catechin gallate	Quercetin 3-O-malonylglucoside
Physico-chemical Properties	MW (g/mol)	442.37 g/mol	550.42 g/mol
	Heavy atoms	32	39
	Aro. Atoms	18	16
	Rotatable bonds	4	8
	H-bond acceptors	10	15
	H-bond donors	7	8
	TPSA (Å^2)	177.14	253.88
Lipophilicity	Log $P_{o/w}$ (Cons)	1.25	−0.67
Water Solubility	Log S (ESOL)	Soluble	Soluble
Pharmacokinetics	GI absorption	Low	Low
	BBB permeant	No	No
	P-GP substrate	No	Yes
Drug likeness	Lipinski	Yes	No
Medi. Chemistry	Synth. accessibility	4.16	5.63

research by Mousavi, 2021 [72]. Thus, Catechin gallate formed chemical bonds with Cys145-His41 catalytic dyad residues and interacted with Met165 and His163, which play an important role in substrate binding.

Polyphenol flavonoids show activity against various human viruses like herpes simplex, hepatitis B, HIV, hepatitis C, influenza, dengue, chikungunya, and zika. Studies were conducted on *Camellia sinensis* (tea) which demonstrated that catechins and gallate derivatives including epicatechin 3-O-caffeoate, etc-pyrrolidinone C, etc-pyrrolidinone D, epigallocatechin gallate, epicatechin gallate, and gallocatechin-3-gallate have inhibition properties against M^{pro} in intracellular and extracellular inhibition assays. Catechin gallate also had stable complex formation with the viral protein due to hydrogen bonds and non-covalent interactions [64,73–75]. Various studies showed that Catechin gallate has better binding energy compared to Michael acceptor inhibitors known as N3, which have a half-maximal inhibitory concentration (IC_{50}) value of 125 μM against SARS-CoV-2 M^{pro}. Its derivatives, like gallo catechin gallate with two carboxy group replacements, also have the potential to decrease infection of SARS-CoV-2 omicron mutant [76].

Similarly, Quercetin 3-O-malonylglucoside formed five conventional hydrogen bonds with residue Asn238, Asp197, Lys137, Thr199, and Tyr239. Leu272 residue formed a carbon-hydrogen bond and Tyr237 formed a pi-pi t-shaped bond with the compound. The majority of chemical interactions were found with domain III of M^{pro} and no bond was formed with the first domain that consists of residue 8–101. Lys137 was the only residue that formed a hydrogen bond with the compound. Quercetin 3-O-malonylglucoside bind at Asp197 within a loop among the domains. This analysis suggests that both compounds have strong and stable binding with M^{pro} because of the multiple numbers of bonds that were formed during docking and Catechin gallate formed bonds with catalytic and binding residues of the target protein.

Molecular dynamics (MD) simulation is vital for post-dock analysis to investigate time-dependent stability and atom movements of biological compounds [77]. MD simulation was performed for 100 ns to understand the structural behavior, binding mechanism, and structural flexibility of M^{pro} [77,78], M^{pro}-Catechin gallate complex, and M^{pro}--Quercetin 3-O-malonylglucoside complex. The comparison between initial and final structural behavior of the complexes has been demonstrated in Fig. 6. The parameters computed and analyzed after 100 ns

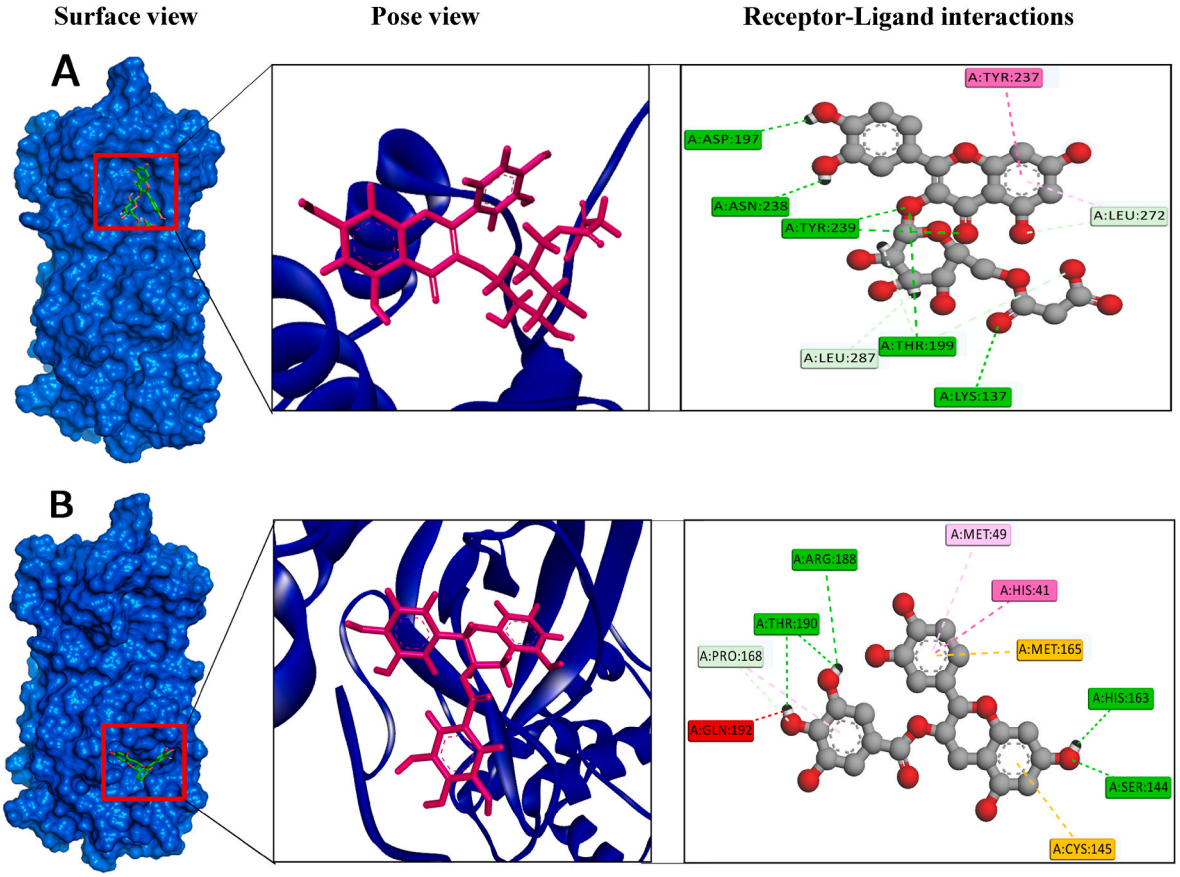

Fig. 4. Binding modes for the selected compounds. The figures illustrate non-bonded interactions of the docked complexes for the top two compounds within the active and catalytic sites of the main protease. (A) Quercetin 3-O-malonylglucoside, and (B) Catechin gallate.

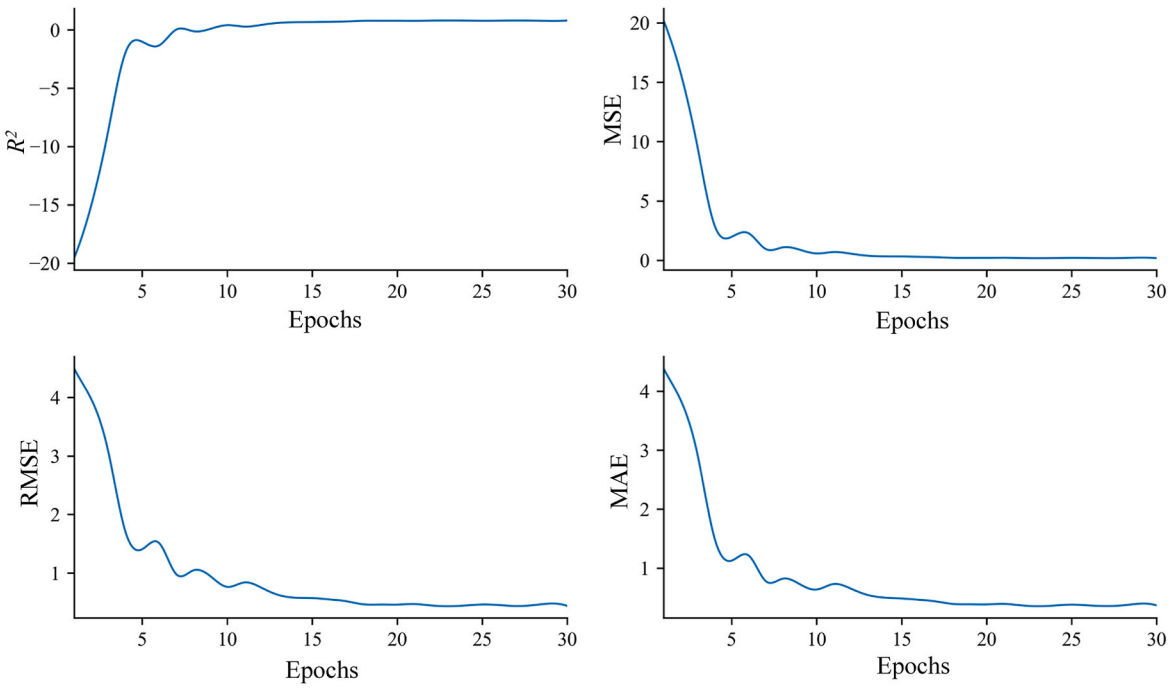

Fig. 5. Regression model accuracy plots showing various statistical parameters of the deep learning model. Upper left plot showed R^2 value of 0.81 at epoch 30. Upper right plot showed MSE value of 0.26 at epoch 30. Lower left plot denoted RMSE value of 0.51 at epoch 30. Lower right plot showed MAE value of 0.41 at epoch 30.

dynamics trajectory were root-mean-square deviation (RMSD), root mean square fluctuation (RMSF), radius of gyration (RG), solvent accessible surface area (SASA) (Fig. 7) as well as hydrogen bonding (Fig. 8), those were measured as because such stabilities are essential to obtain good binding affinities [79]. RMSDs of Cα atoms of these compounds were performed and it was observed that M^{pro} had a minor fluctuation from 45 ns to 67 ns but afterward it remained at an equilibrium state. As for M^{pro}-Catechin gallate complex, 8 ns–35 ns showed a high fluctuation, after that it also reached a stable position. M^{pro}-Quercetin 3-O-malonylglucoside complex showed stability at first but after 45 ns it had several fluctuations. Nevertheless 92 ns afterward it again showed stability. These two complexes showed stability regarding the reference protein at the end.

To better understand the structural compactness, SASA was examined [78]. M^{pro}-Catechin gallate complex showed two higher trajectories at the beginning of the simulation compared to the M^{pro}. However, both showed relatively similar trajectories near the end demonstrating that these complexes had slightly lower compactness. In contrast,

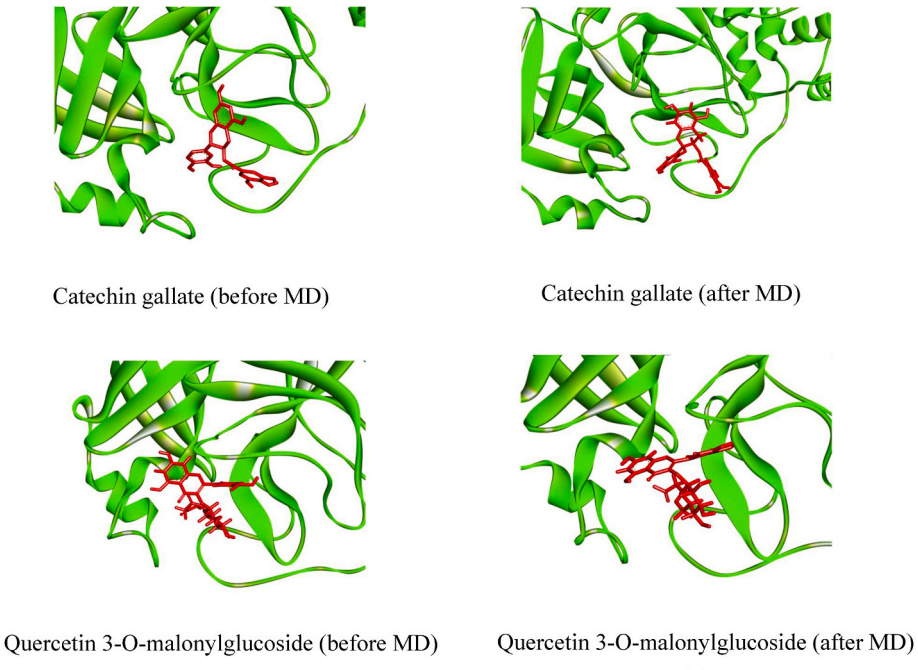

Fig. 6. Comparing between initial and final structures of the complexes acquired from MD simulations.

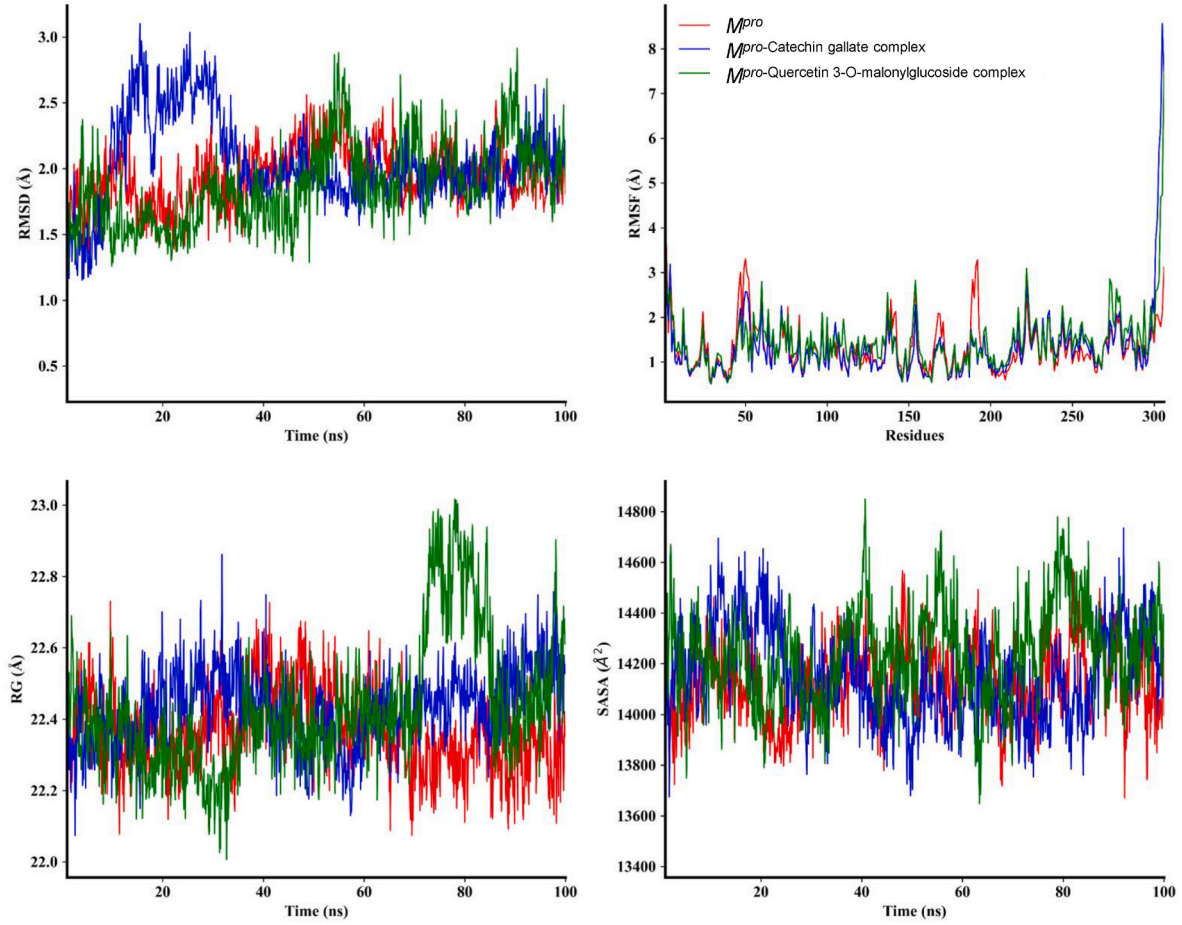

Fig. 7. Analysis of RMSD, RMSF, RG, and SASA of unligated M^{pro} and M^{pro}- Catechin gallate/Quercetin 3-O-malonylglucoside complex. The MD simulations for each system were performed for 100 ns. These MD trajectories were analyzed with the aid of RMSD, RMSF, RG and SASA. RMSD analysis for the alpha carbon atoms, RMSF for the flexibility analysis of amino acids residues, RG for the degree of rigidity and compactness analysis, and SASA for the protein volume with expansion analysis.

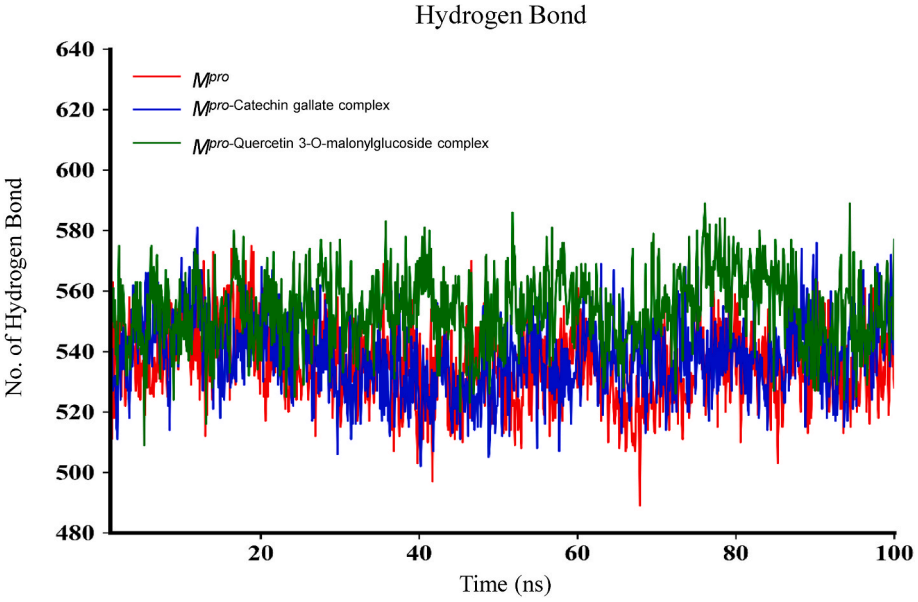

Fig. 8. Determination of Hydrogen bond analysis of unligated M^{pro} and M^{pro}- Catechin gallate/Quercetin 3-O-malonylglucoside complex. The MD simulation for the system was performed for 100 ns.

M^{pro}-Quercetin 3-O-malonylglucoside complex showed higher trajectories after 7 ns for 10 ns (Fig. 7). Despite of this the latter complex showed high RG values between 72 ns and 85 ns, both of them showed comparatively alike plot indicating these complexes somewhat had loosed protein structure packaging [80] that supports SASA result analysis. The reference protein M^{pro} and the other two complexes showed higher fluctuation in RMSF analysis over a similar range of residues, e.g. 45–56, 132–143, 222–230, and 272–279, which strongly suggest homogeneous protein residue flexibility [81]. The hydrogen bonds between intermolecular binding demonstrated that the M^{pro} only and M^{pro}-Catechin gallate had similar hydrogen bond numbers between intra-molecules but M^{pro}-Quercetin 3-O-malonylglucoside complex showed a higher number of hydrogen bindings than these two molecules donating its strong stability (Fig. 8). Analysis from these various results indicates that despite having minor conformational changes due to MD simulation, these two complexes have similar stability like M^{pro} reference protein. The superimposition of the structures between pre- and post-molecular dynamics simulation of the structures of Catechin gallate and Quercetin 3-O-malonylglucoside also was performed to determine changes in the binding cavity (Fig. 9). After the post-dock analysis showed their structural change in every 25 ns, surface view of M^{pro}--Catechin gallate and M^{pro}-Quercetin 3-O-malonylglucoside complexes in MD simulation. The snapshot was taken from 25, 50, 75, and 100 ns, respectively, shown in Figs. 10 and 11.

Biologically active compounds act on proteins, subsequently causing physiological effects directly. The potential biological activities of the compounds were assessed by investigating G protein-coupled receptor (GPCR) ligand activity, ion channel inhibitor, enzyme inhibition, protease inhibition, kinase inhibition, and nuclear receptor ligand activity (Table 3). Catechin gallate exhibited positive biological activity against all of these biological mechanisms, whereas Quercetin 3-O-malonylglucoside showed moderate activity in ion channel, kinase and protease inhibition. Quercetin 3-O-malonylglucoside showed a better result as an enzyme inhibitor, but Catechin gallate outperformed all other biological receptors.

Safety issues, including toxicities and adverse drug effects, are always an essential prerequisite that should be evaluated during drug development. Even though the clinical trial is obligatory, computational predictive models built for various toxicities also facilitated the progress of drug design [82]. Computational methods to investigate toxicities and adverse drug effects have shown significant advantages since they are accurate and can be done before a compound is synthesized [83]. Toxicity assessment and pharmacokinetic characteristics are required to assess the effectiveness and level of indemnification of the lead compounds [84]. These two lead compounds were further analyzed with the pharmacological activity that is an essential indicator for disposition within the human body with ADME/T technology whose intensive application has been reviewed recently [82]. Furthermore, Catechin gallate and Quercetin 3-O-malonylglucoside were evaluated by examining Central Nervous System (CNS) permeability, hepatotoxicity, p-glycoprotein inhibition, and Cytochrome P450 (CYP) inhibition, among several other aspects. CNS permeability refers to the capacity of a drug to penetrate the semipermeable blood-brain barrier, which is meant to protect the CNS from potentially harmful compounds. Permeability of CNS higher than −4 is thought to reflect the ability to pass across the blood-brain barrier. The molecular weights of Quercetin 3-O-malonylglucoside and Catechin gallate were 550.42 and 442.37 g/mol, respectively. None showed inhibitory properties against cytochromes p450 (CYPs) thus, these compounds would be metabolized without having increased side effect risks or possible toxicity [85]. Catechin gallate is in agreement with Lipinski and Ghoses' [86] rule of drug-likeness, however Quercetin 3-O-malonylglucoside fails in their filter. Consequently, Catechin gallate showed better results in physiochemical properties, water solubility, lipophilicity, pharmacokinetics, and toxicity parameters concerning drug-likeness compared to Quercetin 3-O-malonylglucoside.

Moreover, the majority of the potential drug candidates fail clinical trials due to the toxicity they produce which makes it an important filter for the drug design process [87]. Various toxicity properties e.g., hepatotoxicity, carcinogenicity, immunotoxicity, mutagenicity, cytotoxicity, oral rat acute toxicity, LD_{50} value, and toxicity class of the selected two compounds have been assessed (Table 4). These machine learning methods have been applied to construct classifiers and regression models to predict LD_{50} or their toxic categories [88,89]. Both of the compounds were inactive in organ toxicity endpoints. Predicted LD_{50} of Quercetin 3-O-malonylglucoside was 5000 mg kg^{-1}, and for Catechin gallate, it was 1000 mg kg^{-1} making them nearly non-toxic to biological systems. Catechin gallate exhibited comparatively better favorable pharmacokinetics, biological activities, and toxicity properties than Quercetin 3-O-malonylglucoside, which indicates it has a minimal likelihood to fail during clinical trials.

This study provides feature learning and artificial neural network-

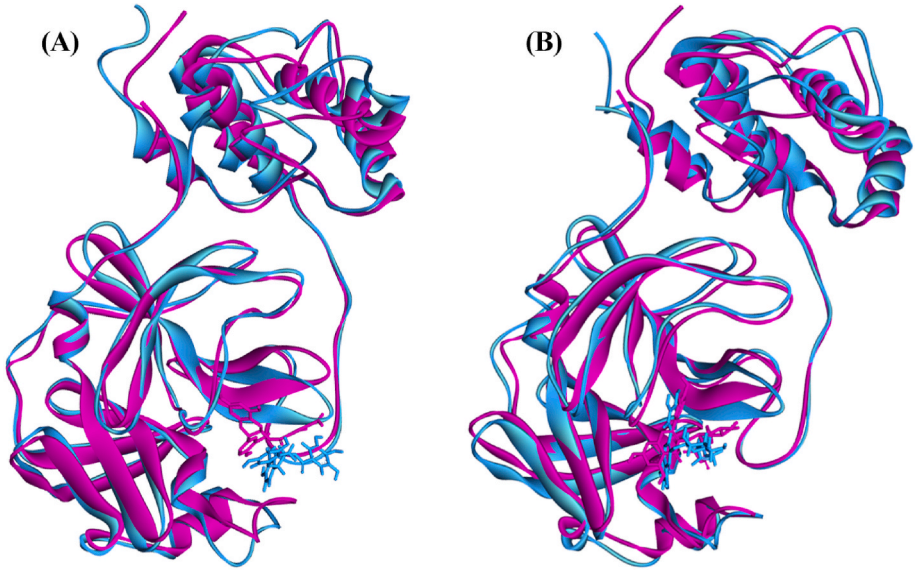

Fig. 9. The superimposed between pre- and post-molecular dynamics simulation structures of (A) Catechin gallate and (B) Quercetin 3-O-malonylglucoside. The pink color denotes the pre-molecular dynamics structure, and the sky color denotes the post-molecular dynamics structure.

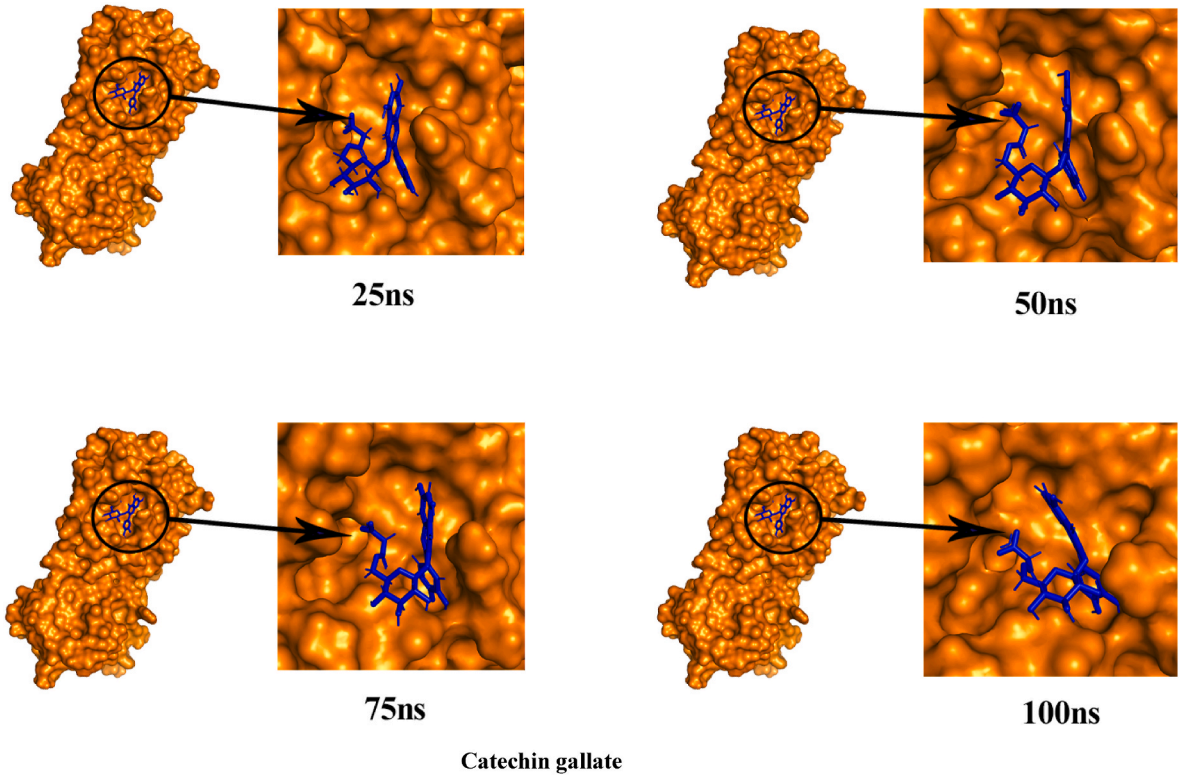

Fig. 10. The surface view of Catechin gallate and M^{pro} complex in MD simulation. The snapshot was taken from 25, 50, 75, and 100 ns, respectively.

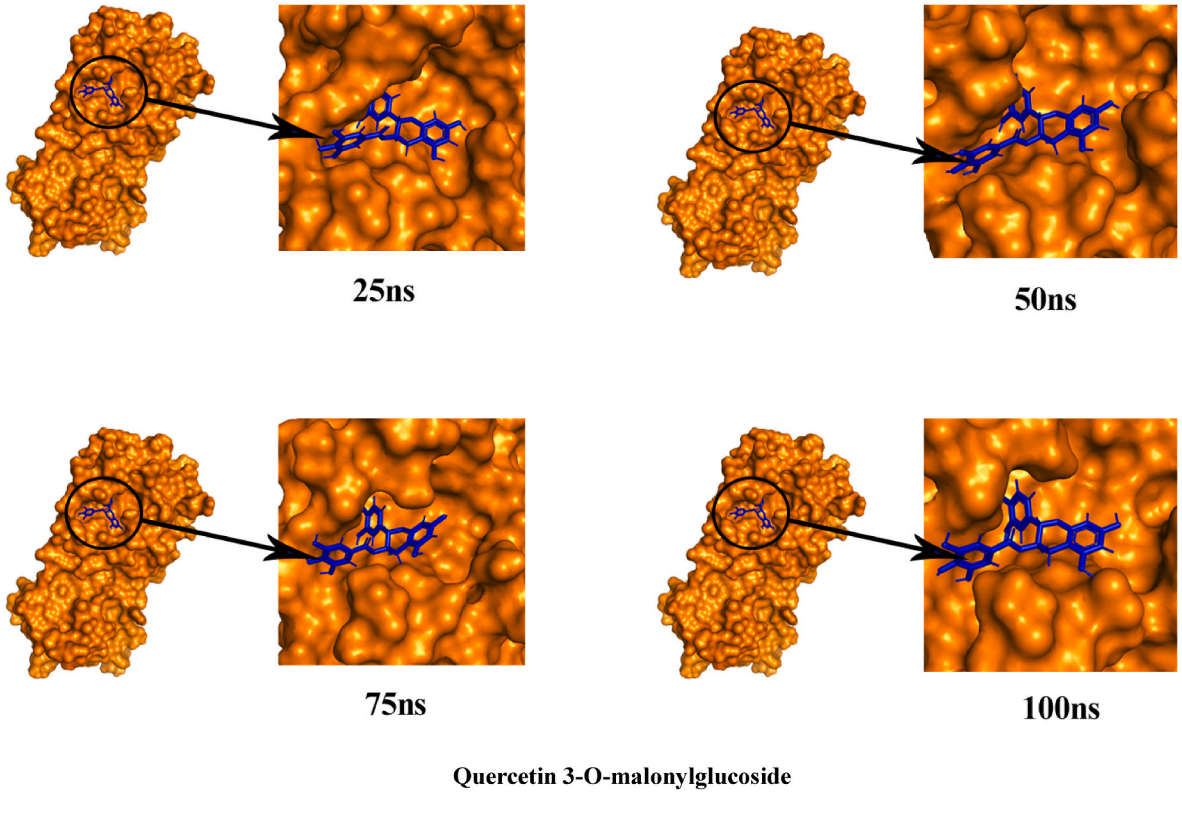

Fig. 11. The surface view of Quercetin 3-O-malonylglucoside and M^{pro} complex in MD simulation. The snapshot was taken from 25, 50, 75, and 100 ns, respectively.

based deep structured learning using artificial intelligence that supports the previous studies on SARS-CoV-2 M^{pro} and Catechin gallate. Cheminformatics-based bioactivity prediction that employed a quantitative structure–activity relationship model on main human enzymes also holds up these findings and data. In our investigation and analysis, we predicted a potential inhibitor, Catechin gallate, against

Table 3
Biological activities of the screened hit phytochemicals.

Compounds	GPCR ligand	Ion channel inhibitor	Kinase inhibitor	Nuclear receptor ligand	Protease inhibitor	Enzyme inhibitor
Catechin gallate	0.17	0.02	0.05	0.34	0.13	0.25
Quercetin 3-O-malonylglucoside	0.04	−0.21	−0.02	0.14	−0.03	0.35

Activities were calculated from the Molinospiration chemoinformatics software package. Here bioactivity score >0.0; biologically active, between −5.0 – 0.0; moderately active, and <-5.0; biologically inactive.

Table 4
Various toxicity properties of the selected two compounds.

Endpoint	Target	Catechin gallate	Quercetin 3-O-malonylglucoside
Organ toxicity and toxicity endpoints	Hepatotoxicity	Inactive	Inactive
	Carcinogenicity	Inactive	Inactive
	Immunotoxicity	Inactive	Inactive
	Mutagenicity	Inactive	Inactive
	Cytotoxicity	Inactive	Inactive
	LD_{50} (mg kg^{-1})	1000	5000
	Toxicity class	4	5

M^{pro} from a large phytochemical library through molecular docking screening and molecular dynamics simulation. We investigated their pharmaceutical profile as a possible drug candidate against SARS-CoV-2. Our findings might help to understand the potential drug design based on their inhibitory activity of a target protein.

4. Conclusion

This study revealed that collected phytochemicals from Korean traditional medicinal plants bind with M^{pro} with different poses. Catechin gallate and quercetin 3-O-malonylglucoside were selected through molecular docking and deep learning method based on binding affinity and inhibitory potency. They were further subjected to post-dock analysis that showed these compounds were stable forming various electrostatic forces like H-bond, pi-bond, and C–H bond, while catechin gallate formed important bonds with catalytic residues. MD simulation method also validated the stability, compactness, flexibility, binding energy, and residue fluctuations of the selected compounds. These were further investigated with ADME/T pharmacokinetic properties and bioactivity properties, denoting that these bioactive dispositions of the compounds inside the body as a chemical would be safe without causing toxicity or carcinogenicity. These analyses demonstrated that Catechin gallate might be a potential drug candidate against SARS-CoV-2. Our identification also might help to advancement of knowledge of Korean traditional medicinal plants.

Author contributions

AH; Conceptualization, methodology, formal analysis, investigation, validation, writing – original draft, **MNM, KN;** Conceptualization, methodology, investigation, validation, writing –review & editing, **MFR;** Validation, funding acquisition, review, editing, **MER, MSR, MOF;** Formal analysis, review & editing.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Declaration of competing interest

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

Acknowledgment

The authors wish to thank the Molecular Genetics Lab. Dept. of Genetic Engineering and Biotechnology, University of Rajshahi, Bangladesh, for providing all facilities during the research.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.compbiomed.2023.106785.

References

[1] M. Cascella, M. Rajnik, A. Aleem, S.C. Dulebohn, R. Di Napoli, Features, Evaluation, and Treatment of Coronavirus (COVID-19), 2022. Statpearls [Internet].
[2] J.F.-W. Chan, K.-H. Kok, Z. Zhu, H. Chu, K.K.-W. To, S. Yuan, K.-Y. Yuen, Genomic characterization of the 2019 novel human-pathogenic coronavirus isolated from a patient with atypical pneumonia after visiting Wuhan, Emerg. Microb. Infect. 9 (2020) 221–236.
[3] F. Wu, S. Zhao, B. Yu, Y.-M. Chen, W. Wang, Z.-G. Song, Y. Hu, Z.-W. Tao, J.-H. Tian, Y.-Y. Pei, A new coronavirus associated with human respiratory disease in China, Nature 579 (2020) 265–269.
[4] R. Lu, X. Zhao, J. Li, P. Niu, B. Yang, H. Wu, W. Wang, H. Song, B. Huang, N. Zhu, Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding, Lancet 395 (2020) 565–574.
[5] A.A.T. Naqvi, K. Fatima, T. Mohammad, U. Fatima, I.K. Singh, A. Singh, S.M. Atif, G. Hariprasad, G.M. Hasan, M.I. Hassan, Insights into SARS-CoV-2 genome, structure, evolution, pathogenesis and therapies: structural genomics approach, Biochim. Biophys. Acta (BBA) - Mol. Basis Dis. 1866 (2020), 165878.
[6] H.-J. Lee, C.-K. Shieh, A.E. Gorbalenya, E. V Koonin, N. La Monica, J. Tuler, A. Bagdzhadzhyan, M.M.C. Lai, The complete sequence (22 kilobases) of murine coronavirus gene 1 encoding the putative proteases and RNA polymerase, Virology 180 (1991) 567–582.
[7] J. Ziebuhr, E.J. Snijder, A.E. Gorbalenya, Virus-encoded proteinases and proteolytic processing in the Nidovirales, J. Gen. Virol. 81 (2000) 853–879.
[8] K. Anand, G.J. Palm, J.R. Mesters, S.G. Siddell, J. Ziebuhr, R. Hilgenfeld, Structure of coronavirus main proteinase reveals combination of a chymotrypsin fold with an extra α-helical domain, EMBO J. 21 (2002) 3213–3224.
[9] K. Anand, J. Ziebuhr, P. Wadhwani, J.R. Mesters, R. Hilgenfeld, Coronavirus main proteinase (3CLpro) structure: basis for design of anti-SARS drugs, Science 300 (2003) 1763–1767.
[10] H. Yang, M. Yang, Y. Ding, Y. Liu, Z. Lou, Z. Zhou, L. Sun, L. Mo, S. Ye, H. Pang, The crystal structures of severe acute respiratory syndrome virus main protease and its complex with an inhibitor, Proc. Natl. Acad. Sci. USA 100 (2003) 13190–13195.
[11] C.-Y. Chou, H.-C. Chang, W.-C. Hsu, T.-Z. Lin, C.-H. Lin, G.-G. Chang, Quaternary structure of the severe acute respiratory syndrome (SARS) coronavirus main protease, Biochemistry 43 (2004) 14958–14970.
[12] M.-F. Hsu, C.-J. Kuo, K.-T. Chang, H.-C. Chang, C.-C. Chou, T.-P. Ko, H.-L. Shr, G.-G. Chang, A.H.-J. Wang, P.-H. Liang, Mechanism of the maturation process of SARS-CoV 3CL protease*[boxs], J. Biol. Chem. 280 (2005) 31257–31266.
[13] J.S. Morse, T. Lalonde, S. Xu, W.R. Liu, Learning from the past: possible urgent prevention and treatment options for severe acute respiratory infections caused by 2019-nCoV, Chembiochem 21 (2020) 730–738.
[14] F.G. Hayden, R.B. Turner, J.M. Gwaltney, K. Chi-Burris, M. Gersten, P. Hsyu, A. K. Patick, G.J. Smith III, L.S. Zalman, Phase II, randomized, double-blind, placebo-controlled studies of ruprintrivir nasal spray 2-percent suspension for prevention and treatment of experimentally induced rhinovirus colds in healthy volunteers, Antimicrob. Agents Chemother. 47 (2003) 3907–3916.
[15] R. Jain, S. Mujwar, Repurposing metocurine as main protease inhibitor to develop novel antiviral therapy for COVID-19, Struct. Chem. 31 (2020) 2487–2499.
[16] V. Thiel, K.A. Ivanov, A. Putics, T. Hertzig, B. Schelle, S. Bayer, B. Weißbrich, E. J. Snijder, H. Rabenau, H.W. Doerr, Mechanisms and enzymes involved in SARS coronavirus genome expression, J. Gen. Virol. 84 (2003) 2305–2315.
[17] H. Su, S. Yao, W. Zhao, M. Li, J. Liu, W. Shang, H. Xie, C. Ke, M. Gao, K. Yu, Discovery of Baicalin and Baicalein as Novel, Natural Product Inhibitors of SARS-CoV-2 3CL Protease in Vitro, BioRxiv, 2020.
[18] H. Yang, W. Xie, X. Xue, K. Yang, J. Ma, W. Liang, Q. Zhao, Z. Zhou, D. Pei, J. Ziebuhr, Design of wide-spectrum inhibitors targeting coronavirus main proteases, PLoS Biol. 3 (2005) e324.

[19] X. Xue, H. Yang, W. Shen, Q. Zhao, J. Li, K. Yang, C. Chen, Y. Jin, M. Bartlam, Z. Rao, Production of authentic SARS-CoV Mpro with enhanced activity: application as a novel tag-cleavage endopeptidase for protein overproduction, J. Mol. Biol. 366 (2007) 965–975.

[20] M.T. Islam, C. Sarkar, D.M. El-Kersh, S. Jamaddar, S.J. Uddin, J.A. Shilpi, M. S. Mubarak, Natural products and their derivatives against coronavirus: a review of the non-clinical and pre-clinical data, Phytother Res. 34 (2020) 2471–2492.

[21] M.A. Shah, A. Rasul, R. Yousaf, M. Haris, H.I. Faheem, A. Hamid, H. Khan, A. H. Khan, M. Aschner, G.E. Batiha, Combination of natural antivirals and potent immune invigorators: a natural remedy to combat COVID-19, Phytother Res. 35 (2021) 6530–6551.

[22] M. Sharma, P. Prasher, M. Mehta, F.C. Zacconi, Y. Singh, D.N. Kapoor, H. Dureja, D.M. Pardhi, M.M. Tambuwala, G. Gupta, Probing 3CL Protease: Rationally Designed Chemical Moieties for COVID-19, Drug Development Research, 2020.

[23] Z. Jin, X. Du, Y. Xu, Y. Deng, M. Liu, Y. Zhao, B. Zhang, X. Li, L. Zhang, C. Peng, Structure of Mpro from SARS-CoV-2 and discovery of its inhibitors, Nature 582 (2020) 289–293.

[24] L. Chen, C. Gui, X. Luo, Q. Yang, S. Günther, E. Scandella, C. Drosten, D. Bai, X. He, B. Ludewig, Cinanserin is an inhibitor of the 3C-like proteinase of severe acute respiratory syndrome coronavirus and strongly reduces virus replication in vitro, J. Virol. 79 (2005) 7095–7103.

[25] J.T. Ortega, M.L. Serrano, B. Jastrzebska, Class AG protein-coupled receptor antagonist famotidine as a therapeutic alternative against SARS-CoV2: an in silico analysis, Biomolecules 10 (2020) 954.

[26] X. Liu, X.-J. Wang, Potential inhibitors against 2019-nCoV coronavirus M protease from clinically approved medicines, J. Genetics Genom. 47 (2020) 119.

[27] R. Hilgenfeld, From SARS to MERS: crystallographic studies on coronaviral proteases enable antiviral drug design, FEBS J. 281 (2014) 4085–4096.

[28] H.M. Mengist, T. Dilnessa, T. Jin, Structural basis of potential inhibitors targeting SARS-CoV-2 main protease, Front. Chem. 9 (2021), 622898.

[29] P.I. Adegbola, B. Semire, O.S. Fadahunsi, A.E. Adegoke, Molecular docking and ADMET studies of Allium cepa, Azadirachta indica and Xylopia aethiopica isolates as potential anti-viral drugs for Covid-19, VirusDisease 32 (2021) 85–97.

[30] E. Pitsillou, J.J. Liang, R.C. Beh, A. Hung, T.C. Karagiannis, Molecular dynamics simulations highlight the altered binding landscape at the spike-ACE2 interface between the Delta and Omicron variants compared to the SARS-CoV-2 original strain, Comput. Biol. Med. 149 (2022), 106035.

[31] Y. Kumar, H. Singh, C.N. Patel, Silico prediction of potential inhibitors for the main protease of SARS-CoV-2 using molecular docking and dynamics simulation based drug-repurposing, J. Infect. Public Health 13 (2020) 1210–1223.

[32] K. Bafna, R.M. Krug, G.T. Montelione, Structural Similarity of SARS-CoV2 Mpro and HCV NS3/4A Proteases Suggests New Approaches for Identifying Existing Drugs Useful as COVID-19 Therapeutics, ChemRxiv, 2020.

[33] Q. Liu, K. Kwan, T. Cao, B. Yan, K. Ganesan, L. Jia, F. Zhang, C. Lim, Y. Wu, Y. Feng, Broad-spectrum antiviral activity of Spatholobus suberectus Dunn against SARS-CoV-2, SARS-CoV-1, H5N1, and other enveloped viruses, Phytother Res. 36 (2022) 3232–3247.

[34] W. Kaplan, T.G. Littlejohn, Swiss-PDB viewer (deep view), Briefings Bioinf. 2 (2001) 195–197.

[35] A.H. Arshia, S. Shadravan, A. Solhjoo, A. Sakhteman, A. Sami, De novo design of novel protease inhibitor candidates in the treatment of SARS-CoV-2 using deep learning, docking, and molecular dynamic simulations, Comput. Biol. Med. 139 (2021), https://doi.org/10.1016/j.compbiomed.2021.104967.

[36] O. Trott, A.J. Olson, AutoDock Vina, Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading, J. Comput. Chem. 31 (2010) 455–461.

[37] S. Murugesan, S. Kottekad, I. Crasta, S. Sreevathsan, D. Usharani, M.K. Perumal, S. N. Mudliar, Targeting COVID-19 (SARS-CoV-2) main protease through active phytocompounds of ayurvedic medicinal plants–Emblica officinalis (Amla), Phyllanthus niruri Linn.(Bhumi Amla) and Tinospora cordifolia (Giloy)–A molecular docking and simulation study, Comput. Biol. Med. 136 (2021), 104683.

[38] Z. Liu, J. Du, J. Fang, Y. Yin, G. Xu, L. Xie, DeepScreening : a Deep Learning-Based Screening Web Server for Accelerating Drug Discovery, 2019, pp. 1–11, https://doi.org/10.1093/database/baz104.

[39] C.W.E.I. Yap, Software News and Update PaDEL-Descriptor : an Open Source Software to Calculate Molecular Descriptors and Fingerprints, 2010, https://doi.org/10.1002/jcc.

[40] T. ting Fu, G. Tu, M. Ping, G. xun Zheng, F. yuan Yang, J. yi Yang, Y. Zhang, X. jun Yao, W. wei Xue, F. Zhu, Subtype-selective mechanisms of negative allosteric modulators binding to group I metabotropic glutamate receptors, Acta Pharmacol. Sin. 42 (2021) 1354–1367, https://doi.org/10.1038/s41401-020-00541-z.

[41] W. Xue, P. Wang, G. Tu, F. Yang, G. Zheng, X. Li, X. Li, Y. Chen, X. Yao, F. Zhu, Computational identification of the binding mechanism of a triple reuptake inhibitor amitifadine for the treatment of major depressive disorder, Phys. Chem. Chem. Phys. 20 (2018) 6606–6616, https://doi.org/10.1039/c7cp07869b.

[42] E. Krieger, T. Darden, S.B. Nabuurs, A. Finkelstein, G. Vriend, Making optimal use of empirical energy functions: force-field parameterization in crystal space, Proteins: Struct., Funct., Bioinf. 57 (2004) 678–683.

[43] D.A. Case, T.E. Cheatham III, T. Darden, H. Gohlke, R. Luo, K.M. Merz Jr., A. Onufriev, C. Simmerling, B. Wang, R.J. Woods, The Amber biomolecular simulation programs, J. Comput. Chem. 26 (2005) 1668–1688.

[44] E. Krieger, J.E. Nielsen, C.A.E.M. Spronk, G. Vriend, Fast empirical pKa prediction by Ewald summation, J. Mol. Graph. Model. 25 (2006) 481–486.

[45] M. Dutta, A.M. Tareq, A. Rakib, S. Mahmud, S.A. Sami, J. Mallick, M.N. Islam, M. Majumder, Z. Uddin, A. Alsubaie, A.S.A. Almalki, M.U. Khandaker, D.A. Bradley, S. Rana, T. Bin Emran, Molecular Dynamics Simulations, (2021).

[46] G. Tu, T. Fu, F. Yang, J. Yang, Z. Zhang, X. Yao, W. Xue, F. Zhu, Understanding the polypharmacological profiles of triple reuptake inhibitors by molecular simulation, ACS Chem. Neurosci. 12 (2021) 2013–2026, https://doi.org/10.1021/acschemneuro.1c00127.

[47] E. Krieger, G. Vriend, New ways to boost molecular dynamics simulations, J. Comput. Chem. 36 (2015) 996–1007.

[48] S. Mahmud, M.R. Parves, Y.M. Riza, K.M. Sujon, S. Ray, F.A. Tithi, Z.F. Zaoti, S. Alam, N. Absar, Exploring the potent inhibitors and binding modes of phospholipase A2 through in silico investigation, J. Biomol. Struct. Dyn. 38 (2020) 4221–4231.

[49] M.A. Khan, S. Mahmud, A.S.M.R.U. Alam, M.E. Rahman, F. Ahmed, M. Rahmatullah, Comparative molecular investigation of the potential inhibitors against SARS-CoV-2 main protease: a molecular docking study, J. Biomol. Struct. Dyn. 39 (2021) 6317–6323.

[50] M.J. Islam, M.R. Parves, S. Mahmud, F.A. Tithi, M.A. Reza, Assessment of structurally and functionally high-risk nsSNPs impacts on human bone morphogenetic protein receptor type IA (BMPR1A) by computational approach, Comput. Biol. Chem. 80 (2019) 31–45.

[51] P.K. Parida, D. Paul, D. Chakravorty, The natural way forward: molecular dynamics simulation analysis of phytochemicals from Indian medicinal plants as potential inhibitors of SARS-CoV-2 targets, Phytother Res. 34 (2020) 3420–3433.

[52] J. Lee, C. Kenward, L.J. Worrall, M. Vuckovic, F. Gentile, A.-T. Ton, M. Ng, A. Cherkasov, N.C.J. Strynadka, M. Paetzel, X-ray crystallographic characterization of the SARS-CoV-2 main protease polyprotein cleavage sites essential for viral processing and maturation, Nat. Commun. 13 (2022) 1–13.

[53] W. Vuong, C. Fischer, M.B. Khan, M.J. van Belkum, T. Lamer, K.D. Willoughby, J. Lu, E. Arutyunova, M.A. Joyce, H.A. Saffran, Improved SARS-CoV-2 Mpro inhibitors based on feline antiviral drug GC376: structural enhancements, increased solubility, and micellar studies, Eur. J. Med. Chem. 222 (2021), 113584.

[54] S. V Stoddard, S.D. Stoddard, B.K. Oelkers, K. Fitts, K. Whalum, K. Whalum, A. D. Hemphill, J. Manikonda, L.M. Martinez, E.G. Riley, Optimization rules for SARS-CoV-2 Mpro antivirals: ensemble docking and exploration of the coronavirus protease active site, Viruses 12 (2020) 942.

[55] Q. Xue, X. Liu, P. Russell, J. Li, W. Pan, J. Fu, A. Zhang, Evaluation of the binding performance of flavonoids to estrogen receptor alpha by Autodock, Autodock Vina and Surflex-Dock, Ecotoxicol. Environ. Saf. 233 (2022), 113323.

[56] J. Lung, Y. Lin, Y. Yang, Y. Chou, L. Shu, Y. Cheng, H. Te Liu, C. Wu, The potential chemical structure of anti-SARS-CoV-2 RNA-dependent RNA polymerase, J. Med. Virol. 92 (2020) 693–697.

[57] L. Zhang, D. Lin, X. Sun, U. Curth, C. Drosten, L. Sauerhering, S. Becker, K. Rox, R. Hilgenfeld, Crystal structure of SARS-CoV-2 main protease provides a basis for design of improved α-ketoamide inhibitors, Science 368 (2020) 409–412.

[58] M.C. Robinson, R.C. Glen, Validating the validation: reanalyzing a large-scale comparison of deep learning and machine learning models for bioactivity prediction, J. Comput. Aided Mol. Des. 34 (2020) 717–730.

[59] W.P. Walters, R. Barzilay, Applications of deep learning in molecule generation and molecular property prediction, Acc. Chem. Res. 54 (2020) 263–270.

[60] J. Kim, S. Park, D. Min, W. Kim, Comprehensive survey of recent drug discovery using deep learning, Int. J. Mol. Sci. 22 (2021) 9983.

[61] T. Joshi, T. Joshi, H. Pundir, P. Sharma, S. Mathpal, S. Chandra, Predictive modeling by deep learning, virtual screening and molecular dynamics study of natural compounds against SARS-CoV-2 main protease, J. Biomol. Struct. Dyn. 39 (2021) 6728–6746.

[62] M. Nand, P. Maiti, T. Joshi, S. Chandra, V. Pande, J.C. Kuniyal, M. A. Ramakrishnan, Virtual screening of anti-HIV1 compounds against SARS-CoV-2: machine learning modeling, chemoinformatics and molecular dynamics simulation based analysis, Sci. Rep. 10 (2020) 1–12.

[63] D. Chicco, M.J. Warrens, G. Jurman, The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation, PeerJ. Comp. Sci. 7 (2021) e623.

[64] R. Ghosh, A. Chakraborty, A. Biswas, S. Chowdhuri, Evaluation of green tea polyphenols as novel corona virus (SARS CoV-2) main protease (Mpro) inhibitors–an in silico docking and molecular dynamics simulation study, J. Biomol. Struct. Dyn. 39 (2021) 4362–4374.

[65] X. Zhou, F. Zhong, C. Lin, X. Hu, Y. Zhang, B. Xiong, X. Yin, J. Fu, W. He, J. Duan, Structure of SARS-CoV-2 main protease in the apo state, Sci. China Life Sci. 64 (2021) 656–659.

[66] D.W. Kneller, G. Phillips, H.M. O'Neill, R. Jedrzejczak, L. Stols, P. Langan, A. Joachimiak, L. Coates, A. Kovalevsky, Structural plasticity of SARS-CoV-2 3CL Mpro active site cavity revealed by room temperature X-ray crystallography, Nat. Commun. 11 (2020) 3202, 2020.

[67] J.C. Ferreira, S. Fadl, A.J. Villanueva, W.M. Rabeh, Catalytic dyad residues His41 and Cys145 impact the catalytic activity and overall conformational fold of the main SARS-CoV-2 protease 3-chymotrypsin-like protease, Front. Chem. 9 (2021) 491.

[68] A. Narayanan, S.A. Toner, J. Jose, Structure-based inhibitor design and repurposing clinical drugs to target SARS-CoV-2 proteases, Biochem. Soc. Trans. 50 (2022) 151–165.

[69] S. Kumar, R. Nyodu, V.K. Maurya, S.K. Saxena, Morphology, genome organization, replication, and pathogenesis of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), in: Coronavirus Disease 2019 (COVID-19), Springer, 2020, pp. 23–31.

[70] S. Khan, Z. Fakhar, A. Hussain, A. Ahmad, D.S. Jairajpuri, M.F. Alajmi, M.I. Hassan, Structure-based identification of potential SARS-CoV-2 main protease inhibitors, J. Biomol. Struct. Dyn. 40 (2022) 3595–3608.

[71] M.T. ul Qamar, S.M. Alqahtani, M.A. Alamri, L.-L. Chen, Structural basis of SARS-CoV-2 3CLpro and anti-COVID-19 drug discovery from medicinal plants, J. Pharmaceut. Anal. 10 (2020) 313–319.
[72] S.S. Mousavi, A. Karami, T.M. Haghighi, S.G. Tumilaar, Fatimawali, Idroes, R. In silico evaluation of Iranian medicinal plant phytoconstituents as inhibitors against main protease and the receptor-binding domain of SARS-CoV-2. Molecules. 26 (2021) 5724.
[73] L. Peterson, COVID-19 and Flavonoids, Silico Molecular Dynamics Docking to the Active Catalytic Site of SARS-CoV and SARS-CoV-2 Main Protease, SSRN, 2020.
[74] S.-Y. Liu, W. Wang, J.-P. Ke, P. Zhang, G.-X. Chu, G.-H. Bao, Discovery of Camellia sinensis catechins as SARS-CoV-2 3CL protease inhibitors through molecular docking, intra and extra cellular assays, Phytomedicine 96 (2022), 153853.
[75] S. Mahmud, S. Biswas, G.K. Paul, M.A. Mita, M.M. Promi, S. Afrose, M.R. Hasan, S. Zaman, M.S. Uddin, K. Dhama, Plant-based phytochemical screening by targeting main protease of SARS-CoV-2 to design effective potent inhibitors, Biology 10 (2021) 589.
[76] T. Hanai, Quantitative in silico analysis of SARS-CoV-2 S-RBD omicron mutant transmissibility, Talanta 240 (2022), 123206, https://doi.org/10.1016/j.talanta.2022.123206.
[77] K.-E. Choi, J.-M. Kim, J. Rhee, A.K. Park, E.-J. Kim, N.S. Kang, Molecular Dynamics Studies on the structural characteristics for the stability prediction of SARS-CoV-2, Int. J. Mol. Sci. 22 (2021) 8714.
[78] V.K. Singh, H. Chaurasia, P. Kumari, A. Som, R. Mishra, R. Srivastava, F. Naaz, A. Singh, R.K. Singh, Design, synthesis, and molecular dynamics simulation studies of quinoline derivatives as protease inhibitors against SARS-CoV-2, J. Biomol. Struct. Dyn. (2021) 1–24.
[79] K. Dutta Dubey, R. Kumar Tiwari, R. Prasad Ojha, Recent advances in protein–ligand interactions: molecular dynamics simulations and binding free energy, Curr. Comput. Aided Drug Des. 9 (2013) 518–531.
[80] F. Mosquera-Yuqui, N. Lopez-Guerra, E.A. Moncayo-Palacio, Targeting the 3CLpro and RdRp of SARS-CoV-2 with phytochemicals from medicinal plants of the Andean Region: molecular docking and molecular dynamics simulations, J. Biomol. Struct. Dyn. 40 (2022) 2010–2023.
[81] R.K. Mohapatra, K. Dhama, A.A. El–Arabey, A.K. Sarangi, R. Tiwari, T. Bin Emran, M. Azam, S.I. Al-Resayes, M.K. Raval, V. Seidel, Repurposing benzimidazole and benzothiazole derivatives as potential inhibitors of SARS-CoV-2: DFT, QSAR, molecular docking, molecular dynamics simulation, and in-silico pharmacokinetic and toxicity studies, J. King Saud Univ. Sci. 33 (2021), 101637.
[82] H. Yang, L. Sun, W. Li, G. Liu, Y. Tang, In silico prediction of chemical toxicity for drug design using machine learning methods and structural alerts, Front. Chem. 6 (2018) 30.
[83] M.D. Segall, C. Barber, Addressing toxicity risk when designing and selecting compounds in early drug discovery, Drug Discov. Today 19 (2014) 688–693.
[84] J.K. Williams, B. Wang, A. Sam, C.L. Hoop, D.A. Case, J. Baum, Molecular dynamics analysis of a flexible loop at the binding interface of the SARS-CoV-2 spike protein receptor-binding domain, Proteins, Struct. Funct. Bioinformat. 90 (2022) 1044–1053.
[85] F. Esteves, J. Rueff, M. Kranendonk, The central role of cytochrome P450 in xenobiotic metabolism—a brief review on a fascinating enzyme family, J. Xenobio. 11 (2021) 94–114.
[86] A.A. El-Banna, R.S. Darwish, D.A. Ghareeb, A.M. Yassin, S.A. Abdulmalek, H. M. Dawood, Metabolic profiling of Lantana camara L. using UPLC-MS/MS and revealing its inflammation-related targets using network pharmacology-based and molecular docking analyses, Sci. Rep. 12 (2022) 1–17.
[87] S.S. Swain, T. Hussain, Role of bioinformatics in early drug discovery: an overview and perspective, Computat. BioInformat.: Multidisc. Appl. (2021) 49–67.
[88] T. Lei, Y. Li, Y. Song, D. Li, H. Sun, T. Hou, ADMET evaluation in drug discovery: 15. Accurate prediction of rat oral acute toxicity using relevance vector machine and consensus modeling, J. Cheminf. 8 (2016) 1–19.
[89] Y. Xu, J. Pei, L. Lai, Deep learning based regression and multiclass models for acute oral toxicity prediction with automatic chemical feature extraction, J. Chem. Inf. Model. 57 (2017) 2672–2685.