Research Article

Natural Products as Inhibitors of COVID-19 Main Protease – A Virtual Screening by Molecular Docking

Marzieh Omrani[1#], Mohammad Bayati[1#], Parvaneh Mehrbod[2*] , Kamal Asmari Bardazard[1], Samad Nejad Ebrahimi[1*]

[1]*Department of Phytochemistry, Medicinal Plants and Drugs Research Institute, Shahid Beheshti University, Tehran, Iran.*
[2]*Influenza and Respiratory Viruses Department, Pasteur Institute of Iran, Tehran, Iran.*

Abstract

Background: The novel coronavirus (2019-nCoV) causes a severe respiratory illness unknown to a human before. Its alarmingly quick transmission to many countries across the world has resulted in a global health emergency. Therefore, an imminent need for drugs to combat this disease has been increased. Worldwide collaborative efforts from scientists are underway to determine a therapy to treat COVID-19 infections and reduce mortality rates. Since herbal medicines and purified natural products have been reported to have antiviral activity against Coronaviruses (CoVs), this *in silico* evaluation was performed for identifying potential natural compounds with promising inhibitory activities against COVID-19.

Methods: In this study, a High Throughput Virtual Screening (HTVS) protocol was used as a fast method for discovering novel drug candidates as potential COVID-19 main protease (M^{pro}) inhibitors. Over 180,000 natural product-based compounds were obtained from the ZINC database and virtually screened against the COVID-19 M^{pro}. In this study, the Glide docking program was applied for high throughput virtual screening. Also, Extra precision (XP) has been used following the induced-fit docking (IFD) approach. The ADME properties of all compounds were analyzed and a final selection was made based on the Lipinski rule of five. Also, molecular dynamics (MD) simulations were conducted for a virtual complex of the best scoring compound with COVID-19 protease.

Results: Nineteen compounds were introduced as new potential inhibitors. Compound ZINC08765174 (1-[3-(1H-indol-3-yl) propanoyl]-N-(4-phenylbutan-2-yl)piperidine-3-carbox-amide showed a strong binding affinity (-11.5 kcal/mol) to the COVID-19 M^{pro} comparing to peramivir (-9.8 kcal/mol) as a positive control.

Conclusion: Based on these findings, nineteen compounds were proposed as possible new COVID-19 inhibitors, of which ZINC08765174 had a high affinity to M^{pro}. Furthermore, the promising ADME properties of the selected compounds emphasize their potential as attractive candidates for the treatments of COVID-19.

Introduction

Unusual viral pneumonia caused by COVID-19, known as the 2019 novel coronavirus (2019-nCoV), broke out in December 2019 in Wuhan, Hubei Province, China.[1] The World Health Organization (WHO) has declared it as a global public health emergency. The 2019-nCoV is spreading all over the world causing a severe respiratory illness unknown to humans before. Its alarmingly quick transmission to many countries around the globe has resulted in a worldwide health emergency. Due to the rapid rate of distribution of COVID-19 worldwide, it was recognized by WHO as a pandemic on 11 March 2020. It has caused a notable percentage of morbidity and mortality.

Globally, there have been more than 91 million confirmed COVID-19 infections and 1.9 million deaths since the beginning of the pandemic to January, 2021.[2]

Belonging to *Coronaviridae*, coronaviruses (CoVs) are enveloped viruses with non-segmented single-stranded positive-sense RNA. The typical genome of CoV includes 5'-cap, 5'-untranslated region (UTR), open reading frames, 3'-UTR, and 3'-poly(A) tail. The first two-thirds of the genome encodes the nonstructural proteins that form the replicase complex. The last third of the genome encodes primarily structural proteins.[3] Two CoVs, including Severe Acute Respiratory Syndrome (SARS) and the Middle East

Respiratory Syndrome (MERS), can result in infections and fatal respiratory diseases in humans.[4,5]

The scientific community has begun an effort to find some candidate drugs with antiviral properties to reduce fatalities caused by COVID-19.[6,7] Herbal medicines and purified natural products play an important role as complementary therapies via modulating the immunomodulatory system of both infected and uninfected individuals. Antiviral effects of some natural products against some critical viral pathogens, including coronavirus (CoV), have been reported.[8,9] Saikosaponins are a group of oleanane derivatives that have been isolated from some medicinal plants, such as *Heteromorpha* spp.[10] Saikosaponins were found to possess antiviral activity on HCoV-229E by inhibiting viral attachment to cells, blocking viral penetration into cells, and interfering with the early stage of viral replication.[11] Myricetin and scutellarein are naturally occurring flavonoids that can be isolated from plants such as tea, berries, fruits, vegetables, and medicinal herbs.[12-14] Myricetin and scutellarein were reported as novel chemical inhibitors of the SARS coronavirus helicase.[12] Significant activity against CoV proteases was demonstrated by certain isolated polyphenols such as broussochalcone A, 4-hydroxyisolonchocarpin, papyriflavonol A, 3'-(3-methylbut-2-enyl)-3',4,7-trihydroxyflavane, kazinol A, kazinol B, broussoflavan A, kazinol F, and kazinol J.[15]

Recently, several computer-based investigates have been undertaken to determine the impact of natural products on various COVID-19 receptors. According to the result of an assay, some terpenoids namely ursolic acid, oleanolic acid, and carvacrol were suggested as potential inhibitors of the M[pro] of COVID-19.[16] Moreover, some natural compounds such as digitoxigenin, β-eudesmol, glycyrrhizin, tryptanthrine, rhein, berberine, and crocin were evaluated and showed inhibitory activity against the M[pro] of COVID-19.[17,18] Hesperidin, rutin, diosmin, and apiin are some flavonoids which were identified to have binding affinity to the M[pro] of COVID-19.[19] Some flavonoids, including naringin and hesperetin, were reported to have the potential to bind to ACE2 and block the entry of 2019-nCoV into host cells.[20] According to the obtained result from an *in silico* study, quercetin, hispidulin, and cirsimaritin showed inhibitory activities against COVID-19 M[pro] active site and ACE2.[21] Also, two alkaloids named echitamine, nicotianamine, and some other phytochemicals including baicalin, scutellarin, hesperetin, 6-α-acetoxygedunin, and glycyrrhizin, seemed to have the potential to block the entry of 2019-nCoV into host cells by binding to ACE2.[22,23]

The M[pro] enzymes of coronaviruses including avian infectious bronchitis virus (IBV), transmissible gastrointestinal virus (TGEV), and murine hepatitis virus (MHV) play an essential role in viral replication. It cleaves the polyproteins into smaller fragments and is considered an absolute requirement for replication.[24] In addition to the typical coronavirus structural proteins and replicase genes, COVID-19 has several currently unidentified nonstructural open reading frames in its genome.[1] Very recently, the crystal structure of the M[pro] of COVID-19 in complex with a peptidomimetic inhibitor has been reported (Figure 1) and makes it possible to investigate the interaction between the compounds and this protein as a target.[25] In this study, we aim to run a HTVS protocol for identifying compounds with potential inhibition against COVID-19 M[pro]. The selected compounds for HTVS were all natural products-derived.

Materials and Methods
Protein preparation
COVID-19 M[pro] crystal structure in a complex with inhibitor N3 was downloaded from the RCSB Protein data bank (PDB ID: 6LU7). The structure of this protein was prepared by removing waters. Minimization of the structure was carried out by using the protein preparation
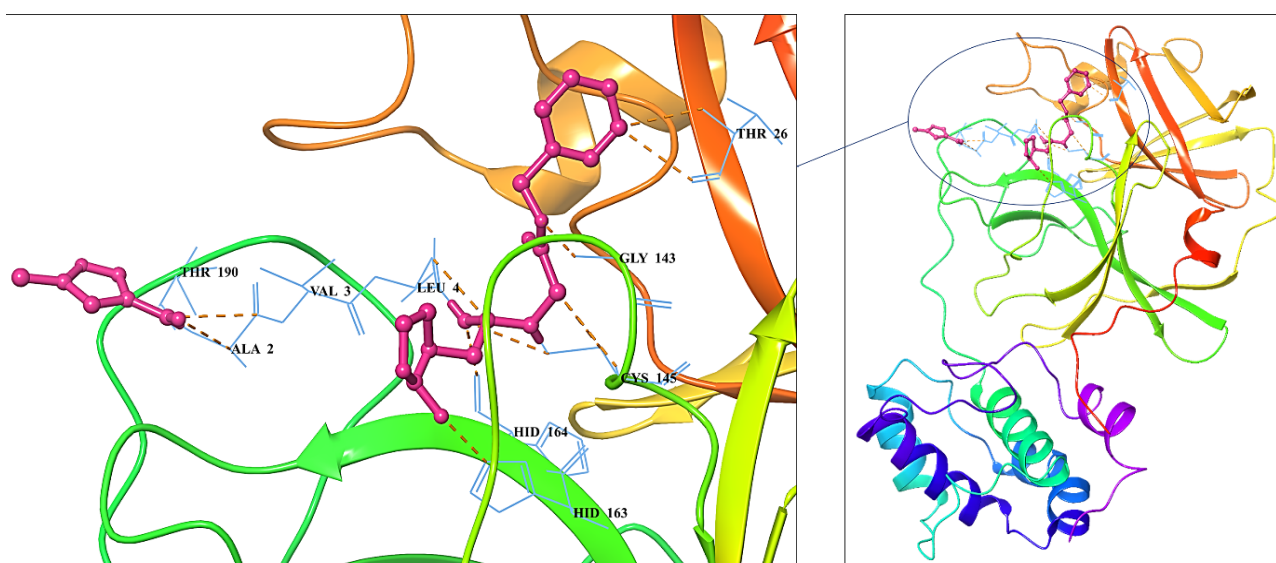
Figure 1. The 3D representation of intermolecular interactions of the co-crystallized N3 inhibitor with COVID-19 main protease (PDB ID: 6LU7).

wizard in the Maestro suite (version 11.8, 2018). Hydrogen atoms were added, disulfide bonds were created, and water molecules beyond 3.00 Å from HET groups were deleted. Missing loops and side chains were filled by using prime. HET states were generated using Epik, protonated states were recognized utilizing PROPKA in pH 7.00 and the other changes were applied by software default. Eventually, the COVID-19 M[pro] structure was optimized and minimized by using the OPLS3 force field.

Ligand preparation

Four natural compound databases, including IBScreenNP database (https://www.ibscreen.com/naturalcompounds), the AnalytiCon Discovery database (https://www.ac-discovery.com), SpecNatural database (https://www.specs.net), as well as ZINC15 database (http://www.zinc.docking.org/browse/catalogs/naturalproducts) were used to download more than (~110.000) required natural products and molecules. The downloaded structures were prepared using the LigPrep application in the Maestro 11.8 suite. The OPLS3 force field was applied to convert 2D into 3D structures and reduce computational errors. Ionization states were applied using Epik at pH 7.00, and at most, four isomers were generated for each ligand.

Grid generation and molecular docking

Receptor grid generation of maestro suite was used to create the active binding site of the protease structure around the residues Glu166, Phe140, His164, Gly143, Cys145, and Ala2. Site map module of maestro Schrödinger suite was applied to validate the grid box, which was generated at (X: -12.06, Y: 13.95, Z: 69.49) and diameter midpoint box of (X, Y, Z: 10 Å), site maps with site score > 1.00 were more valid than the others.

The ligand docking was performed using the Glide of Maestro suite in two precision steps, a large number of ligands were screened quickly employing HTVS, followed by the XP method used to dock the best 10% of poses with excess precision. Flexible ligand sampling was applied to both protocols, and the output result was expressed as a docking score.

Induced-fit docking (IFD)

IFD protocol was used as a redocking experiment to evaluate the selected compounds in a relaxed residue binding pocket. This experiment was performed using the OPLS3e force field and other standard parameters. The receptor grid was centered on the inhibitor N3 co-crystallized ligand around the residues; Leu4, Asn142, and Cys145. Different conformers of ligands were analyzed within a 2.5 kcal/mol energy window, and glide's XP mode was done as a redocking procedure. Eventually, the IFD score was reported in kcal/mol.

Free binding energy calculation

Ligand-protein complex free binding energy was calculated using two equations: molecular mechanics generalized Born surface area (MM-GBSA) and molecular mechanics Poisson–Boltzmann surface area (MM-PBSA). The free binding energy of the docking poses of the references and hit compounds was calculated by utilizing the prime MM-GBSA module of the maestro. The OPLS3e force field and VSGB refinement solvation model were chosen to predict the free binding energy of complexes.

$$\Delta G_{binding} = G_{complex} \text{ (minimized)} - G_{ligand} \text{ (minimized)} - G_{receptor} \text{ (minimized)}$$

Drug-like properties

QikProp application was used to predict pharmacokinetics, ADME properties (Absorption, Distribution, Metabolism, and Excretion), and evaluating Drug-likeness features of all compounds. Hit compounds were selected by applying the Lipinski rule of five (RO5), polar surface area (PSA), central nervous system (CNS) activity, and percent oral absorption.

Molecular dynamics simulation

The highest scoring compound in the complex with the M[pro] of COVID-19 was submitted to MD simulations. The MD simulation was carried out using the Macro Model program from the Schrödinger software package. Moreover, the root means square deviation (RMSD) was calculated to evaluate the structure's optimum energy. The conformation state was considered with a maximum half angstrom of superimposition error in RMSD. Ligand energy minimization was carried out by OPLS3 force field in an aqueous solvent. Using PRSG (Polak- Ribière Conjugate Gradient) methods, the MD of the 6LU7 protein is analyzed using stochastic dynamics mechanisms.

Moreover, energy minimization has been investigated. Dynamic simulation conditions were provided at 300 °K, the time step of 1.5 femtoseconds, the equilibration time of 1 picosecond, and the simulation time of 100 picoseconds. Finally, potential calculations were performed using the OPLS3 force field with solvent water.[26]

Results

Identification of potential Inhibitor of the M[pro] using molecular docking

In this study, the search for new potential inhibitors for coronavirus primary protease (PDB ID: 6lu7) was performed using molecular modeling. HTVS was used for virtual screening. Glide's XP mode was performed to validate molecular docking in the Schrödinger maestro suite (version 11.8, 2018). IFD was chosen to take into account the flexibility of amino acid residues at the active binding site and to avoid false-positive data from the bonding process. Ultimately, the wide time range of motion in the nano and picoseconds scale in the experimental methods of MD simulation was used to study the dynamics of biomolecules such as proteins at the molecular scale. MD simulations are applied for optimization and validation of final complexes. In this way, the characteristics and stability of the best scoring compound in the complex with the M[pro]

were further investigated.

The virtual screening and filtering hit structures out of 183388 compounds were performed in five steps. In the first step, HTVS docking was used to select certain natural products from the whole compounds based on their affinity to the target protein. The results showed 4581 compounds have a good association with docking scores ranged from -8.636 to -3.337 kcal/mol. In the next step, compounds were further selected using the XP docking approach. The affinity of the compounds to bind to a specific target protein was determined more accurately by the application of XP docking mode. Furthermore, the IFD protocol was applied to consider both ligand and receptor flexibility. This process resulted in a preference of 40 ligands (Table S1). The final selection was based on the data obtained from ADME properties, Lipinski rule of five, bioavailability, and rescoring by MM-GBSA. This step resulted in 19 hit compounds out of 40 compounds. The docking scores, pharmacokinetic properties, and MM-GBSA results of hit compounds are shown in Table 1. Also, the structures of hit compounds are represented in Figure 2.

Figure 2. The chemical structure of compounds which showed the best docking scores.
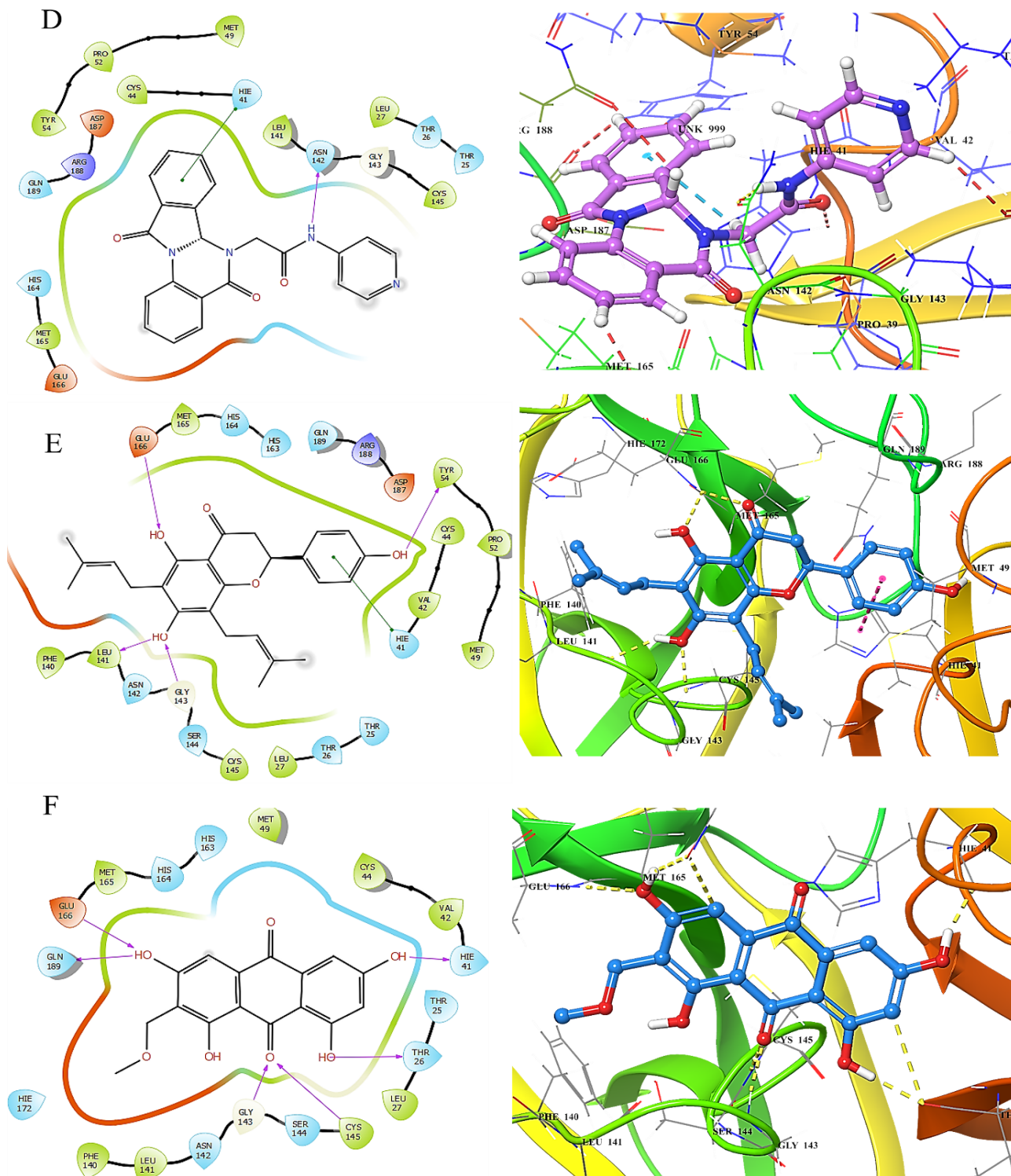
Table 1. Glide docking and IFD scores (kcal/mol), with ADME screening results.

No.	Compound	docking score	glide gscore	IFD Score	ΔG_{bind}	PSA	MW	Donor HB	Accpt HB	QPLog Po/w	%Oral abs.	CNS
NP-1	ZINC03839114	-11.132	-11.132	-674.28	-51.90	149.6	462.455	3	9	1.032	64.756	-2
NP-2	ZINC03841676	-11.068	-11.072	-673.56	-53.07	71.2	394.513	2	6	3.854	100	-1
NP-3	ZINC96114284	-10.743	-10.752	-669.73	-37.30	131.4	404.378	1	7	2.537	75.723	-2
NP-4	ZINC98364422	-10.315	-10.317	-669.05	-46.14	104.2	384.393	1	10	1.917	87.351	-1
NP-5	Lonchocarpol A	-10.234	-10.234	-670.01	-39.96	91.7	408.493	2	4	4.845	100	-2
NP-6	ZINC77257242	-10.102	-10.279	-666.89	-36.90	143.2	316.267	2	7	0.552	58.759	-2
NP-7	Norkurarinone	-9.910	-9.91	-670.69	-18.92	113.5	424.493	3	5	4.199	90.687	-2
NP-8	ZINC72321780	-9.516	-9.516	-669.24	-47.65	96.3	423.388	3	7	3.736	100	-1
NP-9	(-)-Catechin	-9.444	-9.444	-668.69	-37.36	116.8	290.272	5	5	0.448	59.967	-2
NP-10	Altertoxin I	-9.429	-9.429	-664.92	-66.04	133.3	352.343	2	6	1.369	66.233	-2
NP-11	ZINC72321775	-9.356	-9.356	-668.71	-47.28	75.9	379.378	3	6	3.527	100	-1
NP-12	ZINC35458935	-9.309	-9.349	-667.36	-42.9	148.8	428.395	3	8	1.891	72.908	-2
NP-13	ZINC00719192	-9.285	-9.285	-667.72	-38.45	114.5	482.415	1	7	4.416	100	-1
NP-14	ZINC72321774	-9.22	-9.22	-667.92	-40.03	89.2	395.378	3	7	2.982	89.497	-1
NP-15	ZINC31158868	-9.114	-9.115	-668.84	-41.22	121.9	291.303	3	8	1.073	56.981	-2
NP-16	ZINC09033965	-9.069	-9.069	-669.16	-46.67	99.7	433.233	1	9	1.621	81.002	-1
NP-17	ZINC02109515	-8.702	-8.86	-669.36	-41.79	75.5	360.335	3	4	1.854	65.671	-1
NP-18	ZINC15120609	-8.243	-8.243	-667.9	-37.92	95.5	372.417	2	6	3.482	100	-2
NP-19	Isoboldin	-8.049	-8.049	-667.29	-45.22	113.5	424.493	3	5	4.199	90.687	-2
NP-20	ZINC08765174	-11.50	-11.50	-673.4	-27.50	74.1	431.6	2	6	4.8	100	-1
NP-21	Peramivir	-9.763	-9.763	-670.26	-39.74	113.5	424.493	3	5	4.199	90.687	-2
NP-22	Laninamivir	-8.513	-8.513	-671.07	-33.73	113.5	424.493	3	5	4.199	90.687	-2
NP-23	Hydroxychloroquine	-7.407	-7.407	-669.57	-39.41	113.5	424.493	3	5	4.199	90.687	-2
NP-24	Baloxavir marboxil	-7.194	-7.194	-663.34	-40.53	113.5	424.493	3	5	4.199	90.687	-2

Interactions of top scoring compounds with M^{pro}

Interactions between the best scoring compounds and the M^{pro} are presented in Figures 3 and 4. Compound NP-1 (ZINC03839114) possesses a good docking score of -11.132 kcal/mol. The calculated data demonstrated that NP-1 has a strong binding affinity to the active site of COVID-19 M^{pro} with Glide score of -11.132 kcal/mol and IFD scores of -674.28 kcal/mol. The binding of hydrogen atoms of Urea with Gln189, and amide's hydrogen with Leu141, and His163 are the most effective interactions of compound NP-1 with the M^{pro}. Furthermore, the oxygen atom of the carbonyl groups of the best scoring compound established hydrogen bonds with Cys145 and Glu166. Compound NP-2 (ZINC03841676) has computational binding energy of -53.07 kcal/mol and a docking score of -11.068 kcal/mol. Hydrogen binding in this compound, including hydrogen interaction between the pyridine nitrogen atom and Gly143, and the oxygen atom of the carbonyl groups with Gln189. Also, Glu166 interacts with hydroxyl and amine groups. The binding energy of NP-3 (ZINC96114284) as a flavonoid was calculated as -37.3 kcal/mol, and the docking score was estimated at -10.743 kcal/mol. In the NP-3 compound, Gly143 and Cys145 interact with the nitrogen of diazole cycle, and Thr190 binds to the hydroxyl

Figure 3. Ligand-protein 2D and 3D interactions of hit compounds; A) NP-1, including Glu 166, Gln 189, His 163, Leu 141, and Cys 145 B) NP-2, including Glu 166, Gln 189, and Gly 143 C) NP-3, including Cys 145, Gly 143, and Thr 190 hydrogen bond interactions.

Figure 4. Ligand-protein 2D and 3D interactions of hit compounds; D) NP-4, including Hie 41, and Asn 142 E) NP-5, including Hie 41, Glu 166, and Tyr 54 F) NP-6, including Hie 41, Glu 166, Gln 189, Thr 26, Gly 143, and Cys 145 π-π stacking and hydrogen bond interactions.

of the phenolic ring. Also, the obtained docking score and the binding energy of NP-4 (ZINC98364422) were -10.315 kcal/mol and -46.14 kcal/mol, respectively. The benzene ring induced π-π stacking with Hie41, and the nitrogen of amide's binds to Asn142 in NP-4 compound. NP-5 (Lonchocarpol A), a flavonoid, is another hit compound with a good docking score of -10.234 kcal/mol. Based on

our computational study, this compound interacts with the M^pro using hydrogen bonds between the hydroxy groups and the amino acids including Tyr54, Glu166, Gly143, and Leu141. A π-π stacking interaction is shown between the phenolic ring and Hie41 amino acid. The Cys145 and Gly143 interact with the carboxyl and hydroxyl groups, including hydrogen interaction with Glu166, Gln189,

Hie41, and Thr26.

Among all compounds, compound NP-20 (ZINC08765174) (an alkaloid) possessed the best docking score of -11.5 kcal/mol due to hydrogen bond and π-π stacking interactions with the key residues within the active site of COVID-19 M^pro including Gly143, Cys145, Glu146, His41 (Figure 5),[27,28] and Glide score -11.488 kcal/mol, and IFD scores -673.42 kcal/mol against the protease.

Molecular dynamics simulation

According to the obtained results from MD simulation analysis, NP-1 docking performance under the MD condition, approximately -6,685 kcal/mol per 150 steps were obtained, which means the total dynamic simulation time was divided into 30 stages. In each step lasting 20 ns, temperature, energy, and other variable factors were controlled and recorded. Consequently, 150 samples were collected in different steps of the whole experiment. The three fixed hydrogen bonds were observed in amino acids Thr26, Gly143, and Glu166 with a mean distance of 1.7, 1.7, and 1.8 Å. A variable and momentary hydrogen bonding with Asn142 was also observed, making this compound an efficient binder. In the formation of hydrogen bonds at sufficient ligand-protein distances, water molecules have played a significant role. The results of dynamics for the processed compound are represented in Table 2 and the molecular dynamics simulation video (MDS1, supporting information). OPLS3 force field RMSD index for complex estimation was 1.3 Å at its highest and 0.717 Å at its best, confirming the precision of the measurements (The standard RMSD index of the OPLS3 force field is below 2 Å).

Table 2. Molecular dynamics simulation results for NP-1.

Items	Results
Total Energy	-64905.1758 kJ/mol
Stretch	739.2184 kJ/mol
Bend	2718.0793 kJ/mol
Torsion	1932.8824 kJ/mol
Improper Torsion	106.4833 kJ/mol
VDW	-4240.6001 kJ/mol
Electrostatic	-57127.4062 kJ/mol
Explicit Hydrogen Bonds	0
Cross Terms	0
Solvation	-9033.8320 kJ/mol
T.E. for cross-checking	-64905.1758 kJ/mol
Iterations	500 out of 500
Conf 32 E	-64201.418 (1.309)

Discussion

Molecular docking

The nineteen structures, selected based on docking studies, are represented in Figure 2. The compounds belong to different classes of natural products such as alkaloids (NP-19, NP-4, NP-15, NP-20), flavonoids (NP-5, NP-7, NP-9, NP-3), quinones (NP-6, NP-10), coumarins (NP12, NP-13, NP-18), and some other natural products-derived. The obtained docking scores fall in the range from -8.049 to -11.5 kcal/mol.

Among alkaloids, compound NP-20 (ZINC08765174) possessed the best docking score of -11.488 kcal/mol due to hydrogen bond and π-π stacking interactions with Gly143, Cys145, Glu146, His41 (Figure 3), Glide score of -11.488 kcal/mol, and IFD score of -673.42 kcal/mol against the protease. This compound was also the best scoring one among all screened databases. The free binding energy of the best scoring compound was calculated using MM-GBSA. The free binding energy of NP-4 was the lowest (-46.14 kcal/mol) in this class of natural products.

As described before, some flavonoids have been identified as coronavirus inhibitors; four of them are shown in Figure 2. Among the screened flavonoids, NP-3 (ZINC96114284) was the best scoring compound with a docking score of -10.743 kcal/mol. This ligand interacted with the M^pro active site residues including, Cys145, Gly143, and Thr190, through hydrogen bonding with a docking score of -10.743 kcal/mol, Glide score of -10.752 kcal/mol, and IFD score of -669.73 kcal/mol. The free binding energy of 3 compounds in this group was lower than -36 kcal/mol.

Five compounds comprised quinone and coumarin skeletons out of the identified hit compounds. NP-6 (ZINC77257242) is an anthraquinone derivative that showed a docking score of -10.102 kcal/mol due to hydrogen bond interactions of the ligand with Glu166, Gln186, Gly143, Cys145, Thr26, and Hie26 amino acids. The Glide score and IFD score values of this ligand were -10.279 and -666.89 kcal/mol, respectively. The compound NP-10 (ZINC77257242) possessed the lowest free binding energy of -66.04 Kcal/mol and the amount of the free binding energy for other compounds in this class was lower than -36 kcal/mol.

Other derivatives of natural products have been identified as hit components. The compound NP-1 (ZINC03839114) was showing hydrogen bond interactions with Gln189, Glu166, His163, Leu141, and Cys145 amino acids with a docking score of -11.132 kcal/mol, Glide score of -11.132 Kcal/mol, and IFD score of -674.28 Kcal/mol.

The free binding energy was calculated for these compounds and it was lower than -36 Kcal/mol for all of the ligands in this class. The free binding energy of the compound NP-2 (ZINC03841676) was the lowest (-53.07 kcal/mol).

Among the entire compounds, the best values of the free binding energy were for compounds NP-1, NP-2, NP-4, NP-8, NP-10, NP-11, and NP-16. The free binding energy of compound NP-10 (ZINC06092274) was the lowest (-66.04 kcal/mol). Eventually, 21 compounds showed free
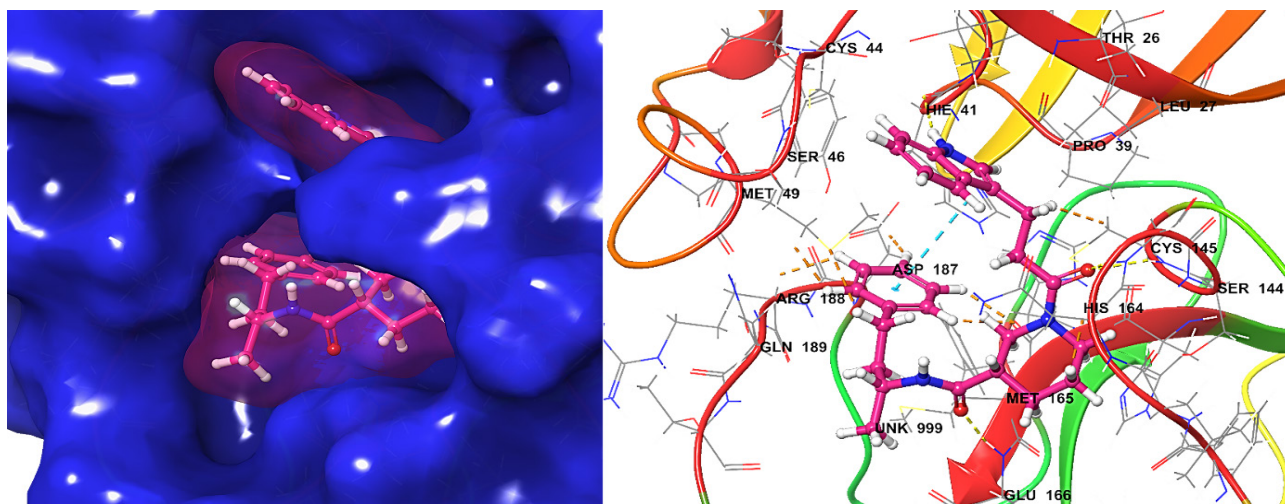
Figure 5. The binding pose of the receptor-binding domain – NP-20 (ZINC08765174) complex, including Gly143, Cys145, Glu146, Hie41 hydrogen bond and π-π stacking interactions.

binding energy lower than -36 kcal/mol. Figure 6 illustrates the molecular docking and ligand-protein interaction results for references such as peramivir, laninamivir, and hydroxychloroquine. Ultimately, the free binding energies of the examined compounds were lower than -35 kcal/mol. Among the suggested hit compounds, some alkaloid derivatives, especially aporphines and isoboldin, have shown antibacterial activity and potent activity against *S. aureus* and *E. coli*, respectively.[29] According to the results presented in Table 1, isoboldin had shown very high free binding energy (-45.22kcal/mol) to SARS-CoV-2 Mpro, and the docking score of this compound is -8.094 kcal/mol.

Catechin is a flavonoid isolated from *Bergenia crassifolia* rhizomes, garden tea leaves (*Camellia sinensis* L), and some other plants.[30,31] Also, it was reported to show antimicrobial and antioxidant activities.[31,32] For this compound, the free binging energy was calculated to be -37.36 kcal/mol and the docking score was -9.444 kcal/mol. Lonchocarpol A, a flavonoid isolated from the stem bark of *Erythrina fusca*, has been reported to show antimicrobial activity.[33,34] Lonchocarpol A showed moderate binding affinity to SARS-CoV-2 Mpro (-39.96 kcal/mol) with a docking score of -10.234 kcal/mol.
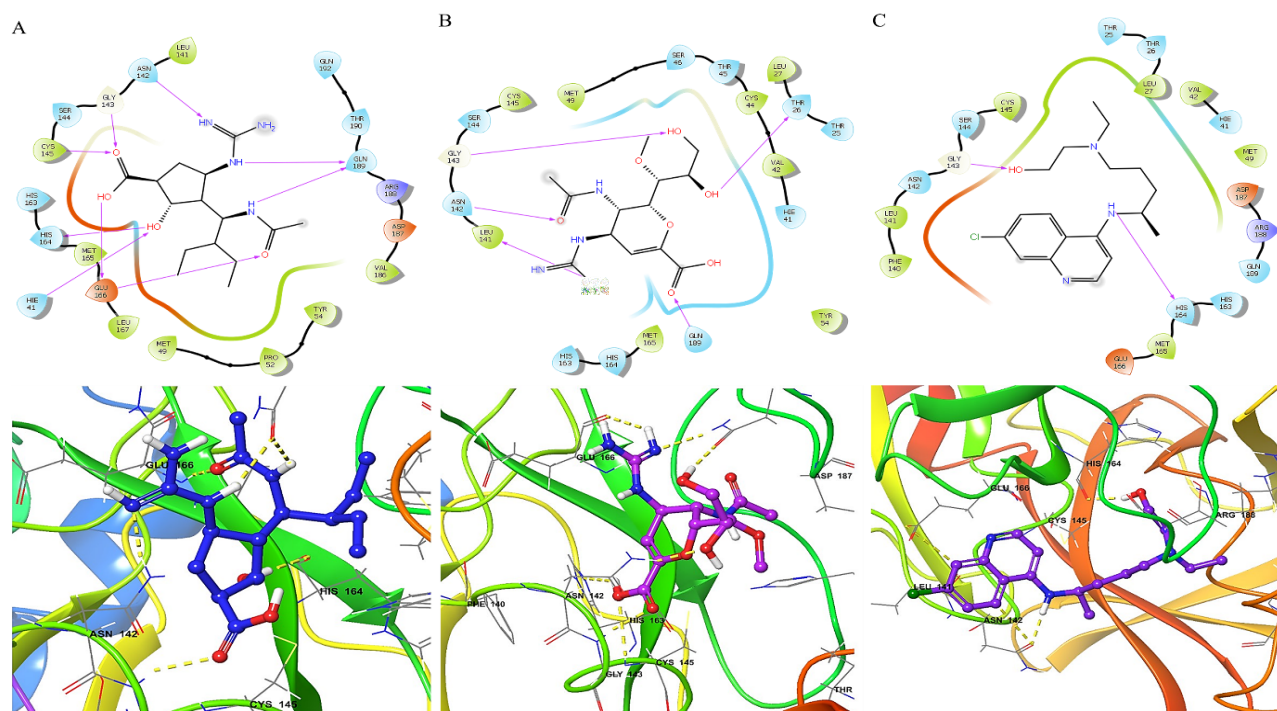
Figure 6. The 2D and 3D ligand-protein interactions of reference compounds with the active site of amino acid residues: A) peramivir, Asn142, Cys145, Gln189 and 166, His164, and Hie41 B) laninamivir, Thr26, Asn142, Leu141, Gly 164, and Glu189 C) hydroxychloroquine, Gly143, and His164 hydrogen bond interaction.

Absorption, Distribution, Metabolism, and Excretion (ADME)

Drug likeness of the initially selected 36 natural products (data of all 36 compounds have been listed in Table S1 in Supplementary Data) was predicted by the QikProp panel of Schrödinger maestro suite. Lipinski's rule of five was used to test bioavailability characteristics (ADME) of these compounds. According to the instruction, compounds with molecular weight ≤ 500, hydrogen bond donors ≤ 5 and acceptors ≤ 10, calculated octanol-water partition coefficient, and log P ≤ 5 possess good oral bioavailability.[35] In this study, the ADME properties of the hit compounds were estimated.

All compounds possess at least one hydrogen bond donor and four hydrogen bond acceptors. Polar surface area (PSA), central nervous system (CNS) activity, and percent oral absorption of hit compounds were predicted. According to the obtained results, these natural products showed good pharmacokinetic properties *in silico*.

The hydrophilicity of all compounds was determined by calculating the log P. It has been suggested that the log P value must be less than five, and high log P results in low absorption. According to this study, the partition coefficient (QPlog Po/w) was within the permissible range for the selected natural products, and the calculated PSA was within the range of 7.0-200.0 Å. CNS activity of these natural products was evaluated. According to the results, all selected natural products were classified as CNS-inactive compounds.

The human oral absorption percentage of 19 compounds was in the appropriate range of 81 to 100%. Compounds NP-2, NP-5, NP-8, NP-11, NP-13, NP-18, and NP-20 showed 100% oral absorption. Lonchocarpol A, a flavonoid with antimicrobial activity, showed 100% oral absorption.[33,34] Also, the entire hit compounds showed over 50% oral absorption. According to Table 1, no violations of Lipinski's rule (polar surface area, molecular weight, number of hydrogen donors, and acceptors) were found for hit compounds, and all mentioned properties were within the allowed range indicating their potential as a drug-like molecule.

Aqueous solubility (log S) is one of the most significant properties in drug discovery. The distributions of hit compounds obtained from HTVS are represented in Figure 7. The distributions are based on the values of blood-brain barrier permeation, octanol/water partition coefficient, aqueous solubility, and percent absorption. Approximately 99% of HTVS results possessed predicted QPlogBB within the acceptable range (-3 to 1.2), 70% of compounds had up to 80% human oral absorption, and about 99% were within the acceptable range of predicted QPlogPo/w (-2 to 5). An increase in the QPlogPo/w values of the compounds
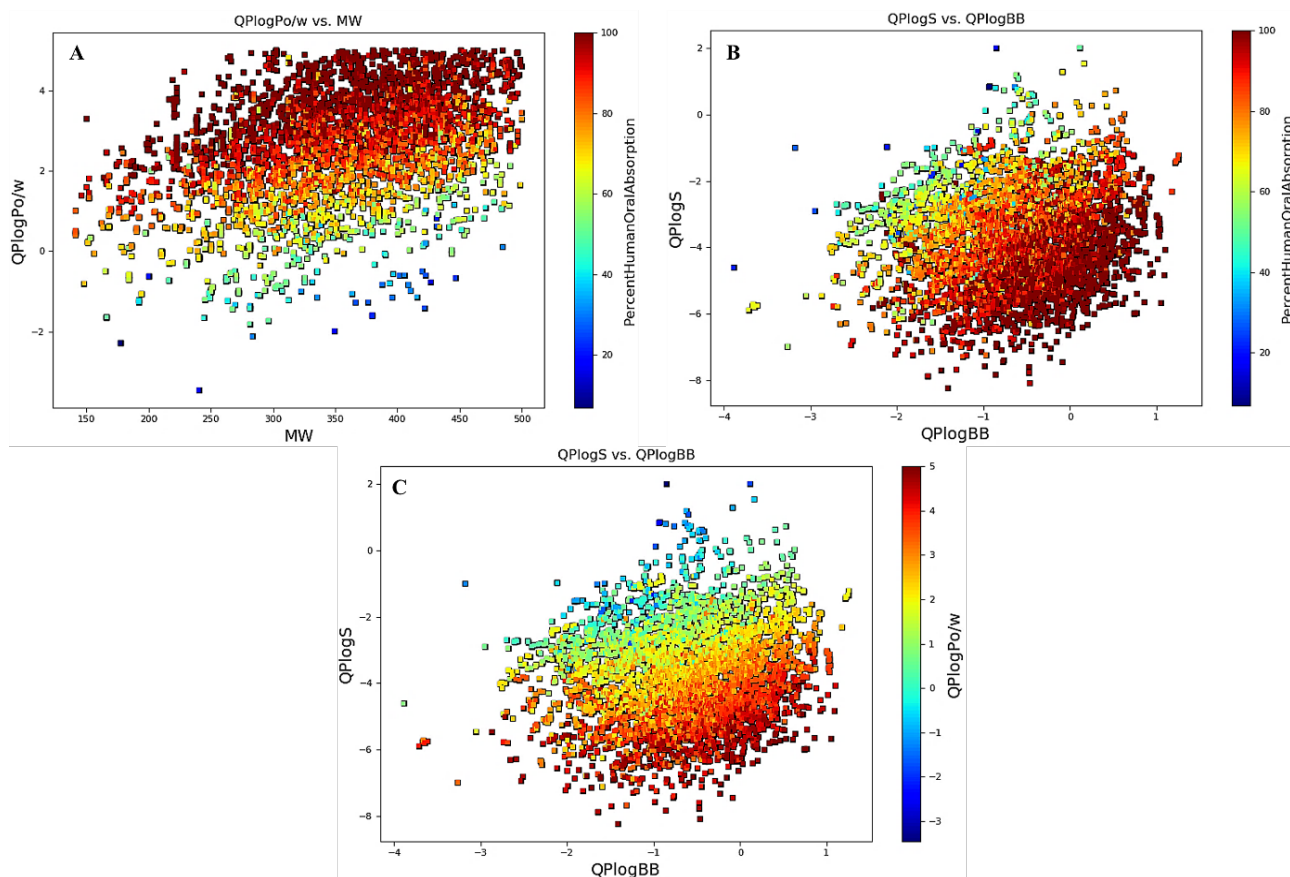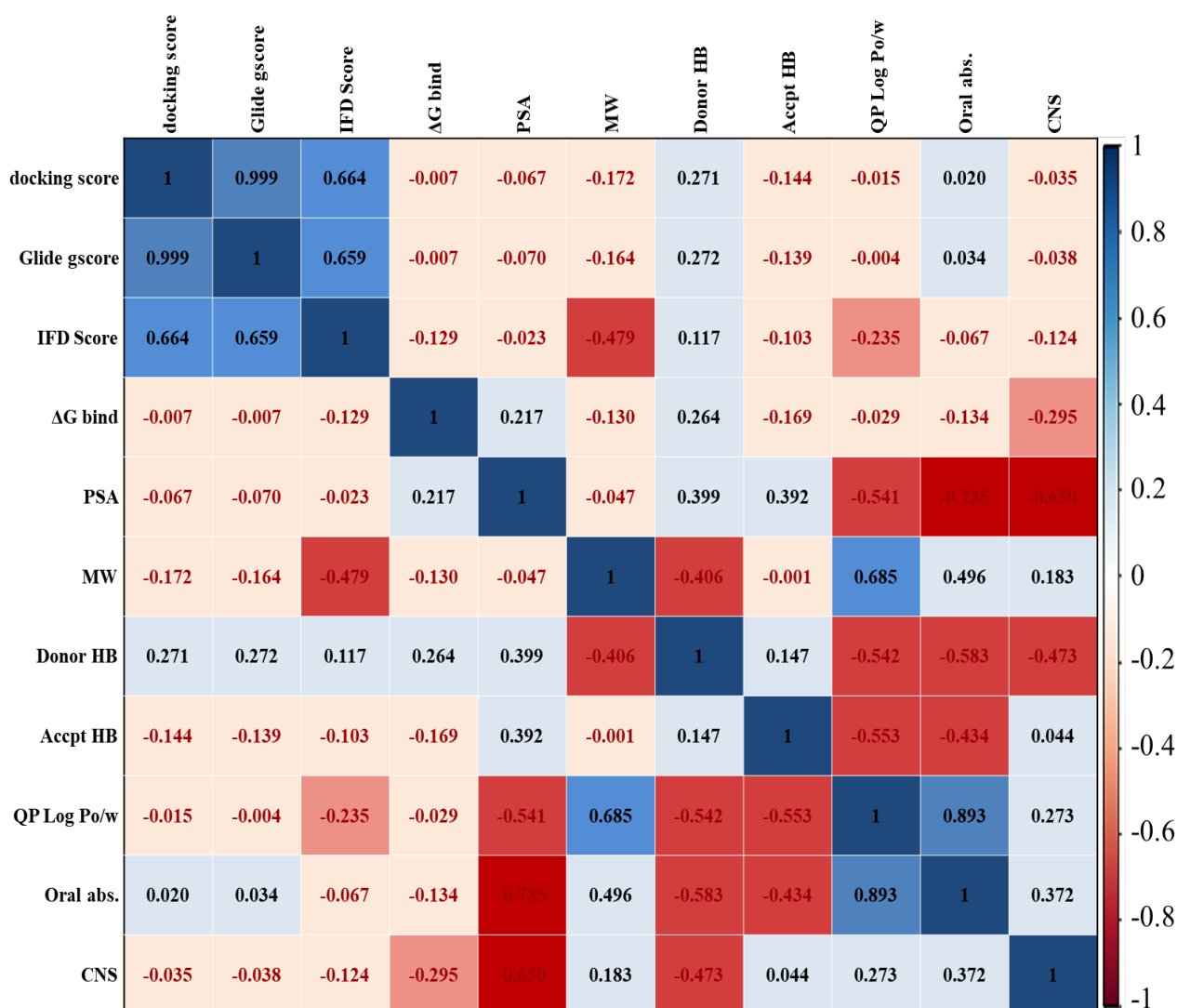

Figure 7. Plots of the distribution of the HTVS compounds A) Compounds distribution of logPo/w vs. MW. B) Compounds distribution of logS vs. logBB. The absorption percentage was colored for both plots. C) Compounds distribution of logS vs. logBB. The logPo/w was colored.

reflects an increase in the absorption percentage. Also, the absorption percentage has a direct relation with QPlogPo/w and QPlogBB. On the contrary, the opposite relation between QPlogPo/w and QPlogS can be observed. The statistical relationship between the variables for the last 40 compounds was investigated using a graph of the correlation matrix shown in Figure 8. According to the matrix, the relationships between the docking score, glide and IFD score, along with IFD and glide score are distinguished. Furthermore, the QPlogPo/w correlation is significantly related to MW and oral absorption.

Analyzing the MD simulation

One of the most critical variable factors in a dynamic system is the temperature, which allows the computational analysis to push through the relative potential and get closer to the global minimum. The criteria to achieve equilibrium and convergence in the dynamic system have been described in the terminal cycles in order to reproduce

comparable energies. Figure 9 shows that the mechanism has reached an appropriate stable equilibrium. In the project, the achievement of a convergence was defined in terms of energy. The convergence threshold was assessed as 0.05 kJ/mol according to the standards. This means that if two or more calculated potential energy values differ by only 0.05 kJ/mol, the system is iteratively optimized. Moreover, if an energy value reaches 46000 kJ/mol in 15 to 20 nanoseconds, this shows that the system has reached energy stability. Furthermore, the low RMSD index (less than 2 Å) indicates that the compounds are very well placed in the active site and that the system has properly identified the active site (Figure S1). The standard deviation in relation to the natural ligand is negligible. The video of this analysis clearly shows that the amino acid glycine number 143 plays a vital role (supplementary material). The binding of the amino acid to the compound leads the compound to the active site. Next, other hydrogen bonds cause persistence of the compound in the active site (Figure 10).

	docking score	Glide gscore	IFD Score	ΔG bind	PSA	MW	Donor HB	Accpt HB	QP Log Po/w	Oral abs.	CNS
docking score	1	0.999	0.664	-0.007	-0.067	-0.172	0.271	-0.144	-0.015	0.020	-0.035
Glide gscore	0.999	1	0.659	-0.007	-0.070	-0.164	0.272	-0.139	-0.004	0.034	-0.038
IFD Score	0.664	0.659	1	-0.129	-0.023	-0.479	0.117	-0.103	-0.235	-0.067	-0.124
ΔG bind	-0.007	-0.007	-0.129	1	0.217	-0.130	0.264	-0.169	-0.029	-0.134	-0.295
PSA	-0.067	-0.070	-0.023	0.217	1	-0.047	0.399	0.392	-0.541	-0.735	-0.650
MW	-0.172	-0.164	-0.479	-0.130	-0.047	1	-0.406	-0.001	0.685	0.496	0.183
Donor HB	0.271	0.272	0.117	0.264	0.399	-0.406	1	0.147	-0.542	-0.583	-0.473
Accpt HB	-0.144	-0.139	-0.103	-0.169	0.392	-0.001	0.147	1	-0.553	-0.434	0.044
QP Log Po/w	-0.015	-0.004	-0.235	-0.029	-0.541	0.685	-0.542	-0.553	1	0.893	0.273
Oral abs.	0.020	0.034	-0.067	-0.134	-0.735	0.496	-0.583	-0.434	0.893	1	0.372
CNS	-0.035	-0.038	-0.124	-0.295	-0.650	0.183	-0.473	0.044	0.273	0.372	1

Figure 8. Visualized correlation matrix chart for variable data. The existence of logical relationships between variables causes a change of the heatmap to blue and a value of 1, whereas its absence leads to negative values and red color.
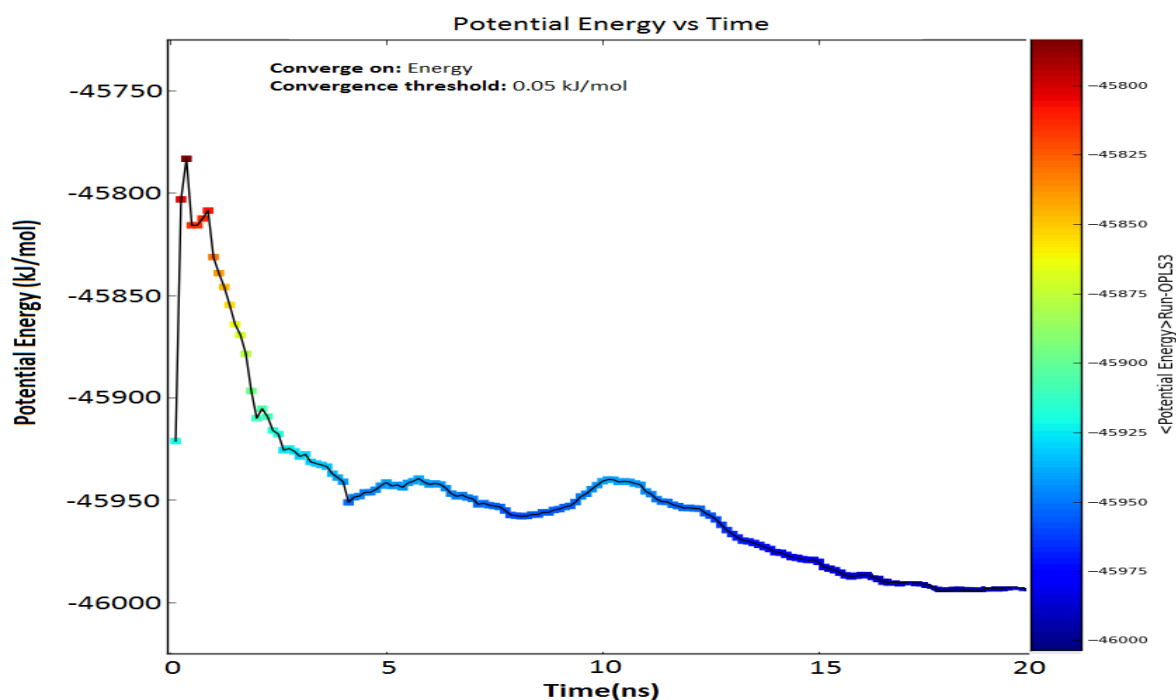
Figure 9. Achieve convergence in terms of energy in OPLS3 force fields at MD: potential energy vs. time for NP-20.
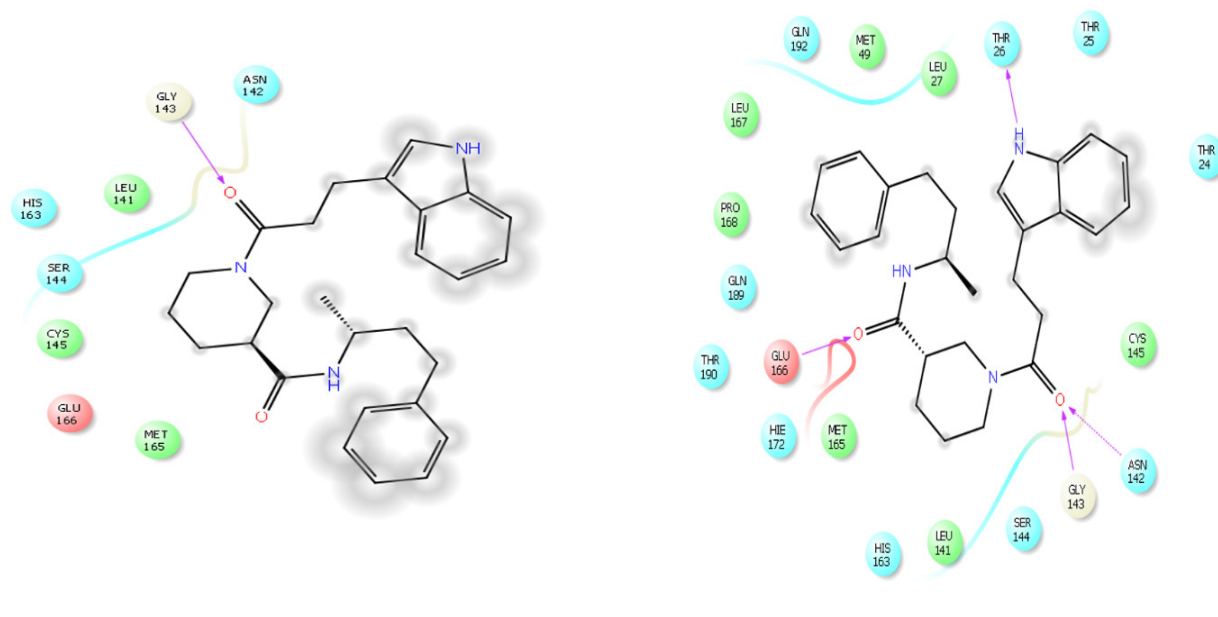

Figure 10. The 2D representation of first contact of the ligand by the protein via the amino acid glycine 143.

Conclusion

The docking scores for selected compounds varied from -8.049 to -11.132 kcal/mol. The PSA, number of hydrogen binding acceptors and donors, molecular weights, and partition coefficient were all within the allowable range for all selected compounds. In conclusion, based on ADME and the free energy binding values of the affected compounds, NP-1 through NP-19 (isoboldin) were selected as potential inhibitors of M[pro] 2019-nCoV. ZINC08765174 (1-[3-(1H-indol-3-yl) propanoyl]-N-(4-phenylbutan-2-yl) piperidine-3-carboxamide) was proposed as a potential compound to treat COVID-19. It showed the highest liaison affinity with the M[pro] of COVID-19 without violation of the Lipinski rules.

Author Contributions
MO and MB performed the molecular modeling part and preparing the draft of the manuscript. KA completed the molecular dynamics of the active ligand and presenting data. SNE and PM supervised the project, finalizing the manuscript. All authors have approved the paper.

Conflict of Interest
The authors declare that there is no competing interest.

Supplementary Data
Glide Docking and DFI scores, with ADME screening results for approximately 40 compounds supplied. The MD simulation film displayed.

References
1. Zhu N, Zhang D, Wang W, Li X, Yang B, Song J, et al. A novel coronavirus from patients with pneumonia in china, 2019. N Engl J Med. 2020;382(8):727-33. doi: 10.1056/NEJMoa2001017
2. World Health Organization. Covid-19 weekly epidemiological update, 1 December 2020.
3. Lo CY, Tsai TL, Lin CN, Lin CH, Wu HY. Interaction of coronavirus nucleocapsid protein with the 5′-and 3′-ends of the coronavirus genome is involved in genome circularization and negative-strand rna synthesis. FEBS J. 2019;286(16):3222-39. doi:10.1111/febs.14863
4. Peiris J, Lai S, Poon L, Guan Y, Yam L, Lim W, et al. Coronavirus as a possible cause of severe acute respiratory syndrome. Lancet. 2003;361(9366):1319-25. doi:10.1016/S0140-6736(03)13077-2
5. Raj VS, Osterhaus AD, Fouchier RA, Haagmans BL. Mers: Emergence of a novel human coronavirus. Curr Opin Virol. 2014;5:58-62. doi:10.1016/j.coviro.2014.01.010
6. Wu F, Zhao S, Yu B, Chen Y-M, Wang W, Song Z-G, et al. A new coronavirus associated with human respiratory disease in china. Nature. 2020;579(7798):265-9. doi:10.1038/s41586-020-2008-3
7. Contini A. Virtual screening of an FDA approved drugs database on two covid-19 coronavirus proteins. Life Sci. 2020; 251:117627. doi:10.1016/j.lfs.2020.117627
8. Ryu YB, Jeong HJ, Kim JH, Kim YM, Park J-Y, Kim D, et al. Biflavonoids from torreya nucifera displaying sars-cov 3clpro inhibition. Bioorg Med Chem. 2010;18(22):7940-7. doi:10.1016/j.bmc.2010.09.035
9. Lau K-M, Lee K-M, Koon C-M, Cheung CS-F, Lau C-P, Ho H-M, et al. Immunomodulatory and anti-SARS activities of *Houttuynia cordata*. J Ethnopharmacol. 2008;118(1):79-85. doi: 10.1016/j.jep.2008.03.018
10. Recio MdC, Just MJ, Giner RM, Mañez S, Rios JL, Hostettmann K. Anti-inflammatory activity of saikosaponins from *Heteromorpha trifoliata*. J Nat Prod. 1995;58(1):140-4. doi: 10.1021/np50115a023
11. Cheng PW, Ng LT, Chiang LC, Lin CC. Antiviral effects of saikosaponins on human coronavirus 229e in vitro. Clin Exp Pharmacol Physiol. 2006;33(7):612-6. doi:10.1111/j.1440-1681.2006.04415.x
12. Yu M-S, Lee J, Lee JM, Kim Y, Chin Y-W, Jee J-G, et al. Identification of myricetin and scutellarein as novel chemical inhibitors of the sars coronavirus helicase, nsp13. Bioorg Med Chem Lett. 2012;22(12):4049-54. doi:10.1016/j.bmcl.2012.04.081
13. Wollenweber E. Flavones and flavonols. The flavonoids advances in research since 1986. Oxfordshire: Routledge; 2017. p. 259-335.
14. Hertog MG, Feskens EJ, Kromhout D, Hollman P, Katan M. Dietary antioxidant flavonoids and risk of coronary heart disease: The zutphen elderly study. Lancet. 1993;342(8878):1007-11. doi:10.1016/0140-6736(93)92876-U
15. Park J-Y, Yuk HJ, Ryu HW, Lim SH, Kim KS, Park KH, et al. Evaluation of polyphenols from *Broussonetia papyrifera* as coronavirus protease inhibitors. J Enzyme Inhib Med Chem. 2017;32(1):504-12. doi:10.1080/14756366.2016.1265519
16. Kumar A, Choudhir G, Shukla SK, Sharma M, Tyagi P, Bhushan A, et al. Identification of phytochemical inhibitors against main protease of covid-19 using molecular modeling approaches. J Biomol Struct Dyn. 2021;39(10):3760-70. doi:10.1080/07391102.2020.1772112
17. Aanouz I, Belhassan A, El-Khatabi K, Lakhlifi T, El-Ldrissi M, Bouachrine M. Moroccan medicinal plants as inhibitors against SARS-CoV-2 main protease: Computational investigations. J Biomol Struct Dyn. 2020;39(8):2971-9. doi:10.1080/07391102.2020.1758790
18. Narkhede RR, Pise AV, Cheke RS, Shinde SD. Recognition of natural products as potential inhibitors of COVID-19 main protease (Mpro): In-silico evidences. Nat Prod Bioprospect. 2020;10(5):297-306. doi:10.1007/s13659-020-00253-1
19. Adem S, Eyupoglu V, Sarfraz I, Rasul A, Ali M. Identification of potent COVID-19 main protease (Mpro) inhibitors from natural polyphenols: An in silico strategy unveils a hope against corona. Preprints. 2020;2020030333. doi:10.20944/preprints202003.0333.v1
20. Cheng L, Zheng W, Li M, Huang J, Bao S, Xu Q, et al. Citrus fruits are rich in flavonoids for immunoregulation and potential targeting ACE2. Preprints. 2020;2020020313.
21. Omar S, Bouziane I, Bouslama Z, Djemel A. In-Silico Identification of potent inhibitors of covid-19 main protease (Mpro) and angiotensin converting enzyme 2 (ACE2) from natural products: quercetin, hispidulin, and cirsimaritin exhibited better potential inhibition than hydroxy-chloroquine against covid-19 main protease active site and ACE2. ChemRxiv. 2020. doi:

10.26434/chemrxiv.12181404.v1

22. Chen H, Du Q. Potential natural compounds for preventing SARS-CoV-2 (2019-nCoV) infection. Preprints. 2020. 2020010358. doi:10.20944/preprints202001.0358.v3

23. Khandelwal A, Sharma T. Computational screening of phytochemicals from medicinal plants as covid-19 inhibitors. ChemRxiv. 2020. doi:10.26434/chemrxiv.12320273.v1

24. Lai L, Han X, Chen H, Wei P, Huang C, Liu S, et al. Quaternary structure, substrate selectivity and inhibitor design for SARS 3c-like proteinase. Curr Pharm Des. 2006;12(35):4555-64. doi:10.2174/138161206779010396

25. Jin Z, Du X, Xu Y, Deng Y, Liu M, Zhao Y, et al. Structure of m(pro) from -2 and discovery of its inhibitors. Nature. 2020;582(7811):289-93. doi:10.1038/s41586-020-2223-y

26. Duan L, Guo X, Cong Y, Feng G, Li Y, Zhang JZ. Accelerated molecular dynamics simulation for helical proteins folding in explicit water. Front Chem. 2019;7:540. doi:10.3389/fchem.2019.00540

27. Al-Shar'i NA. Tackling COVID-19: Identification of potential main protease inhibitors via structural analysis, virtual screening, molecular docking and MM-PBSA calculations. J Biomol Struct Dyn. 2020. doi:10.1080/07391102.2020.1800514

28. 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. J Biomol Struct Dyn. 2021;39(10):3662-80. doi:10.1080/07391102.2020.1768151

29. Abbasoglu U, Sener B, Günay Y, Temizer H. Antimicrobial activity of some isoquinoline alkaloids. Arch Pharm. 1991;324(6):379-80. doi:10.1002/ardp.19913240612

30. Gadkari PV, Kadimi US, Balaraman M. Catechin concentrates of garden tea leaves (*Camellia sinensis* L.): Extraction/isolation and evaluation of chemical composition. J Sci Food Agric. 2014;94(14):2921-8. doi:10.1002/jsfa.6633

31. Katalinić V, Milos M, Modun D, Musić I, Boban M. Antioxidant effectiveness of selected wines in comparison with (+)-catechin. Food Chem. 2004;86(4):593-600. doi: 10.1016/j.foodchem.2003.10.007

32. Bais HP, Walker TS, Stermitz FR, Hufbauer RA, Vivanco JM. Enantiomeric-dependent phytotoxic and antimicrobial activity of (±)-catechin. A rhizosecreted racemic mixture from spotted knapweed. Plant Physiol. 2002;128(4):1173-9. doi:10.1104/pp.011019

33. Salvatore MJ, King AB, Graham AC, Onishi HR, Bartizal KF, Abruzzo GK, et al. Antibacterial activity of lonchocarpol A. J Nat Prod. 1998;61(5):640-2. doi:10.1021/np9703961

34. Khaomek P, Ichino C, Ishiyama A, Sekiguchi H, Namatame M, Ruangrungsi N, et al. In vitro antimalarial activity of prenylated flavonoids from erythrina fusca. J Nat Med. 2008;62(2):217-20. doi:10.1007/s11418-007-0214-z

35. Proudfoot JR. Drugs, leads, and drug-likeness: An analysis of some recently launched drugs. Bioorg Med Chem Lett. 2002;12(12):1647-50. doi:10.1016/S0960-894X(02)00244-5