Screening of Natural Product and Natural Product like Molecules against SARS–CoV–2 Main Protease Using Molecular Modeling Methods

Ismail Hakki Akgün[1*]

[1]*Bioengineering Department, Faculty of Engineering, Ege University, 35100, Bornova, İzmir, Türkiye.*

Author contribution

The sole author designed, analysed, interpreted and prepared the manuscript.

Original Research Article

ABSTRACT

Objective: To determine possible M^{Pro} enzyme inhibitors by using structure-based virtual screening methods, in the ZINC Biogenic Data Set containing natural products and natural product-like molecules.
Materials and Methods: QVina, an AutoDockVina derivative, was used in virtual screening operations, GROMACS in molecular dynamics studies and SwissAdme server in ADME (Absorption, Distribution, Metabolism, and Excretion) calculations. KNIME (Konstanz Information Miner) and ChemAxon software were used for filtering data and creating three-dimensional structures of the molecules.
Results: Seven out of totally screened 51535 natural products or natural products like molecules were identified as possible candidate to be used as SARS–CoV–2 Main Protease (MPro) enzyme inhibitors based on the results obtained from structure based virtual screening and ADME models.

Corresponding author: E-mail: ismail.hakki.akgun@ege.edu.tr;

Conclusion: Among the seven potent molecules, two of them (ZINC000604382012 and ZINC000514288074) were selected as candidate molecules for further studies according to the results obtained from g_mmpbsa simulations and synthetic accessibility models. In addition, a workflow has been established to identify novel or potent M^{pro} enzyme inhibitors.

1. INTRODUCTION

The SARS-CoV-2 (COVID-19) outbreak emerged in Wuhan, China, towards the end of 2019 and soon turned into a pandemic. Although the numbers are not certain, more than 85 million people have been affected by the disease and more than 1.5 million people have died so far. [1] Significant progress has been made in developing vaccines to prevent the infection of SARS-CoV-2 and mass vaccination studies have started in some countries. However, there is still no significant drug developed for use in treatment. [2] Multiple approaches are being evaluated in the treatment of the virus, and studies are still ongoing. Preventing the entry of the virus into the cell, preventing its replication, slowing the autophagy of the host cell, etc. are the examples of these approaches. Antibodies, peptides, proteins and small molecules can be used to block the entry of the virus into the cell. Proteases such as RNA-dependent RNA polymerase (RdRp), papain-like cysteine protease (PL^{pro}) or main protease (M^{Pro}) can be targeted to circumvent the virus replication. [3] In order to inhibit these proteases, studies have been carried out with molecules developed or under development for some other viruses. For example, Choy et al. studied the inhibition of the replication of SARS-CoV-2 in Vero-E6 cells with 16 antivirals, including the well-known HIV M^{pro} (3CL^{pro}) inhibitor, lopinavir EC$_{50}$ value of remdesivir was found to be 23.15 µM and EC$_{50}$ value of lopinavir was found to be 26.63 µM in the in-vitro assay used in the research. In the same study, the EC$_{50}$ value of emetine, which is an anti-protozoal molecule, was determined as 0.46 µM. [4] However, the expected clinical results were not obtained from the trail performed with the aforementioned lopinavir and its companion ritonavir.[5]

Similar unsatisfactory results from such well-known antivirals have prompted researchers to conduct studies for the discovery of new antivirals. As the first attempts, drug repurposing approach was carried out with the existing drug molecules. [6-8] Although there are many reports about the fact that quite a large number of drug molecules can find various uses at different points in the life cycle of the virus, an accepted protocol has still not been reached. Because of this situation, it has become important to screen novel molecules that may be lead bioactive / drug molecules and to carry out studies on using them to cure viral infection.

Virtual screening methods are one of the first and frequently used methods to identify and develop such novel bioactive / drug molecule or molecules. [9] This method can be used alone to identify lead molecules that can be used in studies, as well as a complement to high-throughput screening (HTS) studies. [10] There are many successful examples of discovering a novel bioactive, lead-like molecule using this approach. [11] For example, Vangrevelinghe and colleagues analyzed 400000 molecules using virtual screening methods and determined potent and selective CK2 inhibitor molecules among the prioritized molecules using docking scores. [12] Another example of the successful application of this method was the determination of Dipeptidyl Peptidase IV Inhibitors by Ward and colleagues. As a result of their virtual screening studies using 800000 molecules, they tested the bioactivity of 4000 molecules and discovered approximately 50 inhibitors. [13] Various virtual screening studies have been conducted in which not only inhibitors but also receptor agonists were discovered. Schapira and colleagues have discovered two new nuclear hormone receptor agonists using rational virtual screening methods. [14] While it is possible to increase these and similar examples, we can recommend that those who want to get more information about the subject should read the chapter prepared by Matter and Sotriffer. [15]

From the past to the present, natural products have been quite common resources in the discovery of bioactive / drug molecules. Many of them are still used as medicine. Newman and Cragg's review article published in 2020 contains very important information about the use of natural products as drugs. [16] According to the

article, 71 (5.1%) of 1394 small molecule class drugs accepted as drugs by the FDA between 1981 and 2020 were natural products, 14 (1%) were natural products of plant origin, 356 (27.5%) were natural products or natural product derivatives, 434 (30.5%) are molecules that mimic natural productsalthough they are synthetic. For the mentioned molecules, the total reaches to 64.1 percent, indicating that more than half of the accepted molecules are natural products, natural products analogues or derivatives. [16] This rather high number is an indication that natural and its derivatives continue their existence as very important resources in the discovery of new bioactive / drug molecules. Considering the above-mentioned issues, we carried out virtual screening operations with natural products and natural product-like molecules deployed in ZINC "Biogenic Dataset" which can be developed as new SARS-CoV-2 M^{Pro} inhibitor or inhibitors.

2. MATERIALS AND METHODS

2.1 Materials

The ZINC "Biogenic Dataset" containing the molecules used in this study was downloaded on April 10, 2020. [17] The storage and processing of data were carried out in the Ubuntu 18.04 installed workstation with an i7 processor, Nvidia GTX 960 and GTX 1050 graphics card with 16 GB RAM capacity. Other materials and software used are described in the methods section in the relevant places.

2.2 Methods

2.2.1 Preparation of molecules

The downloaded dataset was transferred to KNIME (Konstanz Information Miner, KNIME AG, Zurich, Switzerland) and it was put into docking simulations after various filtration processes. [18] As the first step, molecules existing in the "Biogenic Dataset" catalog were loaded to KNIME workspace and smiles of molecules were transformed into chemical structures. Following this process, molecules marked as "in-vitro biogenic" were selected from the molecules in the catalog and the others were removed. This process was carried out to focus on working with secondary metabolites (natural products) or similar (natural products like) molecules. Upon this selection MannholdLogP, Hydrogen Bond Acceptors, Hydrogen Bond Donors, Rotatable Bonds Count, Lipinski's Rule of Five, Topological Polar Surface Area (A), Molecular Weight (g/mol - Da), XLogP, SP3 Character, and Rotatable

Bonds Count (non terminal) parameters were calculated using the "CDK Molecular Properties" node in order to determine the molecules that comply with the Lipinski rule of 5 rule (Ro5) and investigate distribution of this parameters of the molecules. [19] The compatibility of molecules with Ro5 was tested with the node "CDK Lipinski's Rule of Five" and compatible molecules were selected. Molecules passing through the Ro5 filter were transformed into a format that RDKit software can read with the "RDKitFrom Molecule" node, then hydrogens were added and their three-dimensional structures were optimized by using the MMFF94 force filed with a maximum of 1000 iterations with the help of the "RDKit Optimized Geometry" node. Molecules whose three-dimensional structure had not been optimized have been optimized with the help of ChemAxon Marvin Sketch. [20] All of the molecules were saved in sdf format and converted into mol2 and then pdbqt formats using OpenBabel. [21]

2.2.2 Screening binding affinities of molecules with docking

Docking simulations were performed on TRUBA servers with QVina 2.1, a derivative of AutoDockVina. [22, 23] The coordinates of the binding site were determined with the aid of the inhibitor N3 in the pdb file coded as 6LU7. [24] Accordingly, the x, y, z coordinates of the center of the box were defined as -10.83, 12.58 and 68.73, respectively. The dimensions of the box in x, y, z axes were determined as 18.75, 33.75 and 22.5 Å, respectively in order to cover residues in the binding pocket of the ligand N3. The docking simulation of each molecule was repeated once, requiring a maximum of five poses to be created and the difference in simulated binding energies between poses to be 0.5 kcal/mol (energy_range = 0.5). In addition, the parameter exhaustiveness (time taken to find a better binding poses) of each molecule was set to 64. The simulated bonding energies obtained from docking poses were compiled with the Python script written in house. The top 250 of the molecules with the lowest simulated binding energies were transferred to "SwissAdme Server" in order to calculate ADME (Absorption, Distribution, Metabolism, and Excretion) related properties. [25]

2.2.3 Calculation of ADME properties of the top 250 molecule

As we mentioned in the previous section, ADME properties of 250 molecules with the best

simulated binding energy were obtained using SwissAdme Server. Accordingly, the structures of the molecules were converted into smiles format and uploaded to the server. The results from the server were saved in csv format and processed (in) KNIME workspace. Molecules with high gastrointestinal absorption, not labeled as possible inhibitor of CYP isoenzymes, and with no PAINS and BRENK alerted were selected. In addition, by examining the common properties of drug molecules such as Lipinski, Ghose, Veber, Egan, Muegge, molecules that are compatible with the filters were selected and the filtering process at this step was completed. [26-31]

2.2.4 Screening binding affinities of molecules with second round docking

With the same target structure (6LU7) and same search space parameter second round of docking screenings were performed using molecules which passed ADME filters, with higher exhaustiveness value (128, default value is 8) to search possible better binding pose or poses and for triplicate to ensure about reproducibility of the docking simulation results.

2.2.5 Molecular dynamics and MM-PBSA studies

Molecular dynamics and MM-PBSA studies were performed using GROMACS 5.1.4 and g_mmpbsa software. [32,33] Data formats compatible with molecular dynamics studies of the molecules we screened were prepared using Acpype. [34] In the preparation of the topology files compatible with GROMACS of molecules, Acpype default settings were used and the obtained parameter files were used in MD simulations.The AMBER99SB force field was used to prepare the topology and coordinates of the receptor. [35] After combining the coordinate and topology files of the receptor and the molecule, the resulting system is placed in a dodecahedron water cube whose edges were set to be at least 1 nm away from the system created. The ion concentration of the cube formed was adjusted to 0.15 M using sodium and chloride ions. The energy minimization of the created system was carried out by using the steepest descent minimization method in a maximum of 50000 steps, when the maximum force falls below 10 kJ /mol. The equilibrium process of the energy minimized system was carried out in two steps by using NVT and NPT

ensembles. The positions of the proteins and molecules are fixed during the equilibrium simulations. NTV equilibrium process was continued for 100 ps with time step 2 fs, the temperature of the system was set to 300 K, modified Berendsen thermostat (V-rescale) was used as a thermostat. In NPT equilibrium process, time step 2 fs and simulation time was determined as 100 ps similar to NVT. Also, similar to NVT, proteins and molecules are fixed in their places. Brendensen was used as barostat in NPT and 1 bar was used as reference pressure. After these two equilibrium steps, the production MD step, in which proteins and molecules are released and their interactions are examined, was carried out. This step was advanced 2 ns (1,000,000 steps) in total by using 2 fs time steps, long range interactions were calculated with PME, neighbor searching Verlet cutoff – scheme. The leap-frog integrator was used as integrator in the production MD operations. As a result of the production MD operations, the shifts and rotations in the system were corrected, and then the RMSD values of the protein and molecules compared to the initial coordinates were calculated using GROMACS. Following these procedures, MM-PBSA method was applied with the help of g_mmpbsa module and binding free energies of the molecules were calculated. MM-PBSA calculations were performed by taking a sample at 200 ps (11 samples in total) using production MD simulations between 0 and 2 ns using default settings.

3. RESULTS AND DISCUSSION

After loading the 307814 molecules in the "Biogenic Dataset" to KNIME, and molecules marked as "in-vitro biogenic" in the catalog have been selected for further studies. The number of these molecules were determined as 85553.Table 1 shows the upper and lower limits of the parameters of these molecules before and after the Ro5 filter was applied. After the Ro5 filtration, the number of molecules decreased to 51542. The filtered molecules were transferred to "RDKit to Molecule" node and structure of the 51535 was generated correctly. Three-dimensional optimization was applied successfully to 50572 molecules and unsuccessfully to 963 molecules. The unsuccessful attempts were completed using ChemAxon software as mentioned in the method section.

Table 1. The upper and lower limits of the parameters of these molecules before and after the Ro5 filter was applied

	Before Ro5 filter applied (n = 85553)		After Ro5 filter applied (n = 51542)	
	Lower limit	Upper limit	Lower limit	Upper limit
MannholdLogP	-1.07	8.94	0.80	4.54
Hydrogen bond acceptors	0	31	0	10
Hydrogen bond donors	0	20	0	5
Rotatablebonds count	0	69	0	20
Topological polar surface area (Å^2)	0.00	506.13	0.00	242.31
Molecular weight (g/mol - Da)	58.005	1062.423	58.005	499.341
XLogP	-9.597	84.587	-5.235	5.000
SP3 character	0.000	0.444	0.000	0.392
Rotatablebonds count (non terminal)	0	62	0	10

When the simulated binding energies obtained from the docking simulation were examined, it was determined that the lowest value was -10.9 (best binding), the highest value was -2.5 (worst binding), and -7.105 mean value was the median was -7.2 kcal/mol. The histogram plot of the simulated binding energies is shown in Fig. 1. The best simulated binding energy results of the docking processes were in molecules coded as ZINC000015675941 and ZINC000247722436 with a value of -10.9 kcal/mol. (Fig. 2). The worst results were obtained in molecules coded as ZINC000033830853 and ZINC000085530484 with a value of -2.5 kcal/mol. (Fig. b). Binding poses of the worst two molecules were shown in Fig. 3. When the top 250 molecules were examined, it was determined that the lowest binding energy was -10.9 and the highest binding energy was -9.1 kcal/mol. In addition, it was determined that the average binding energy of this group of molecules was - 9.33 and the median value was -9.2 kcal/mol.
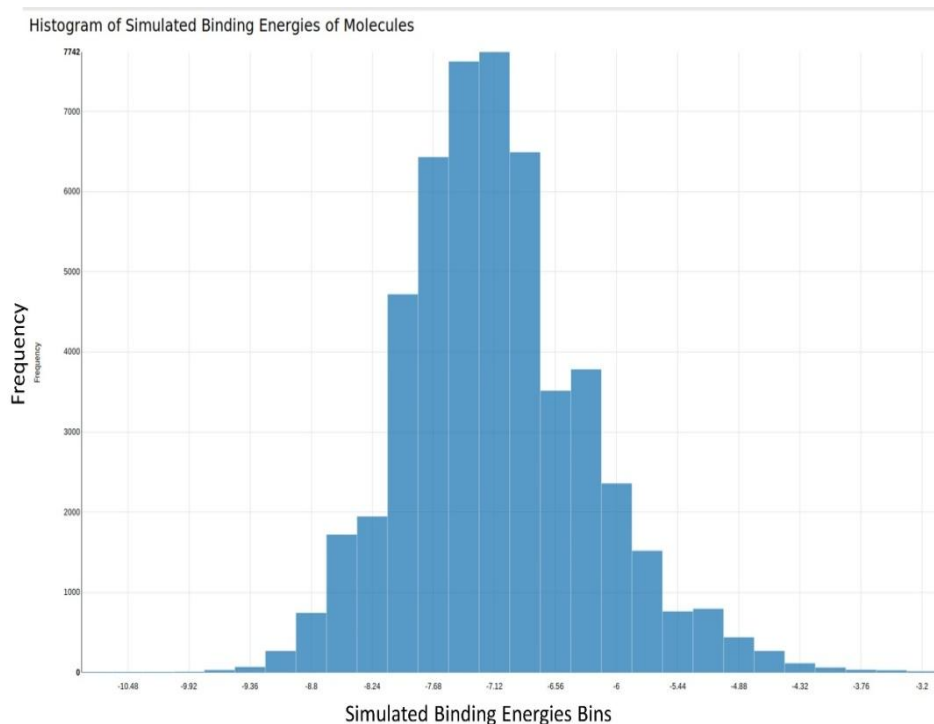
Fig. 1. The histogram plot of the simulated binding energies of 51535 molecules
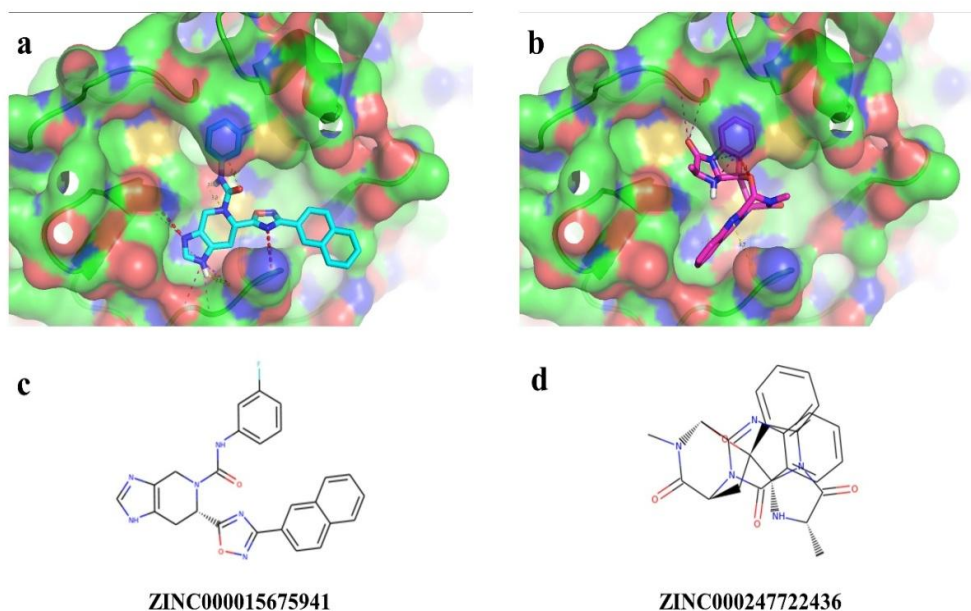
Fig. 2. Docking poses of ZINC000015675941 and ZINC000247722436 with a value of -10.9 kcal/mol
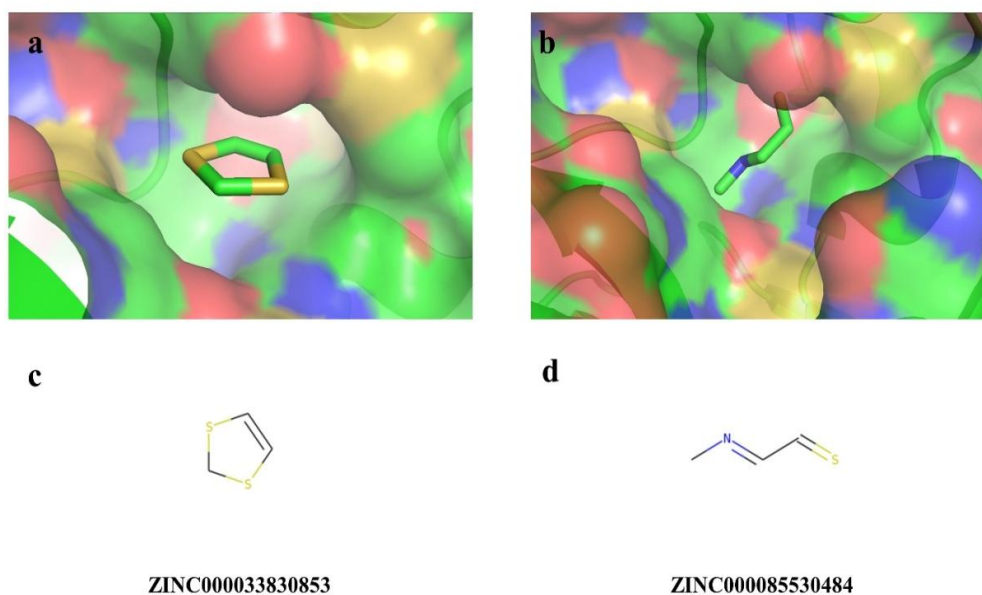

Fig. 3. Docking poses of ZINC000033830853 and ZINC000085530484 with a value of -2.5 kcal/mol

It is very important that the molecules to be developed or studied as drugs must be adsorbed in the gastrointestinal (GI) system. With the effect of the Ro5 filter we applied in the first step, 234 out of 250 molecules were marked as having high GI absorption values. CYP isoenzymes are responsible for metabolizing drug molecules and making them suitable for excretion from the body. Possible inhibition of these enzymes may prevent the working molecule from being developed as a drug or cause the project to be terminated due to the toxicity problem in the later

stages of the studies. The results obtained from the SwissAdme Server were examined and it was observed that 18 out of 234 molecules with high GI absorption properties were not marked as possible inhibitors of any CYP isoenzyme. In the study carried out by Beall and Holloway (j), it was determined that some functional groups caused false positive results in bioactivity screening studies (PAINS Filter). When 18 molecules that were not labeled as CYP inhibitors were examined, it was determined that anyof them did not contain such a functional group. Some functional groups may add unwanted ADME properties to the studied molecules. In the study conducted by Brenk et al., by evaluating a large molecule library, such functional groups were determined and adapted to in silico studies. When 18 molecules passing through the PAINS filter were examined in terms of the mentioned functional groups, it was determined that 14 of them do not contain any of

these. When Lipinski, Ghose, Veber, Egon and Mugge filters which were created by examining the physicochemical properties of drug molecules are applied together, it is possible to create a consensus about the molecule being developed as bioactive / drug molecule would have drug – like properties. When the remaining 14 molecules were examined using these filters, it was determined that 7 of them were suitable for all filters. As a result, when the data obtained using the SwissAdme Server were evaluated, it was determined that 7 of the 250 molecules had good or favorable ADME properties and these molecules are suitable for bioactive / drug development studies (Fig. 4, Table 2). The simulated binding energies of the remaining 7 molecules were found to be between -10.9 kcal/mol and -9.1 kcal/mol. and these molecules were subjected to the second round of docking process in order to make more detailed investigations about binding poses.
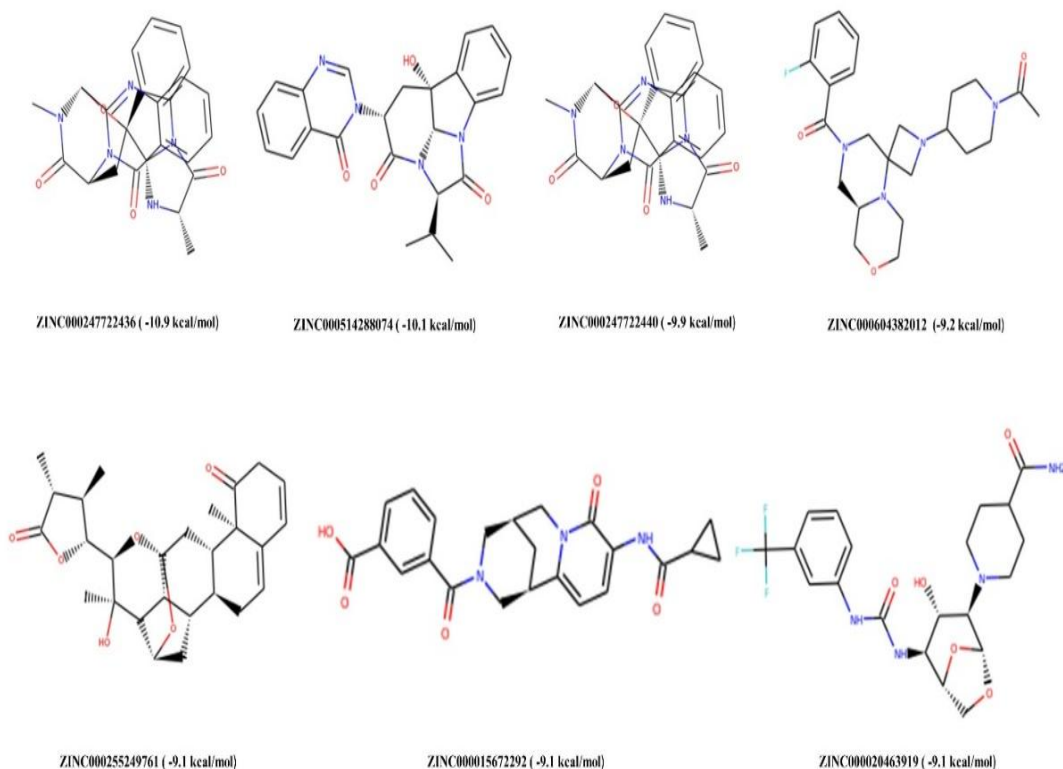

Fig. 4. Structures and ZINC codes of selected seven molecules

Table 2. Some physicochemical properties of top seven molecules

	Formula	MW (g/mol - Da)	MannholdLogP	HBA	HBD	RBC	T PSA (Å^2)	XLogP	SP3 Character	RBC(nt)
ZINC000247722436	$C_{24}H_{21}N_5O_4$	443.159	3.11	7	1	2	96.77	1.617	0.148	0
ZINC000514288074 (isochaetominine C)	$C_{24}H_{22}N_4O_4$	430.164	3.22	6	1	5	95.74	1.693	0.148	2
ZINC000247722440	$C_{24}H_{21}N_5O_4$	443.159	3.11	7	1	2	96.77	1.617	0.148	0
ZINC000604382012	$C_{23}H_{31}FN_4O_3$	430.238	3.11	7	0	4	56.33	0.869	0.242	2
ZINC000255249761	$C_{27}H_{34}O_6$	454.236	3.77	6	1	6	82.06	2.910	0.313	1
ZINC000015672292	$C_{23}H_{23}N_3O_5$	421.164	3.11	7	0	4	108.71	2.501	0.167	4
ZINC000020463919	$C_{20}H_{25}F_3N_4O_5$	458.178	2.34	9	4	9	126.15	0.124	0.211	5

(MW = Molecular Weight ,HBA = Hydrogen Bond Acceptors, HBD = Hydrogen Bond Donors, RBC = Rotatable Bonds Count, TPSA = Topological Polar Surface Area (Å^2), RBC(nt) = Rotatable Bonds Count (non terminal))

The second round of docking simulations were repeated three times for each molecule. The average and standard deviations of the binding energies of the poses and the RMSD values of the structures were calculated from these simulations. The average binding energies obtained and the results obtained from the first round of docking simulation were compared and shown in Table 2. Standard deviations of the binding energies obtained from the docking simulations in the second round were 0.000 for each of the molecule. The average RMSD values of the binding poses were examined and it was observed that they varied between 0.001 and 0.643, and the standard deviation values varied between 0.000 and 0.236. These values show that reproducible results were obtained in the second round docking process. The results obtained in the first docking round were compared with those obtained in the second round and they were quite close to each other. ΔΔG values for the same molecules of these two rounds were 0.0 for each of the molecules. The average of the RMSD values of the poses obtained in the first round and the poses obtained in the second round varied between 0.001 and 0.601. Although the upper limit of this

comparison is a little bit high, other values are at acceptable levels (Table 3). As it was seen that reproducible results were obtained in the second round of docking simulation, the poses obtained were used as starting structures in MD simulations.

MD simulations of the 7 molecules, the known inhibitor N3 and the receptor (protein structure of M^{pro}enzyme) free of ligand were completed without any problem and the obtained information was processed. RMSD plots of receptors obtained for each molecule, the known inhibitor N3 and free receptor as a result of MD simulations are shown in Fig. 5. When the graph was examined, it was observed that the RMSD values of the protein structure for most of the molecules were lower than 0.25 nm (2.5 Å) and they formed relatively stable complexes during the simulation. When the RMSD values of the molecules were examined, it was observed that some of them remained quite stable during the simulation (ZINC000247722436), while others were more mobile (ZINC000020463919) (Fig. 6). When the RMSF values of the C$_a$ of proteins were examined, it was observed that they are not very different from each other (Fig. 7).

Table 3. Results of first and second round of docking simulations

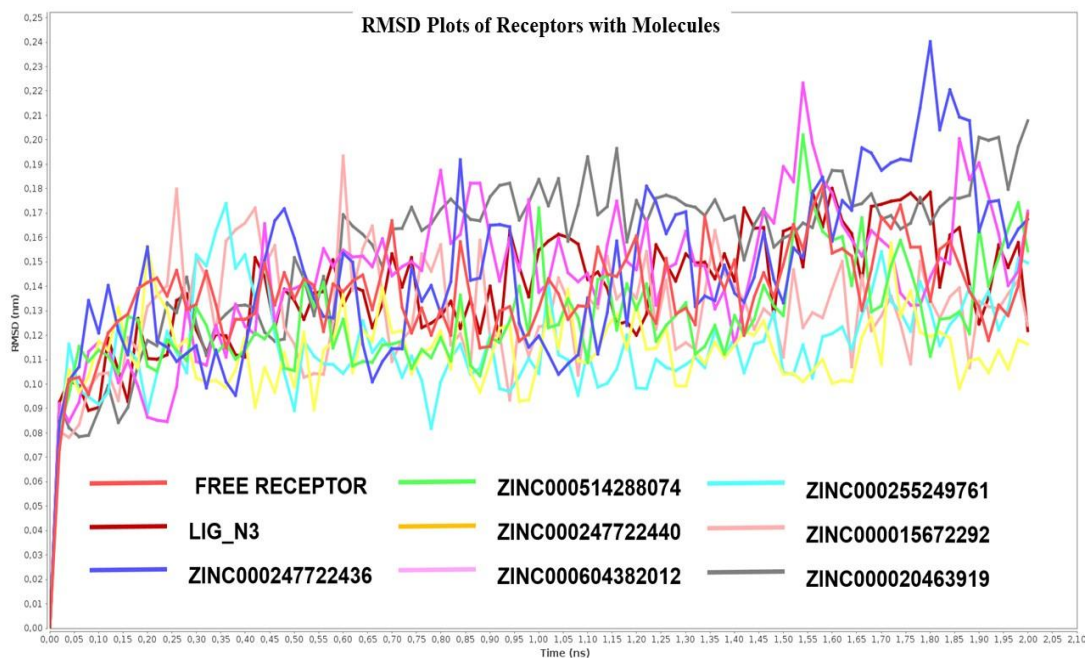|	Simulated Binding Energy in the First Round Docking (-kcal/mol)	Average of Simulated Binding Energy in the Seconda Round Docking (-kcal/mol)	Standard Deviation of Simulated Binding Energy in the Second Round Docking	ΔΔG of First Round and Second Round Docking (-kcal/mol)	Average RMSD vales of First Round Doking Pose and Second Round Docking Poses	Standard Deviation of RMSD vales of First Round Doking Pose and Second Round Docking Poses	Average RMSD vales of Second Round Docking Poses	Standard Deviation of RMSD vales of Second Round Docking Poses
ZINC000247722436	-10.9	-10.9	0.0	0.0	0.001	0.001	0.001	0.000
ZINC000514288074	-10.1	-10.1	0.0	0.0	0.153	0.031	0.100	0.024
ZINC000247722440	-9.9	-9.9	0.0	0.0	0.008	0.004	0.007	0.004
ZINC000604382012	-9.2	-9.2	0.0	0.0	0.029	0.021	0.388	0.005
ZINC000255249761	-9.1	-9.1	0.0	0.0	0.314	0.097	0.193	0.085
ZINC000015672292	-9.1	-9.1	0.0	0.0	0.467	0.434	0.156	0.124
ZINC000020463919	-9.1	-9.1	0.0	0.0	0.601	0.079	0.643	0.236

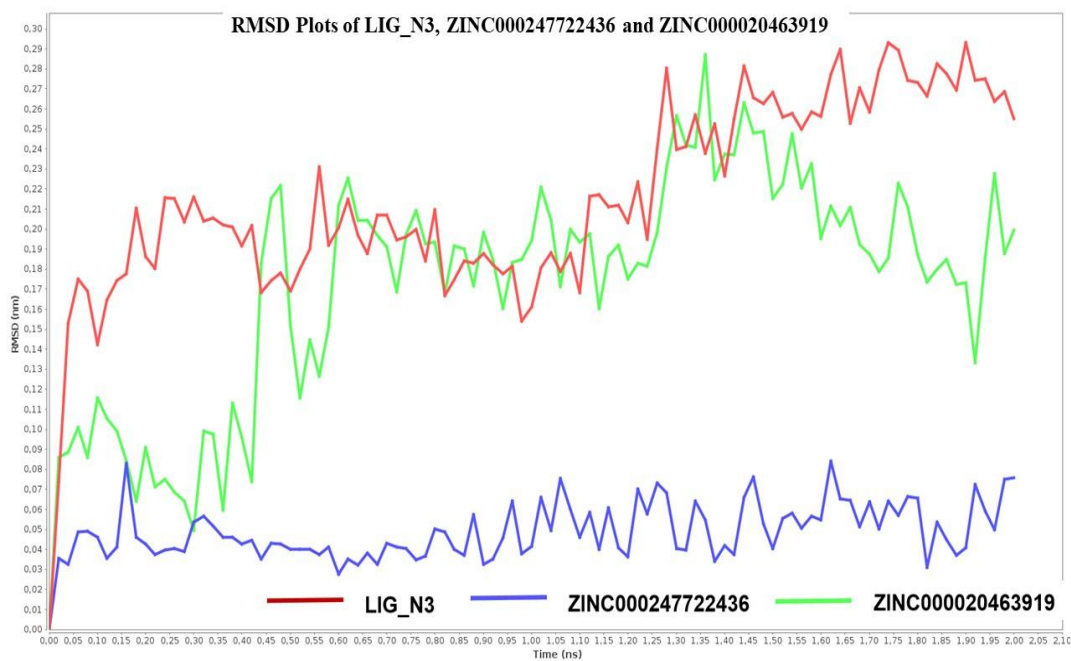**Fig. 5. RMSD plots of receptors with molecules**

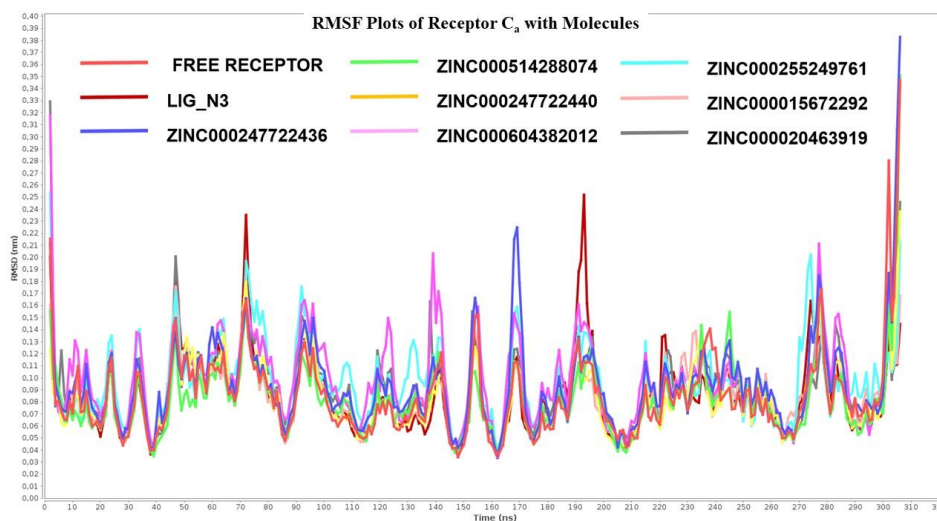**Fig. 6. RMSD plots of LIG_N3, ZINC0000247722436 and ZINC000020463919**

Fig. 7. RMSF plots of the C$_a$ of receptors with molecules

Before calculating the binding free energies, the partial charges assigned to the atoms in the molecules by Acpype were calculated and grouped according to the total charge obtained. It was emphasized that the software we use (g_mmpbsa) can create variable results for molecules and proteins with different charges in the user mail groups. The sum of the charges assigned by Acpype on the atoms was 0 for ZINC000514288074, ZINC000255249761, ZINC000015672292, 1 for LIG-N3, ZINC000247722436, ZINC000247722440, ZINC000020463919 and 2 for ZINC000604382012. The molecules grouped according to their charge and their calculated binding free energies are shown in Table 4. When Table 4 is examined, it is observed that as the total charge on the molecules increases, the contribution of electrostatic interactions to the binding free energy increases. It is highly expected to observe increase in the electrostatic interaction between molecules and protein as the partial charges of increase. Of the 7 molecules we examined, 6 were found to have higher binding free energies (worse binding), while four ZINC000604 (-325.588 +/- 16.754 kJ/mol), had lower binding free energies than the known inhibitor LIG_N3 (-305.174 +/- 26.247 kJ/mol). When the molecule we mentioned was examined, it was observed that it has total charge of 2. As expected, its electrostatic energies are higher than other molecules. Among those with 0 total charge, the best

binding free energy was found to belong to ZINC000514288074 with -112.657 +/- 13.933 kJ /mol and those with 1 total charge, the best binding free energy was found to belong to ZINC000514288074 with -242.320 +/- 12.248 kJ /mol.

We also assessed the synthesis difficulties of potential molecules in this project where we are in search for a novel natural product or a natural product like molecules as M^{pro} inhibitors. For this process, the SwissAdme webserver was used because of its high prediction accuracy. The SwissAdme server scores the synthetic accessibility (SA) of the molecules between 1 (easy to synthesize) and 10 (difficult to synthesize). According to the model's results, the lowest SA value was 4.43 to the molecule ZINC000514288074 and the highest was 6.97 that belonged to the molecule ZINC000255249761. The SA value of ZINC000604382012, the molecule with the lowest binding free energy, was determined as 4.64, which is a relatively high value by the model but fair among the studied selected molecules.

When all the above-mentioned features are combined, it was observed that there are no significant differences between the docking scores of the molecules we study, the reproducibility of the docking poses, docking energies and the ADME properties.

Table 4. Binding free energies, sum of partial charges of atoms and synthetic accessibility of the selected molecules

	Van der Waal Energy (kj/mol)	Electrostatic energy (kj/mol)	Polar solvation energy (kj/mol)	SASA energy (kj/mol)	Binding free energy (kj/mol)	Sum of Partial Charge of Atoms	Synthetic Accessibility (SA)
LIG_N3	-277.324 +/- 20.874	-213.144 +/- 27.830	214.402 +/-55.890	-29.107 +/-1.708	-305.174 +/-26.247	1	-
ZINC000247722436	-167.076 +/- 17.654	207.883 +/- 17.096	160.466 +/-25.161	-16.466 +/-1.265	-231.073 +/-9.852	1	5.49
ZINC000514288074	-197.852 +/- 6.142	-17.787 +/- 8.602	111.904 +/-9.323	-18.922 +/-0.373	-122.657 +/-13.933	0	4.43
ZINC000247722440	-193.068 +/- 6.583	-244.571 +/- 20.006	215.566 +/-14.094	-20.247 +/-0.917	-242.320 +/-12.248	1	5.49
ZINC000604382012	-154.234 +/- 18.219	-409.287 +/- 70.447	254.785 +/-83.241	-16.851 +/-1.231	-325.588 +/-16.754	2	4.64
ZINC000255249761	-142.191 +/- 5.329	-15.648 +/- 6.027-	82.648 +/-11.915	-16.932 +/-1.183	-92.123 +/8.763-	0	6.97
ZINC000015672292	-177.012 +/-19.807	-19.712 +/- 4.894	108.060 +/-12.474	-17.274 +/-1.027	-105.939 +/-14.016	0	4.53
ZINC000020463919	-170.388 +/- 14.971	-200.160 +/-20.247	182.558 +/-20.028	-19.553 +/-1.797	-207.543 +/-13.860	1	5.79
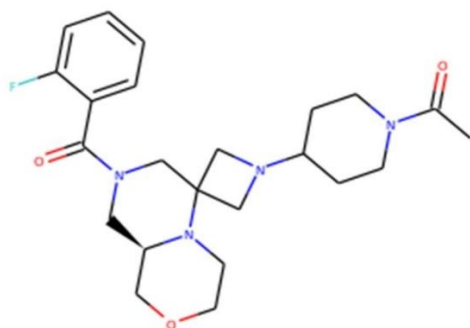
ZINC000604382012
Docking Binding Energy : - 9.2 kcal/mol
Binding Free Energy -325.588 +/-16.754 kJ/mol
Synthetic Accessibility : 4.64
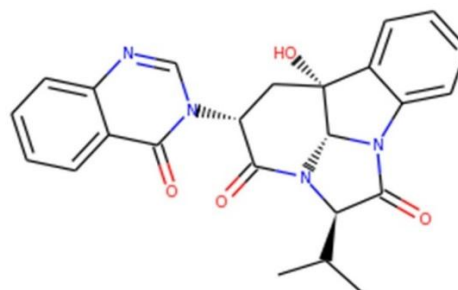
ZINC000514288074
Docking Binding Energy : - 10.1 kcal/mol
Binding Free Energy -122.657 +/- 13.933 kJ/mol
Synthetic Accessibility : 4.43

Fig. 8. Selected candidate molecules ZINC000604382012 and ZINC000514288074

It was observed that the calculated binding free energy values increased with the increasing total partial charges of the molecules we studied. In this case, molecules with similar charges have been compared among themselves. Since we aimed to work with natural products or natural product like molecules that have relatively complex structures, it is obvious that the synthetic accessibility (SA) values of the molecules will also be important. After evaluating all these conditions, it is appropriate to select ZINC000604382012 (-325.588+/-16.754 kJ/mol) because of having lower simulated binding free energy values than LIG_N3 (-305.174 +/- 26.247 kJ/mol) and ZINC000514288074 (-122.657+/- 16.754 kJ/mol) because of having lowest SA values of 4.43 among filtered seven molecules. (Fig. 8)

4. DISCUSSION

In this study, we worked with 85553 molecules classified as "in-vitro biogenic" in the ZINC "Biogenic Dataset" containing natural products, natural product derivatives and natural products like and some other molecule groups like FDA approved drug etc.. One of the most frequently used filters in medicinal chemistry studies, Ro5, was applied to these molecules. After Ro5 filter, the number dropped to 51422. Preparing these molecules to the docking simulations, KNIME software and various add-ons were used. During this process, a KNIME workflow has been created that can be used in other projects. There are some studies in the literature about preparing

molecules for virtual screening processes using KNIME software. One of these is VSPrep, which was prepared by Gally and his colleagues. [36] This workflow contains some filtrations and structure preparation nodes like our workflow. However, it has a more general application, as it only deals with library preparation for virtual screening operations unlike our workflow. It was observed that the simulated binding energies varied between - 10.9 and - 2.5 kcal/mol as a result of the docking virtual screen performed using the QVina 2.1 (AutoDockVina derivative) software with 51535 molecules whose three-dimensional structure was optimized. According to the docking study performed by Keretsu et al using AutoDockVina, the simulated binding energy of the known inhibitor N3 is -7.8 kcal/mol. [37] In our study, the number of molecules with a simulated binding energy lower than -7.8 kcal/mol is 9525. Considering the 250 molecules with the best binding energy in our project in terms of efficient use of the available resources, we have examined only 2.62 percent of the possible active molecules in detail. As a result of evaluating the ADME properties of the best 250 molecules, we found that only 7 of them meet the criteria we set. There are some studies in the literature that have used the SwissAdme server in virtual screening studies to examine the ADME properties of the screened molecules against SARS–CoV-2 M^{Pro} enzyme. For example, Sepay et al. used SwissAdme in their study on the evaluation of chromone-derived molecules as M^{Pro} inhibitors to predict ADME properties. According to their evaluations, they mentioned

48

that some of the molecules they think could be inhibitors of one CYP isoenzyme, and some of them could be more than one CYP isoenzyme inhibitor, so the related molecules possibly have toxic effects. [38] Compared with these results, since the molecules we focused on in our study were not marked as possible CYP isoenzyme inhibitors, the possibility of encountering toxicity problems is lower than their molecules. However, the SA values of the molecules examined by Sepay and colleagues are lower than the natural product and natural product-like molecules we examined in our study. This shows that the aforementioned molecules of them will be easier to synthesize. Joshi et al. evaluated the potential usage of natural products isolated from lichens as SARS-CoV-2 M[pro] inhibitors using the methods similar to our study. [39] After screening the molecules isolated from 412 lichens with the help of AutoDockVina, they evaluated the ADME properties. They carried out molecular dynamics and MM-PBSA studies of the molecules that passed the filters they applied. Among the natural products they screened, there are some with better simulated binding energy compared to our molecules. For example, the docking score of the molecule named as Rugulosin had -13.2 kcal/mol. However, this molecule did not pass the ADME filters they applied. Of the 27 molecules with better binding energy than X77, the ligand of enzyme of interest, only four were able to pass the Ro5, PAINS, and Drug-Likeness filters. Two of these four molecules failed to pass filters associated with toxicity. The binding energies of these two molecules calculated by AutoDockVina were -8.7 for rhizocarpon acid and -8.4 kcal/mol for calycin. The binding free energies of the known inhibitors X77, rhizocarpic acid and calycin were also calculated by the authors. Accordingly, it was determined that the known inhibitor X77 had the values –of 91.78 ± 11.09, calycin - 42.42 ± 9.21 and rhizocarpic acid - 57.85 ± 8.89 kJ/mol. These numbers are lower than the numbers we obtained from both docking and MM-PBSA studies.

Among the candidate molecules, ZINC000604382012 is marked as a natural product-like synthetic molecule in the ZINC database. In the searches we made in the ZINC database, it was observed that this molecule did not have a biological activity study indexed in CHEMBL. In addition, there are no clinical studies on ZINC000604382012. ZINC000514288074 (isochaetominine C) is an alkaloidal metabolite isolated from a Marine-Derived *Aspergillus sp*.[40] For example, the MIC

values of this molecule against *Staphylococcus aureus, Bacillus subtilis, Micrococcus luteus, Salmonella typhimunium, Proteus hauseri, Escherichia coli* microorganisms are above 100 µg / ml concentration. The cytotoxicity of ZINC000514288074 on some cell lines was also examined. The IC50 values of the molecule on A549 and K562 cell lines were determined as 18.55 and 14.40 µM, respectively. The last reported biological activity of ZINC000514288074 is the Na^+ / K^+ ATPase inhibition value, which is reported as 37.35 µM. [40] According to the ZINC database, there are no clinical studies on ZINC000514288074.

5. CONCLUSION

As a result, the molecules we have mentioned as candidate lead-molecules by considering many conditions are advantageous compared to the examples mentioned in the literature. In addition, it has been observed that it is possible to find more candidate molecules among the molecules that have not been examined in detail, especially since they were not listed among the top 250 molecules after the first docking step when compared to the studies in the literature related with this target.

CONSENT

It's not applicable.

ETHICAL APPROVAL

It's not applicable.

ACKNOWLEDGEMENT

COMPETING INTEREST

Author has declared that no competing interests exist.

REFERENCES

1. Available:https://www.aa.com.tr/en/p/coron avirus-infection-map
2. Frecer V, Miertus S. Antiviral agents against COVID-19: Structure-based design of specific peptidomimetic inhibitors of SARS-CoV-2 main protease. RSC Advances. 2020;10(66):40244-63.

3. Shyr ZA, Gorshkov K, Chen CZ, Zheng W. Drug discovery strategies for SARS-CoV-2. J PharmacolExpTher. 2020;375(1):127-38.

4. Choy K-T, Wong AY-L, Kaewpreedee P, Sia S-F, Chen D, Hui KPY et al. Remdesivir, lopinavir, emetine, and homoharringtonine inhibit SARS-CoV-2 replication *in vitro*. Antivir Res. 2020;104786.

5. Baden LR, Rubin EJ. Covid-19—The search for effective therapy. N Engl J Med. 2020;382(19):1851-2.

6. Singh TU, Parida S, Lingaraju MC, Kesavan M, Kumar D, Singh RK. Drug repurposing approach to fight COVID-19. Pharmacol Rep. 2020;1-30.

7. Dotolo S, Marabotti A, Facchiano A, Tagliaferri R. A review on drug repurposing applicable to COVID-19. Brief Bioinform; 2020.

8. Sultana J, Crisafulli S, Gabbay F, Lynn E, Shakir S, Trifirò G. Challenges for Drug Repurposing in the COVID-19 Pandemic Era. Front Pharmacol. 2020;11(1657).

9. Ghosh S, Nie A, An J, Huang Z. Structure-based virtual screening of chemical libraries for drug discovery. CurrOpinChem Biol. 2006;10(3):194-202.

10. Bajorath J. Integration of virtual and high-throughput screening. Nat Rev Drug Discov. 2002;1(11):882-94.

11. Lionta E, Spyrou G, K Vassilatis D, Cournia Z. Structure-based virtual screening for drug discovery: Principles applications and recent advances. Curr Top Med Chem. 2014;14(16):1923-38.

12. Vangrevelinghe E, Zimmermann K, Schoepfer J, Portmann R, Fabbro D, Furet P. Discovery of a potent and selective protein kinase CK2 inhibitor by high-throughput docking. J Med Chem. 2003;46(13):2656-62.

13. Ward RA, Perkins TD, Stafford J. Structure-based virtual screening for low molecular weight chemical starting points for dipeptidyl peptidase IV inhibitors. J Med Chem. 2005;48(22):6991-6.

14. Schapira M, Raaka BM, Samuels HH, Abagyan R. Rational discovery of novel nuclear hormone receptor antagonists. Proc Natl AcadSci U S A. 2000;97(3):1008-13.

15. Matter H, Sotriffer C. Applications and success stories in virtual screening. Virtual screening: Principles, challenges, and practical guidelines. 2011:319-58.

16. Newman DJ, Cragg GM. Natural products as sources of new drugs over the nearly four decades from 01/1981 to 09/2019. J Nat Prod. 2020;83(3):770-803.

17. Sterling T, Irwin JJ. ZINC 15–ligand discovery for everyone. J ChemInf Model. 2015;55(11):2324-37.

18. Berthold MR, Cebron N, Dill F, Gabriel TR, Kötter T, Meinl T, et al. KNIME-the Konstanz information miner: version 2.0 and beyond. SIGKDD Explor. 2009;11(1):26-31.

19. Lipinski CA. Lead-and drug-like compounds: The rule-of-five revolution. Drug Discov Today Technol. 2004;1(4):337-41.

20. Marvin 2019. ChemAxon; 2019. Available:http://www.chemaxon.com

21. O'Boyle NM, Banck M, James CA, Morley C, Vandermeersch T, Hutchison GR. Open Babel: An open chemical toolbox. J Cheminformatics. 2011;3(1):33.

22. The numerical calculations reported in this paper were partially performed at TUBITAK ULAKBIM, High Performance and Grid Computing Center (TRUBA resources); 2003.

23. Alhossary A, Handoko SD, Mu Y, Kwoh C-K. Fast, accurate and reliable molecular docking with QuickVina 2. Bioinformatics. 2015;31(13):2214-6.

24. Jin Z, Du X, Xu Y, Deng Y, Liu M, Zhao Y et al. Structure of M pro from SARS-CoV-2 and discovery of its inhibitors. Nature. 2020;1-5.

25. Daina A, Michielin O, Zoete V. SwissADME: A free web tool to evaluate pharmacokinetics, drug-likeness and medicinal chemistry friendliness of small molecules. Sci Rep. 2017;7:42717.

26. Baell JB, Holloway GA. New substructure filters for removal of pan assay interference compounds (PAINS) from screening libraries and for their exclusion in bioassays. J Med Chem. 2010;53(7):2719-40.

27. Brenk R, Schipani A, James D, Krasowski A, Gilbert IH, Frearson Jet al. Lessons learnt from assembling screening libraries for drug discovery for neglected diseases. ChemMedChem. 2008;3(3):435.

28. Ghose AK, Viswanadhan VN, Wendoloski JJ. A knowledge-based approach in designing combinatorial or medicinal chemistry libraries for drug discovery. 1. A qualitative and quantitative

characterization of known drug databases. J Comb Chem. 1999;1(1):55-68.

29. Veber DF, Johnson SR, Cheng H-Y, Smith BR, Ward KW, Kopple KD. Molecular properties that influence the oral bioavailability of drug candidates. J Med Chem. 2002;45(12):2615-23.

30. Egan WJ, Merz KM, Baldwin JJ. Prediction of drug absorption using multivariate statistics. J Med Chem. 2000;43(21):3867-77.

31. Muegge I, Heald SL, Brittelli D. Simple selection criteria for drug-like chemical matter. JMedChem. 2001;44(12):1841-6.

32. Abraham MJ, Murtola T, Schulz R, Páll S, Smith JC, Hess B, et al. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. SoftwareX. 2015;1:19-25.

33. Kumari R, Kumar R, Consortium OSDD, Lynn A. g_mmpbsa□ A GROMACS tool for high-throughput MM-PBSA calculations. J ChemInf Model. 2014;54(7):1951-62.

34. Da Silva AWS, Vranken WF. ACPYPE-Antechamber python parser interface. BBMC Res Notes. 2012;5(1):367.

35. Hornak V, Abel R, Okur A, Strockbine B, Roitberg A, Simmerling C. Comparison of multiple Amber force fields and development of improved protein backbone parameters. PROTEINS. 2006;65(3):712-25.

36. Gally JM, Bourg S, Do QT, Aci-Sèche S, Bonnet P. VSPrep: a general KNIME workflow for the Preparation of molecules for virtual screening. Mol Inform. 2017;36(10):1700023.

37. Keretsu S, Bhujbal SP, Cho SJ. Rational approach toward COVID-19 main protease inhibitors via molecular docking, molecular dynamics simulation and free energy calculation. Sci Rep. 2020;10(1):1-14.

38. Sepay N, Sepay N, Al Hoque A, Mondal R, Halder UC, Muddassir M. In silicofight against novel coronavirus by finding chromone derivatives as inhibitor of coronavirus main proteases enzyme. J Struct Chem. 2020;1.

39. Joshi T, Sharma P, Joshi T, Pundir H, Mathpal S, Chandra S. Structure-based screening of novel lichen compounds against SARS Coronavirus main protease (Mpro) as potentials inhibitors of COVID-19. Mol Divers. 2020;1-13.

40. Liao L, You M, Chung BK, Oh D-C, Oh K-B, Shin J. Alkaloidal metabolites from a marine-derived Aspergillus sp. fungus. 2015;78(3):349-54.

Peer-review history:
The peer review history for this paper can be accessed here:
http://www.sdiarticle4.com/review-history/65683