Virtual screening and molecular docking studies of certain lead-like compounds from ZINC15 database against COVID-19 Mpro enzyme

Potential lead-like compounds against COVID-19 Mpro

Erman Salih İstifli
Department of Biology, Cukurova University, Faculty of Science and Literature, Adana, Turkey

Abstract
Aim: The COVID-19 epidemic, which first emerged in Wuhan in December 2019, rapidly affected the globe in the form of a pandemic, and currently millions of people are classified as active cases in terms of this infection, while tens of thousands of people are in serious or critical conditions.

Material and Methods: In this study, 100 lead-like chemical entities were downloaded from the ZINC15 database and subjected to structure-based virtual screening (SBVS) in terms of their potential to be used in combating COVID-19. In our molecular docking study, the Mpro (3CLpro) enzyme encoded by the COVID-19 genome was selected as the receptor molecule that could be targeted in the treatment.

Results: A total of 3 small molecules (ZINC000000011271, ZINC000000026220 and ZINC000000006645) were identified as potential inhibitors due to their high binding affinities to this enzyme. Physicochemical profiles of the top-ranked ligands, determined in our study, showed that these compounds were safe and had drug-like features. Furthermore, in molecular dynamics (MD) simulations, the top-ranked compound ZINC000000011271 was found to strongly bind to the Mpro enzyme and preserved its hydrogen bonding stability throughout the whole trajectory.

Discussion: In this study, three orphaned drugs (ZINC000000011271, ZINC000000026220 and ZINC000000006645) were found to strongly inhibit Mpro in molecular simulations. These compounds also displayed hit features according to the SwissADME database. Therefore, these hits defined in this study may be used in the development and advanced optimization of potent COVID-19 inhibitors.

Keywords
COVID-19; Structure-based virtual screening (SBVS); Molecular docking; ADMET; Molecular dynamics
Introduction
Coronavirus disease (SARS-CoV-2, COVID-19) continues to be a major burden on global health since December 2019. Although 18 million active cases have been reported worldwide as of 2 December 2020, the disease has spread to 218 countries and territories in total according to real time world statistics (Worldometer). Vaccines targeting different protein structures of the virus have been developed in different countries, however, there is still no known effective cure for the disease, and the incidence of new cases continues to rise with daily and linear increases. Therefore, urgent intervention is needed to reduce the worldwide spread of the virus and to develop a universally effective anti-COVID-19 agent. In some early and recent studies targeting the coronavirus main protease (Mpro, 3CLpro), various novel inhibitors have been developed against this protein, or known drugs have been repurposed against this target [1-4].

The SARS-CoV-2 genome size is approximately 30,000 nucleotides in length, and the viral replicate gene encodes two overlapping polyproteins (pp1a and pp1b), which are functional in viral replication and transcription [5, 6].

The conversion of these polyproteins to functional polypeptides is predominantly performed by the 33.8-kDa Mpro (3CLpro) protein. The Mpro cleaves these polyproteins at least at 11 evolutionarily conserved sites, and this process starts with an autolytic cleavage of pp1a and pp1ab regions that reside on the protein sequence of the enzyme [7].

The fact that Mpro has such a functional importance and that it does not have a close homolog in humans makes Mpro an important target in the design of antiviral drugs or repurposing licensed drugs against this enzyme. Thus, by inhibiting viral replication and transcription, the recovery process can be effectively accelerated in patients suffering from COVID-19 infection. The crystal structure of COVID-19 main protease in complex with its specific inhibitor (N3) (PDB ID: 6LU7), uploaded to Protein Data Bank in February 2020, can be used as a suitable target in virtual screening of libraries of small molecules in order to stop viral replication and transcription. Since the emergence of the SARS-CoV-2 infection, molecular docking and target-based virtual screening studies as the computer-aided drug design tools are progressing at an even faster pace [8].

In our study, the potential of 100 lead-like molecules to inhibit SARS-CoV-2 main protease was investigated through virtual screening, molecular docking and ADMET profile analysis. Instead of randomly filtering quite a large number of small molecules from the ZINC15 database, in this study, it is envisaged to accelerate the drug discovery process by scanning a small molecule library with only lead-like features.

Material and Methods
Bioinformatic methods
In this study, one hundred (100) small molecules, which meet the following lead-likeness criteria, were downloaded from the ZINC15 (https://zinc15.docking.org/) database in order to speed up the drug discovery process: i. 250 ≤ MW ≤350, ii. Log P ≤3.5, iii. The number of rotatable bonds was ≤7 [9]. The ligands downloaded from the ZINC15 database had a molecular charge of -1 to 0, were set up to have the dominate form at pH 7.4 and the highest reactivity.

Protein retrieval and ligand preparation
In our study, the crystal structure of COVID-19 main protease in complex with the inhibitor N3 (resolution: 2.16 Å) downloaded from Protein Data Bank (PDB ID: 6LU7) was used as receptor molecule. This enzyme consists of a single chain (chain A) and 306 amino acids. The Mpro enzyme, ligands, and docking parameters were prepared using AutoDockTools-1.5.6 [10]. Water molecules and other heteroatoms were removed from the receptor structure. One hundred (100) lead-like small molecules (ligands) in sdf format downloaded from the ZINC15 database were converted to pdb format using Open Babel [11]. The energy of ligands in pdb format was minimized using the ‘steepest descent algorithm’ (2500 steps). Finally, the geometrically optimized ligands in pdb format were converted to pdbqt format using a specific script.

Structure-based virtual screening
Structure-based virtual screening (SBVS) or target-based virtual screening (TBVS) technique serves to predict the best interaction conformation of the complex that will occur between the ligand and its molecular target (receptor).

As a result, the ligands tested are ranked according to their binding affinity with the target molecule [12]. AutoDock Vina has significant advantages in binding affinity and binding mode prediction compared to AutoDock, and is also two orders of magnitude faster than AutoDock 4 in binding affinity and binding mode calculations [13]. The parameters used in the docking experiments of the Mpro are given in Table 1. To calculate the binding energy of the inhibitor (N3) with the crystal structure of Mpro, AutoDock Vina was used in our study. The value obtained as a result of this calculation was used to compare the binding energies of the lead-like molecules downloaded from the ZINC15 database. The docking score of N3 inhibitor against Mpro was found to be -10.04 kcal/mol, and this value was used as the ‘binding energy threshold’ in docking experiments with other candidate ligands. Thus, the search space for candidate ligands was considerably narrowed. As a result, this cut-off value (ΔG° ≤ -10.04 kcal/mol) that we determined gave a total of 3 molecules, and these showed the highest binding affinity against Mpro.

Table 1. Grid box coordinates and number of grid points in xyz dimensions set up in AutoDock Vina and AutoDock 4.2.6

<table>
<thead>
<tr>
<th>Mpro</th>
<th>Centor (Å)</th>
<th>Number of grid points (Å)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>x: -14.087</td>
<td>x: 52</td>
</tr>
<tr>
<td></td>
<td>y: 14.07</td>
<td>y: 44</td>
</tr>
<tr>
<td></td>
<td>z: 68.663</td>
<td>z: 58</td>
</tr>
</tbody>
</table>
**Molecular Docking using AutoDock 4.2.6.**

The top-ranked ligands (hit compounds) obtained by us as a result of the SBVS study were subjected to further molecular docking against their respective receptor, Mpro, using AutoDock 4.2.6. The parameters used in the adjustment of the grid box are given in Table 1. The grid box dimensions were adjusted to cover all amino acids in the ligand binding pocket (Figure 1a). For each ligand, 2,500,000 energy evaluations and a total of 100 genetic algorithm runs were implemented. Lamarckian Genetic Algorithm (LGA) was used as the search parameter, and Gasteiger partial charges were added to the receptor.

**Drug-likeness analysis**

In the pharmaceutical industry, clarifying the drug-likeness properties of hit compounds is a crucial step in reducing the side effects of these agents. Additionally, the drug-likeness concept is functional in optimizing the pharmacokinetic and pharmaceutical properties of drug candidate molecules such as solubility, chemical stability, bioavailability and distribution profiles [14]. In our study, a web-based SwissADME tool was used to determine the drug-likeness properties of top-ranked ligands downloaded from the ZINC15 database [15-18].

**Molecular dynamics simulation of the top-ranked receptor-ligand complex**

Molecular dynamics (MD) is a computer simulation used to analyze the physical movements of atoms and molecules in a system and their conformational changes within a certain time interval. Atoms and molecules are allowed to interact within this specified time interval and thus provide information about the dynamic ‘evolution’ of the system. To confirm the binding mode of the top-ranked ligand (ZINC0000000011271), the molecular dynamics (MD) simulations were performed using the GROMACS 2020.1 version based on the docked conformation of ZINC0000000011271 and the Mpro protein [19]. The hydrogen atoms of the ligand were added explicitly using the Avogadro program. The CHARMM36 force field was used to create the topology for the protein and CHARMM General Force Field (CgenFF, https://cgenff.umaryland.edu/) was used to build the topology for our top-ranked ligand, ZINC0000000011271. The topologies and coordinates of receptor and ligand were then combined to construct the topology and coordinate of the protein-ligand complex. Four Na+ ions, corresponding to physiological concentration, were added to neutralize the system, and then the protein-ligand complex was solvated in a dodecahedral box with SPC water molecules [20]. Before the MD simulation, the complexes were subjected to 50,000 steps of energy minimization to relieve any unfavorable interactions in the initial configuration of the system. Equilibration simulations were carried out in two steps: equilibration using NVT for 100 ps consisting of an energy minimization of 50,000 steps, and an NPT equilibration for 100 ps with 50,000 steps. After the completion of two equilibration phases, a 10-ns MD simulation was performed. During the MD run, the trajectories were collected every 10 ps to analyze protein-ligand interactions and dynamics. The trajectories were analyzed to quantify the structural stability (RMSD and RMSF) of the protein, ligand and the protein-ligand complex as well as to detect the number of intermolecular H-bonds formed between the complex during the simulation. Excel 2013 program was used to graphically visualize the generated results obtained using the in-built tools of GROMACS.

**Results**

Top-ranked ligands with their calculated binding energies and estimated inhibition constants against Mpro in our study are given in Table 2. In this study, ZINC0000000011271, ZINC0000000026220 and ZINC000000006645 showed the highest binding affinities against Mpro among 100 lead-like small molecules downloaded from ZINC15 database. The binding energy of ZINC0000000011271 to Mpro was found to be -10.91 kcal/mol with an estimated inhibition constant of 10.09 nM. ZINC0000000026220 formed conventional H-bonds with Leu141, Gly143, Ser144, Cys145 and Glu166, carbon-hydrogen bonds with Phe140, His164 and His172, showed a n-sigma interaction with His41 and a n -alkyl interaction with Met165. Furthermore, this compound showed four van der Waals interactions with Asn142, His163, Arg188 and Gln189, respectively (Figure 1b). The binding energy of ZINC0000000026220 to Mpro was found to be -10.46 kcal/mol with an estimated inhibition constant of 21.39 nM. ZINC0000000026220 showed three n-alkyl interactions with Met49, Leu167 and Pro168, one n-sulfur interaction with Met165, and three conventional H-bonds with His164, Gln189 and Gln192. This compound also shared many van der Waals interactions with several residues of Mpro (Figure 1c). The third compound ZINC000000006645 showed a binding energy of -10.43 kcal/mol and an estimated inhibition constant of 22.50 nM in its interaction with Mpro. ZINC000000006645 showed a n-alkyl and a n-n T-shaped interaction with the Met165 and His41 residues of the Mpro, respectively. In addition, this compound formed a non-classical carbon-hydrogen bond with Pro168, and three classical hydrogen bonds with Glu166, Arg188, and Thr190 (Figure 1d).

**Figure 1.** Catalytic amino acid residues of the ligand binding pocket of Mpro enzyme (a) and 2D ligand interaction diagram of the top-ranked Mpro-complexes; (b) ZINC0000000011271, (c) ZINC0000000026220, (d) ZINC000000006645
The physicochemical properties of ZINC000000026220, ZINC000000006645 and ZINC000000024786 ligands determined as hit compounds in our study, have been subjected to stringent rules determined by Lobell et al. based on the principle of Ro5 (Lipinski’s rule of 5) [18]. In accordance with this rule, the 3 hit compounds we determined showed physicochemical properties. As an exception, only the topological polar surface area (TPSA) of ZINC000000026220 was found slightly above the specified upper limit of 120 Å2 (Table 3).

In this study, 10-ns MD simulation was carried out to evaluate the conformational changes, stability and the dynamic evolution of the Mpro-ZINC000000011271 complex. To determine how much the ligand binding pose has changed over the 10-ns simulation period, the root mean square deviation (RMSD) of protein and protein-ligand complex was computed (Figure 2). In Figure 2, the RMSD of unbounded Mpro enzyme remained stable between 1.4 ns and 10 ns time period at 0.15 nm. The RMSD of Mpro-ZINC000000011271 remained stable at 0.2 nm at 0.3 ns until 1.7 ns, then slightly increased to 0.32 nm between 1.8 ns - 3.4 ns, and finally reached its steady state at 0.65 nm between 3.5 ns and 10 ns (Figure 2). RMSF of the unbounded Mpro showed that the residues located in the active site (Phe140, Gly143, Cys145, His164, Glu166, His172, Cln189, Thr190 and Glu 192) of the enzyme remained stable between 0.05 nm and 0.1 nm (Figure 3a). The other residues of the enzyme also showed a rigid character in terms of RMSF which was between 0.05 nm and 0.21 nm. The RMSF of ZINC0000000011271 during MD simulation (Figure 3b) showed that the ligand atoms had low fluctuations between 0.02 nm and 0.20 nm, indicating that the atom positions of the ligand did not significantly changed during the interactions with the binding pocket residues. Although the 29., 30. and 31. atoms of the ligand displayed high RMSF values, their fluctuations were limited to an acceptable value of 0.2 nm (Figure 3b). The H-bond plotting showed an average of 3.66 hydrogen bond interactions between Mpro and ZINC000000011271 throughout the 10-ns MD simulation with a maximum of 13 hydrogen bonds (Figure 3c). The resulting dynamic H-bond formation pattern in Figure 3c well illustrates the tight binding mechanism of ZINC0000000011271 to Mpro and the underlying reason for its inhibition.
Discussion
In this study, ZINC000000011271, ZINC000000026220 and ZINC000000006645 showed the highest binding affinities against Mpro among 100 lead-like small molecules downloaded from ZINC15 database. The binding energy of ZINC000000011271 to Mpro was found to be -10.91 kcal/mol with an estimated inhibition constant of 10.09 nM. ZINC000000011271 formed conventional H-bonds with Leu141, Gly143, Ser144, Cys145 and Glu166, carbon-hydrogen bonds with Phe140, His164 and His172, showed a n-sigma interaction with His41 and a n -alkyl interaction with Met165. Furthermore, this compound showed four van der Waals interactions with Asn142, His163, Arg188 and Gln189, respectively. ZINC000000011271 is an analogue of 3’-[4-Aryl-(1,2,3-triazol-1-yl)]-3’-deoxythymidine and it has been reported to inhibit thymidine kinase and deoxynucleoside kinase enzymes in Herpes virus 1 (HSV1), Homo sapiens and Drosophila melanogaster [21]. The binding energy of ZINC000000026220 to Mpro was found to be -10.46 kcal/mol with an estimated inhibition constant of 21.39 nM. ZINC000000026220 showed three n-alkyl interactions with Met49, Leu167 and Pro168, one n-sulfur interaction with Met165, and three conventional H-bonds with His164, Gln189 and Gln192. This compound also shared many van der Waals interactions with several residues of Mpro. This compound is an analogue of 2′-deoxyguanosine, and 2′-deoxyguanosine has been proved to be inhibitory against thymidine kinases (TKs) encoded by Herpes simplex viruses type 1 (HSV1) and type 2 (HSV2)[22, 23]. The third compound ZINC000000006645 showed a binding energy of -10.43 kcal/mol and an estimated inhibition constant of 22.50 nM in its interaction with Mpro. ZINC000000006645 showed a n-alkyl and a n-n T-shaped interaction with the Met165 and His41 residues of the Mpro, respectively. In addition, this compound formed a non-classical carbon-hydrogen bond with Pro168, and three classical hydrogen bonds with Gln166, Arg188 and Thr190. The compound ZINC000000006645 is an isomer of the compound 2-Anilino-6a-(hydroxymethyl)-4,5,6,9a-tetrahydro-3aH-cyclopenta[d][1,3]oxazole-4,5,6-triol and it has been reported to strongly inhibit lysosomal α-glucosidase, β-galactosidase and β-glucocerebrosidase enzymes in vitro [24].

In this study, 10-ns MD simulation was carried out to evaluate the conformational changes, stability and the dynamic evolution of the Mpro-ZINC000000011271 complex. To determine how much the ligand binding pose has changed over the 10-ns simulation period, the root mean square deviation (RMSD) of protein and protein-ligand complex was computed. The RMSD of unbounded Mpro enzyme remained stable between 1.4 ns and 10 ns time period at 0.15 nm. The RMSD of Mpro-ZINC000000011271 remained stable at 0.2 nm at 0.3 ns until 1.7 ns, then slightly increased to 0.32 nm between 1.8 ns - 3.4 ns, and finally reached its steady state at 0.65 nm between 3.5 ns and 10 ns. The root mean square fluctuation (RMSF) captures the fluctuation from the average position for each atom, and therefore gives information about the flexibility of regions of the protein or ligand molecules. RMSF of the unbounded Mpro showed that the residues located in the active site (Phe140, Gly143, Cys145, His164, Glu166, His172, Gln189, Thr190 and Glu 192) of the enzyme remained stable between 0.05 nm and 0.1 nm. The other residues of the enzyme also showed a rigid character in terms of RMSF, which ranged from 0.03 nm to 0.21 nm. The RMSF of ZINC000000011271 during MD simulation showed that the ligand atoms had low fluctuations between 0.02 nm and 0.20 nm, indicating that the atom positions of the ligand did not significantly changed during the interactions with the binding pocket residues. Although the 29., 30. and 31. atoms of the ligand displayed high RMSF values, their fluctuations were limited to an acceptable value of 0.2 nm. The H-bond plotting showed an average of 3.66 hydrogen bond interactions between Mpro and ZINC000000011271 throughout the 10-ns MD simulation with a maximum of 13 hydrogen bonds. The resulting dynamic H-bond formation pattern well illustrated the tight binding mechanism of ZINC000000011271 to Mpro and the underlying reason for its inhibition.

One of the targets, which is important in the treatment of Coronavirus (COVID-19) and whose inhibition can be quite functional, is the Mpro enzyme. In this study, the structure-based virtual screening revealed the potential of three orphaned drugs (ZINC000000011271, ZINC000000026220 and ZINC000000006645) to inhibit Mpro. Two of these entities (ZINC000000006645, ZINC000000026220) have thymidine kinase inhibitory properties, while the third entity (ZINC000000006645) inhibits glucosidase, galactosidase and glucocerebrosidase enzymes. These compounds displaying hit features are drug-like, ADMET profiles were found to be harmless, and according to the SwissADME database (http://www.swissadme.ch/index.php#), their synthetic accessibility (ease of synthesis) scores fall within the acceptable range. Furthermore, among these three compounds, ZINC000000011271 proved its tight binding with Mpro in 10-ns molecular dynamics simulation and preserved its stable hydrogen bond formation until the end of the trajectory. For this reason, these hits defined in our study may be used in the development and advanced optimization of potent COVID-19 inhibitors.

Scientific Responsibility Statement
The authors declare that they are responsible for the article’s scientific content including study design, data collection, analysis and interpretation, writing, some of the main line, or all of the preparation and scientific review of the contents and approval of the final version of the article.

Animal and human rights statement
All procedures performed in this study were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards. No animal or human studies were carried out by the authors for this article.

Funding: None

Conflict of interest
None of the authors received any type of financial support that could be considered potential conflict of interest regarding the manuscript or its submission.

References

How to cite this article: