A computational biology approach for the identification of potential SARS-CoV-2 main protease inhibitors from natural essential oil compounds. [version 2; peer review: 1 approved]

Rizone Al Hasib\(^1\), Md. Chayan Ali\(^1\), Md. Shahedur Rahman\(^2,3\), Md. Mafizur Rahman\(^1\), Fee Faysal Ahmed\(^4\), Md. Abdullah Al Mashud\(^{1,5}\), Md. Azizul Islam\(^1\), Mohammad Abu Hena Mostofa Jamal\(^1\)

\(^1\)Biotechnology and Genetic Engineering, Islamic University, Kushtia, Khulna, 7003, Bangladesh
\(^2\)Department of Genetic Engineering and Biotechnology, Jashore University of Science and Technology, Jeshore, Khulna, Bangladesh
\(^3\)Bioinformatics and Microbial Biotechnology Laboratory, Department of Genetic Engineering and Biotechnology, Jashore University of Science and Technology, Jashore, 7408, Bangladesh
\(^4\)Department of Mathematics, Faculty of Science, Jashore University of Science and Technology, Jeshore, Khulna, 7408, Bangladesh
\(^5\)Department of Electrical and Electronic Engineering, Islamic University, Kushtia, Bangladesh

Abstract

**Background:** Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) has fomented a climate of fear worldwide due to its rapidly spreading nature, and high mortality rate. The World Health Organization declared it a global pandemic on 11 March 2020. Many endeavors have been made to find appropriate medications to restrain the SARS-CoV-2 infection from spreading but there is no specific antiviral therapy to date. However, a computer-aided drug design approach can be an alternative to identify probable drug candidates within a short time. SARS-CoV-2 main protease is a proven drug target, and it plays a pivotal role in viral replication and transcription.

**Methods:** In this study, we identified a total of 114 essential oil compounds as a feasible anti-SARS-CoV-2 agent from several online reservoirs. These compounds were screened by incorporating absorption, distribution, metabolism, excretion, and toxicity (ADMET) profiling, molecular docking, and 50 ns of molecular dynamics simulation to identify potential drug candidates. The crystallized SARS-CoV-2 main protease structure was collected from the Research Collaboratory for Structural Bioinformatics Protein Data Bank database (Protein Data Bank ID 6LU7).

**Results:** According to the results of the ADMET study, none of the compounds have any side effects that could reduce their druglikeness or pharmacokinetic properties. Among 114 compounds, we selected bisabololoxide B, eremanthin, and leptospermine as top drug
candidates based on their higher binding affinity scores, and strong interaction with the Cys 145-His 41 catalytic dyad. Finally, the molecular dynamics simulation was implemented to evaluate the structural stability of the ligand-receptor complex. Molecular dynamics simulation disclosed that all the hits showed conformational stability compared to the positive control α-ketoamide.

**Conclusions:** Our study showed that the top three hits might work as potential anti-SARS-CoV-2 agents, which can pave the way for discovering new drugs, but further in vivo trials will require for experimental validation.

**Keywords**
SARS-CoV-2, Main protease, ADMET analysis, Molecular docking, Molecular dynamics simulation, Essential oil.
Introduction
A new strain of coronavirus termed severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) is responsible for a variety of respiratory diseases that emerged in Wuhan, Hubei province, China towards the end of 2019, and since then has spread globally. SARS-CoV-2 is a positive sense, enveloped RNA virus that belongs to the β genus of the coronaviridae family. Live animals like bats sold at the local Huanan seafood wholesale market are thought to be a possible point of origin of SARS-CoV-2 as most of the fatalities are inhabited nearby it. Patients with SARS-CoV-2 infection suffer from a fever at the preliminary stage of the disease as a clinical symptom with some other symptoms including headache, dry cough, difficulty breathing, and pneumonia. Progressive respiratory failure due to alveolar damage with the initiation of the disease can even cause the death of the patient. Considering the potential health risks associated with this disease, a therapeutic strategy to reduce the transmission rate is urgently needed, but no specific drugs have yet been discovered. Several medications like ritonavir, lopinavir, oseltamivir, and ganciclovir have been assayed as yet for the inhibition of SARS-CoV-2. All such drugs are repurposed, and though they may have shown some effectiveness against SARS-CoV-2, but their adverse effects and poor success rate against emerging diseases cannot be neglected. Simultaneously, lopinavir-ritonavir-based treatment in hospitalized adult patients with severe SARS-CoV-2 has not shown any significant effect. Combined doses of hydroxychloroquine and azithromycin exhibited more satisfactory results in the removal of viral load of SARS-CoV-2 than hydroxychloroquine alone, as claimed by Gautret et al., in their study but was refuted by another study mentioning the clinical failure of these two drugs in severe cases.

A protease enzyme called main protease (Mpro) or chymotrypsin-like protease (3CLPro) and a papain-like protease act in the maneuvering of two polyproteins (1a and 1ab) into non-structural proteins (nsp5), which are liable for viral replication. The SARS-CoV-2 main protease specifically cleaves polypeptide sequences after a residue of glutamine as like Leu-Gln (Ser, Ala, Gly) (indicates the cleavage site), which represents the main protease as a suitable drug target since no human host-cell proteases are identified with such substrate specificity. SARS-CoV-2 main protease maintains an amino acid sequence analogy of 96% in contrast to SARS-CoV-1 and varies merely at 12 out of 303 places in the amino acid sequence. Moreover, SARS-CoV-2 main protease comprises three major domains, namely chymotrypsin-like domain I (residues 10–99), picornavirus 3C protease-like domain II (residues 100–184); these two catalytic domains contain a six-stranded antiparallel β-barrel and the α-helical domain III (residues 201–303); necessary for dimerization. Domain II and domain III are linked together through an extended loop region between residues 185–200. A cleft between domains I and II holds the substrate-binding site, and the Cys145 and His41 catalytic dyad residues are also included here.

As mentioned earlier, there is no appropriate medication for SARS-CoV-2; however, several previous studies indicated that phytochemicals and essential oils secreted from medicinal plants may have been used to prevent viral replication. Essential oils have been reported to exhibit antiviral effects along with their antibacterial and antifungal activity. Compounds derived from essential oils have been proven efficient against the herpes simplex virus (HSV-1 and HSV-II) and SARS-CoV by numerous experiments. An in-silico investigation against SARS-CoV-2 reported that several antiviral agents or other alleviation of SARS-CoV-2 symptoms could be amplified by essential oils. So, through this research, we hope to provide evidence that our selected essential oil compounds could be effectively used as a treatment against the deadly SARS-CoV-2 infection. The goal of this study was to investigate the effects of essential oils and phytochemicals obtained from different plants against SARS-CoV-2 main protease using absorption, distribution, metabolism, elimination, and toxicity (ADMET) profiling, molecular docking, and molecular dynamics simulation.

Methods
Essential oil library generation
In our study, we considered essential oil compounds as a possible anti-SARS-CoV-2 agent. In August 2020, we conducted a randomized study to identify relevant essential oil compounds for the current study. From previous research and observations, we formulated a repository of the bioactive compounds of essential oils. Through checking relevant studies in PubMed, PubChem, Google Scholar, the Web of Science, and Scopus databases, we built a catalog of biologically active essential oil molecules. As inclusion criteria, the compounds were evaluated for their antimicrobial and antiviral properties. Our research included the compounds that had previously shown these two characteristics.
Compounds were excluded from the study if they did not have these two key features. The 3D conformation in spatial data file (SDF) format, InChI key, and canonical simplified molecular input line entry systems (SMILES) of all the compounds were collected from the PubChem database once identified from the search databases. This research identified a total of 114 essential oil compounds, which were then investigated to see whether they had antiviral properties against SARS-CoV-2 (See extended data file for plant information of all the compounds including their binding affinity scores).

ADMET analysis
Based on their ADMET properties, all the essential oil compounds were filtered. Qikprop determined the absorption, distribution, metabolism, and elimination (ADME) properties of the compounds. Physicochemical properties, pharmacokinetics, lipophilicity, water-solubility, and medicinal chemistry are some of the ADME properties estimated to analyze drug profiles of top drug candidates. In this study, the rule of five, rule of three, solvent accessible surface area (SASA), aqueous solubility (QPlogS), human oral absorption, Caco-2 cell permeability, and other pharmacological and druglikeness properties like predicted octanol/water partition coefficient (QPlogPo/w), predicted water/gas partition coefficient (QPlogPw), and conformation-independent predicted aqueous solubility (CIQPlogS) were analyzed using Qikprop and the outputs were compared to the threshold values. The compounds were filtered in a stepwise process. The ligands were initially preprocessed for ADME profiling using the LigPrep module of the Schrödinger suite 2020-3 (https://www.schrodinger.com/products/ligprep). The processed ligands were then directly inserted as an input file (.maegz) into Qikprop (https://www.schrodinger.com/products/qikprop). For the pharmacokinetic properties mentioned above, an output file (.csv) was generated automatically.

First, compounds that pursued the rule of five and the rule of three were included. Following that, these compounds that passed the rule of five and three tests were assessed based on their maximum oral absorption rate (the cut-off value is equal to three). In addition, the toxicity of the compounds was analyzed via the online platform admetSAR (http://lmmd.ecust.edu.cn/admetSAR). admetSAR is a free tool that offers a user-friendly interface to search for ADMET properties. It is estimated that there are more than 210,000 ADMET specified datasets for more than 96,000 specific compounds in admetSAR, with 45 different kinds of ADMET-associated properties, which were collated from a wide range of previous studies. The toxicity assessment results were generated using the Canonical SMILES of the compounds obtained from PubChem as input data into admetSAR. Here, only negative outcomes in the hERG inhibition, AMES test, and carcinogenic test were taken into account. In summary, the following parameters were used to filter the compounds from the initial 114 identified: Rule of Five = 0 violations, Rule of Three = 0 violations, Human Oral Absorption = 3, hERG inhibition = non-inhibitor, AMES Toxicity = non-AMES Toxic, and Carcinogenic Property = non-carcinogen.

Preparation of receptor
The crystallized structure of the SARS-CoV-2 main protease (PDB ID 6LU7) was compiled from the reserves of the Research Collaboratory for Structural Bioinformatics Protein Data Bank (RCSB PDB) (https://www.rcsb.org). The RCSB PDB is the world’s first open-access digital resource for experimentally determined 3D structures of biological macromolecules and their complexes. Using the search, visualization, and analysis tools of RCSB PDB, users can easily explore chemical interactions that stabilize macromolecules and play a significant role in their interactions and functions. During receptor preparation, the water molecules and ligands had been eliminated from it, and the steepest descent and conjugate gradient techniques were utilized for energy minimization. The final receptor was configured using GROMACS 96 43B1 algorithm in Swiss-PdbViewer (version 4.0.4) (https://spdbv.vital-it.ch) and Chimera (Amber Force field, version 1.14) (https://spdbv.vital-it.ch). Swiss-PdbViewer is a user-friendly platform that combines the functions for protein structure modeling, assessment, and manipulation. It has been improved with numerous features that allow users to identify and search for general structural motifs in a set of structures. In addition, UCSF Chimera is an interactive tool for visualizing molecular structures and is completely free for non-commercial purposes.

Ligand preparation and molecular docking
The ligand preparation was carried out by utilizing the open-source virtual screening tool PyRx (https://pyrx.sourceforge.io). To minimize the energy of the ligands, the mmff94 (Merck molecular force field) force field, which is available in PyRx, was employed. For ligand optimization, the steepest descent technique was used, with a total number of 200 steps. After that, the ligands were transformed into a protein data bank (PDB) format by implementing the open-source chemical toolbox Open Babel (http://openbabel.org/wiki/Main_Page). For this, the SDF format of the ligands retrieved from the PubChem database was used as inputs in Open Babel. Then the autodock protein data bank (PDB), partial charge (Q), and atom type (T) [PDBQT] format of the ligands was created via PyRx for further study and analysis. The grid box generation and molecular docking study were performed by utilizing the autodock wizard of PyRx, version 0.8 as the free version (https://pyrx.sourceforge.io). PyRx provides an autodock wizard with a simple user interface, making it a useful tool for computer-aided drug design (CADD). During the molecular docking study, ligands were considered flexible, and the protein was assumed to be rigid. The grid parameter configuration file was initiated utilizing the auto grid engine in autodock (grid
Molecular dynamics simulation

To obtain detailed information about the predicted molecular interactions, the best-docked complexes were analyzed for 50 ns of MD simulations utilizing the GROMOS96 43a1 force field of the ‘GROMACS’ macromolecular simulation system (https://simlab.uams.edu/ProteinWithLigand/protein_with_ligand.html). In the beginning, topology parameters for the selected ligands were obtained from the ‘PRODRG’ webserver (http://davapc1.bioch.dundee.ac.uk/cgi-bin/prodrg). PRODRG delivers high-throughput protein-ligand crystallography with quick, automated, and consistent access to small-molecule topologies and parameters. PRODRG produced topologies are often of higher quality than topologies generated by other methods and are acceptable for MD simulation in the GROMACS environment.

In this study, Mpro-ligand complexes were solvated in a cubic box (0.15 M) with the SPC (spc216) water model. Cubic boxes are generally utilized in molecular dynamics simulations for the sake of geometric simplicity. Despite considering polarization, the SPC water model is one of the most reliable three-center water models. It provides accurate water density and dielectric permittivity in molecular dynamics simulations.

For each simulation method, the steepest gradient technique (5000 cycles) was implemented to minimize the energy. Under periodic boundary conditions with a predefined pressure of 1 bar, the equilibration period was accomplished for 100 ps at a constant temperature of 300 K. Finally, the simulation outputs like root mean square deviation (RMSD), root mean square fluctuations (RMSF), radius of gyration (Rg), solvent accessible surface area (SASA), and the number of hydrogen bonds were calculated employing the built-in features of ‘GROMACS’. The lower RMSD, Rg, and SASA values represent the compactness of the system whereas the high RMSF value indicates the structural flexibility. Likewise, intermolecular hydrogen bonds formed during protein-ligand MD simulation are crucial for the conformational stability of the complex.

Results

ADMET profiling

The propensity of a drug to acquire pharmacologically active concentration at targeted therapies can be evaluated by a set of ADME parameters. Evaluation of ADME may be efficient in combination with toxicity appraisement in silico models. The ADMET properties and drug-likeness of each essential oil compound were determined in this study (See underlying data file, ADME.xlsx). We did not consider the ability of the compounds to cross the blood-brain barrier (BBB) because SARS-Cov-2 is not directly linked with our central nervous system. Out of 114 compounds tested, 41 were found to violate the rule of five, rule of three, or both. Just one of the remaining 73 compounds had an oral absorption rate of less than three. As a result, the remaining 72 compounds were screened based on the hERG inhibition, AMES, and carcinogenic tests (See underlying data file, Toxicity.xlsx). Finally, 53 essential oil compounds met the druglikeness and ADMET profiles requirements outlined in the methods. Following that, a molecular docking approach was then used to examine these 53 filtered compounds.

Molecular docking

Since there is no specific medication to date for the treatment of SARS-CoV-2, it is very urgent to identify feasible biological compounds which block SARS-CoV-2 main protease and act as potential anti-SARS-CoV-2 agents. In this investigation, we docked 53 essential oil molecules against SARS-CoV-2 main protease (PDB ID: 6lu7) that passed the ADMET criteria and evaluated the binding affinity of all these probable inhibitor compounds to the receptor-binding site (See underlying data file for all binding affinity scores). We chose the top three potential compounds based on their higher binding affinity scores. The PubChem Id, molecular formula, molecular weight, IUPAC name, 2D structures, and binding affinity of the selected hits including the positive control are shown in Table 1.

Among the compounds, bisabololoxide B (obtained from Matricaria recutita) disclosed the highest binding affinity of –6.6 kcal/mol, and both eremanthin (obtained from Laurus nobilis) and leptospermone (obtained from Leptospermum scoparium) exhibited a binding affinity of –6.3 kcal/mol during docking analysis. Zhang et al. have shown in a recent analysis that the positive inhibitor α-ketoamide interacts with the main protease residues His41, Gly143, Ser144, Cys145, His163, His164, Glu166, Pro168, and Gln189. The Cys-His catalytic dyad (Cys-145 and His-41) is the target region for the inhibitors to bind in. Our selected three hit compounds interact with either Cys-145 or His-41 or both by forming different types of bonds.
Table 1. PubChem Id, molecular formula, molecular weight, IUPAC name, 2D structure, and binding affinity of selected compounds and the control.

<table>
<thead>
<tr>
<th>Selected compounds</th>
<th>PubChem Id</th>
<th>Molecular formula</th>
<th>Molecular weight (Da)</th>
<th>IUPAC name</th>
<th>2D structure</th>
<th>Binding affinity (kcal/mol)</th>
</tr>
</thead>
<tbody>
<tr>
<td>α-ketoamide</td>
<td>6482451</td>
<td>C_{36}H_{53}N_{5}O_{11}</td>
<td>731.8</td>
<td>(2S)-2-[[2-[[3-[[2S)-2-((2-methylpropoxy)carboxy)amino]acetyl]amino]-3-(1,3-dioxan-2-yl)propanoyl]amino]-2-oxohexanoyl]amino]acetyl]amino]-2-phenylacetic acid</td>
<td><img src="image" alt="2D structure" /></td>
<td>7.3</td>
</tr>
<tr>
<td>Bisabololoxide B</td>
<td>117301</td>
<td>C_{15}H_{26}O_{2}</td>
<td>238.37</td>
<td>2-[5-methyl-5-(4-methylcyclohex-3-en-1-yl)oxolan-2-yl][propan-2-ol</td>
<td><img src="image" alt="2D structure" /></td>
<td>6.6</td>
</tr>
<tr>
<td>Eremanthin</td>
<td>100572</td>
<td>C_{15}H_{18}O_{2}</td>
<td>230.30</td>
<td>(3α,5αR,9αR,9βS)-6-methyl-3,9-dimethylidene-4,6α,7,8,9,9β-hexahydro-3αH-azuleno[4,5-b]furan-2-one</td>
<td><img src="image" alt="2D structure" /></td>
<td>6.3</td>
</tr>
<tr>
<td>Leptospermone</td>
<td>3083632</td>
<td>C_{15}H_{22}O_{4}</td>
<td>266.33</td>
<td>2,2,4,4-tetramethyl-6-(3-methylbutanoyl)cyclohexane-1,3,5-trione</td>
<td><img src="image" alt="2D structure" /></td>
<td>6.3</td>
</tr>
</tbody>
</table>
Figure 1. Molecular interactions between selected essential oil compounds and SARS-Cov-2 main protease (M\textsuperscript{pro}). (a) Molecular interactions of positive control α-ketoamide with M\textsuperscript{pro}, (b) interactions between Bisabololoxide B and M\textsuperscript{pro}, (c) between Eremanthin and M\textsuperscript{pro}, and (d) between Leptospermine and M\textsuperscript{pro}.
Bisabololoxide B is demonstrated in Figure 1(b), it interacted with A: HIS-41 and A: MET-49 by forming alkyl bonds. Eremanthin, demonstrated in Figure 1(c), formed conventional hydrogen bonds with A: SER-144 and A: GLY-143. At the same time, eremanthin interacted with A: CYS-145 by forming both conventional hydrogen and alkyl bonds, whereas it formed only alkyl bonds with A: HIS-163 and A: HIS-172. Leptospermone, demonstrated in Figure 1(d), interacted with A: CYS-145, A: SER-144, and A: GLY-143 through strong hydrogen bonds. It also interacted with A: HIS-41 and A: MET-49 by forming alkyl bonds and with A: MET-165 by carbon-hydrogen bond. Such findings indicated that our selected essential oil compounds could be effective in treating SARS-CoV-2.

**Molecular dynamics simulation**

The molecular interactions can describe biological systems in a dynamic fashion, but usually, molecular docking defines only a single whiff of protein-ligand interaction. Hence, we simulated the dynamic nature of all the top three complexes to capture their different conformations attained in the solvated condition. Five parameters, RMSD, RMSF, Rg, SASA, and the number of hydrogen bonds were utilized in our study to analyze the dynamic behavior of the top three protein-ligand complexes. The average RMSD values for α-ketoamide-Mpro, bisabololoxide B-Mpro, eremanthin-Mpro, and leptospermone-Mpro complex were 0.34 Å, 0.33 Å, 0.32 Å, and 0.23 Å, respectively where leptospermone showed the lowest RMSD value (shown in Table 2). The positive control α-ketoamide had a stable conformation from 1 to 20 ns, but from 21 ns to the rest of the simulation, there was significant structural distortion. From 1 ns to 20 ns and 40 ns to 50 ns, the bisabololoxide B complex was relatively stable; however, it fluctuated noticeably from 21 ns to 40 ns. Eremanthin remained almost stable throughout the simulation process; only a small amount of structural drifting was found at 8 ns to 10 ns and from 48 ns to 50 ns. The leptospermone complex remained stable from 15 ns to 20 ns and 35 ns to 45 ns but exhibited significant structural instability during the rest of the time. Furthermore, to evaluate the local residual transition, the RMSF benchmark was used. There is a proportional relationship between the RMSF value and the flexibility of a system. If the RMSF value is greater, the system’s flexibility will be greater. In comparison to the positive control

![Figure 2](image-url)
α-ketoamide-Mpro complex (0.21 Å), all three hits had a lower average RMSF value. Eremanthin-Mpro complex showed the lowest average RMSF value (0.15 Å) compared to the other systems. Besides this, both bisabololoxide B-Mpro and leptospermone-Mpro complex displayed an average RMSF value of 0.16 Å (shown in Table 2). Consequently, the lower Rg value evaluates the compactness of the protein structure. In this study, bisabololoxide B-Mpro and eremanthin-Mpro complex showed the lowest Rg value (2.13 Å). On the other hand, leptospermone-Mpro complex exhibited the same Rg value (2.16 Å) as the positive control α-ketoamide-Mpro complex (shown in Table 2). A minimal fluctuation rate is always expected during a simulation process. Here also, we calculated the SASA for all systems to determine the compactness of the system (shown in Table 2). Both eremanthin-Mpro (134.31 nm²) and bisabololoxide B-Mpro complex (134.44 nm²) showed a lower average SASA value than the positive control α-ketoamide-Mpro complex (139.03 nm²). However, the average SASA value of leptospermone-Mpro complex was 139.75 nm² which is almost similar to the positive control α-ketoamide-Mpro complex (shown in Table 2). Furthermore, we calculated the total average number of hydrogen bonds for Mpro-ligand complexes, as intermolecular hydrogen bonds between protein-ligand complex play a vital role in their conformational stability. Here, the positive control α-ketoamide-Mpro complex formed the highest average number of hydrogen bonds (213.79) during the simulation process. Similarly, the eremanthin-Mpro complex produced almost the same number (213.45) of hydrogen bonds as the positive control α-ketoamide-Mpro complex, whereas bisabololoxide B-Mpro and leptospermone-Mpro complex formed an average of 211.03 and 207.13 hydrogen bonds respectively (shown in Table 2).

Discussion
In reality, SARS-CoV-2 is a big concern for the scientific community due to its high infectivity. On 30 December 2019 and 30 January 2020, the World Health Organization (WHO) raised warnings and announced this viral infection a ‘public health emergency of international concern’. Since the first report of a coronavirus-related pneumonia outbreak in December 2019, the virus SARS-CoV-2 that causes the infection has emerged into a pandemic, with >100 million individuals infected in over 210 countries and around two million people dying from COVID-19 as of today.

Nevertheless, no sufficient breakthrough in action against SARS-CoV-2 has been developed yet, but researchers are profoundly focused on implementing vaccines and methodologies to eradicate the disease. However, clinical trials of many candidates are still going on, and some of them pose contentious issues. For instance, an adenosine analogue named Remdesivir which inhibits viral RNA polymerases was known to be an advantageous antiviral agent against a diverse species of RNA viruses along with Ebola, SARS-CoV, and MERS-CoV and exhibited remedial activity in non-clinical models of these coronaviruses but had several undesirable effects in case of SARS-CoV-2 and failed in phase three clinical trials. Failure of hydroxychloroquine in preventing illness of patients with high or moderate-risk exposure to SARS-CoV-2 was also reported previously. Therefore, we desperately need potential anti-SARS-CoV-2 drugs to keep people’s lives safe. New plant-based therapies compensate for around 60% of total Western medicines, and now they are being extensively studied to develop a cure for combating SARS-CoV-2. This study assayed several plant essential oil compounds through computational approaches to find out effective SARS-CoV-2 Mpro inhibitors.

Since prehistoric periods, plants and their derivatives have been utilized to treat diseases in humans, and they were reported to be clinically safe for humans. Plant essential oils are considered to have a range of useful antimicrobial, antibacterial, antiviral, antiparasitic, and insecticidal properties, and are now widely studied as a consequential source of drug development for their safety profiles and advantages over synthetic drugs. Some essential oils have shown to be effective against multiple RNA, and DNA viruses, including avian influenza A virus, herpes simplex virus, dengue virus, and poliovirus, as well as are being successful in the treatment of both infectious and chronic diseases.

The protocol of new drug development is complicated, time-intensive, costly, and involves both preclinical and clinical stages. With the aid of computational biology and bioinformatics, manufacturing operations of new drugs are altered and optimized, and hence development costs are minimized. This approach focuses on ligands interacting with their target proteins to evaluate the clinical efficacy of specific compounds.
The current study can help to explain the attributes of druglikeness and medicinal chemistry of the key metabolites against SARS-CoV-2. ADME properties estimated by the computational approach provide crucial perceptions into how the body treats or accepts a drug. Computational methods perform a vital role in predicting possible issues with ADME and toxicity and reducing the number of experiments that demand animal trials. Thus, supreme drug candidates were utilized to estimate their drug profiles with ADME analysis. In this experiment, all 114 compounds were firstly filtered based on their different ADMET properties. The rule of five and the rule of three was perfectly implemented in 72 compounds out of 114. There were no unfavorable outcomes discovered that could lead to a decrease in their ADME properties and druglikeness. The druglikeness property is assessed by Lipinski’s rule of five, whereas Jorgensen’s rule of three is more likely to suggest a compound’s bioavailability.

Furthermore, when developing new drugs for oral administration, solubility is one of the most significant parameters influencing absorption, and it improves many drug development operations. The 72 filtered compounds are all water-soluble and have a high rate of oral absorption. Consequently, toxicity tests showed that 53 of the 72 compounds were non-hERG inhibitors and found to be non-Ames toxic and non-carcinogenic in the Ames test and carcinogenic test, respectively.

The 53 screened compounds were then evaluated by molecular docking study. We considered the compounds as a set of ligands in our current analysis and selected the top three hits (Bisabololoxide B, Eremanthin, and Leptospermone) based on their higher binding affinity scores over the other compounds. Results revealed that bisabololoxide B showcased the maximum binding affinity score of -6.6 and the remaining two metabolites (eremanthin and leptospermone) exhibited a binding affinity score of -6.3 kcal/mol, respectively. Bisabololoxide B possesses anti-inflammatory properties, and its pharmacological significance was previously noted. Antioxidant properties of eremathin were previously assayed in STZ-induced diabetic rats. In an earlier study, eremathin exhibited notable cytotoxic effects on HeLa cervical cancer cells and hence acted as a promising anti-cancer agent against human cervical cancer. Leptospermone is an active compound of the β-triketone portion of manuka oil that exhibited good antiviral effects against HSV-1 and HSV-2 prior to infection. It works by interrupting the structure of the virion envelope and mask viral particles requires for adsorption. Leptospermone also showed antimicrobial effects against two bacterial strains, namely C. difficile and C. perfringens in an earlier study.

Moreover, molecular dynamics simulation was employed to define the exact physical movement and flexibility of ligand-receptor complexes to further evaluate the molecular docking performance. RMSD indicates the stability of ligand-protein complexes, where residual flexibility is characterized by RMSF. All of our hit compounds showed lower RMSD and RMSF values compared to the positive α-ketoamide control, which depicts their stable conformation with SARS-CoV-2 main protease. The lowest RMSD value was asserted by leptospermone-M<sup>rpo</sup> complex, which represents its structural compactness. Compared to other systems, eremathin-M<sup>rpo</sup> complex displayed a lower RMSF value that denotes its conformational stability. In addition, we also computed the Rg value for each system to determine their compactness. Both bisabololoxide B-M<sup>rpo</sup> and eremathin-M<sup>rpo</sup> complex showed a lower Rg value (2.13 Å), however, the leptospermone-M<sup>rpo</sup> complex depicted the same Rg value (2.16 Å) as the positive control α-ketoamide-M<sup>rpo</sup> complex. These outputs suggested that bisabololoxide B-M<sup>rpo</sup> and eremathin-M<sup>rpo</sup> complex were more compact than the positive control α-ketoamide-M<sup>rpo</sup> complex. Likewise, to evaluate the nature of protein expansion, we estimated the SASA values for all complexes. Here, the eremathin-M<sup>rpo</sup> (134.31 nm<sup>2</sup>) and bisabololoxide B-M<sup>rpo</sup> complex (134.44 nm<sup>2</sup>) disclosed a smaller SASA value compared to the positive control α-ketoamide-M<sup>rpo</sup> complex (139.03 nm<sup>2</sup>). Consequently, the leptospermone-M<sup>rpo</sup> complex exhibited a nearly identical but a bit higher SASA value (139.75 nm<sup>2</sup>) to that of the positive control α-ketoamide-M<sup>rpo</sup> complex, indicating that the M<sup>rpo</sup> expanded slightly when it interacted with leptospermone. Furthermore, we calculated the average mean value of total intermolecular hydrogen bonds generated during the simulation process to analyze the structural consistency of the complexes. The bisabololoxide B-M<sup>rpo</sup> and leptospermone-M<sup>rpo</sup> complex formed 211.03 and 207.13 hydrogen bonds during the molecular dynamics simulation, while the number of intermolecular hydrogen bonds for eremathin-M<sup>rpo</sup> complex (213.45) is nearly equivalent to the positive control α-ketoamide-M<sup>rpo</sup> complex (213.79). All such outcomes obtained from molecular dynamics simulation indicate the conformational stability and compactness of our selected target compounds with the SARS-CoV-2 main protease.

Throughout the current investigation, we unveiled the molecular interactions of the leading drug candidates with SARS-CoV-2 main protease. The catalytic domain of SARS-CoV-2 main protease owned the ligand-binding sites for each ligand, and His41 and Cys145 residues serve as a site for substrate recognition among the other active residues as it comprises the catalytic dyad. All of the top three drug candidates were well adjusted into the pocket of SARS-CoV-2 main protease in which several hydrophobic amino acid residues such as Met-49, Gly-143, His-41, Ser-144, His-163, Cys-145, and Met-165 constitute a fairly hydrophobic atmosphere to maintain conformational stabilization of the compounds. Thus, based on the findings of this study, we can claim that each of our lead compounds has the potential...
to act as an inhibitor of SARS-CoV-2, and we hope that the study will pave the way for the development of novel pharmaceutical drugs.

Conclusion
The rapid spreading of novel coronavirus threatens human life terribly, which needs to be stopped for reducing the mortality rate. The main objective of this study was to identify novel inhibitors against the SARS-CoV-2 main protease. To discover novel Mpro inhibitors derived from natural essential oil compounds, ADMET analysis, molecular docking, and MD simulation were successfully conducted here. A total number of 114 essential oil compounds were screened by ADMET profiling and molecular docking study, and the efficacy of the top three compounds was validated by MD simulation. During the 50ns simulation, all the hits (bisabololoxide B, eremanthin, and leptospermone) appeared stable and possessed high affinities against SARS-CoV-2 main protease but further in vivo studies are needed for experimental validation. We conclude that the findings of this study could lead to the development of promising natural pharmaceutical agents for SARS-CoV-2 in the future.

Data availability
Underlying data
Figshare: Underlying data for ‘A computational biology approach for the identification of potential SARS-CoV-2 main protease inhibitors from natural essential oil compounds.’

https://doi.org/10.6084/m9.figshare.16879777.v1.56

This project contains the following underlying data:

- ADME.xlsx (spreadsheet of ADME properties for all the 114 compounds).
- Toxicity.xlsx (spreadsheet of toxicity for screened 72 compounds).
- Binding Affinity.xlsx (spreadsheet of binding affinity for screened 53 compounds).

Data are available under the terms of the Creative Commons Attribution 4.0 International license (CC-BY 4.0).

Extended data
Figshare: Underlying data for ‘A computational biology approach for the identification of potential SARS-CoV-2 main protease inhibitors from natural essential oil compounds.’

https://doi.org/10.6084/m9.figshare.16879777.v1.56

This project contains the following extended data:

- Extended Data 1.docx (document file for the plant source with their corresponding binding affinity of all the 114 compounds).

Data are available under the terms of the Creative Commons Attribution 4.0 International license (CC-BY 4.0).

Authors contribution
RAH and MCA conceived the plan of this research. RAH wrote the manuscript. RAH analyzed the data and made the figures. MSR, MMR, MAI, and FFA edited the manuscript. MCA did the molecular dynamics simulation, and MAHMJ supervised the whole project. All authors revised and approved the manuscript for final submission.

References


Thi Thanh Binh Nguyen
Department of Pediatrics, Hue University of Medicine and Pharmacy, Hue University, Hue, Vietnam

Article "A computational biology approach for the identification of potential SARS-CoV-2 main protease inhibitors from natural essential oil compounds". The manuscript is clearly presented and well substantiated and I recommend it for indexing in F1000Research with the following minor changes.

Specific Comments:
1. What is the reason behind choosing the PDB ID 6LU7?

2. *Physicochemical properties, pharmacokinetics, lipophilicity, water-solubility, and medicinal chemistry are some of the ADME properties estimated to analyze drug profiles of top drug candidates.* Rewrite the sentence like this.

3. *‘Despite considering polarization, the SPC water model is one the most reliable three-center waters models.’* Fix the typological error in this sentence.

4. “The propensity of a drug to acquire pharmacologically active concentration at targeted therapies can be evaluated by a set of ADME parameters"– please add at least two recent references.

5. “The protocol of new drug development is complicated, time-intensive, costly, and involves both preclinical and clinical stages"– please add references.

6. There are some spelling mistakes in the manuscript. And please avoid acronyms in the abstract. Please check and correct.

Is the work clearly and accurately presented and does it cite the current literature?  
Yes

Is the study design appropriate and is the work technically sound?
Yes

Are sufficient details of methods and analysis provided to allow replication by others?
Yes

If applicable, is the statistical analysis and its interpretation appropriate?
Yes

Are all the source data underlying the results available to ensure full reproducibility?
Yes

Are the conclusions drawn adequately supported by the results?
Yes

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: medicine, molecular in pharmacy, clinical study

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Author Response 11 Jun 2022

Mohammad Abu Hena Mostofa Jamal, Islamic University, Kushtia, Bangladesh

Thank you for your insightful feedback to improve our manuscript. Regarding your valuable feedback, we have addressed all the comments individually as follows:

Comment 1: What is the reason behind choosing the PDB ID 6LU7?
Response: Thanks for your valuable comment. The structure of PDB ID 6LU7 was obtained using the X-RAY Diffraction method and includes a built-in N3 inhibitor. Furthermore, several studies have previously utilized this PDB ID. For instance,


Moreover, the resolution of the main protease structure of PDB ID 6LU7 is 2.16 Å, indicating the high quality of the structure. For these reasons, we choose the PDB ID 6LU7 for retrieving the main protease structure.

Comment 2: Physicochemical properties, pharmacokinetics, lipophilicity, water-
solubility, and medicinal chemistry are some of the ADME properties estimated to analyze drug profiles of top drug candidates. Rewrite the sentence like this.

**Response:** We appreciate your feedback. We have rewritten the sentence.

**Comment 3:** 'Despite considering polarization, the SPC water model is one the most reliable three-center waters models.' Fix the typological error in this sentence.

**Response:** Thanks for your concern. We already have fixed the typological error in the latest version of the manuscript.

**Comment 4:** The propensity of a drug to acquire pharmacologically active concentration at targeted therapies can be evaluated by a set of ADME parameters"- please add at least two recent references.

**Response:** Thanks for your valuable feedback. We have added two recent references accordingly.

**Comment 5:** “The protocol of new drug development is complicated, time-intensive, costly, and involves both preclinical and clinical stages”- please add references.

**Response:** We appreciate your effort. Two new references have been added to the new version of the article.

**Comment 6:** There are some spelling mistakes in the manuscript. And please avoid acronyms in the abstract. Please check and correct.

**Response:** Thanks for your concern. We have checked and fixed all the spelling mistakes in the previous version of the manuscript and all the acronyms are avoided in the abstract section of the latest manuscript.

**Competing Interests:** No competing interests were disclosed.
The benefits of publishing with F1000Research:

- Your article is published within days, with no editorial bias
- You can publish traditional articles, null/negative results, case reports, data notes and more
- The peer review process is transparent and collaborative
- Your article is indexed in PubMed after passing peer review
- Dedicated customer support at every stage

For pre-submission enquiries, contact research@f1000.com