In silico studies of natural products from medicinal plants to identify potential inhibitors for SARS-CoV-2 3C-like protease

Nguyen Tan Khanh¹, Tran Thanh Hoa²*  
¹Office of Scientific Research, Dong A University, 33 Xo Viet Nghe Tinh, Hai Chau, Da Nang 55000, Viet Nam  
²VN-UK Institute for Research and Executive Education, The University of Danang, 158A Le Loi, Hai Chau, Da Nang 55000, Viet Nam

Submitted August 18, 2020; Revised August 30, 2021; Accepted September 5, 2021

Abstract

The coronavirus outbreak of 2019 (COVID-19) has been seen as a global health concern. The 3-chymotripsin-like protease or SARS-CoV-2 main protease, whose ability to control the coronavirus replication, has been a promising target for developing drugs against coronavirus infection. In this study, we used the molecular docking approach to screen natural compounds from traditional medicinal plants in Vietnam to find the possible inhibitors of SARS-CoV-2 main protease. The docking results revealed that among 170 compounds, the top 15 compounds with the high binding affinity to 3CL protease (PDB ID: 6WNP) might act as promising anti-SARS-CoV-2 molecules. Among these compounds, Gracillin and Proanthocyanidins showed the highest binding affinity with a docking score of -9.2 Kcal/mol. However, further experiments are required for developing possible natural therapeutic medications for combating this coronavirus.

Keywords. COVID-19, coronavirus, 3CL protease, natural compound, molecular docking.

1. INTRODUCTION

Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2), which is a positive-stranded RNA virus belonging to the Coronaviridae family, causes the current pandemic of Coronavirus disease 2019 (COVID-19). This virus was first identified in December 2019 in China and has quickly spread across the world.[1] This virus is able to cause multiple typical symptoms such as fever, cough, nausea, pneumonia and fatigue. The patients who have chronic diseases, including cardiovascular, kidney failure, blood pressure, diabetes and delayed care for acute emergencies, have a high mortality rate. To June 22 2021, SARS-CoV-2 has four variants of concern (B.1.1.7, B.1.351, B.1.617.2, P.1), reached almost nations worldwide with more than 177 million cases and 3.857.974 deaths.[2]

At present, the standard treatment of COVID-19 is based on supportive treatments. Many countries and pharmaceutical corporations are urgently researching medicines for coronavirus-patients treatment. However, only one specific drug (Remdesivir) has been approved by FDA to treat this disease.[3] Some scientific evidence showed that therapies apply antiviral drugs combined with traditional medicines have the ability to combat coronavirus effectively.[4] Therefore, natural compounds extracted from medicinal plants can be potential solutions for COVID-19 treatment. In this direction, computational techniques like molecular docking and machine learning have proved valuable in the screening of large-scale several medical plants based on traditional medicine background to find out potential compounds to be effective in inhibiting coronavirus. This in silico process was a prerequisite for later further experimental analyses to discover new medication to the treatment of SARS-CoV-2.

Based on sequence analysis, the SARS-CoV-2 genome is suggested to contain 14 open reading frames. The major open reading frame ORF1ab encodes for two polyproteins (PP1a and PP1ab), which are cleaved to generate 16 non-structural proteins that are essential for viral propagation.[5] The 3-chymotripsin-like protease (3CL protease or SARS-CoV-2 main protease) and the papain-like protease are two proteases responsible for this proteolytic processing. In addition, 3CL protease cleavages proteins at a cleavage site with a glutamine in position P1, a feature that has not been discovered in human proteases.[6] Therefore, this protease is considered to be a potential selective target for anti-coronavirus drug development. In this research, the natural compounds from traditional medicinal plants
were screened by virtual docking with the 3CL protease to discover potential inhibitors of coronavirus infection.

2. MATERIALS AND METHODS

2.1. Protein and ligand preparation

The crystal structure of coronavirus 3CL$^{\text{pro}}$ (PDB ID: 6WNP) was downloaded from Research Collaboratory for Structural Bioinformatics Protein Data Bank (RCSB PDB). Next, the crystal structure of the 3CL$^{\text{pro}}$ was processed by removing existing ligand, water molecules. After that, polar hydrogens and Kollman charges were added to the protein using the Autodock tools version 1.5.6. Finally, the macromolecule was exported into a dockable pdbqt format for molecular docking.

The structures of natural compounds from medicinal plants in central Vietnam were retrieved from the PubChem database or prepared by MarvinSketch (ChemAxon, USA). After that, the structures of natural compounds were converted into 3D structures by “Clean in 3D” function in MarvinSketch and save in the structure-data file (sdf) format. In the last step, all ligands were converted to dockable pdbqt format utilizing Open babel.

2.2. Molecular docking

Virtual screening of the bioactive compounds on coronavirus 3CL$^{\text{pro}}$ were carried out using AutoDock Vina. The center of grid box was determined based on the active site of coronavirus 3CLpro and the binding site of Boceprevir in crystal structure (PDB ID: 6WNP). After identifying the center site, the grid box was generated with parameters: center $x = 8.24$, center $y = 26.90$, center $z = 23.69$, size $x = 25$, size $y = 25$, size $z = 25$. Docking scores are reported in kcal/mol, and the results were classified by compounds docking scores. Finally, the molecular interactions between 3CL$^{\text{pro}}$ and selected ligands were visualized by Discovery Studio Visualizer.

2.3. Solubility and toxicity prediction

The solubility of compounds was calculated using MarvinSketch (ChemAxon, USA). The acute oral toxicity was predicted using DL-AOT prediction server.[7]

3. RESULTS AND DISCUSSION

From published literature, the information of 170 compounds from 15 medicinal plants in the central of Vietnam was collected. 3D structures of these compounds were constructed and used for the docking process with the target 3CL protease of SARS-CoV-2. These compounds were docked to the catalytic site of 3CL protease.

To demonstrate that the selected protein structure (PDB ID: 6WNP) and docking protocol are reliable, we re-docked Boceprevir, the ligand in the crystal structure, to 3CL protease and analyzed the root mean square deviation (RMSD) of docking pose against the crystal pose. The result showed that the docking pose has a similar conformation and orientation as the crystal pose with RMSD $\approx 1.35$ Å (figure 1). Previous studies have considered an RMSD within 2 Å to be good and valuable for docking experiments,[8] so this docking protocol can be used in our study. In addition, Boceprevir was also shown activities to inhibit 3CL protease of SARS-CoV-2.[9] Therefore, we decided to choose Boceprevir as a reference inhibitor.

![Figure 1: Docking pose (orange) and crystal pose (grey) of Boceprevir at the crystal structures of 3CL protease](image-url)

After docking, the results revealed 27 compounds that have better docking scores than -7.7 Kcal/mol, a docking score of Boceprevir. For selecting bioactive compounds that have a good binding affinity with 3CL protease, a cut-off of -8.2 Kcal/mol was set. Fifteen compounds surpassed this threshold and the results of these top-ranked compounds were represented in table 1.

Structurally, these 15 compounds are mostly in the polyphenol group (7 flavonoids, 3 phenolic compounds) and steroidal saponins (3 compounds). Flavonoid is well-known to be a group of substances with good biological effects because of the antiviral and antibacterial activities of phenol radicals.[10,11] The flavonoid group has also been docking screened for inhibitors of influenza H1N1[12] and SARS-CoV.[13] Saponins, including steroid or triterpenoid glycosides, are commonly found in a wide range of plants. Saponins also have a broad spectrum of
Table 1: Docking results of 15 top-ranked natural compounds

| No. | Bioactive Compound | Plant source          | Structure | Docking score (kcal/mol) | Solubility | Toxicity       |
|-----|--------------------|-----------------------|-----------|--------------------------|------------|---------------|
| 1   | Gracillin          | *Paris vietnamensis*  |           | -9.2                     | High       | Non-toxic     |
| 2   | Proanthocyanidins  | *Cinnamomum sp.*      |           | -9.2                     | High       | Non-toxic     |
| 3   | Salvianolic B      | *Orthosiphon aristatus* |          | -8.8                     | High       | Non-toxic     |
| 4   | Geraniin           | *Phyllanthus urinaria* |           | -8.6                     | High       | Non-toxic     |
| 5   | Rutin              | *Coscinium fenestratum* |         | -8.5                     | High       | Non-toxic     |
| 6   | Taccasuboside D    | *Tcca integrifolia*   |           | -8.5                     | High       | Non-toxic     |
| 7   | 7-O-galloyltricetiflavan | *Eurycoma longifolia* |           | -8.5                     | High       | Non-toxic     |
| 8   | Polyphyllin C      | *Paris vietnamensis*  |           | -8.4                     | Low        | Non-toxic     |
| 9   | Torvoside J        | *Solanum procumbens*  |           | -8.4                     | High       | Non-toxic     |
pharmacological characteristics, especially antiviral activity. Extensive study has revealed that saponins have a biological effect on HCV replication to become a potent therapeutic agent for HCV.\textsuperscript{[14]} Therefore, the compounds having the polyphenol and steroid core structure might have potential properties of combat coronavirus.

Among these compounds, Gracillin and Proanthocyanidins (steroidal saponin and flavonoid, respectively) showed the highest binding affinity to 3CL protease of SARS-CoV-2 with a docking score of -9.2 Kcal/mol. The interaction analysis revealed that Gracillin and Proanthocyanidins form several significant interactions with 3CL protease. Moreover, both of them have a noticeable interaction with the catalytic dyad (Cys145 and His41), two key residues for the enzymatic activity of 3CL protease.\textsuperscript{[15]} As shown in figure 3, Gracillin forms six hydrogen bonds with Leu141, Asn142, Gly143, Ser144, Cys145, Gly166. Similarly, Proanthocyanidins forms five hydrogen bonds with Tyr54, His163, His164, Glu166, Gln189 and three hydrophobic interactions with His41, Met165, Asp187 (figure 4). These interactions are quite similar to the interactions between Boceprevir and 3CL protease. Boceprevir was also found to interact with 3CLpro of coronaviruses at Ser-144, Cys-145, Gly-143, His-164, Glu-166, Met-165, His-41 and Met-49 with six hydrogen bonds and two hydrophobic interactions (figure 2).

The solubility and toxicity of the 15 top-ranked compounds were also calculated by MarvinSketch and DL-AOT prediction server. The results showed that most of these compounds are non-toxic and have high solubility (table 1). Two remarkable compounds with the highest binding affinity Gracillin and Proanthocyanidins, also showed high solubility and non-toxic characteristics. These results suggested that bioactive compounds identified in this study can be the potential candidates for combating SARS-CoV-2.
infection.

![Figure 3: Gracillin - SARS-CoV-2 3C-like protease interactions](image)

4. CONCLUSION

Molecular docking has been seen as a remarkable role in search of potential medications that contained the bioactive compounds. Although the docking results of the interaction between natural ligands and 3CL protease are not guaranteed to have activity in a clinical test, this process is the first step in drug discovery applied computational tools to reduce time and cost to combat COVID-19. By using the molecular docking method, our study revealed several natural agents extracted from medicinal plants in central Vietnam are capable of inhibiting 3CL protease, an essential factor that has a critical role in SARS-CoV-2 replication. Gracillin and Proanthocyanidins are the top 2 compounds with the highest docking score (-9.2 Kcal/mol), indicating these compounds can be the high potential inhibitors of 3CL protease and the promising candidates for anti-coronavirus drug development. However, further

\textit{in vitro} and \textit{in vivo} studies are needed to develop clinical therapies for the treatment of COVID-19 from these promising bioactive compounds.

\textbf{Conflicts of interest.} There are no conflicts of interest.

\textbf{REFERENCES}

1. Z. J. Cheng, J. Shan. 2019 Novel coronavirus: where we are and what we know, \textit{Infection}, \textbf{2020}, 48(2), 155-163.

2. World Health Organization. COVID-19 Weekly Epidemiological Update 45, \textbf{2021}, No. June.

3. J. Cohen, K. Kupferschmidt. The ‘very, very bad look’ of remdesivir, \textit{Science}, \textbf{2020}, 370(6517), 642-643.

4. J. ling Ren, A. H. Zhang, X. J. Wang. Traditional Chinese medicine for COVID-19 treatment, \textit{Pharmacol. Res.}, \textbf{2020}, 155, 104743.

5. D. E. Gordon, G. M. Jang, M. Bouhaddou, J. Xu, K. Obernier, \textit{et al}. A SARS-CoV-2 protein interaction map reveals targets for drug repurposing, \textit{Nature}, \textbf{2020}, 583, 459-468.

6. R. Hilgenfeld. From SARS to MERS: crystallographic studies on coronavirus proteases enable antiviral drug design, \textit{FEBS J.}, \textbf{2014}, 281(18), 4085-4096.

7. Y. Xu, J. Pei, L. Lai. Deep learning based regression and multiclass models for acute oral toxicity prediction with automatic chemical feature extraction, \textit{J. Chem. Inf. Model.}, \textbf{2017}, 57(11), 2672-2685.

8. D. Ramirez, J. Caballero. Is It Reliable to Take the Molecular Docking Top Scoring Position as the Best Solution without Considering Available Structural Data?, \textit{Molecules}, \textbf{2018}, 23(5), 1038.

9. C. Ma, M. D. Sacco, B. Hurst, J. A. Townsend, Y. Hu, \textit{et al}. Boceprevir, GC-376, and calpain inhibitors II, XII inhibit SARS-CoV-2 viral replication by targeting the viral main protease, \textit{Cell Res.}, \textbf{2020}, 30(8), 678-692.

10. T. N. Kaul, E. Middleton, P. L. Ogra. Antiviral effect of flavonoids on human viruses. \textit{J. Med. Virol.}, \textbf{1985}, 15(1), 71-79.

11. W. Dai, J. Bi, F. Li, S. Wang, X. Huang et al. Antiviral efficacy of flavonoids against enterovirus 71 infection in vitro and in newborn mice, \textit{Viruses} \textbf{2019}, 11(7), 625.

12. S. M. Sadati, N. Gheibi, S. Ranjbar, M. S. Hashemzadeh. Docking study of flavonoid derivatives as potent inhibitors of influenza H1N1 virus neuraminidas, \textit{Biomed. Rep.}, \textbf{2019}, 10(4), 33-38.

13. S. Jo, S. Kim, D. H. Shin, M. S. Kim. Inhibition of SARS-CoV 3CL protease by flavonoids, \textit{J. Enzyme Inhib. Med. Chem.}, \textbf{2020}, 35(1), 145-151.
14. J. Lee, S. Lim, S. M. Kang, S. Min, K. Son et al. Saponin inhibits hepatitis C virus propagation by up-regulating suppressor of cytokine signaling 2, *PLoS One*, **2012**, *7*(6), e39366.

15. C. A. Ramos-Guzmán, J. J. Ruiz-Pernía, I. Tuñón. Unraveling the SARS-CoV-2 Main Protease Mechanism Using Multiscale Methods, *ACS Catal.*, **2020**, *10*(21), 12544-12554.

**Corresponding author:** Tran Thanh Hoa  
VN-UK Institute for Research and Executive Education  
The University of Danang  
158A Le Loi, Hai Chau, Da Nang 55000, Viet Nam  
Email: hoa.tran@vnuk.edu.vn.