Preprint Title	In silico approach: Anthocyanin derivatives as potential inhibitors of the COVID-19 main protease
Authors	Muhammad I. Abdjan, Khusna A. Rakhman, Sri Handayani, Chairil Anjasmara R. Putra and Imam Siswanto
Publication Date	15 Sep. 2020
Article Type	Full Research Paper
Supporting Information File 1	Supporting Information.docx; 460.3 KB
ORCID® iDs	Muhammad I. Abdjan - https://orcid.org/0000-0003-0783-5791

In silico approach: Anthocyanin derivatives as potential inhibitors of the COVID-19 main protease

Muhammad Ikhlas Abdjan[1], Khusna Arif Rakhman[2], Sri Handayani[3], Chairil Anjasmara Robo Putra[4], Imam Siswanto[1]*

[1]Department of Chemistry, Faculty of Science and Technology, Universitas Airlangga, kampus C Mulyorejo, Surabaya 60115, Indonesia

[2]Department of Chemistry Education, Universitas Khairun, Ternate, Indonesia

[3]Department of Chemical Education, Faculty of Mathematics and Natural Sciences, State University of Yogyakarta, Yogyakarta, Indonesia

[4]Post Graduate School, Universitas Airlangga, Jl. Dharmawangsa 30, Surabaya 60286, Indonesia

Email: imamsiswanto@fst.unair.ac.id

* Corresponding author

Abstract

In silico approach has been carried out for the determination of drug candidates from anthocyanin derivative as inhibitors of the COVID-19 main protease. Geometry optimization has performed using the DFT/B3LYP/6-31G(d,p) method as an initial step to prepare candidate ligand. The results of molecular docking showed that candidates C5 and C6 had promising results with a grid score smaller than the ligand reference (X77)

with a flexible conformation type. Studies on absorption, distribution, metabolism, excretion, and toxicity of C5 and C6 candidates were conducted to study the physicochemical properties of drug candidates and to show good predictive results as drugs. Molecular dynamic simulation uses the ffSB14 force field for 200 ns to study the interaction between ligand and receptor, the system stability, solvent accessibility, energy interactions, and hydrogen bonds. The results show good interaction stability on the C5 complex compared to the reference ligand which is characterized by the binding free energy value of C5 was -42.77 ± 0.37 kcal/mol and X77 was -42.37 ± 0.41 kcal/mol.

Keywords

anthocyanin; COVID-19 main protease; molecular docking; molecular dynamic simulation

Introduction

The new generation coronavirus (SARS-Cov-2) which infects the respiratory tract was first detected in December 2019 in Wuhan, China [1–3]. Recorded on 11 March 2020 the World Health Organization (WHO) established the coronavirus outbreak as a COVID-19 pandemic with a total of 118,000 cases from 110 countries [4,5]. The spread of this virus occurs until the world division of various countries has increased very rapidly in the past 9 months. Besides, the latest data in the last 9 months of July 2020 shows that the spread of the COVID-19 virus has increased quite rapidly, especially in countries in the Americas and Europe. Where, WHO has confirmed the grand total of COVID-19 cases in the world on september 09, 2020, showing a total of 27,486,960 cases with total death cases were 894,983 cases [6]. The spread of the virus is increasing and uncontrolled making research

groups in various countries make contributions to prevent and develop drugs and vaccines to reduce the increasing number of COVID-19 cases [7].

The SARS-CoV-2 is a single-stranded positive-RNA virus that belongs to the beta-corona virus group but is different from MERS-CoV and SARS-CoV [8]. Where the coronavirus is covered by a positive RNA strand with the largest RNA genome of 30-32 kb [9]. In addition, recent studies have shown that CoV contains at least six ROFs where ORF1 and ORF2 are responsible for producing two pp1a and pp1ab polypeptides [10]. Then the polypeptide is processed by the main protease (M^{Pro} or $3CL^{Pro}$) which is responsible for the coronavirus replication sequence [11]. Therefore, this enzyme is often used as one of the very promising drug targets in inhibiting coronavirus activity [12,13].

The development of drug candidates from natural ingredients has promising potential as an antiviral, including anthocyanin derivatives [14]. Around 17 Anthocyanin derivatives have been found in nature and only 6 major anthocyanin derivatives are widely distributed, namely pelargonidin, delphinidin, petunidin, cyanidin, peonidin, and malvidin [15]. The placement of different functional groups in the basic structure of anthocyanin will determine the activity and characteristics of anthocyanin compounds [16]. Some studies report that anthocyanin in the flavonoid group has biological activity as antiviral [17,18]. Recent studies have shown that flavonoid derivatives with the same basic structure as anthocyanins have the potential as SARS-CoV antivirals against COVID-19 main protease [19]. Several anthocyanidin derivatives that bind to carbohydrate groups such as rhoifolin and pectolinarin have been known to have SARS-Cov $3CL^{pro}$ inhibitory

activity with IC_{50} values of 27.45 μM and 37.78 μM, respectively [19]. Additionally, anthocyanin-derived compounds that have a hydroxy group in their structure are expected to be able to inhibit the activity of the virus by binding to the target protein at the molecular level. These have been reported by previous studies of adding hydroxy groups to candidates giving effective results against the toxicity and inhibition of SARS-CoV2 [20,21].

This article reports of studies of anthocyanin activity as an anti-SARS-Cov-2 used in silico approach. The study of antiviral activity using the in silico approach is one of the most effective and efficient alternatives in studying and predicting the interaction of drug candidates with protein targets [22–26]. In addition, the in silico approach is able to predict the physicochemical and biological activity of a drug candidate before a wet laboratory test is carried out [27–29]. The combination of several computational techniques used such as molecular docking, and molecular dynamic simulation is expected to be able to provide accurate prediction results through complex calculations [30]. Besides, the quantum mechanics approach using density functional theory (DFT) on small molecules especially anthocyanin derivative molecules shows promising results in molecular modeling [31,32]. Therefore, the selection of an in silico approach is able to offer fast, effective, and accurate alternatives in finding a SARS-CoV2 drug or vaccine.

Materials and Method

Computational Resource and Data Set

The operating systems and hardware used in this study are Windows operating systems (Intel Core i5-9300H, 2.40 GHz, GPU Nvidia GTX 1650, and 8.0 GB RAM) and Linux operating systems (Intel Core i7-8700, 32 GB RAM, NVidia GPUs GTX 1080 Ti 11GB and SSD 500GB SATA). Where the Windows operating system is used for preparation and analysis. Meanwhile, the Linux operating system is used to perform heavy computational chemical calculations. The software used in this study, namely Gaussian 09W, ChemOffice 2016, Chimera 1.13, Open Bubble GUI, Putty, WinSCP, Dock6, AMBER18, and Discovery Studio 2019. The target protein used in this study as a receptor is COVID-19 main protease (PDB: 6W63) obtained from the Protein Data Bank website (http://www.rcsb.org/structure/6W63). Where the target protein is included in the classification of viral protein with a resolution of 2.10 Å using the XRD method. The candidate compounds modeled in this study were 6 anthocyanin derivatives, namely pelargonidin, delphinidin, petunidin, cyanidin, cyanidin-3-rutinoside, and cyanidin 3,5-O-diglucoside.

Modeling and Preparation of Data Set

Modeling anthocyanin derivatives using the DFT/B3LYP/6-31G(d,p) method to determine the geometry optimization using Gaussian 09W (Gaussian 09) [33]. Anthocyanin geometry optimization using density functional theory [34] with Becke 3-parameter Lee-Yang-Parr [35] hybrid functional based on considerations from previous studies that showed good correlation results with experimental results on spectroscopic modeling and chemical properties [36]. The energy calculation from the process of optimizing the geometry of molecular data sets, namely anthocyanin derivatives aims to obtain optimal

geometrical conformation [37]. Meanwhile, receptor preparation and reference ligands are carried out using the ff14SB force field with AM1-BCC ligand charge through the Chimera 1.13 package. Besides, some handlers of the force field and charge aims to prevent charge imbalances in the system when performing molecular dynamic simulations.

Study of Molecular Docking

Molecular docking is done using the dock6 package through the calculation command in parallel using the Linux operating system. The molecular docking stage includes the validation stage and the candidate docking stage into the active side of the receptor. The selection of cluster spheres at the receptor is carried out within a 10 Å radius with a grid spacing of 0.3 in determining the grid box system used. Additionally, the type of conformation used in the docking validation process uses two types of conformation, namely rigid conformation and flexible conformational. The conformation with the smallest grid score will be selected as the reference ligand at the candidate docking stage. Also, the ligand validation stage was stated to comply with the criteria if the reference ligand had RMSD value ≤ 2.0 Å [38]. The candidate docking stage is run with a grid-based score function with a fast calculation which aims to obtain the initial coordinates of each candidate on the active side of the receptor.

Molecular Dynamic Simulation

The initial coordinates of the reference ligands and candidate obtained from the results of molecular docking continued as a basis for building the topology of each complex using tleap. Where the force field used the ff14SB force field and a solvated box used the TIP3P

box with a distance of 12 Å. The stages of MD simulation are carried out in several stages such as minimization, heat, density, equilibration, and production. All stages are carried out for 200 ns which is used as the basis for the analysis of several bonding variables between ligands and receptors. Analysis of several variables measured in this study uses output files and trajectory files generated during the simulation process. Besides, the analysis using the MM-GBSA method is also used to calculate the binding free energy of each complex based on the number of frames produced during the simulation when the system has reached stability [39].

Result and Discussion

Molecular Docking Validation

The molecular docking validation stage is performed using a grid-based cluster sphere selection as a score function (Figure 1A). The validation process is carried out to determine the active site or ligand-binding site of the receptor [40]. The validation process by redocking the reference ligand (code: X77) aims to find the initial coordinates of the ligand. The results show that the reference ligand pose is very promising with an RMSD value ≤ 2.0 Å for each conformation, namely XRD: yellow, rigid: orange, and green: flexible (Figure 1B). Based on the value of RMSD shows that flexible conformation has a smaller RMSD value compared to rigid conformation. Thus, flexible conformation is chosen as the reference ligand in determining the initial coordinates of molecular docking. The active site of the receptor shows that there are amino acid residues responsible for

binding to the reference ligand such as Thr26, Leu27, His41, Met49, Leu141, Asn142, Gly143, Cys145, Met165, Glu166, and Pro168 with different types of interactions (Figure 1C and Figure 1D).

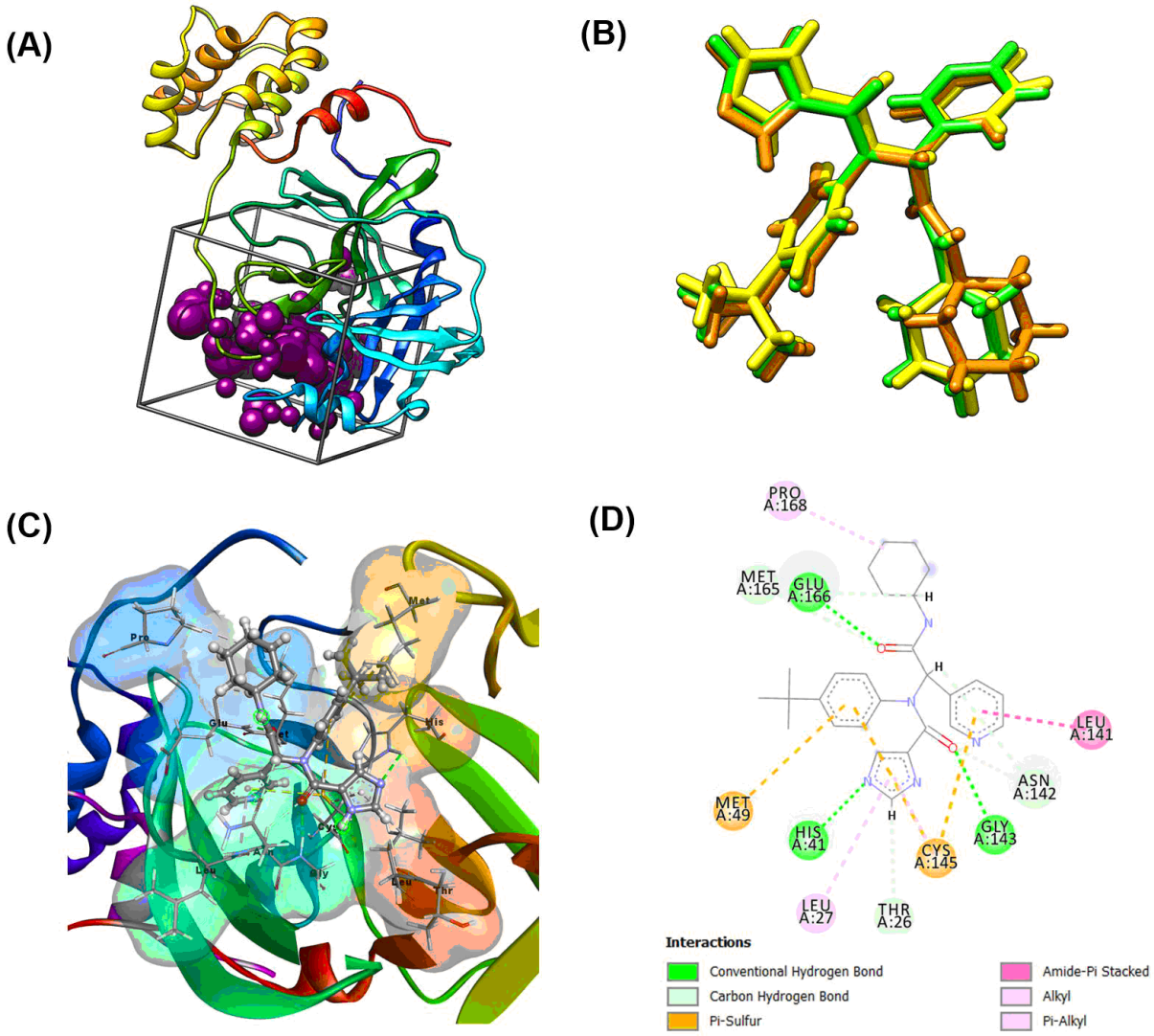

Figure 1: Visualization of molecular docking validation: (A) Cluster sphere selected (PDB code: 6W63), (B) Pose of reference ligand (X77), (C) Fleksibel conformation: 3D interaction between X77 and residues, and (D) Fleksibel conformation: 2D interaction between X77 and residues.

The results of molecular docking validation of the two types of conformation also show grid score values that are not too much different from the difference of ~ 0.90 kcal/mol. Several variables are also generated in molecular docking using the dock6 package, namely Van der Waals energy (E_{VDW}), electrostatic energy (E_{es}), and internal energy repulsive (E_{inter}). However, flexible conformation shows more promising results because it has a smaller grid score compared to rigid conformation (Table 1). Thus, consideration of selection as a reference ligand is getting stronger. This consideration is caused by two key parameters that hold the key to molecular docking success such as a smaller RMSD value and a smaller grid score [41,42].

Table 1: Molecular docking validation of reference ligand (code: X77).

Type	RMSD (Å)	H-bonds (Å)	Grid score (kcal/mol)	E_{VDW} (kcal/mol)	E_{es} (kcal/mol)	E_{Inter} (kcal/mol)
Rigid	0.79	Gly143: 2.14 Glu166: 2.09	-78.00	-74.82	-3.18	24.97
Flexible	0.42	His41: 3.08 Gly143: 2.22 Glu166: 1.84	-78.90	-75.00	-3.90	22.61

Molecular Level Interactions of Anthocyanin Derivatives with COVID-19 Main Protease

The anthocyanin derivatives modeled in this study are secondary metabolites and derivative compounds that give a blackish-purple color to some tropical fruits [43,44]. Variations in the structure of different anthocyanin derivatives are expected to be able to provide different activities as antivirals, especially in inhibiting the expression of the COVID-19 main protease (Figure 2). The hydroxy group contained in each anthocyanin derivative compound is expected to provide good results in binding to amino acid residues on the active site of the receptor as a hydrogen bond donor. Geometry optimization of anthocyanins using the DFT/B3LYP/6-31G(d,p) method was expected to provide optimal geometry results as shown by previous studies [45].

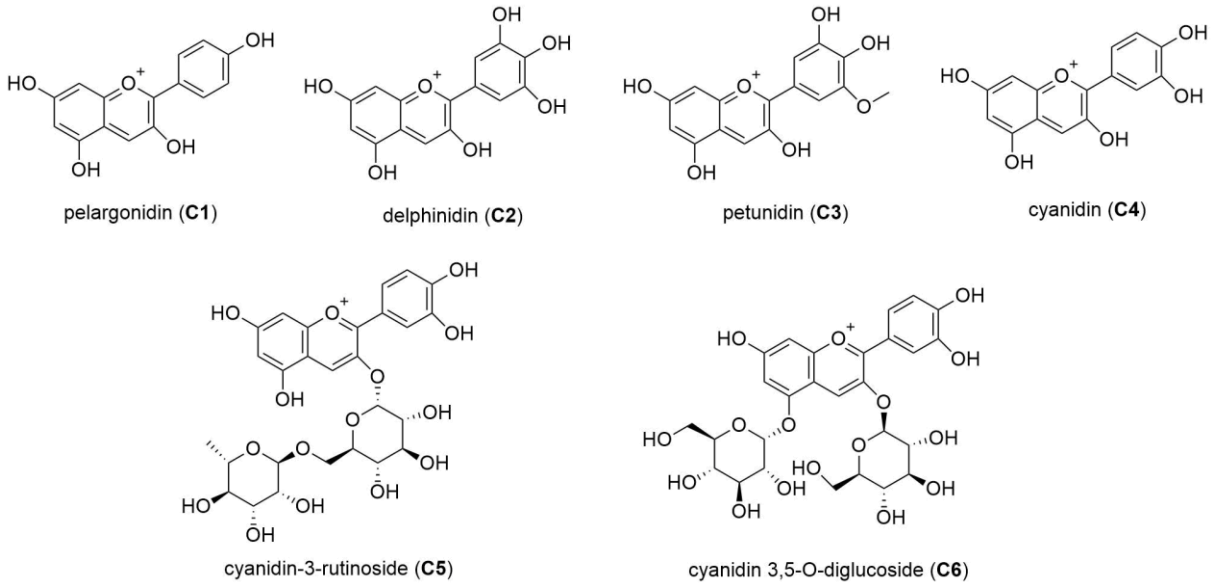

Figure 2: Anthocyanin derivatives as inhibitor COVID-19 main protease (M^{Pro}).

The candidate docking stage is carried out using reference ligands from the flexible conformations through some of the considerations previously mentioned. The results show that anthocyanin derivatives occupy the active site of the receptor very well (Figure

3). Where the candidate's position shows interactions with several amino acid residues that interact with the reference ligand. Additionally, the results of docking using the grid score function showed promising results on two anthocyanin derivatives, specifically at C5 (-83.25 kcal/mol) and C6 (-82.43 kcal/mol) which had a smaller grid score compared to the reference ligand (-78.90 kcal/mol). The other candidates such as C1, C2, C3, and C4 showed poor results with grid score values greater than reference ligands. This is because C5 and C6 candidates have carbohydrate groups at position 3 (C5) and position 3,5 (C6) which have greater potential to bind to residues on the active site of the receptor. Several variables such as Van der Waals's energy and electrostatic energy play an important role in determining the interaction energy between ligands and receptors expressed by grid scores [46]. The results show that the smaller the Van der Waals energy and electrostatic energy values, the smaller the grid score (Table 2).

Table 2: Molecular docking result of best candidates used flexible conformation.

Code	Grid score (kcal/mol)	E_{VDW} (kcal/mol)	E_{es} (kcal/mol)	E_{Inter} (kcal/mol)
C1	-50.92	-43.59	-7.43	2.55
C2	-57.31	-44.69	-12.62	2.34
C3	-57.97	-49.07	-8.89	3.04
C4	-54.44	-42.67	-11.87	2.65
C5	-83.25	-75.24	-8.01	30.17
C6	-82.43	-76.66	-5.77	17.79

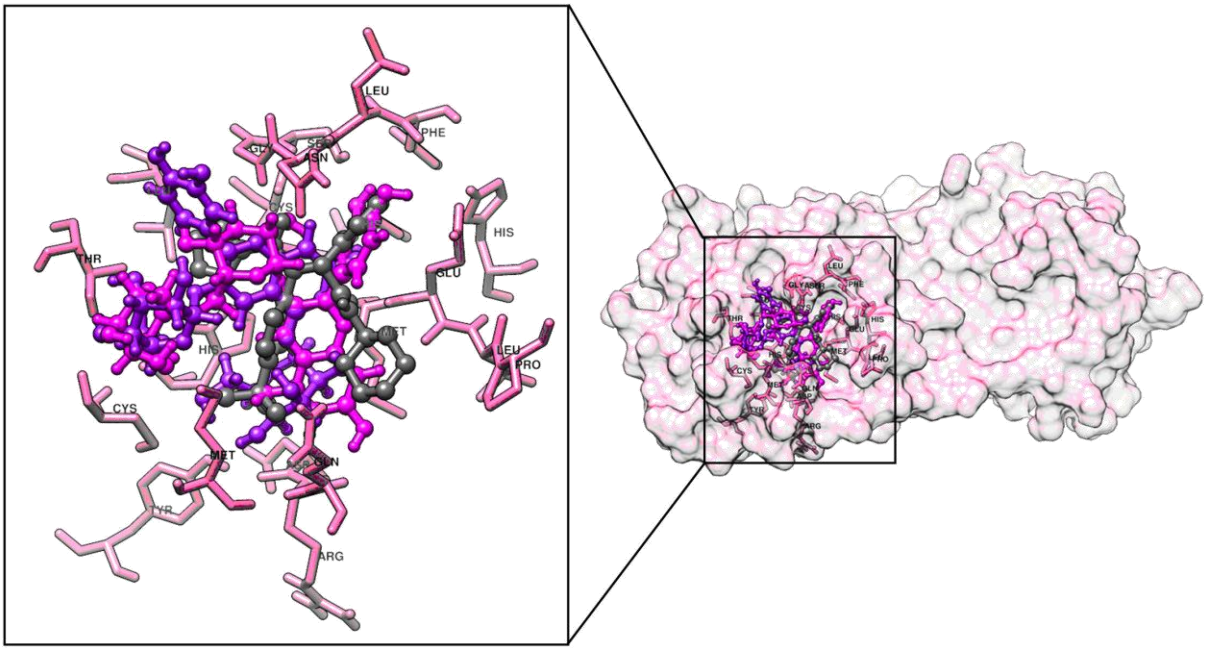

Figure 3: Visualization pocket area of active site: X77 (dim gray), C5 (magenta), and C6 (purple).

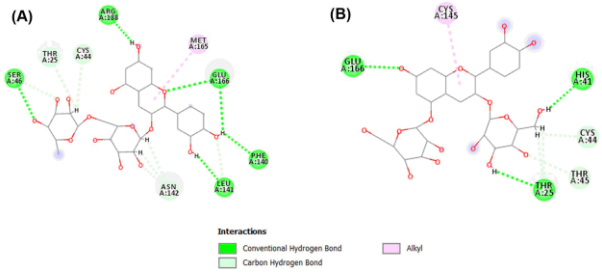

Figure 4: Interaction between candidate and amino acid residues (2D-diagram): (A) C5 and (B) C6.

Anthocyanin derivatives C5 and C6 show promising results at the candidate docking stage to study the interaction of them with amino acids on the receptor active side. The interaction between the two candidates with the amino acid on the receptor was H-bond

interaction and Alkyl interaction (Figure 4). The C5 candidate who interact with 4 amino acid residues are the same as the reference ligand (X77), namely amino acid residues Leu141, Asn142, Met165, and Glu166. Meanwhile, The C6 candidate who interact with 3 amino acid residues are the same as the reference ligand (X77), namely the amino acid residues His41, Cys145, and Glu166. Additionally, the number of hydrogen bonds produced by each complex shows good interaction results. Where, C5 has 6 hydrogen bonds (green line) with the type of bond in the form of a hydrogen bond donor (Phe140, Leu141, Glu166, and Arg168) and hydrogen bond acceptor (Ser46 and Glu166). Meanwhile, C6 has 3 hydrogen bonds (green line) with a type of hydrogen bond donor (Thr25, His41, and Glu166). Consideration regarding the number of hydrogen bonds strongly influences the strength of interactions between ligands and receptors [47]. Thus, based on the docking results of C5 candidate shows good result in the perspective of its binding with the COVID-19 main protease receptor. Additionally, the hydrophobicity property of anthocyanin derivatives C5 and C6 was modeling based on the surface area of the hydrophobicity on the receptor site (Figure 5). The results show that C5 and C6 have many hydroxy groups that are polar so that they decrease their hydrophobicity, especially on the carbohydrate groups of each candidate [48].

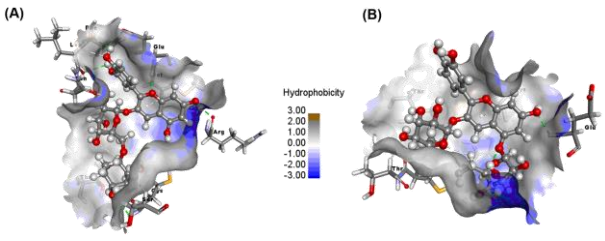

Figure 5: Visualization of hydrophobic surface area: (A) C5 and (B) C6.

Molecular Dynamic Analysis: Stability, Solvent Accessibility, and Energy Interactions

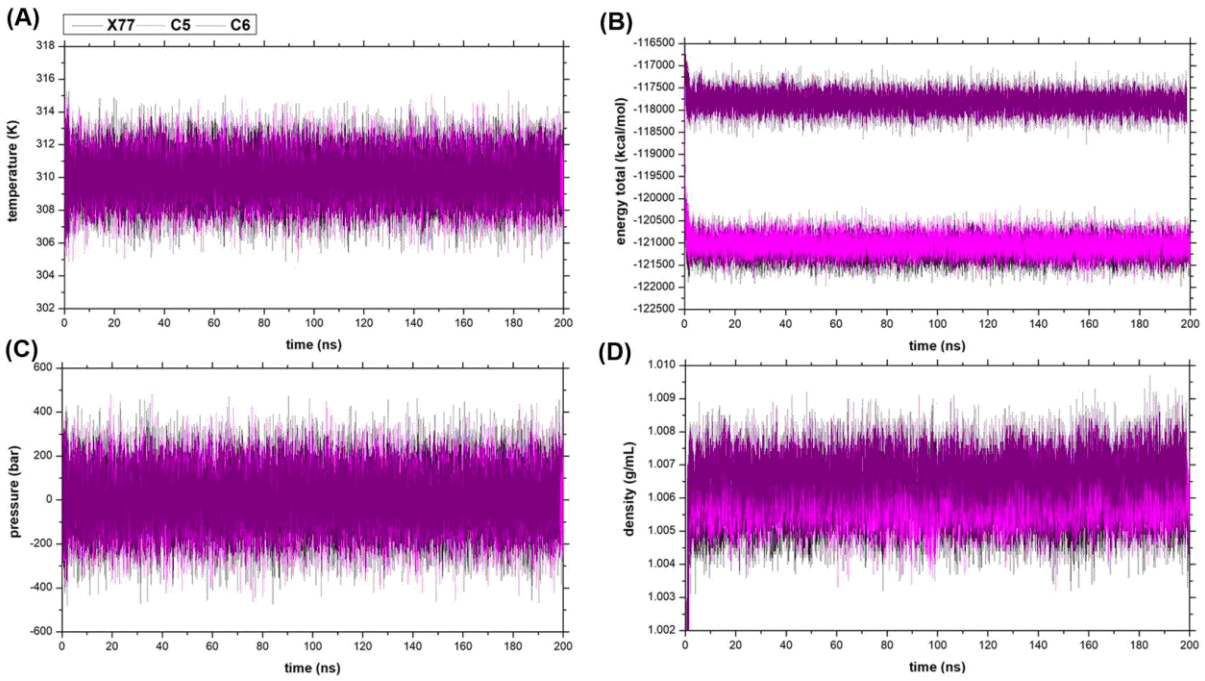

Figure 6: The summary of output files during the simulation 200 ns: (A) Temperature, (B) Energy total, (C) Pressure, (D) Density.

The initial stages of MD simulation are carried out to see the system stability of each complex when given a different treatment. Some of the effects of the simulated parameters are minimization, heat, density, and equilibrium. Where the effect of the parameter result then becomes a strong consideration to see the stability of each system. MD simulation results show that temperature, total energy, pressure, and density show good graphs with no significant fluctuation changes (Figure 6). This indicates that during

the simulation time the stability of temperature, total energy, pressure, and density was achieved in each system. So, the stability at each stage shows good results so that it can be continued on the trajectory analysis of each system.

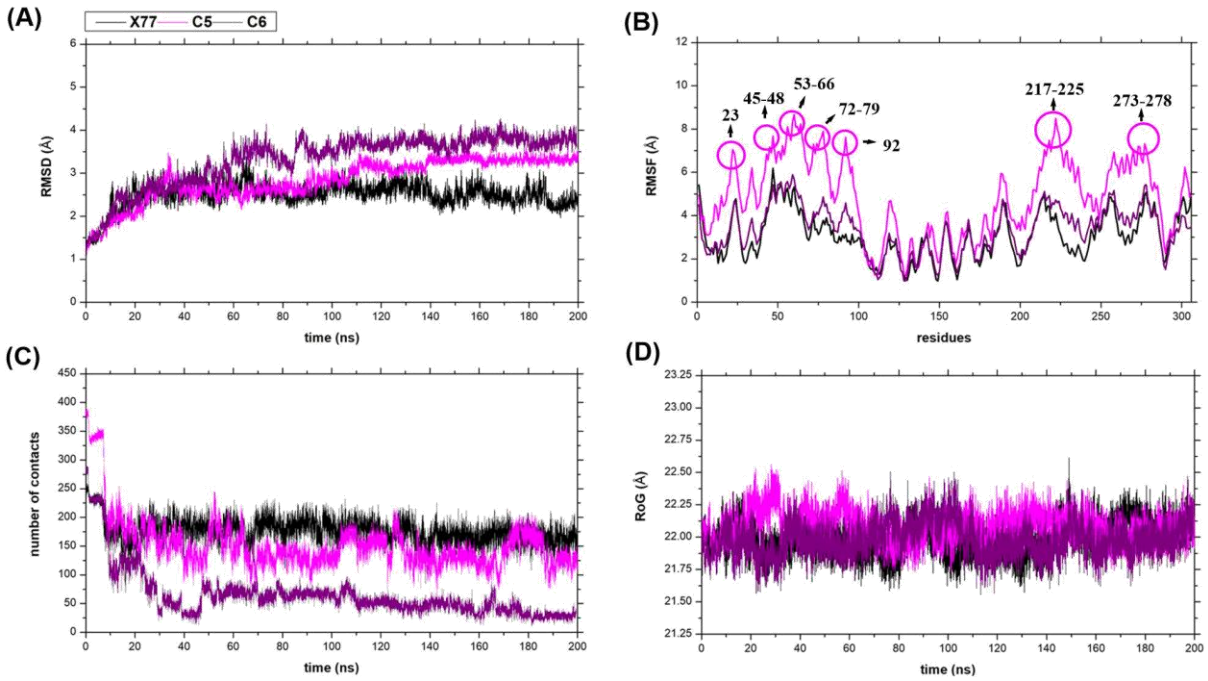

Figure 7: Trajectories analysis of stability system: (A) Root mean square displacement of complex, (B) Root mean square fluctuation of complex, (C) Number of contacts, (D) Radius of gyration.

The simulation stage is carried out for 200 ns on each system to measure several variables including RMSD of complex, RMSF of backbone, SASA, and energy interactions. The results of the trajectory analysis of each complex showed that the RMSD of the complex has good stability (Figure 7A). Where the graph shows the stability of the system in each complex reached during the simulation at trajectories 100 ns until 200 ns.

The stability of the RMSD plot is an important parameter for analyzing other parameters by looking at the stability of the system which is characterized by RMSD deviations that are not too significant [49]. Particularly, the C5 system showed very good stability with no significant deviation. Meanwhile, the C6 system began to reach its stability point at trajectory 100 ns until 200 ns with no significant deviation.

Root mean square fluctuation of complex (RMSF) analysis is also performed on each complex by looking at the fluctuations that occur in the backbone receptor (Figure 7B). The RMSF of each complex was carried out during the last 20 ns when the system stability was achieved. This aims to save calculation time but not sacrifice calculation accuracy. Fluctuations that occur in each complex show the greatest fluctuations in C5 and the lowest fluctuations occur in C6. This identifies that the bonds that occur between receptors and C6 are more flexible than the bonds between C5 and receptors during simulation time when system stability is achieved.

The interaction between the ligands to the receptors in each system showed a decrease in contact during the initial simulation of 1 ns until 6 ns dramatically (Figure 7C). The results show that X77 as a reference ligand has the number of contacts 163 ± 14 over the last 20 ns. Meanwhile, anthocyanin derivatives showed that C5 (135 ± 24) had a greater number of contacts than C6 (28 ± 5) over the last 20 ns. The radius of gyration is performed to analyze the compactness of bonds between ligands and receptors in each system. It is indicated by the fixed value and there is no significant change in value. The results show that each complex formed during the simulation time (over the last 20 ns)

shows a fixed value for X77: 22.08 ± 0.08 Å, C5: 22.05 ± 0.08 Å, and C6: 22.01 ± 0.10 Å. This indicates that there is a stable folding of the protein when it binds to the ligand in each complex (Figure 7D).

Solvent accessibility surface area analysis of COVID-19 M^{pro} active site was conducted to see the role of water molecules in playing the stability of each system during molecular dynamics simulations. This was done to see the interaction between water molecules with each complex during the 200 ns simulation time (Figure S1). The analysis was carried out on amino acids which were located on the active site of the receptor that had interactions with each ligand. The results show the average value of the X77 complex SASA as a reference ligand of 981.05 ± 74.47 Å with a surface area change that is not too significant over the last 20 ns. Meanwhile, C5 and C6 showed an average SASA value of 898.74 ± 90.48 Å and 1133.68 ± 100.59 Å over the last 20 ns.

Table 3: Energy calculation of each complex used MM-GBSA method. Data are shown as mean ± standard error of mean.

Energy component	X77	C5	C6
E_{VDW} (kcal/mol)	-52.17 ± 0.30	-61.35 ± 0.35	-48.17 ± 0.42
E_{Elec} (kcal/mol)	-29.82 ± 0.46	-136.07 ± 1.36	-127.59 ± 1.56
E_{GB} (kcal/mol)	45.65 ± 0.31	162.37 ± 1.26	158.51 ± 1.42
E_{Surf} (kcal/mol)	-6.03 ± 0.02	-7.72 ± 0.02	-6.27 ± 0.04
Δ_{Gas} (kcal/mol)	-81.99 ± 0.57	-197.42 ± 1.32	-175.76 ± 1.54

Δ_{Solv} (kcal/mol)	39.62 ± 0.30	154.65 ± 1.24	152.23 ± 1.41
ΔG (kcal/mol)	-42.37 ± 0.41	-42.77 ± 0.37	-23.52 ± 0.47

Energy interactions analysis was carried out over the last 20 ns trajectory using MM-GBSA (Molecular Mechanics-Generalized Born Surface Area) method where the energy calculations carried out included Van der Waals energy (E_{VDW}), electrostatic energy (E_{Elec}), Generalized Born energy (E_{GB}), energy of contribution nonpolar of solvent (E_{Surf}), free energy gas (Δ_{Gas}), free energy solvent (Δ_{Solv}), and binding free energy (ΔG) [50]. Data shows that each energy component shows a close relationship in its contribution to each complex in the system during the simulation. Where the role of solvents in the system is indeed a very crucial role. This is due to the fact that Generalized Born energy and energy of contribution of nonpolar of solvent contribute to free energy solvent [51]. Meanwhile, Van der Waals energy and electrostatic energy contribute to the value of free energy gas. Therefore, the contribution of each energy component will affect the final result of total energy or free binding energy which is an important parameter in determining the energy of interaction between ligands and receptors. The results show that the free binding energy of C5 is more promising compared to C6. It can be seen that the binding free energy value of C5: -42.77 ± 0.37 kcal/mol is smaller than C6: -23.52 ± 0.47 kcal/mol (Table 3). Additionally, C5 can be considered as the COVID-19 main protease inhibitor because it has a binding free energy value that is smaller than the value of the binding free energy reference ligand: -42.37 ± 0.41 kcal/mol with a deviation: ~ 0.40 kcal/mol.

Energy calculation using the MM-GBSA method can see the energy distribution of each residue through an analysis of the energy decomposition of each complex (Figure 9). The results show that amino acid residues that have values below -1.00 kcal/mol that interact with ligands are residues that are present at the active receptor site [52]. Specifically, C5 has more amino acid residue interactions (12 amino acid residues) with energy values below -1.00 kcal/mol. The energy decomposition results show good suitability with the results of free binding energy. Therefore, the binding free energy value of C5 is smaller than C6 and X77.

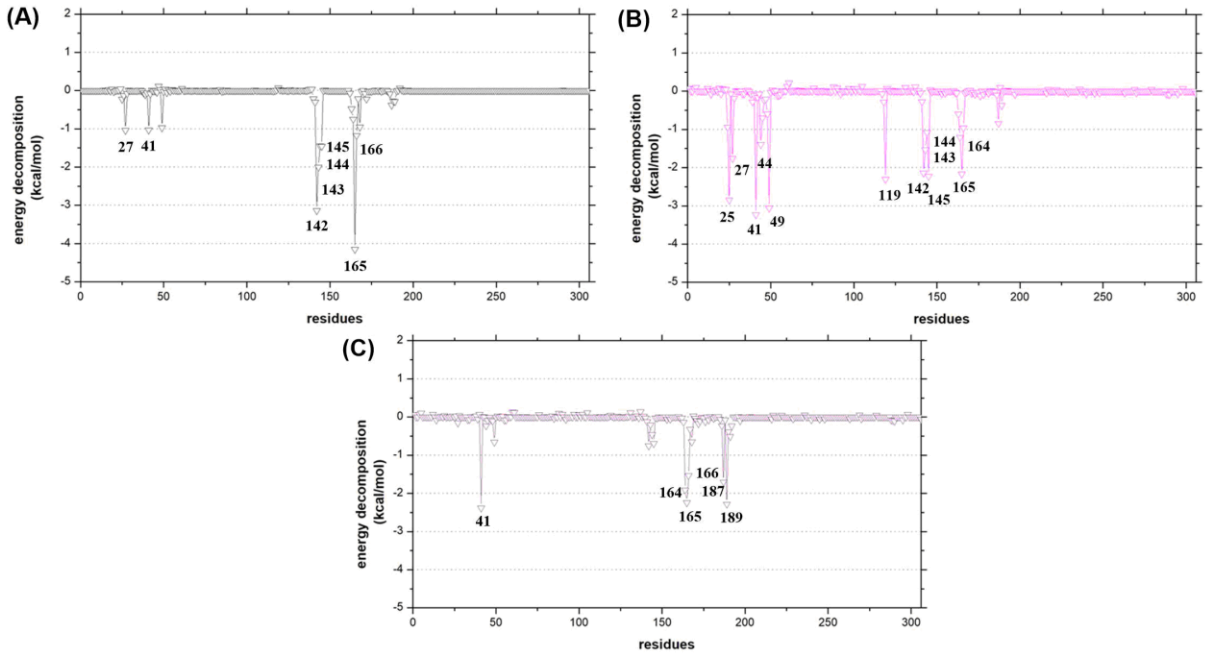

Figure 9: Energy decomposition analysis over the last 20 ns of each complex: (A) X77, (B) C5, and (C) C6.

Hydrogen Bonds Analysis

Hydrogen bonds (H-bonds) plays an important role in the interaction between ligands and receptors at the molecular level [53]. Where one important parameter is the distance of the H-bonds interaction that determines the strength of the bond in the interaction of each complex. The H-bonds analysis is performed using molecular docking and MD simulations to see the properties of the bonds. The combination of the two computational techniques is done to reduce the cost of computational calculations.

Table 4: Comparison of H-bonds result between molecular dynamic simulation and molecular docking.

Code	Frames	Fraction	Ligand-amino acid residues		Distance (Å)	
			Acceptor	Donor	MDS	Docking
X77	16775	0.8065	X77307: O13	Glu166: H	2.89	1.84
	14810	0.7120	X77307: O01	Gly143: H	2.93	2.22
	ND	ND	X77307: N32	His41: H	ND	3.08
C5	3162	0.1520	Glu166: OE1	Unk307:H71	2.62	2.43
	1376	0.0662	Arg188: O	Unk307:H66	2.78	1.93
	835	0.0401	Leu141: O	Unk307:H72	2.89	2.91
	566	0.0272	Unk307:O32	Glu166: H	2.99	2.48
	460	0.0221	Phe140: O	Unk307:H72	2.86	2.78
	83	0.0040	Unk307: O19	Ser46: HG	2.89	2.62
C6	6573	0.3170	Glu166: OE1	Unk307:H69	2.64	2.48
	527	0.0254	Thr25:OG1	Unk307:H67	2.89	2.42
	6	0.0003	His41: ND1	Unk307:H65	3.04	2.28

Molecular docking aims to find the initial coordinates of the H-bonds and amino acid residue responsible for that bond. Furthermore, to study the properties of H-bonds that are formed further such as the bond quality and quantity during a certain time, MD simulation is performed. The results show that in each complex several parameters are measured by making comparisons of the two computational techniques used (Table 4).

The study of the number of H-bonds, H-bonds interactions, and H-bond lifetime is presented clearly in this article through molecular dynamics simulations (Figure 10). Further studies on H-bond interactions using molecular dynamics aim to look at the contribution of amino acid residues that have interactions with ligands based on the initial coordinates obtained from molecular docking. Complex X77 shows the results of docking there are three measured H-Bond. Meanwhile, during the simulation process, 200 ns only showed two bonds with the amino acid residues responsible for Glu166 and Gly143. This is caused by molecular docking taking only the best poses of H-Bond interactions that occur between ligands and receptors. When H-bonds His41 amino acid residue is treated with variations in several variables such as temperature, pressure, and density during the simulation time the bond is not detected (ND). In addition, residues have a long period of occupancy with a fraction of 80.65% (frame: 16775) and 71.20% (frame: 14810) which shows a very good level of bonding. The same thing is also shown by the complex candidates C5 and C6 which show the comparison of distances between molecular docking and MD simulation. Meanwhile, C5 and C6 all amino acid residues detected by molecular docking were also detected during the simulation time. Additionally, only one

amino acid residue had an H-bond fraction presentation> 10%, namely the Glu166 residue in candidates C5 and C6. This is supported by the H-bonds lifetime time that is so fast with a decrease during the simulation on each complex dramatically. This is because the residues analyzed are specific residues. Where the residue analyzed is the residues obtained from the initial coordinates when performing molecular docking. However, it is hoped that the results of the analysis provide a clearer picture of hydrogen bond properties.

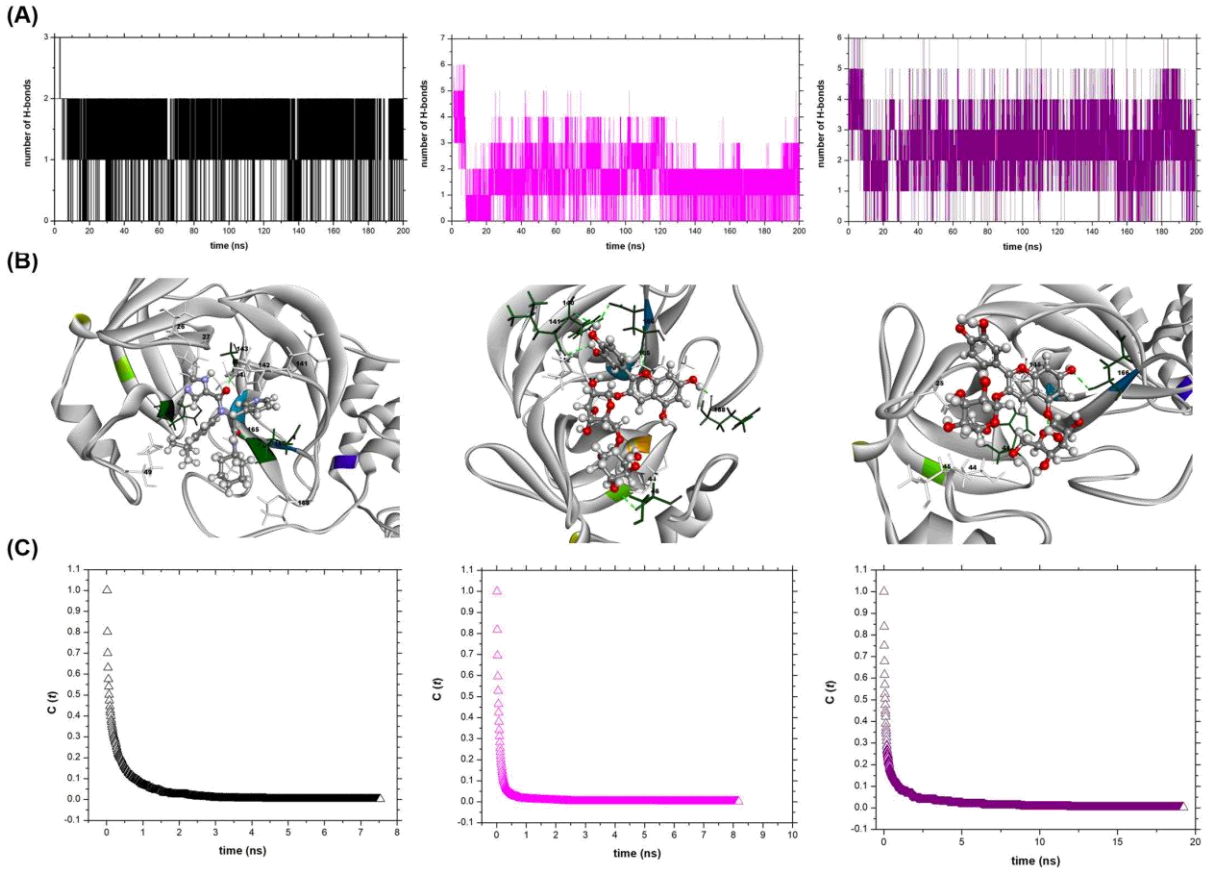

Figure 10: Hydrogen bonds analysis: (A) Number of H-bonds (B) 3D Visualization of H-bonds interaction between ligand-receptor, and (C) Lifetime of H-bonds.

Study of Bioavailability and Drug-Likeness Screening

ADMET prediction (absorption, distribution, metabolism, excretion, and toxicity) used the admetSAR service website (http://lmmd.ecust.edu.cn/admetsar1/predict/). Prediction of several important variables of the candidate's biological activity as a drug in the body needs to be done as preliminary data on bioavailability and drug-likeness [54]. The absorption prediction results show that C5 and C6 do not penetrate the blood-brain barrier which, identifies that the candidate will not affect the central nervous system especially the brain [55]. Besides, variables such as human intestinal absorption show good results can be absorbed well. One important indication in absorption is the solubility of each candidate which shows very good results, namely LogS C5: -2.63 and LogS C6: -2.10. The hydroxy group in each candidate will be increasing the nature of its solubility. The metabolism stage shows that both candidates are non-inhibitors and non-substrates of cytochrome isoenzymes (CYP). This identified that both candidates were very promising as drugs because they did not inhibit or interfere with the activity of the enzyme which is the main enzyme in the process of metabolism [56]. As a result, C5 and C6 are expected to have no side effects on the body.

Variable toxicity calculated in this study showed good results for each candidate. Where, it can be said that C5 and C6 are not toxic because the results show Weak inhibitor of the human-related gene, non-AMES toxic and non-carcinogens. Thus, over all the results show that each candidate has ADMET properties that are promising to be considered as drug candidates that can be considered (Table S1).

Conclusions

The combination of computational chemistry techniques between molecular docking and molecular dynamic simulation is very promising in predicting drug candidates in silico by considering the efficiency and effectiveness of calculations without compromising prediction accuracy. An understanding of the COVID-19 main protease pathway needs to be well understood to determine the pathway of inhibition and interaction of drug candidates with the target protein. Anthocyanin derivatives have the potential to be considered as drug candidates. The results of molecular docking indicate that the validation stage is feasible to be used as an initial preparation in finding the active site of the target protein. The results show that flexible conformation produces better criteria than rigid conformation with RMSD value was 0.42 Å and the grid score was -78.90 kcal/mol. Meanwhile, C5 and C6 showed promising docking results with a grid score smaller than the reference ligand of -83.25 kcal/mol and -82.43 kcal/mol respectively. This stage is a fundamental consideration to be studied more deeply about the bioavailability and drug-likeness properties of C5 and C6 candidates. Over all the results show that each candidate has ADMET properties that promise to be used as a drug candidate that can be considered. The MD simulation studies were also continued to study ligand-receptor interactions and free binding energy during the 200 ns simulation time using the ffSB14 force field. The results show C5 candidate has the binding free energy value lower than the reference ligand with an energy deviation of ~ 0.40 kcal/mol. Overall, the C5 candidate showed good predictive results using the in silico approach and deserves to be considered as a COVID-19 main protease inhibitor to be continued experimentally in the future.

References

(1) Ou, X.; Liu, Y.; Lei, X.; Li, P.; Mi, D.; Ren, L.; Guo, L.; Guo, R.; Chen, T.; Hu, J.; Xiang, Z.; Mu, Z.; Chen, X.; Chen, J.; Hu, K.; Jin, Q.; Wang, J.; Qian, Z. *Nat. Commun.* **2020**, *11*, 1620. doi:10.1038/s41467-020-15562-9.

(2) Li, X.; Geng, M.; Peng, Y.; Meng, L.; Lu, S. *J. Pharm. Anal.* **2020**, *19*, 1–7. doi:10.1016/j.jpha.2020.03.001.

(3) Xu, X.; Chen, P.; Wang, J.; Feng, J.; Zhou, H.; Li, X.; Zhong, W.; Hao, P. *Sci. China Life Sci.* **2020**, *63*, 457–460. doi:10.1007/s11427-020-1637-5.

(4) Hussein, J. *Sex. Reprod. Heal. Matters* **2020**, *28*, 1-4. doi:10.1080/26410397.2020.1746065.

(5) Andersen, K. G.; Rambaut, A.; Lipkin, W. I.; Holmes, E. C.; Garry, R. F. *Nat. Med.* **2020**, *26*, 450–452. doi:10.1038/s41591-020-0820-9.

(6) Coronavirus disease (COVID-19) 9 september 2020. https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports (accessed september 13, 2020).

(7) Zhou, Y.; Hou, Y.; Shen, J.; Huang, Y.; Martin, W.; Cheng, F. *Cell Discov.* **2020**, *6*, 1-18. doi:10.1038/s41421-020-0153-3

(8) Yi, Y.; Lagniton, P. N. P.; Ye, S.; Li, E.; Xu, R. *Int. J. Biol. Sci.* **2020**, *16*, 2753-1766, doi:10.7150/ijbs.45134

(9) Gennaro, F. Di; Pizzol, D.; Marotta, C.; Antunes, M.; Racalbuto, V.; Veronese, N.; Smith, L. *International Journal of Environmental Research and Public Health*, **2020***, 17,* 1-11. https://doi.org/10.3390/ijerph17082690

(10) Tan, Y.; Lim, S. G.; Hong, W. *Antiviral Research*, **2005**, *65*, 69-78. doi:10.1016/j.antiviral.2004.10.001

(11) Zhang, L.; Lin, D.; Sun, X.; Curth, U.; Drosten, C.; Sauerhering, L.; Becker, S.; Rox, K.; Hilgenfeld, R. *Science.* **2020**, *368*, 409-412. doi:10.1126/science.abb3405

(12) Mengist, H. M. *Signal Transduct. Target. Ther.* **2020**, *5*, 1-2. doi:10.1038/s41392-020-0178-y

(13) Xue, X.; Yu, H.; Yang, H.; Xue, F.; Wu, Z.; Shen, W.; Li, J.; Zhou, Z.; Ding, Y.; Zhao, Q.; Zhang, X. C.; Liao, M.; Bartlam, M.; Rao, Z. *J. Virol.* **2008**, *82*, 2515–2527. doi:10.1128/jvi.02114-07

(14) Pour, P. M.; Fakhri, S.; Asgary, S.; Farzaei, M. H.; Echeverría, J. *Front. Pharmacol.* **2019**, *10*, 1–23. doi:10.3389/fphar.2019.01207

(15) Prior, R. L.; Wu, X. *Free Radic. Res.* **2006**, *40* (10), 1014–1028. doi:10.1080/10715760600758522

(16) Rakhman, K. A.; Khadijah; Abdjan, M.I.; Kumendong, N.; Puspitasari, S. D.; *JOTCSA*, **2019**, *5*, 1287–1294. doi:10.18596/jotcsa.452558.

(17) Panche, A. N.; Diwan, A. D.; Chandra, S. R. *J. Nutr. Sci.* **2016**, *5,* 1-15. doi:10.1017/jns.2016.41

(18) Nikolaeva-Glomb, L.; Mukova, L.; Nikolova, N.; Badjakov, I.; Dincheva, I.; Kondakova, V.; Doumanova, L.; Galabov, A. S. *Nat. Prod. Commun.* **2014**, *9*, 51–54. doi:10.1177/1934578x1400900116

(19) Jo, S.; Kim, S.; Shin, D. H.; Kim, M. S. *J. Enzyme Inhib. Med. Chem.* **2020**, *35*, 145–151. doi:10.1080/14756366.2019.1690480

(20) Liu, J.; Cao, R.; Xu, M.; Wang, X.; Zhang, H.; Hu, H.; Li, Y.; Hu, Z.; Zhong, W.; Wang, M. *Cell Discov.* **2020**, *6*, 6–9. doi:10.1038/s41421-020-0156-0

(21) Devaux, C. A.; Rolain, J.-M.; Colson, P.; Raoult, D. *Int. J. Antimicrob. Agents* **2020**, *5*, 1-6. doi:10.1016/j.ijantimicag.2020.105938

(22) Leelananda, S. P.; Lindert, S. *Beilstein J. Org. Chem.* **2016**, *12*, 2694–2718. doi:10.3762/bjoc.12.267

(23) Yi, F.; Li, L.; Xu, L. jia; Meng, H.; Dong, Y. mao; Liu, H. bo; Xiao, P. gen. *Chinese Med. (United Kingdom)* **2018**, *13*, 1–20. doi:10.1186/s13020-018-0190-0

(24) Jain, N.; Gupta, S.; Sapre, N.; Sapre, N. S. *RSC Adv.* **2015**, *5*, 14814–14827. doi:10.1039/c4ra15478a

(25) Gao, Y.; Chen, Y.; Tian, Y.; Zhao, Y.; Wu, F.; Luo, X.; Ju, X.; Liu, G. *New J. Chem.* **2019**, *43*, 17004–17017. doi:10.1039/c9nj03353j

(26) Lin, X.; Li, X.; Lin, X. *Molecules* **2020**, *25*, 1–17. doi:10.3390/molecules25061375

(27) Wang, Y.; Xing, J.; Xu, Y.; Zhou, N.; Peng, J.; Xiong, Z.; Liu, X.; Luo, X.; Luo, C.; Chen, K.; Zheng, M.; Jiang, H. *Q. Rev. Biophys.* **2015**, *48*, 488–515. doi:10.1017/S0033583515000190

(28) Zang, Q.; Mansouri, K.; Williams, A. J.; Judson, R. S.; Allen, D. G.; Casey, W. M.; Kleinstreuer, N. C. *J. Chem. Inf. Model.* **2017**, *57*, 36–49. doi:10.1021/acs.jcim.6b00625

(29) Raies, A. B.; Bajic, V. B. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2016**, *6*, 147–172. doi:10.1002/wcms.1240

(30) Nutho, B.; Khuntawee, W.; Rungnim, C.; Pongsawasdi, P.; Wolschann, P.; Karpfen, A.; Kungwan, N.; Rungrotmongkol, T. *Beilstein J. Org. Chem.* **2014**, *10*,

2789–2799. doi:10.3762/bjoc.10.296

(31) He, J.; Li, X.; Silva, G. T. M.; Quina, F. H.; Aquino, A. J. A. *J. Braz. Chem. Soc.* **2019**, *30*, 492–498. doi:10.21577/0103-5053.20180233

(32) Márquez-Rodríguez, A. S.; Grajeda-Iglesias, C.; Sánchez-Bojorge, N. A.; Figueroa-Espinoza, M. C.; Rodríguez-Valdez, L. M.; Fuentes-Montero, M. E.; Salas, E. *Molecules* **2018**, *23*, 1–16. doi:10.3390/molecules23071587

(33) *Gaussian 09*, revision A.01; Gaussian, Inc.: Wallingford, CT, 2009.

(34) Becke, A. D. *J. Chem. Phys.* **1997**, *107* (20), 8554–8560. doi:10.1063/1.475007

(35) Lee, C. T.; Yang, W.; Parr, R. G. *Phys. Rev. B.* **1988**, *37*, 785-789

(36) Prima, E. C.; Nuruddin, A.; Yuliarto, B.; Kawamura, G.; Matsuda, A. *New J. Chem.* **2018**, *42*, 11616–11628. doi:10.1039/c8nj01202d

(37) Rakhman, K. A.; Saraha, A. R.; Umar, S.; Zainuddin, R.; Abdjan, M. I. *Asian J. Chem.* **2020**, *32*, 941–944. doi:10.14233/ajchem.2020.22231

(38) Hoffer, L.; Chira, C.; Marcou, G.; Varnek, A.; Horvath, D. *Molecules* **2015**, *20,* 8997-9028. doi:10.3390/molecules20058997

(39) Wang, E.; Sun, H.; Wang, J.; Wang, Z.; Liu, H.; Zhang, J. Z. H.; Hou, T. *Chem. Rev.* **2019**, *119*, 9478–9508. doi:10.1021/acs.chemrev.9b00055

(40) Ferreira, L. G.; Dos Santos, R. N.; Oliva, G.; Andricopulo, A. D. *Molecules* **2015**, *20,* 13384-13421. doi:10.3390/molecules200713384

(41) Guedes, I. A.; de Magalhães, C. S.; Dardenne, L. E. *Biophys. Rev.* **2014**, *6*, 75–87. doi:10.1007/s12551-013-0130-2

(42) Rao, S. N.; Head, M. S.; Kulkarni, A.; LaLonde, J. M. *J. Chem. Inf. Model.* **2007**, *47*, 2159–2171. doi:10.1021/ci6004299

(43) Yang, L.; Wen, K. S.; Ruan, X.; Zhao, Y. X.; Wei, F.; Wang, Q. *Molecules* **2018**, *23*, 1–26. doi:10.3390/molecules23040762

(44) Khoo, H. E.; Azlan, A.; Tang, S. T.; Lim, S. M. *Food Nutr. Res.* **2017**, *61*. doi:10.1080/16546628.2017.1361779

(45) Gil, M.; Avila-Salas, F.; Santos, L. S.; Iturmendi, N.; Moine, V.; Cheynier, V.; Saucier, C. *J. Agric. Food Chem.* **2017**, *65*, 10591–10597. doi:10.1021/acs.jafc.7b04461

(46) Kitchen, D. B.; Decornez, H.; Furr, J. R.; Bajorath, J. *Nat. Rev. Drug Discov.* **2004**, *3*, 935–949. doi:10.1038/nrd1549

(47) Chen, D.; Oezguen, N.; Urvil, P.; Ferguson, C.; Dann, S. M.; Savidge, T. C. *Sci. Adv.* **2016**, *2*. doi:10.1126/sciadv.1501240

(48) Henkel, S.; Misuraca, M. C.; Troselj, P.; Davidson, J.; Hunter, C. A. *Chem. Sci.* **2017**, *9*, 88–99. doi:10.1039/c7sc04890d

(49) Mohamed, T. K.; Batran, R. Z.; Elseginy, S. A.; Ali, M. M.; Mahmoud, A. E. *Bioorg. Chem.* **2019**, *85*, 253–273. doi:10.1016/j.bioorg.2018.12.040

(50) Mahalapbutr, P.; Thitinanthavet, K.; Kedkham, T.; Nguyen, H.; Theu, L. thi ha; Dokmaisrijan, S.; Huynh, L.; Kungwan, N.; Rungrotmongkol, T. *J. Mol. Struct.* **2019**, *1180*, 480–490. doi:10.1016/j.molstruc.2018.12.025

(51) Onufriev, A. V.; Case, D. A. *Annu. Rev. Biophys.* **2019**, *48*, 275–296. doi:10.1146/annurev-biophys-052118-115325

(52) Nutho, B.; Mahalapbutr, P.; Hengphasatporn, K.; Pattaranggoon, N. C.; Simanon, N.; Shigeta, Y.; Hannongbua, S.; Rungrotmongkol, T. *Biochemistry* **2020**, *59*, 1769–1779. doi:10.1021/acs.biochem.0c00160

(53) Nittinger, E.; Inhester, T.; Bietz, S.; Meyder, A.; Schomburg, K. T.; Lange, G.; Klein, R.; Rarey, M. *J. Med. Chem.* **2017**, *60*, 4245–4257. doi:10.1021/acs.jmedchem.7b00101

(54) Shen, J.; Cheng, F.; Xu, Y.; Li, W.; Tang, Y. *J. Chem. Inf. Model.* **2010**, *50*, 1034–1041. doi:10.1021/ci100104j

(55) Nau, R.; Sörgel, F.; Eiffert, H. *Clin. Microbiol. Rev.* **2010**, *23*, 858–883. doi:10.1128/CMR.00007-10

(56) Tyzack, J. D.; Kirchmair, J. *Chem. Biol. Drug Des.* **2019**, *93* (4), 377–386. doi:10.1111/cbdd.13445