



## MALAYSIAN JOURNAL OF BIOCHEMISTRY & MOLECULAR BIOLOGY

The Official Publication of The Malaysian Society For Biochemistry & Molecular Biology (MSBMB)

<http://mjbmb.org>

### COMPUTATIONAL DRUG DISCOVERY TARGETING SARS-CoV-2 MAIN PROTEASE TOWARDS A TREATMENT FOR COVID-19

Panida Boontawee<sup>a,c</sup>, Kulisara Kittivibul<sup>a,b</sup>, Punyaporn Pongsuwan<sup>a,b</sup>, Suriya Tateing<sup>a,b</sup>, Nuttee Suree<sup>\*a,c</sup>

<sup>a</sup>*Division of Biochemistry and Biochemical Technology, Department of Chemistry, Faculty of Science, Chiang Mai University, Chiang Mai 50200, Thailand*

<sup>b</sup>*Graduate Program in Biotechnology, The Graduate School, Chiang Mai University, Chiang Mai 50200, Thailand*

<sup>c</sup>*Department of Chemistry, Faculty of Science, Chiang Mai University, Chiang Mai 50200, Thailand*

\*Corresponding Author: [nuttee.suree@cmu.ac.th](mailto:nuttee.suree@cmu.ac.th)

#### History

Received: 12 April 2022

Accepted: 4 November 2022

#### Keywords:

*SARS-CoV-2 main protease, Natural compounds, Drug-likeness, MD simulations, MMPBSA calculations*

#### Abstract

The serious pandemic of coronavirus disease 2019 (COVID-19), which started in Wuhan City, China, in late December 2019, has undoubtedly been a global health emergency that severely affects the world population. The disease is caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). One of the crucial proteins of SARS-CoV-2 is the main protease (M<sup>pro</sup>), a cysteine protease that plays an important role in viral replication. Therefore, M<sup>pro</sup> has become an attractive drug target for chemotherapeutic intervention. In this study, an *in silico* screening of natural compounds from the ZINC database was performed in order to identify inhibitors targeting the active site of SARS-CoV-2 M<sup>pro</sup>. Using combined computational methods including docking-based virtual screening, drug-likeness evaluation, molecular dynamics (MD) simulations, and MMPBSA calculations, the screening platform could identify three promising compounds. These include ZINC253412009, ZINC65297929, and ZINC65298044 which exhibited satisfactorily low free binding energy levels of  $-28.56 \pm 15.71$ ,  $-28.55 \pm 11.78$ , and  $-28.20 \pm 12.66$  kcal/mol, respectively. These compounds show significant interactions with key residues lining the M<sup>pro</sup> active site, warranting their high potential to be developed further. Future *in vitro* confirmation and viral challenge experiments are also needed to obtain a more detailed pharmacological profile of the candidate compounds.

#### INTRODUCTION

The pandemic of coronavirus disease 2019 (COVID-19) has been a severe health problem greatly impacting the world population. Since its first outbreak in Wuhan City of China in late December 2019, COVID-19 has caused a global health emergency, prompting urgent developments of vaccines and therapeutic treatments [1, 2]. As of September 2021, the total cumulative infection cases are more than 230 million worldwide, with the reported deaths exceeding 4.7 million [3], rendering COVID-19 one of the most devastating pandemics in the history of mankind. Symptoms observed in infected patients can vary among different age groups which could include fever, dry cough, fatigue, loss of

taste or smell, nasal congestion, headache, and difficulty breathing or shortness of breath [4]. For the treatment of COVID-19 to date, the US FDA has currently approved only one therapeutic drug Veklury (Remdesivir) as antiviral treatment for use in adults and pediatric patients requiring hospitalization [5].

COVID-19 is caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), which is classified in the order *Nidovirales*, family *Coronaviridae*, subfamily *Coronavirinae*, and genus *Betacoronavirus*. The virus can cause respiratory, digestive, and nervous system diseases in humans and several other animals [6]. The coronavirus genetic material consists of a single-stranded and positive-sense genomic RNA ranging from 26 to 32 kilobases in

length [7]. Similarly to other coronaviruses, SARS-CoV-2 viral particle comprises four structural proteins, including membrane protein (M), nucleocapsid (N), spike protein (S), and envelop protein (E) [8, 9]. For the drug development targeting SARS-CoV-2, however, the main protease ( $M^{pro}$ ) has been a crucial drug target as it plays a critical role in the viral replication.  $M^{pro}$  is a cysteine protease involved in the cleavage and maturation of polyproteins translated from the viral RNA [1, 10]. This key protease enzyme consists of three domains, including domain I (residues 8–101), domain II (residues 102–184), and domain III (residues 201–303). The active site of SARS-CoV-2  $M^{pro}$  is located within a cleft formed between the domains I and II, enclosing its catalytic dyad residues His41 and Cys145 [11]. For catalysis, Cys145 acts as a nucleophile during the first step of hydrolysis reaction, assisted by His41 as a base catalyst [12]. Small molecules targeting these highly conserved residues can thus inhibit the proteolytic reaction mediated by  $M^{pro}$ .

Previous drug discovery studies have screened several antiviral agent groups, with different levels of success, for their inhibitory potency against SARS-CoV-2  $M^{pro}$  catalytic activity. These include RNA-dependent RNA polymerases (RdRps) inhibitors, HIV inhibitors, anti-retroviral agents, and anti-influenza drugs [13–15]. However, as many mutated variants of the virus emerges, more therapeutic options for SARS-CoV-2 are urgently needed to combat the ever-increasing infection rate. Therefore, we aimed to employ computational methods to help accelerate the drug discovery process. In this study, we virtually screened a large pool of natural product compounds for their potential inhibitory activity against SARS-CoV-2  $M^{pro}$ . Drug-likeness following the Lipinski's rules was also used as a selection criterion for the screening. The selected compounds were subsequently subjected to additional analyses of their binding mechanisms via molecular dynamics simulations and molecular mechanics calculations. The final three candidate compounds discovered from this platform can thus be tested further towards becoming SARS-CoV-2 antiviral agents and COVID-19 chemotherapeutic treatment.

## MATERIALS AND METHODS

### Protein Preparation

Three-dimensional structure of SARS-CoV-2  $M^{pro}$  complex with baicalein ligand bound in the active site (PDB ID: 6M2N), determined by x-ray crystallography, was retrieved from the RSCB Protein Data Bank (<https://www.rcsb.org/>) in a pdb file format and used in the docking experiment. The  $M^{pro}$  structure is as of good structural quality with a resolution of 2.20 Å, and Ramachandran outliers is 0.1% (not more than 5%), indicating that the structure adequately fits the expected main-chain torsion angle distribution [16]. The coordinates of baicalein molecule was removed from the protein structure by using BIOVIA Discovery Studio Visualizer v16.1.0.15350. Water molecules were deleted

from the protein, hydrogen atoms were later added to the structure using AutoDock Tools version 1.5.7rc 1 [17] before being saved in a pdbqt file format. Subsequently, for preparing the protein structure prior to the molecular dynamics (MD) simulations, the protonation types of histidines (HID, HIE, and HIP) [18] were predicted from the APBS-PDB2PQR software suite (<https://server.poissonboltzmann.org/>) and then annotated in the structure.

### Ligand Preparation

For this study, baicalein molecule that was originally co-crystallized with SARS-CoV-2 main protease [16] was used as a reference ligand, as well as for the validation of docking. Baicalein molecule was extracted from the protein structure and saved in a pdb file format. The pdb file was then converted into a pdbqt file format by using MGLTools. Moreover, 279,970 natural product compounds to be virtually screened were retrieved from the ZINC database (<https://zinc.docking.org/>) in mol2 file format. The mol2 files were then converted to pdbqt files using AutoDock Tools (ADT) package via a Python script (run\_prepareligand).

### Molecular Docking of Protein-Ligand Complexes

Autodock Vina [19] on a Linux operating system was employed for our docking experiment. The aforementioned pdbqt files of protein and ligands were subjected to the docking protocol. A grid box was defined on the active site of  $M^{pro}$ , which covers the two crucial residues, His41 and Cys145 (supplementary figure S1). The grid box center (x, y, z) was set at -18.967, -41.985, and 29.234, respectively. Size of the grid box (x, y, z) was set as 18, 16, and 22, respectively. Subsequently, the exhaustiveness, energy range, and num\_mode was set as 30, 3, and 10, respectively. Of note, the active site of  $M^{pro}$  is highly conserved as there is no evidence, to date, showing that the major SARS-CoV-2 viral variants, such as Alpha, Beta, and Delta, have any mutation within the active site.

### Drug-likeness Evaluation

From the docking-based virtual screening result, natural product compounds showing better binding affinities (kcal/mol) than that of the reference ligand were analyzed further for their drug-like properties according to the Lipinski's rules of five [20] using the ADMETlab web server.

### Molecular Dynamics (MD) Simulations

Top compounds selected from both the docking-based screening and the drug-likeness screening were then formed into a complex with  $M^{pro}$  (via docking) and subjected to an

MD simulation by using GROMACS 4.6.3 program [21]. The MD production run for all simulating system was set as 100 ns. The Antechamber module was employed for the ligand topology preparation with a general AMBER force field (GAFF). A semi-empirical model (AM1) bond charge correction (BCC) charge model, or AM1-BCC, was used to generate the atomic charges [22]. The complex systems, coordinates, and parameters were prepared via Amber ff99SB force field [23] for enhancing the dynamics and the protein secondary structures. TIP3P water model with a cubic box of 10.0 Å was used to refine the structure for simulating and calculating thermodynamic parameters. Sodium and chloride ions were added to neutralize the charge of the system [24]. For energy minimization, the steepest algorithms were used to minimize the systems at a maximum force under 10.0 kJ/mol and a maximum number of 50,000 steps. All equilibrium of MD simulation systems were completed using a leap-frog integrator at 2 fs, temperature coupling of Berendsen thermostat at 300K [25], and isotropic Parrinello-Rahman for pressure control at 1 atm [26]. Bond constraints were accounted throughout the simulations by using a linear constraint solvent (LINCS) algorithm [27]. The Particle-Mesh Ewald (PME) method was employed for effective long-range electrostatic calculations [25]. Subsequently, the output files were analyzed by the GROMACS utilities.

### MM-PBSA Energy Calculations

Molecular trajectories of the protein-ligand complexes resulted from the MD simulations were employed in order to calculate Gibb's free energy of binding using molecular mechanics Poisson-Boltzmann surface area (MM-PBSA) approach [28, 29], which was achieved by a `g_mmpbsa` tool [30, 31] available within the APBS and GROMACS software packages. Energy parameters generated from the MM-PBSA approach was described by the following equations:

$$\Delta G_{\text{bind}} = \Delta E_{\text{MM}} + \Delta G_{\text{solv}} - T\Delta S$$

, where  $\Delta G_{\text{bind}}$  is total binding free energy.  $\Delta E_{\text{MM}}$  is calculated as a sum of the gas-phase energy,  $\Delta G_{\text{solv}}$  is solvation energy correlated to the transition from the gas-phase to the solvated state, and  $T\Delta S$  is conformational change entropy related to the ligand binding process. Of note, the entropic term is not considered in this work as its calculation require significant amount of computational effort.  $\Delta E_{\text{MM}}$  is also calculated as a sum of the gas-phase energetic terms, e.g. internal energy ( $\Delta E_{\text{int}}$ ), electrostatic interactions ( $\Delta E_{\text{elec}}$ ), and van der Waals interaction energies ( $\Delta E_{\text{vdW}}$ ) as following:

$$\Delta E_{\text{MM}} = \Delta E_{\text{int}} + \Delta E_{\text{elec}} + \Delta E_{\text{vdW}}$$

Additionally,  $\Delta G_{\text{solv}}$  is calculated as a sum of polar and non-polar components. The Poisson-Boltzmann implicit solvent model ( $\Delta G_{\text{PB}}$ ) was employed to estimate the polar contribution. On the other hand, solvent accessible surface area ( $\Delta G_{\text{SA}}$ ) was employed to calculate the non-polar contribution for free energy of solvation as follows:

$$\Delta G_{\text{solv}} = \Delta G_{\text{PB}} + \Delta G_{\text{SA}}$$

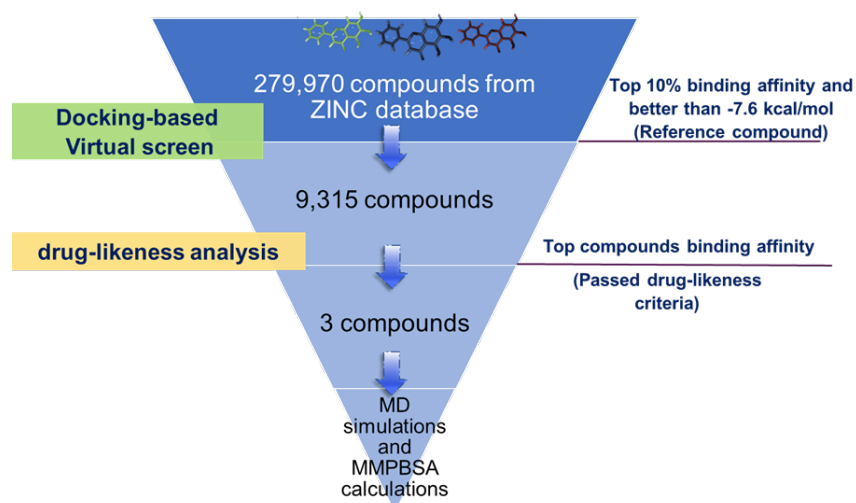
In this study, snapshots for the MM-PBSA calculation were collected every 2 ns for the last 20 ns (during 80-100 ns) of the MD runs. Additionally, in order to confirm if the structures found during the last 20 ns were good representatives of all conformations, we performed a structure clustering study. The GROMACS tool (`gmx cluster`) was utilized to evaluate structural clustering for identifying similar structures obtained during MD simulation of all  $M^{\text{pro}}$ -ligand complexes. An RMSD cut-off of 0.15 nm was set for this step, which revealed three, five, three, and three clusters for the  $M^{\text{pro}}$  complexes with C1, C2, C3, and reference compounds, respectively. The top clusters represent 90%, 58%, 96% and 96% of structures of complexes with C1, C2, C3, and reference compounds, respectively. This structure clustering analysis also confirmed that the major conformations are mainly within a duration time of 80-100 ns of the MD simulations.

## RESULTS AND DISCUSSION

### Molecular Docking and Structure-based Virtual Screening

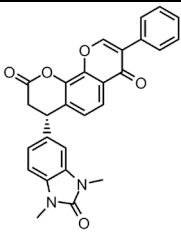
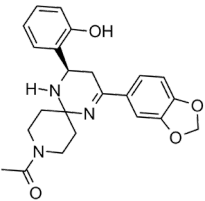
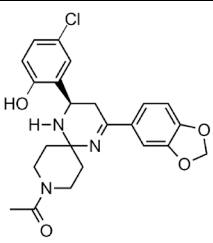
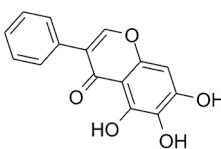
In order to obtain an optimized molecular docking protocol suitable for our virtual screening experiment, molecular docking was first validated using the co-crystal structure of SARS-CoV-2  $M^{\text{pro}}$  bound with baicalein, which was designated as our reference compound. To do this, baicalein molecule was removed from the protein-ligand complex structure, and then re-docked into the active site of the protein. In this study, we focused our docked region on the  $M^{\text{pro}}$  active site, particularly around the location of catalytic dyad, His41 and Cys145. The docked complex structure was then overlaid onto the original co-crystallized complex structure in order to compare the structural coordinates and thus determine the accuracy of the docking protocol. Root-mean-square derivation (RMSD) between the re-docked ligand and the original co-crystallized ligand structures was 0.7 Å (supplementary figure S2), indicating that the docking protocol could yield an accurate prediction of the ligand binding conformation.

For the virtual screening, 279,970 natural product compound structures were docked onto the active site and subsequently ranked based upon their calculated binding affinities (Figure 1). From this step, we selected the top 10%



**Figure 1.** Schematic diagram of the virtual screening process and the selection criteria employed in this study, which include ranking based on the docking results (binding affinity), the presence of specific interactions with the catalytic dyad residues, and the calculated drug-likeness

**Table 1.** Chemical structures and binding affinity values (determined by molecular docking) of the top three candidate compounds and the reference compound

No.	Structure	Compound	IUPAC name	Binding affinity (kcal/mol)
1		ZINC253412009 (C1)	10-(1,3-dimethyl-2-oxo-2,3-dihydro-1H benzo[d] imidazol -5-yl)-3-phenyl-9,10 dihydropyrano[2,3-f]chromene-4,8-dione	-10.5
2		ZINC65297929 (C2)	1-[2-(2H-1,3-benzodioxol-5-yl)-4-(2-hydroxyphenyl)-1,5,9-tri - azaspiro[5.5]undec-1-en-9-yl] ethan-1-one	-9.4
3		ZINC65298044 (C3)	1-[2-(2H-1,3-benzodioxol-5-yl)-4-(5-chloro-2-hydroxyphenyl)-1,5,9-triazaspiro[5.5]undec-1-en -9-yl]ethan-1-one	-9.4
4		Baicalein (Reference)	5,6,7-Trihydroxy-2-phenyl-4H-1-benzopyran-4-one	-7.6

of the ranked compounds and screened further using a cut-off binding affinity of -7.6 kcal/mol, which was the level from the docking of the reference compound (baicalein). Compounds with binding affinities lower (better) than the reference were inspected further for their specific interactions with the active site's catalytic dyad residues His41 and Cys145. As a result, 9,315 compounds passed our structure-based virtual screening criteria and were then subjected to the drug-likeness evaluation step (*vide infra*). Notably, as summarized in Table 1, the final top three candidate compounds, namely ZINC253412009 (C1), ZINC65297929 (C2), and ZINC65298044 (C3), have better binding affinities (-10.5, -9.4, and -9.4 kcal/mol, respectively) than that of the reference compound (-7.6 kcal/mol).

### Drug-likeness Evaluation

Drug-likeness analysis was employed in order to assess the molecular properties possibly affecting pharmacokinetics of

the compounds, which was also considered as one of our criteria to further screen the large pool of compounds. In this study, the compounds were analyzed with the Lipinski's rule of five, which includes: 1) number of hydrogen bond acceptors not more than 10; 2) number of hydrogen bond donors not more than 5; 3) molecular weight less than 500 g/mol; 4) Log *P* not greater than 5 (lipophilicity); and 5) polar surface area (PSA) not more than 140 Å [32]. From the top 10% of compounds selected by the binding affinity ranking and interactions with catalytic dyad residues, we selected three compounds that passed all five criteria of the Lipinski's rule of five, the results of which were summarized in Table 2. Of note, the reference compound, baicalein, also meets all of the drug-likeness requirements of Lipinski's rule of five. Therefore, all four compounds were then subjected to MD simulations and MMPBSA calculations in order to evaluate the dynamics behaviors, energetic contribution, as well as the binding mechanisms towards the active site of M<sup>pro</sup>.

**Table 2.** The properties of drug-likeness analysis of top compounds and reference compound

Compound	Mw	HB Acceptor	HB Donor	Log <i>P</i>	PSA (Å)
C1	452.47	7	0	4.091	83.44
C2	407.47	6	2	2.983	83.39
C3	441.92	6	2	3.637	83.39
Reference	270.24	5	3	2.577	90.90

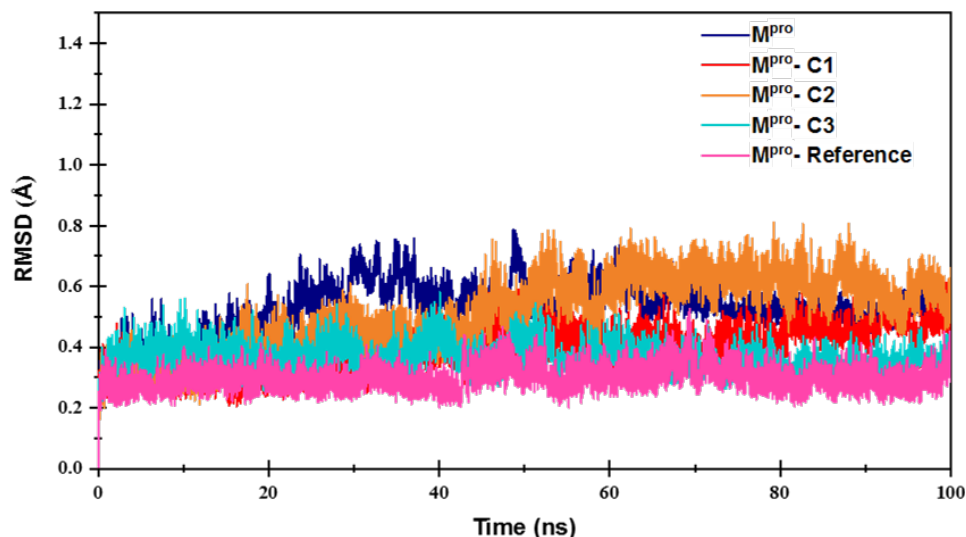
### Molecular Dynamics Simulations Study

To investigate the binding modes of the candidate compounds in the aspect of conformational motions, MD simulations were employed to analyze the complexes formed with SARS-CoV-2 M<sup>pro</sup>. Observation of atomic and molecular movements could lead to an explanation of their high binding affinity, which will help guiding future optimization of the inhibitor ligands. For a comparison, a complex formed with the reference ligand baicalein was also investigated in this experiment. All protein-ligand complexes were run for 100 ns to assess the local and regional flexibility and stability of the molecules. Outputs from the MD simulations were assessed for the common parameters generated, including root-mean-square deviation (RMSD) and root-mean-square fluctuation (RMSF). In addition to the molecular details of motions, intermolecular interactions between the candidate ligands and the protein could also be obtained from the MD snapshots, and the binding mechanisms could thus be derived based on the evidence of these intermolecular contact formation.

In order to track structural changes in complex trajectories during the MD runs, backbone heavy atom

coordinates of each protein amino acid in the molecular systems at collected time point were compared to those at the initial time point, the discrepancy of which can be presented as root-mean-square deviation (RMSD) values. Magnitude of fluctuations in the RMSD values during the MD run could infer to the formation stability of the complex structure [33]. RMSD trends of the protein coordinates in each complex were plotted in Figure 2.

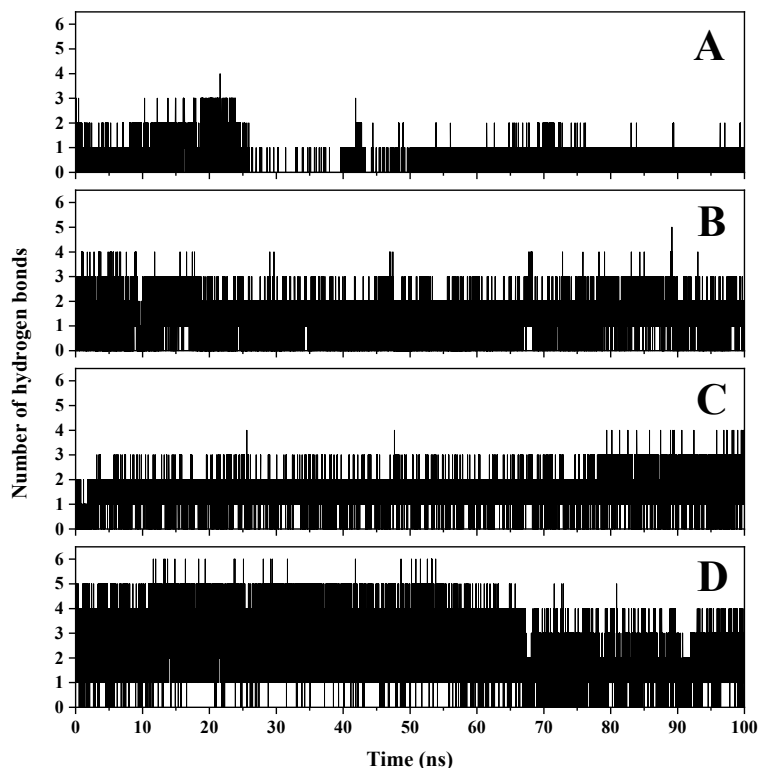
RMSD trends from the MD simulations of the four molecular systems were compared for differences in structural deviation. Results revealed that the protein backbone of M<sup>pro</sup> with no ligand bound showed some fluctuation during 20 – 40 ns period but became relatively stabilized for the remaining progression. However, the level of fluctuation was not significantly different from those found in other complexes. Interestingly, among all complexes, the M<sup>pro</sup>-C2 complex exhibited the most fluctuated RMSD's, suggesting a relatively less stable complex formation. On the other hand, M<sup>pro</sup>-C1, M<sup>pro</sup>-C3, and M<sup>pro</sup>-reference complexes appeared very stable with an average RMSD's lower than 0.5 Å, indicating a quickly stabilized complex formed between the protein and the ligands.



**Figure 2.** Root-mean-square deviation (RMSD) trends of structural coordinates comparing free  $M^{\text{pro}}$  protein (blue),  $M^{\text{pro}}$ -C1 complex (red),  $M^{\text{pro}}$ -C2 complex (orange),  $M^{\text{pro}}$ -C3 complex (cyan), and  $M^{\text{pro}}$ -reference complex (pink)

Notably, we also confirmed the reliability of the MD protocol by repeating the simulations of the  $M^{\text{pro}}$ -reference complex and subjected the trajectory data to the MMPBSA calculation in order to obtain the quantitative parameters. The results from the repeating MD production were similar, within the error ranges, to the first MD simulation, indicating that the MD protocol and the simulation data were reliable (supplementary table S1).

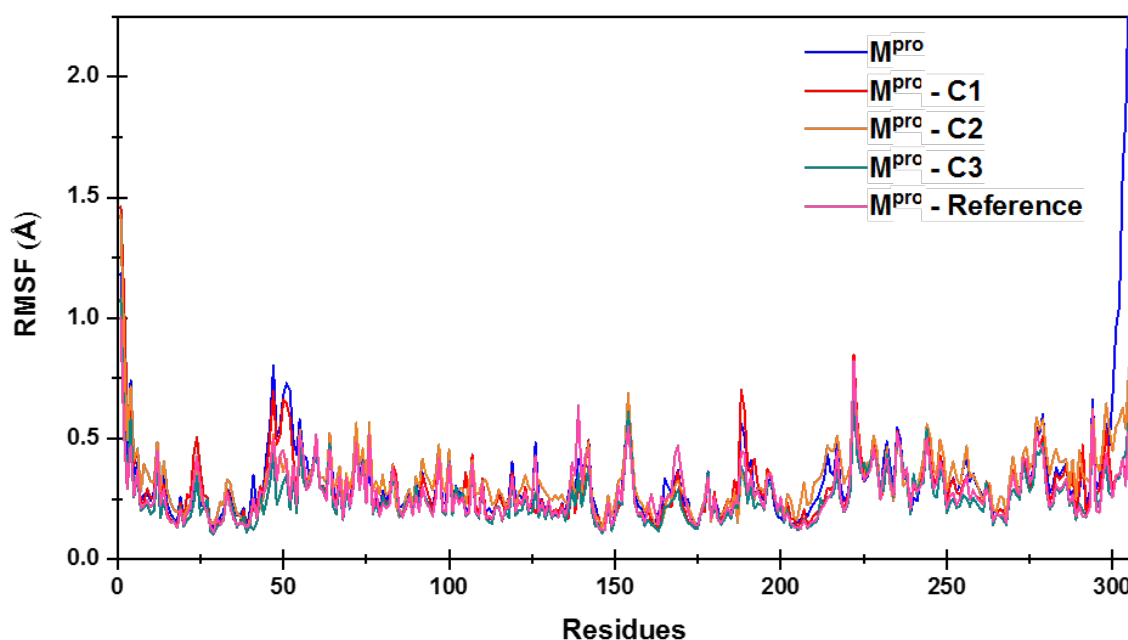
Hydrogen bond analysis was performed to further demonstrate the polar interaction of  $M^{\text{pro}}$ -ligand complexes. Figure 3 depicts the number of hydrogen bonds determined from the GROMACS utilities (gmx hbond tool). During the last 20 ns of the MD trajectories, the number of hydrogen bonds formed within the  $M^{\text{pro}}$  complexes with C1, C2, C3, and reference compounds were, by average, two, three, three, and four hydrogen bonds, respectively.



**Figure 3.** Hydrogen bond occupancy of  $M^{\text{pro}}$ -C1 complex (A),  $M^{\text{pro}}$ -C2 complex (B),  $M^{\text{pro}}$ -C3 complex (C), and  $M^{\text{pro}}$ -reference complex (D)

Additionally, we investigated local structural fluctuations of amino acid backbone along the peptide chain, as averaged for the entire MD run. To this end, root-mean-square fluctuation (RMSF) values could be calculated and presented per-residue in order to evaluate local flexibility of the peptide chain as affected by the binding of various ligands. This could be achieved by comparing the protein RMSF data from the ligand-bound and -unbound forms. The RMSF results were illustrated in Figure 4. The data showed

no significant difference in the local structural flexibility observed in bound or unbound forms of the protein. This observation also applies to the catalytic dyad residues within the active site of M<sup>pro</sup>. This could suggest that all of the ligands examined may employ a similar binding mechanism, or the binding process does not significantly perturb the local flexibility of the binding site. Notably, the entire chain of the protein is not highly fluctuated, as most of the regions exhibited low RMSF values within the level of 0.75 Å.

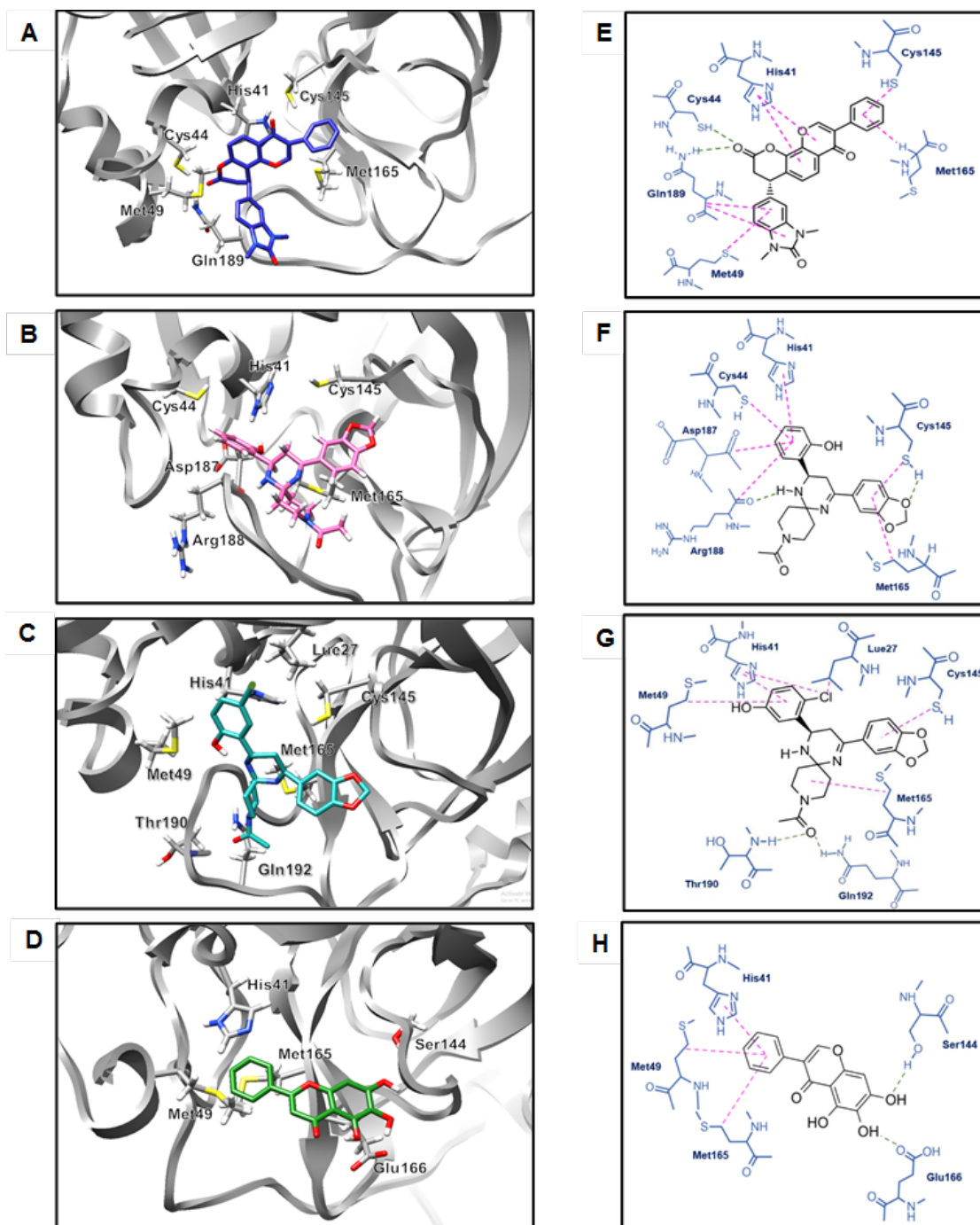


**Figure 4.** Root-mean-square fluctuations (RMSF) of M<sup>pro</sup> free protein (blue), M<sup>pro</sup>-C1 (red), M<sup>pro</sup>-C2 complex (orange), M<sup>pro</sup>-C3 complex (cyan), and M<sup>pro</sup>-reference complex (pink). Topology diagram of the M<sup>pro</sup> secondary structures is also illustrated atop of the graph indicating the residue ranges of alpha helices (pink) and beta sheets (cyan). Locations of the catalytic dyad residues His41 and Cys145 in the active site are indicated

In order to better understand how the protein and ligands interact, snapshots of the last structural trajectories were obtained and compared for their intermolecular contacts. Interactions within the active site cleft were focused, which include hydrogen bonding (H-bonds) and hydrophobic interactions such as  $\pi$ - $\pi$  stacking,  $\pi$ -sigma,  $\pi$ -sulfur, or  $\pi$ -alkyl intermolecular patterns. It was found that C1 compound could form hydrogen bonds with Cys44 and Gln189. For hydrophobic interactions, C1 formed  $\pi$ - $\pi$  stackings with His41 and Gln189,  $\pi$ -sulfur with Met49 and Cys145, and  $\pi$ -sigma with Met165. C2 compound, on the other hand, could form H-bonds with Cys145 and Arg188, as well as several hydrophobic interactions including  $\pi$ - $\pi$  stacking with His41,  $\pi$ -sulfur with Cys44 and Cys145, amide- $\pi$  with Asp187 and Arg188, and  $\pi$ -alkyl with Met165. C3 compounds could also form H-bonds with Thr190 and Gln192, and  $\pi$ - $\pi$  stacking hydrophobic interactions with

His41,  $\pi$ -sulfur with Cys145,  $\pi$ -alkyl with Lue27, His41, Met49, and Met165. To our surprise, the M<sup>pro</sup>-reference complex revealed fewer interactions than those found in other complexes, which might influence the binding free energy and the stabilizing of the molecular system. Though baicalein could form H-bonds with Ser144 and Glu166, it could form limited hydrophobic interactions only through  $\pi$ - $\pi$  stacking with His41, and  $\pi$ -alkyl with Met49 and Met165. This observation agrees well with the free energy of binding calculated from the M<sup>pro</sup>-reference complex, which will be discussed in the next section. Nevertheless, the interaction analysis could indicate that all of the candidate compounds exhibited strong intermolecular contacts with the residues lining the active site of M<sup>pro</sup>, suggesting a tight binding toward the target protein, which is one of the desired properties commonly possessed by a potent inhibitor.





**Figure 5.** Binding modes and intermolecular interactions observed within the protein-ligand complexes obtained from the last snapshot structures of the MD simulations. (A-D) three-dimensional conformations and (E-H) two-dimensional representation of interactions. Complexes include M<sup>pro</sup>-C1 (A and E), M<sup>pro</sup>-C2 (B and F), M<sup>pro</sup>-C3 (C and G), and M<sup>pro</sup>-reference (D and H). Hydrogen bonds are highlighted in green dotted lines, while hydrophobic interactions are indicated in pink dotted lines



### Binding Free Energy Calculations

Although the binding affinity values calculated from the molecular docking could provide some information regarding the trend of ligand binding efficacy, they might not be sufficiently accurate to predict the binding of ligands with similar levels of potency. To achieve higher accuracy of binding prediction, a more sophisticated calculation method such as molecular mechanics Poisson-Boltzmann surface area (MM-PBSA) could be employed to provide rich details of binding, especially the energetic contributions of the protein-ligand recognition. To do this, the trajectories of all molecular systems from our MD simulations were subsequently subjected to the MMPBSA calculations to predict the binding free energy ( $\Delta G_{\text{bind}}$ ). Additionally, the calculation could provide a set of energetic parameters contributing to the  $\Delta G_{\text{bind}}$ , which could help us analyze further by comparing with the modes of binding observed by the MD simulations. The results from MMPBSA

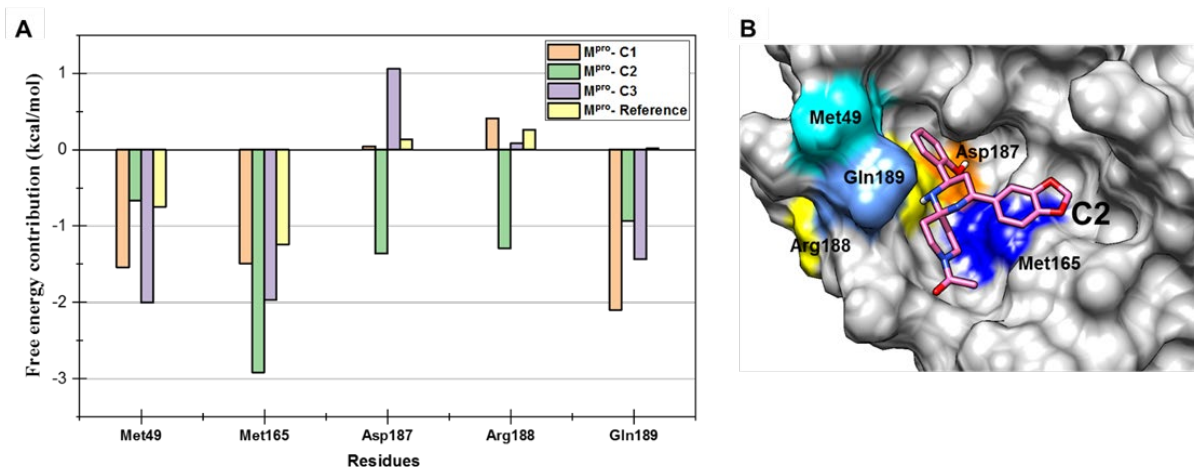
calculations are summarized in Table 3, which describes all contributing energy terms amounted to the binding of each M<sup>Pro</sup>-ligand complexes. Interestingly, all of the complexes with candidate compounds exhibited equally low  $\Delta G_{\text{bind}}$  when compared with the M<sup>Pro</sup>-reference complex, indicating stronger binding than that of the reference ligand. The levels of contributing energies are also similar among the candidate ligands. However, the reference ligand binding exhibited lower (more negative) electrostatic energy ( $\Delta E_{\text{elec}}$ ) and higher  $\Delta G_{\text{nonpolar}}$ , which could be due to the extensive delocalization of pi-electrons in its connected aromatic rings.  $\Delta E_{\text{vdw}}$  of the reference ligand binding is also poorer (less negative), which could be the result of having a significantly smaller molecular size than those of the candidate compounds. The stark difference in the total binding free energies between the reference and the candidate ligands agrees well with the number of intermolecular contacts observed from the MD snapshots as described in the previous section.

**Table 3.** Binding free energy and contributing energetic parameters of the complexes between M<sup>Pro</sup> and either candidate or reference ligands calculated from MMPBSA

Compounds	Contributing Energy (kcal/mol)				
	$\Delta E_{\text{vdw}}$	$\Delta E_{\text{elec}}$	$\Delta G_{\text{nonpolar}}$	$\Delta G_{\text{polar}}$	$\Delta G_{\text{binding}}$
C1	-197.99 ± 12.86	-39.51 ± 9.59	-17.44 ± 1.01	135.38 ± 14.76	-28.56 ± 15.71
C2	-220.17 ± 10.21	-37.47 ± 8.71	-18.81 ± 0.81	156.92 ± 9.54	-28.55 ± 11.79
C3	-233.57 ± 11.54	-38.54 ± 8.39	-20.60 ± 0.91	174.63 ± 13.33	-28.20 ± 12.66
Reference	-122.10 ± 20.50	-62.06 ± 32.16	-12.79 ± 1.25	136.23 ± 25.92	-14.50 ± 13.63

Additionally, in order to further delineate individual contribution of each particular interaction with the M<sup>Pro</sup> active site, free energy decomposition per amino acid residue was also calculated (Figure 6). Data from this decomposition method could also identify key amino acid residues playing important role in the recognition of tight-binding ligands. Furthermore, information from this experiment could help guiding our future optimization of the inhibitor by chemically modifying the compounds to accommodate more contacts with the key residues, thereby enhancing the binding efficacy. From the calculation, we found five amino acids, including Met49, Met165, Asp187, Arg188, and Gln189, that could interact significantly with the ligands, exhibiting free energy decomposition lower than -1.0 kcal/mol per-residue. From the trend of energetic

contribution, it was apparent that the candidate ligands interact with these key residues better than the reference ligand. By average, Met165 residue exhibited the most significant level of free energy contribution, indicating its importance in determining the ligand binding robustness. Previous research studies [34, 35] have also revealed that Met165 contributed the most significant binding free energy toward several inhibitors. However, in the cases of Asp187 and Arg188, only C2 compound can interact favorably with these two residues, which coincide well with the result from our MD snapshot (Figure 5F) showing that these particular interactions are exclusive to only C2 ligand. This key piece of information could thus help direct our future ligand optimization effort to include these specific interactions in the molecular design.



**Figure 6.** (A) Free energy decomposition per-residue for the binding between M<sup>pro</sup> and C1 (orange), C2 (green), C3 (purple), and reference (yellow) ligands. (B) Three-dimensional surface model of the M<sup>pro</sup> active site when bound with a representative ligand highlighting all key amino acid residues significantly contributing to the ligand binding

## CONCLUSION

The current COVID-19 pandemic severely affecting the world population has prompted biomedical scientists around the globe to discover an effective therapeutic treatment for the disease, both for the preventative vaccine and the drug treating infected patients. Additionally, as the mutated variants of the virus constantly emerge, the development process has to be accelerated to respond quickly to the outbreaks. To this end, we have employed *in silico* methods to discover novel inhibitor candidates from a large pool of natural product ligands available in the compound database. From our virtual screening, key interactions with catalytic dyad, and drug-likeness screening, three potential candidates were identified. Additional MD simulations, MMPBSA calculations, and decomposition energy could also describe the binding of the candidate compounds and identify several key contacts with the amino acid residues lining the active site of M<sup>pro</sup>. These key intermolecular interactions could also help us to better understand the binding mechanism of potent inhibitors as well as to guide future endeavors in optimizing the chemical structure of the ligand. Future *in vitro* and *in vivo* testing is also required in order to confirm the result and evaluate the possibility towards becoming an actual treatment for COVID-19.

## ACKNOWLEDGEMENTS

We would like to thank the Drug Discovery Research Laboratory, the Teaching Assistant and Research Assistant Scholarship Program, and the Department of Chemistry, Faculty of Science, Chiang Mai University, for supporting this research study.

## CONFLICT OF INTEREST

The authors declare that there is no conflict of interest regarding the publication of this manuscript.

## REFERENCES

- Dai, W., Zhang, B., Jiang, X.M., Su, H., Li, J., Zhao, Y., Xie, X., Jin, Z., Peng, J. and Liu, F. (2020) Structure-based design of antiviral drug candidates targeting the SARS-CoV-2 main protease. *Science* **368**(6497), 1331-1335.
- Zhou, P., Yang, X.L., Wang, X.G., Hu, B., Zhang, L., Zhang, W., Si, H.R., Zhu, Y., Li, B. and Huang, C.L. (2020) A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature* **579**(7798), 270-273.
- World Health Organization. (2021, July 19). Coronavirus disease 2019 (COVID-19) situation report. <https://covid19.who.int/table>
- World Health Organization. (2021, February 11). Coronavirus disease 2019 (COVID-19) situation report. <https://www.who.int/emergencies/diseases/novel-coronavirus-2019/question-and-answers-hub/q-a-detail/coronavirus-disease-covid-19>
- ADMINISTRATION, U.S.F.D. (2021, September 28). coronavirus-covid-19-drugs 2021. <https://www.fda.gov/drugs/emergency-preparedness-drugs/coronavirus-covid-19-drugs>
- Yang, Y., Xiao, Z., Ye, K., He, X., Sun, B., Qin, Z., Yu, J., Yao, J., Wu, Q. and Bao, Z. (2020) SARS-CoV-2: characteristics and current advances in research. *Virol. J.* **17**(1), 1-17.
- Lu, R., Zhao, X., Li, J., Niu, P., Yang, B., Wu, H., Wang, W., Song, H., Huang, B. and Zhu, N. (2020) Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding. *Lancet* **395**(10224), 565-574.
- Boopathi, S., Poma, A.B. and Kolandaivel, P. (2021) Novel 2019 coronavirus structure, mechanism of action, antiviral drug promises and rule out against its treatment. *J. Biomol. Struct. Dyn.* **39**(9), 3409-3418.

9. Haque, S.M., Ashwaq, O., Sarief, A. and Mohamed, A.K.A.J. (2020) A comprehensive review about SARS-CoV-2. *Future Virol.* **15**(9), 625-648.
10. Zhang, L., Lin, D., Sun, X., Curth, U., Drosten, C., Sauerhering, L., Becker, S., Rox, K. and Hilgenfeld, R. (2020) Crystal structure of SARS-CoV-2 main protease provides a basis for design of improved  $\alpha$ -ketoamide inhibitors. *Science* **368**(6489), 409-412.
11. Amin, S.A., Banerjee, S., Ghosh, K., Gayen, S. and Jha, T. (2021) Protease targeted COVID-19 drug discovery and its challenges: Insight into viral main protease ( $M^{pro}$ ) and papain-like protease ( $PL^{pro}$ ) inhibitors. *Bioorg. Med. Chem.* **29**, 115860.
12. Suárez, D. and Díaz, N. (2020) SARS-CoV-2 main protease: A molecular dynamics study. *J. Chem. Inf. Model.* **60**(12), 5815-5831.
13. Delang, L., Abdelnabi, R. and Neyts, J. (2018) Favipiravir as a potential countermeasure against neglected and emerging RNA viruses. *Antivir. Res.* **153**, 85-94.
14. Croom, K. F., Dhillon, S. and Keam, S. J. (2009). Atazanavir: a review of its use in the management of HIV-1 infection. *Drugs* **69**(8), 1107-1140.
15. Von Hentig, N. (2008) Atazanavir/ritonavir: a review of its use in HIV therapy. *Drugs Today* **44**(2), 103-132.
16. Su, H.X., Yao, S., Zhao, W.F., Li, M.J., Liu, J., Shang, W.J., Xie, H., Ke, C.Q., Hu, H.C. and Gao, M.N. (2020) Anti-SARS-CoV-2 activities in vitro of Shuanghuanglian preparations and bioactive ingredients. *Acta Pharmacol. Sin.* **41**(9), 1167-1177.
17. Morris, G.M., Huey, R., Lindstrom, W., Sanner, M.F., Belew, R.K., Goodsell, D.S. and Olson, A.J. (2009) AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. *J. Comput. Chem.* **30**(16), 2785-2791.
18. Olsson, M.H., Søndergaard, C.R., Rostkowski, M. and Jensen, J.H. (2011) PROPKA3: consistent treatment of internal and surface residues in empirical pKa predictions. *J. Chem. Theory Comput.* **7**(2), 525-537.
19. Trott, O. and Olson, A. J. (2010) AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J. Comput. Chem.* **31**(2), 455-461
20. Jia, C.Y., Li, J.Y., Hao, G.F. and Yang, G.F. (2020) A drug-likeness toolbox facilitates ADMET study in drug discovery. *Drugs Today* **25**(1), 248-258.
21. Berendsen, H.J., van der Spoel, D. and van Drunen, R. (1995) GROMACS: A message-passing parallel molecular dynamics implementation. *Comput. Phys. Commun.* **91**(1-3), 43-56.
22. He, X., Man, V.H., Yang, W., Lee, T.S. and Wang, J. (2020) A fast and high-quality charge model for the next generation general AMBER force field. *Chem. Phys.* **153**(11), 114502.
23. Maier, J.A., Martinez, C., Kasavajhala, K., Wickstrom, L., Hauser, K. E. and Simmerling, C. (2015) ff14SB: improving the accuracy of protein side chain and backbone parameters from ff99SB. *J. Chem. Theory Comput.* **11**(8), 3696-3713.
24. Jorgensen, W.L., Chandrasekhar, J., Madura, J.D., Impey, R.W. and Klein, M.L. (1983) Comparison of simple potential functions for simulating liquid water. *Chem. Phys.* **79**(2), 926-935.
25. Berendsen, H.J., Postma, J.V., Van Gunsteren, W.F., DiNola, A., and Haak, J.R. (1984) Molecular dynamics with coupling to an external bath. *Chem. Phys.* **81**(8), 3684-3690.
26. Parrinello, M. and Rahman, A. (1981) Polymorphic transitions in single crystals: A new molecular dynamics method. *J. Appl. Phys.* **52**(12), 7182-7190.
27. Hess, B. (2008) P-LINCS: A parallel linear constraint solver for molecular simulation. *J. Chem. Theory Comput.* **4**(1), 116-122.
28. Genheden, S. and Ryde, U. (2015) The MM/PBSA and MM/GBSA methods to estimate ligand-binding affinities. *Expert Opin. Drug Discov.* **10**(5), 449-461.
29. Huang, K., Luo, S., Cong, Y., Zhong, S., Zhang, J.Z. and Duan, L. (2020) An accurate free energy estimator: based on MM/PBSA combined with interaction entropy for protein-ligand binding affinity. *Nanoscale* **12**(19), 10737-10750.
30. Kumari, R., Kumar, R., Consortium, O.S.D.D. and Lynn, A. (2014) g\_mmpbsa A GROMACS tool for high-throughput MM-PBSA calculations. *J. Chem. Inf. Model.* **54**(7), 1951-1962
31. Paissoni, C., Spiliotopoulos, D., Musco, G. and Spitaleri, A. (2015) GMXPBSA 2.1: A GROMACS tool to perform MM/PBSA and computational alanine scanning. *Comput. Phys. Commun* **186**, 105-107.
32. Pollastri, M.P. (2010) Overview on the Rule of Five. *Curr. Protoc. Pharmacol.* **49**(1), 9-12.
33. Sargsyan, K., Grauffel, C. and Lim, C. (2017) How molecular size impacts RMSD applications in molecular dynamics simulations. *J. Chem. Theory Comput.* **13**(4), 1518-1524.
34. Joshi, T., Sharma, P., Joshi, T., Pundir, H., Mathpal, S. and Chandra, S. (2021) Structure-based screening of novel lichen compounds against SARS Coronavirus main protease ( $M^{pro}$ ) as potentials inhibitors of COVID-19. *Mol. Divers.* **25**(3), 1665-1677.
35. Sk, M.F., Roy, R., Jonniya, N.A., Poddar, S. and Kar, P. (2021) Elucidating biophysical basis of binding of inhibitors to SARS-CoV-2 main protease by using molecular dynamics simulations and free energy calculations. *J. Biomol. Struct. Dyn.* **39**(10), 3649-3661.