ORIGINAL RESEARCH

Exploration of chalcones as 3-chymotrypsin-like protease (3CL^pro) inhibitors of SARS-CoV-2 using computational approaches

Thua-Phong Lam[1] · Dac-Nhan Nguyen[1] · Tan Thanh Mai[1] · Thanh-Dao Tran[1] · Minh-Tri Le[1,2] ·
Phuong Nguyen Hoai Huynh[1] · Duc-Tuan Nguyen[1] · Viet-Hung Tran[3] · Dieu-Thuong Thi Trinh[4] · Phuong Truong[1] ·
Cam-Van T. Vo[1] · Khac-Minh Thai[1]

Abstract

The main protease 3CL^pro is one of the potential targets against coronavirus. Inhibiting this enzyme leads to the interruption of viral replication. Chalcone and its derivatives were reported to possess the ability to bind to 3CL^pro protease in the binding pocket. This study explored an in-house database of 269 chalcones as 3CL^pro inhibitors using in silico screening models, including molecular docking, molecular dynamics simulation, binding free energy calculation, and ADME prediction. **C264** and **C235** stand out as the two most potential structures. The top hit compound **C264** was with the Jamda score of −2.8329 and the MM/GBSA binding energy mean value of −28.23 ± 3.53 kcal/mol, which was lower than the reference ligand. Despite the lower mean binding energy (−22.07 ± 3.39 kcal/mol), in-depth analysis of binding interaction suggested **C235** could be another potential candidate. Further, in vitro and in vivo experiments are required to confirm the inhibitory ability.

Keywords SARS-CoV-2 · Chalcone derivatives · 3CL^pro protease · In silico screening

Introduction

The COVID-19 pandemic caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) virus was first reported in Wuhan City, Hubei Province, China, in late 2019 [1]. Until March 7th, 2022, there have been over 445 million confirmed cases of COVID-19, including nearly 6.0 million deaths, reported to the WHO [2]. Four primary

Thua-Phong Lam and Dac-Nhan Nguyen contributed equally to this work and are co-first authors.

✉ Khac-Minh Thai
  thaikhacminh@ump.edu.vn; thaikhacminh@gmail.com

1 Department of Medicinal Chemistry, Faculty of Pharmacy, University of Medicine and Pharmacy at Ho Chi Minh City, Ho Chi Minh City 700000, Vietnam

2 School of Medicine, Vietnam National University Ho Chi Minh City, Ho Chi Minh City 700000, Vietnam

3 Institute of Drug Quality Control Ho Chi Minh City, Ho Chi Minh City 700000, Vietnam

4 Faculty of Traditional Medicine, University of Medicine and Pharmacy at Ho Chi Minh City, 700000 Ho Chi Minh City, Vietnam

treatments, including repositioning antiviral drugs, convalescent plasma, vaccines, and other adjuvant therapies are currently applied to deal with the pandemic. Vaccines are the most effective preventative therapy, with more than 10.7 billion doses have been administered. Although the numbers of antiviral agents have been developed in the early stage of the pandemic, molnupiravir was the first oral medicine approved by the FDA since December 2021 [3]. Paxlovid—a combination of nirmatrelvir and ritonavir by Pfizer, has recently been reported to reduce the risk of hospitalization or death by 89% compared to placebo in non-hospitalized high-risk adults and being assessed by the FDA for the emergency use authorization (EUA). Nirmatrelvir was specifically designed as a covalent 3CL^pro inhibitor [4].

The main protease 3CLpro (also called 3-chymotrypsin-like protease or nsp5) is one of the potential medicine targets against coronavirus. The complex of 3CLpro and papain-like protease plays a vital role in the post-translation. Blocking 3CLpro could interrupt the viral replication [5]. 3CLpro is a dimer protease with two identical protomers. Each comprises three domains: domain I (residues 10 to 99), domain II (residues 100 to 182), and domain III (residues 198 to 303). Two crucial residues at the substrate-binding site (catalytic dyad)

are Cys145 and His41 [6]. Many studies focused on virtual screening for SARS-CoV-2 3CL-protease inhibitors [7–10]. Some reported that flavonoids could strongly bind to the protein target [9, 11, 12]. However, the bioactivities of most top hit candidates have not yet been confirmed experimentally.

Flavonoids are major natural structures with a wide range of bioactivities. The flavonoid family consisted of different subgroups such as chalcone, flavone, and flavonol [13]. Chalcone structures withdraw much interest from researchers due to their broad spectrum of biological activity. Different substituents on ring A and ring B of chalcone derivatives could potentially form many hydrogen bonds, inducing the binding of chalcone toward 3CLpro. This study aimed to screen for 3CLpro inhibitors from the in-house chalcone database with in silico models, including molecular docking, molecular dynamics simulation, binding free energy calculation, and ADME prediction. The study of Mathpal et al. shared a similar screening method with various in silico filters from 3000 compounds having chalcone scaffold in the PubChem library [12]. The dataset of 269 in-house chalcone derivatives used in our study is the one of readily synthesized structures. Therefore, further in vitro experiments on the top hit structures can be performed easily.

Materials and methods

Preparation of chalcone structures

The 2D molecular structures of 269 in-house chalcone derivatives were created by ChemDraw 18.1 [14], converted into 3D-structure, molecular energy minimized by Molecular Operating Environment (MOE) 2019.01 software [15]. In the energy-minimized process, the forcefield AMBER10:EHT was used with 0.0001 RMS kcal/mol/A^2 gradient. The individual chalcone structure was saved in *.sdf format and then transformed to *.pdbqt format using OpenBabel 2.4.1 software [3].

Preparation of 3CL^{pro} structure of SARS-CoV-2

Up to December 2021, there are 392 SARS-CoV-2 3CL^{pro} structures in the Protein Data Bank using the 6LU7 FASTA sequence. The first crystal structure of the SARS-CoV-2 3CL^{pro} was with the N3 covalent inhibitor. With the aim to determine the non-covalent inhibitor of the SARS-CoV-2 3CL^{pro}, the crystal structure of SARS-CoV-2 3CL^{pro} (PDB ID: 7L14) with the co-crystallized ligand (2-(3-(3-chloro-5-(cyclopropylmethoxy)phenyl)-2-oxo(2H-(1,3′-bipyridine))-5-yl)benzonitrile) (**Compound 26**) and the resolution of 1.80 Å (R-free, 0.204, and R-work, 0.176) was selected. The IC$_{50}$ value of **Compound 26 (Cmpd 26)** was 0.170 nM [16]. Screening models in this study were generated based

on the interaction between **Cmpd 26** and 3CL^{pro}. After the protein structure was downloaded from the PDB website, water molecules were removed. Non-polar hydrogens were subsequently added and charged using AutoDock tools 1.5.7. Only chain A of the homodimer protein was used for virtual screening; chain B and their corresponding ligand were eliminated.

In silico screening approaches

The workflow of this study was illustrated in Fig. 1. At first, 269 in-house chalcones were docked into 3CL protease (PDB ID: 7L14) by AutoDock Vina 1.1.2 [17]. The outcome was then optimized and rescored by JamdaScorer [18]. The chalcones with the top 10 Jamda score were subjected to molecular dynamics simulation using GROMACS 2021.6. The stability of the protein–ligand complex was evaluated through RMSD and RMSF values. The interaction of protein and ligand was estimated by hydrogen bond frequency and MM/GBSA binding free energy calculation. The pharmacodynamic properties of potential candidates were predicted by ADME assessment (SwissADME) [19].

Molecular docking

The docking process was performed by AutoDock Vina 1.1.2 [17], in which the binding pocket was defined to cover the whole of the reference ligand **Cmpd 26** and the interacted residues. When docking, the top 1 pose of each ligand was recorded. The molecular docking results were then optimized and rescored using JamdaScorer, as the JAMDA scoring function is a novel empirical scoring function built from easily optimizable individual terms. It showed excellent pose prediction performance in the CASF-2016 docking power benchmark [18]. The potential substances were selected based on the Jamda score and protein–ligand interaction model, including hydrogen bonds, ionic bonds, arene–arene, arene–H, arene–cationic, and van der Waals contacts.

Molecular dynamics simulation

The molecular dynamics simulations (MDs) of apoprotein, reference complex, and top 10 chalcone complexes were implemented by GROMACS 2021.6 [20]. The ligand topology was created by the Swissparam tool (http://www.swissparam.ch) [21] in the CHARMM-27 force field and then merged with the protein's coordinate to generate the complex topology file in *.gro format. This study used a dodecahedron grid box with an edge of 10 Å to perform MDs, in which the system was solvated in the TIP3P water model and neutralized by adding Na$^+$ and Cl$^-$. The energy minimization took place in 100 ps to eliminate the system's spatial conflicts or geometrical mismatches. The system
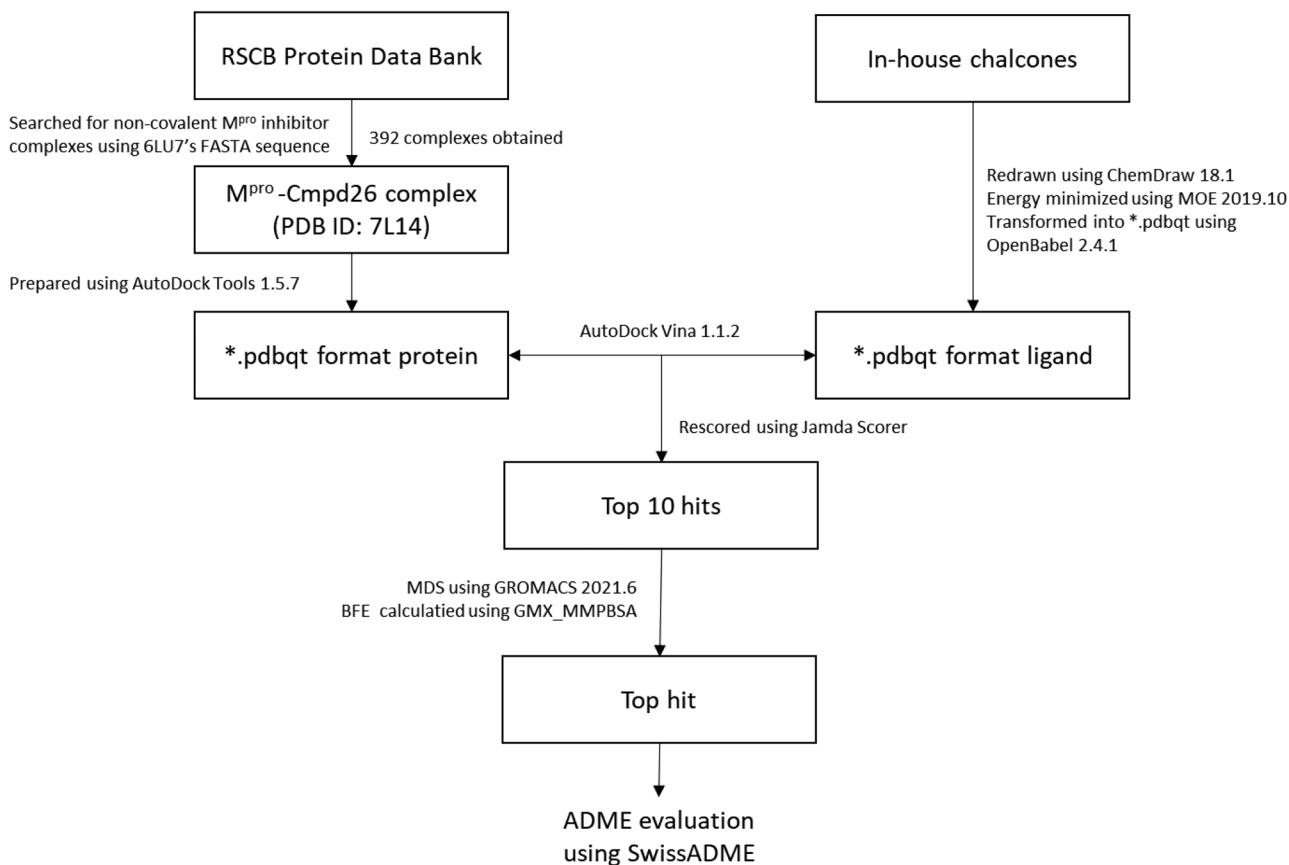
Fig. 1 In silico flow chart techniques used in this research for the evaluation of the in-house chalcone for the ability to inhibit SARS-CoV-2 M^pro

was also equilibrated in "isothermal–isochoric" and "isothermal–isobaric" conditions, with two equilibrium phases, namely, NVT and NPT ensembles. Each equilibration was conducted in 100 ps, with a temperature increased to 300 K in the NVT step using the Berendsen thermostat and a stable 1 bar (0.987 atm) pressure in the NPT step using the Parrinello–Rahman pressurization unit. After reaching equilibrium, the system was simulated in 50 ns, at a temperature of 300 K, and a pressure of 1 bar. The simulation results were recorded frame-by-frame every 0.01 ns. Simultaneously, the output trajectory was then transformed into pdb format for further analysis, such as protein–ligand interactions using the PLIF tool in MOE 2019.01 and binding free energy estimation using MM/GBSA method integrated with the GMX_MMPBSA program [22].

The stability of the protein–ligand complex was evaluated through RMSD and RMSF values. RMSD values indicated the stability of the complex during simulation by analyzing the equilibration of MD trajectories (Eq. (1)). RMSF values indicated fluctuation of specific amino acids in the interactive region (Eq. (2)). A stable structure should have an RMSD value under 3 Å [23] and an RMSF value less than 2 Å [24]:

$$RMSD(t) = \sqrt{\frac{1}{M}\sum_{i=1}^{N} m_i \left| r_i(t) - r_i^{ref} \right|^2} \tag{1}$$

$$RMSF_i = \sqrt{\frac{1}{T}\sum_{t=1}^{T} \left| r_i(t) - r_i^{ref} \right|^2} \tag{2}$$

Binding free energy calculation

The binding free energy of the ligand–protein complex was calculated based on molecular mechanical surface area (MM/GBSA) using the GMX_MMPBSA program [22]. The free energy of each system was calculated based on vacuum molecular mechanical potential (E_{MM}), solvation energy ($G_{solvation}$), and entropy (ΔS) in Eq. (4). The binding free energy was the difference between the protein–ligand complex free energy and the total free energy of apoprotein and ligand in an aqueous solvent (Eq. (3)):

$$\Delta G_{bind} = G_{complex} - (G_{protein} + G_{ligand}) \tag{3}$$

$$G_x = E_{MM} + G_{solvation} - T\Delta S \tag{4}$$

ADME prediction

The ADME properties of the top hit after the binding free energy estimation was assessed using the SwissADME tool available at http://www.swissadme.ch/ [19]. SwissADME is an online tool predicting the physicochemical properties, pharmacokinetics, drug-likeness, and accessibility to pharmaceutical chemistry with various reliable and quickly predictive models (Table S3). The BOILED-Egg (Brain Or Intestinal Estimated Permeation) model predicts whether a substance could passively absorb through the intestinal epithelium or blood–brain barrier. Meanwhile, the possibility of oral medicine was tested by five laws, including Lipinski et al. [25], Ghose et al. [26], Veber et al. [27], Egan et al. [28], and Muegge et al. [29]. The lead-likeness substances were filtered by Brenk et al. criteria [30], and these substances were allowed to interact with cytochrome P450. Since the SwissADME server did not provide any screening criteria for a "good" or "bad" compound, our criteria were set to assume that the best compound would not violate any of the assessed models.

Result and discussion

Database

A total of 269 in-house chalcone derivatives were redrawn, minimized, and gathered in a database. The compound **C1** was the basic chalcone structure. The remaining database consisted of different substituents on the two aromatic rings that were assumed to facilitate the stability of chalcones in the protease binding pocket via multiple hydrogen bonds and multiple hydrophobic interaction types. The entire database of these chalcones with ID, SMILE structure, and other molecular properties can be found in Table S1.

Molecular docking results

The grid box for the virtual screening model was set based on the co-crystallized ligand **Cmpd 26** and the corresponding interacted residues. In short, the grid box center was set at ($x = -21.376$, $y = -4.398$, $z = -28.478$), with the size of $18 \times 14 \times 18$ Å. The docking model was validated throughout
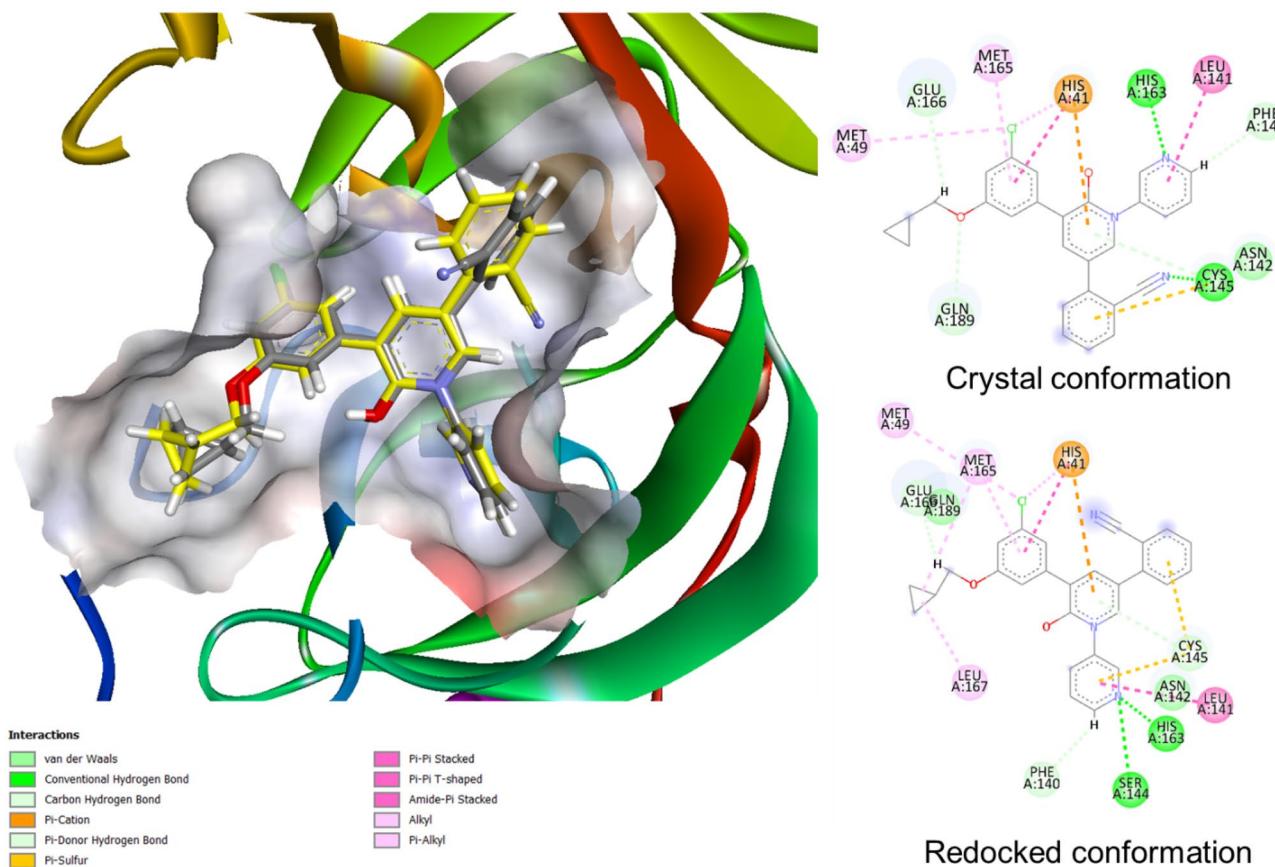
Fig. 2 Structure of SARS-CoV-2 Mpro in complex with co-crystallized **Compound 26** (yellow sticks) (PDB ID: 7L14) and the pose after the redocking process (gray sticks) and 2D-diagram of interactions between the **Compound 26** with SARS-CoV-2 Mpro before and after the redocking process

**Table 1** Top 10 chalcones having the best Jamda score after optimization
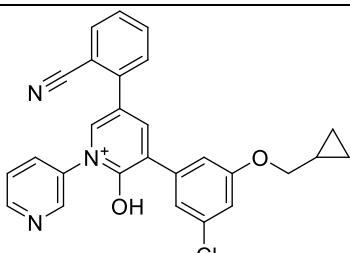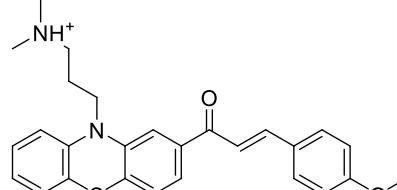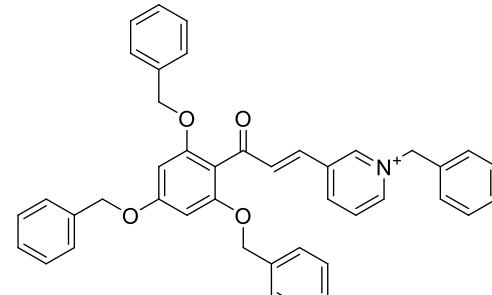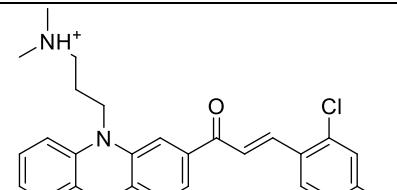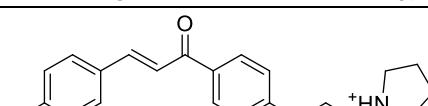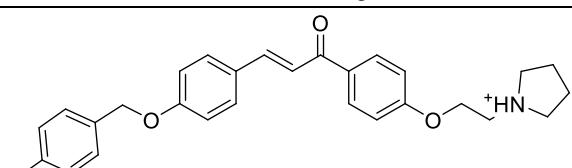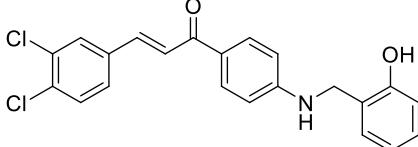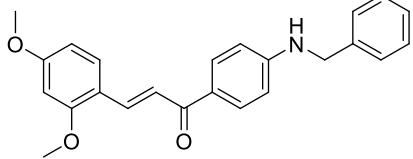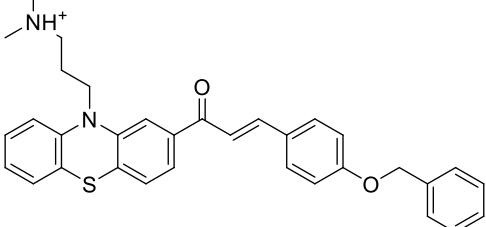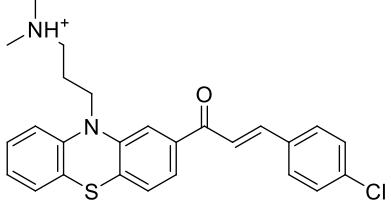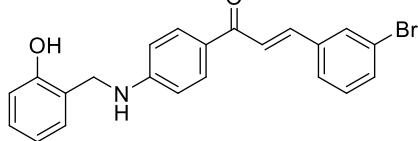
| No.	Chalcone ID	2D structure	Jamda Score	
			Before optimization	After optimization
0	Compound 26 (Cmpd 26)		-2.98	
1	C250		-2.27	-2.94
2	C269		-2.33	-2.92
3	C264		-2.30	-2.83
4	C200		-2.24	-2.81
5	C251		-2.33	-2.81

Table 1 (continued)

No.	Chalcone ID	2D structure	Jamda Score	
			Before optimization	After optimization
6	C233		-2.35	-2.78
7	C214		-2.45	-2.77
8	C267		-2.35	-2.75
9	C254		-2.34	-2.75
10	C235		-2.30	-2.74

the redocking process, with the result described in Fig. 2. As can be seen, the RMSD value is 1.637 Å, which reveals that the re-docked pose was highly similar to the original crystal pose, and the interaction between the co-crystallize ligand and the protease was highly conserved. Thus, the docking model is reliable and suitable for virtual screening.

All 269 chalcone structures were successfully docked into the active site of 3CL^{pro} with negative AutoDock Vina docking scores (Table S2). After re-scoring and optimizing by JamdaScorer software [18], the Jamda score (JS) of the chalcone derivatives ranges from −1.38298 to −2.94106 JS, in which the co-crystallized **Cmpd 26** took a value of −2.9843 JS. The top 10 chalcones with the lowest JS score are listed in Table 1, which could be divided into four chalcone structural framework groups, namely, phenothiazine ring, O-(2-(1-pyrrolidinyl)ethyl)phenol ring, N-benzylaniline ring, and tribenzyloxybenzene ring.

Chalcone with phenothiazine ring

C250, **C254**, **C264**, and **C267** were 4 chalcones containing phenothiazine rings that could bind strongly to 3CL^{pro} (Fig. 3). The Jamda scores of these four chalcones were, respectively, −2.9411, −2.7456, −2.8330, and −2.7474. These substances were bound to 3CL^{pro} via similar residues as the interaction between co-crystallized ligand **Cmpd 26** and 3CL^{pro}. All four chalcone derivatives generated van der Waals or π–sulfur interaction toward the catalytic residue Cys145. The main carbonyl and the −NH$_4^+$ side chain of the **C267** chalcone contributed attractive charge interactions toward Glu166. Meanwhile, another catalytic residue His41 connected with three out of four chalcones (**C250**, **C254**, **C264**) through π–cation interaction, similar to the binding between 3CL^{pro} and the reference ligand. In addition,
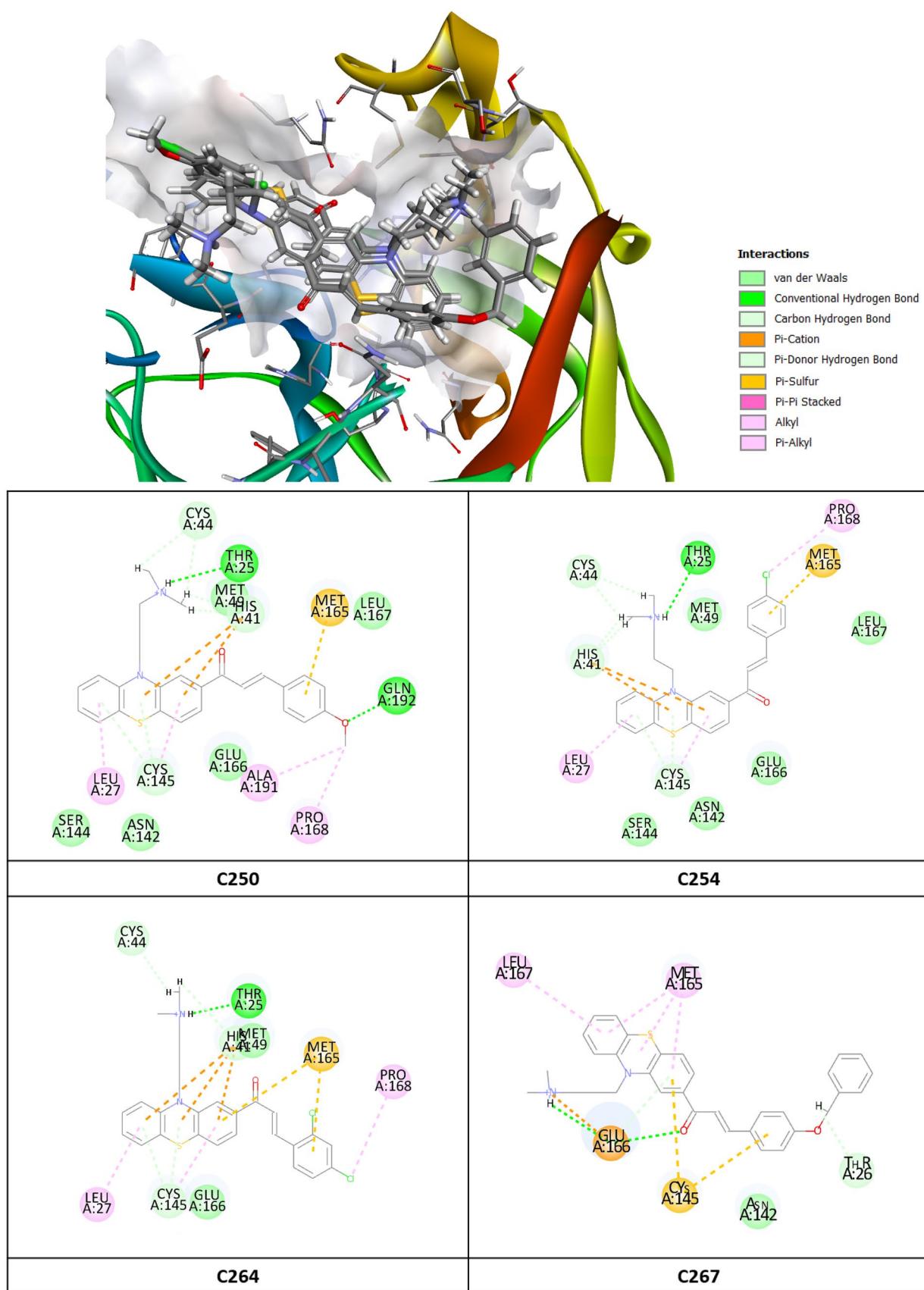
Fig. 3 Binding mode of 4 chalcones which having ring A as phenothiazine with SARS-CoV-2 M^pro
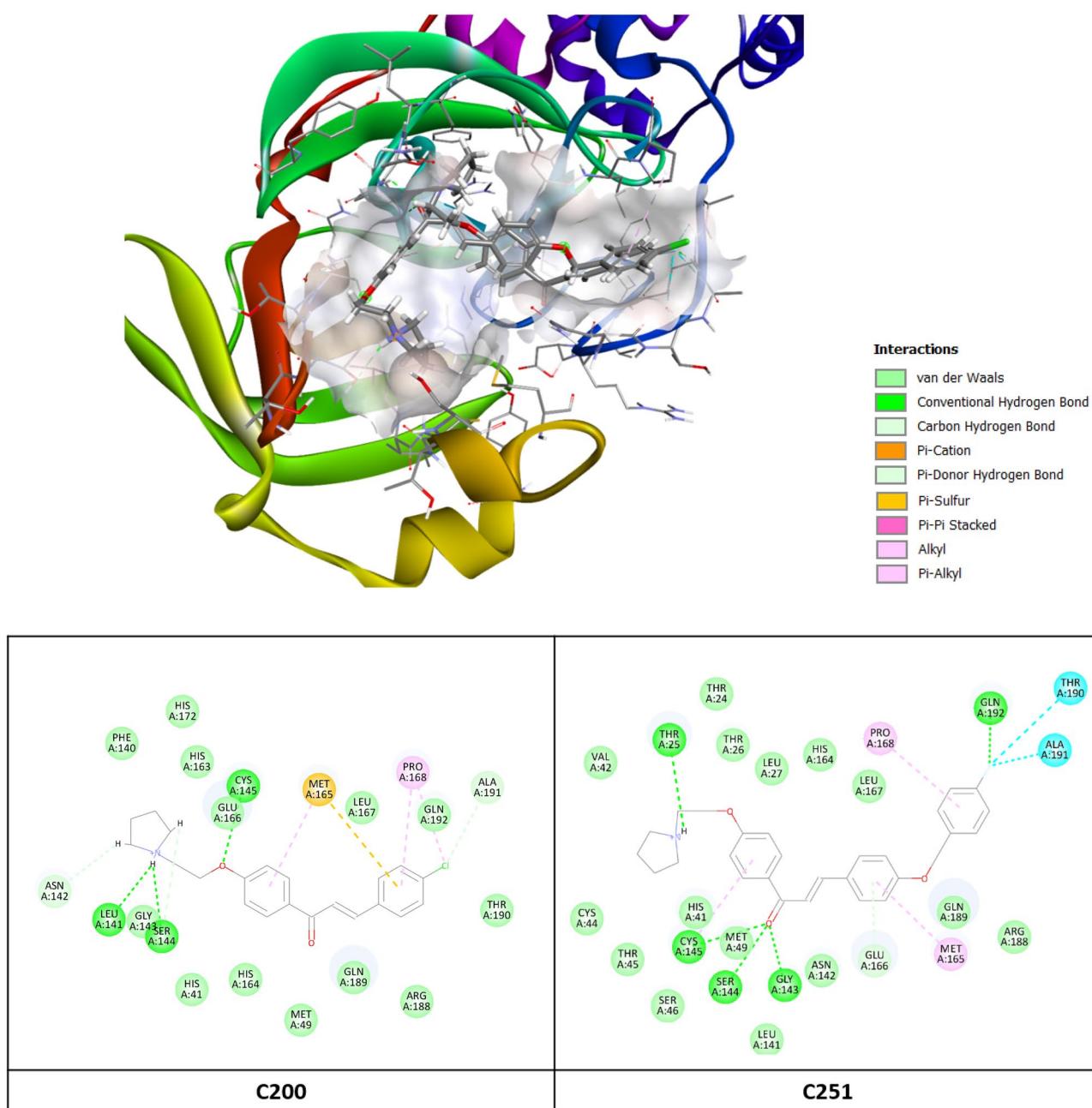
Fig. 4 Binding mode of 2 chalcones which having ring A as O-(2-(1-pyrrolidinyl)ethyl)phenol with SARS-CoV-2 M^pro

the three chalcones **C250**, **C254**, and **C264** also formed a hydrogen bond with Thr25 in the ammonium side chain. In brief, the **C250** structure, whose ring B formed an additional hydrogen bond with Gln192 through a 3-methoxy group, was the most potent substance in this phenothiazine group.

Chalcones with O-(2-(1-pyrrolidinyl)ethyl)phenol ring

C200 and **C251** were two chalcone derivatives with *O*-(2-(1-pyrrolidinyl)ethyl)phenol ring that showed the most

potent binding ability, with the Jamda scores of −2.8097 and −2.8056, respectively. These two ligands interacted with 3CL^pro through hydrogen bonds connection (Fig. 4). The ketone group of the **C251** chalcone formed three hydrogen bonds with Gly143, Ser144, and the catalytic Cys145, while Gln192 formed a hydrogen bond with fluor substituent. Meanwhile, the **C200** ligand generated a hydrogen bond with Cys145 to the O atom of the ethoxy group and two hydrogen bonds with Leu141 and Ser144 in the N atom of the pyrrolidine ring.
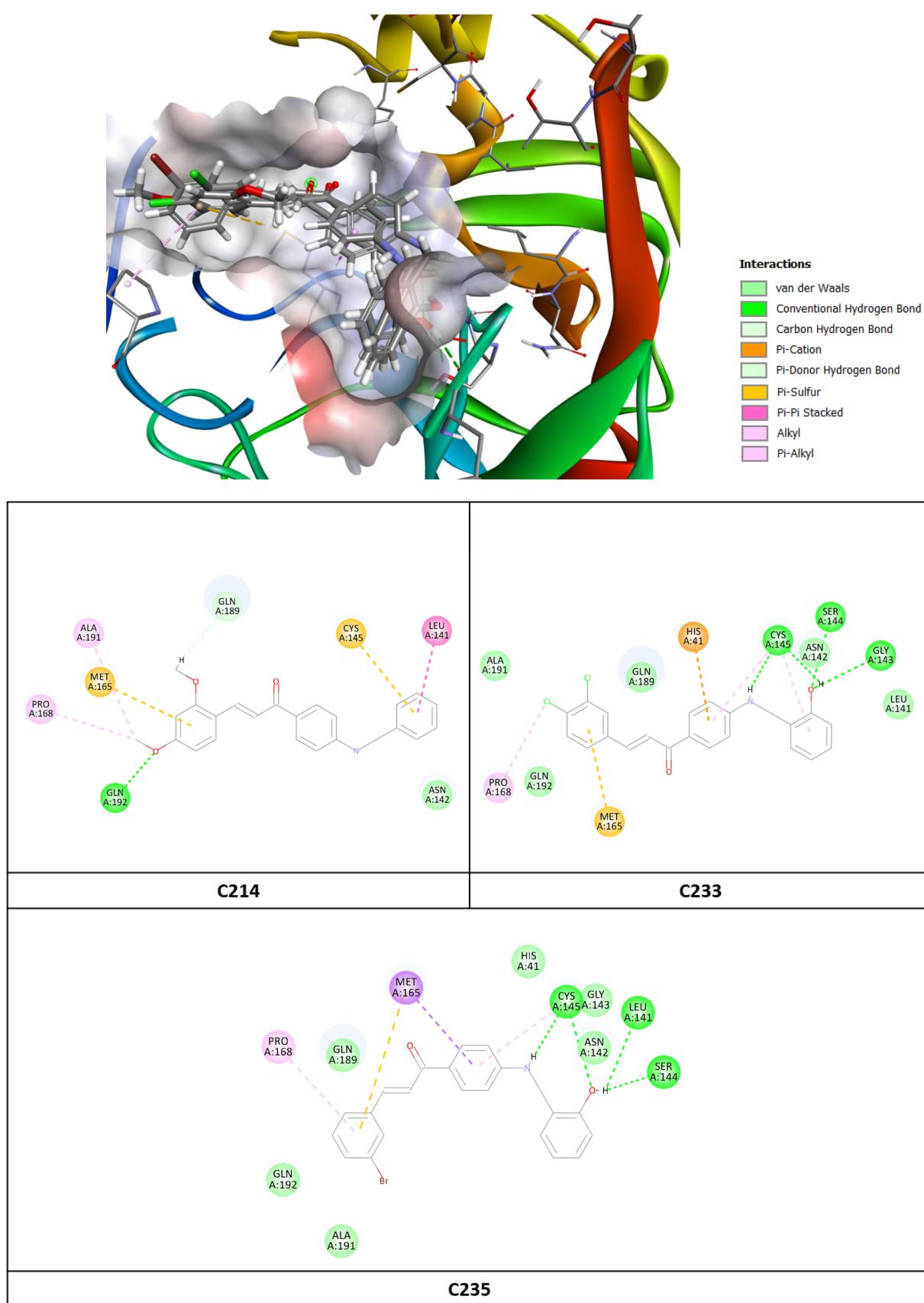
**Fig. 5** Binding mode of 3 chalcones which having ring A as N-benzylaniline with SARS-CoV-2 M^pro
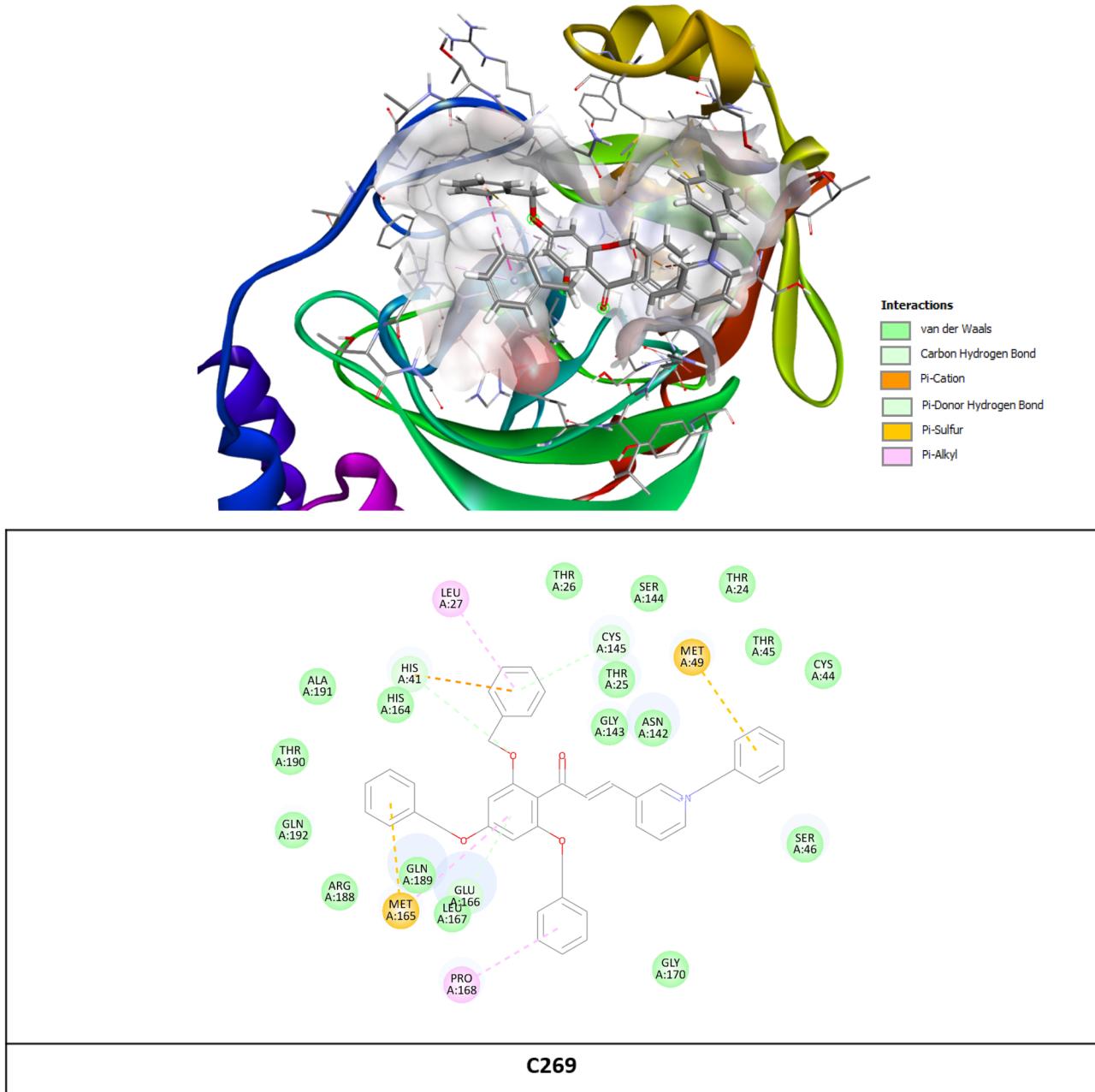
Fig. 6 Binding mode of a chalcone having ring A as tribenzyloxybenzene with SARS-CoV-2 M^pro

Chalcones with N-benzylaniline ring

The *N*-benzylaniline category included three chalcones (**C214**, **C233**, and **C235**). These chalcones shared similar binding models in the docking complex with 3CL^pro (Fig. 5), with Jamda scores of −2.7672, −2.7819, and −2.7421, respectively. The aromatic ring of the three mentioned chalcones formed π–sulfur interaction with the Met165, while *N'*-aniline and *O'*-phenol of the **C233** and **C235** formed hydrogen bonds with Cys145. Besides, the hydroxy substituent in the benzyl ring contributed additional hydrogen bonds in the three complexes. In addition, **C214** was stabilized furthermore by the hydrogen bond interaction between the methoxy substituent of the ring B and Gln192.
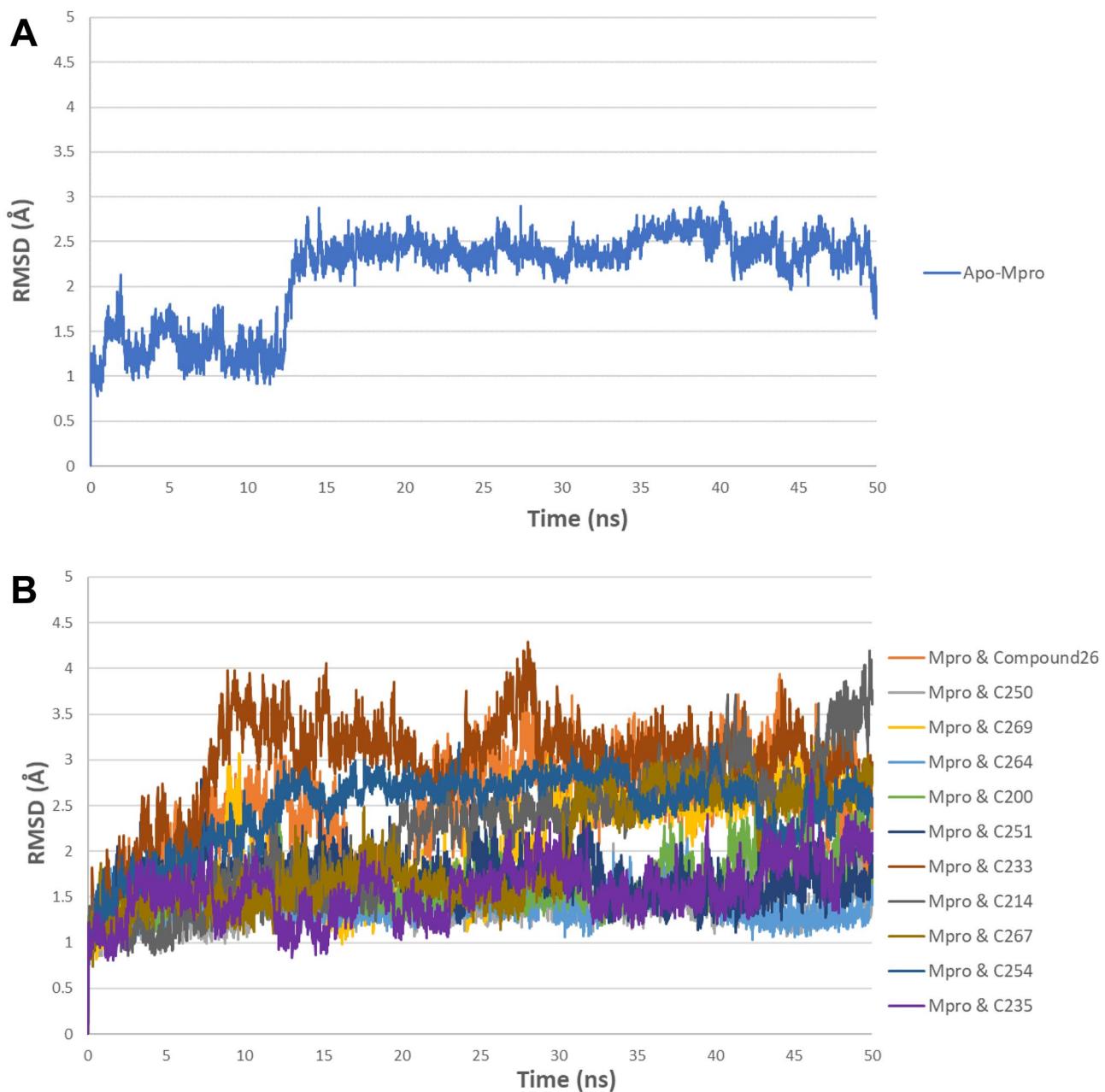
Fig. 7 Carbon backbone RMSD values of M^{pro} in apoprotein form (**A**) and in complexes with the 10 chalcones and the reference inhibitor **Compound 26** (**B**) in 50 ns of molecular dynamics simulations

Chalcones with tribenzyloxybenzene ring

C269 was the only chalcone in the top 10 substances with a tribenzyloxybenzene ring. The bulkiness of **C269** might help this ligand bind strongly to the catalytic surface of 3CL^{pro}, resulting in a good Jamda score of −2.9229 (Fig. 6). This ligand had a highly hydrophobic surface, so it was bound to the 3CL^{pro} mainly via hydrophobic connection, including van der Waals, π–sulfur, and π–cation interactions with Met165, His41, Leu27, Pro168, and Met49 residues.

Molecular dynamics simulation

Protein stability

The stability of 3CL^{pro} apoprotein and its complexes with top 10 chalcone derivatives during 50-ns simulation time was evaluated through the RMSD value (Fig. 7). Nine chalcone complexes almost reached equilibrium after 15 ns, which was earlier than apoprotein and the complex of co-crystalized ligand **Cmpd 26**. The complex with

Complex	RMSD of protein $C_{backbone}$ (Å)	SASA (nm^2)	Rg (Å)	RMSD of ligand heavy atom (Å)
Apoprotein	2.43 ± 0.16	169.57 ± 1.25	22.09 ± 0.13	
Compound 26	2.61 ± 0.52	169.45 ± 1.47	22.45 ± 0.17	1.18 ± 0.34
C250	1.43 ± 0.15	168.87 ± 1.21	22.19 ± 0.1	1.13 ± 0.2
C269	2.22 ± 0.54	169.37 ± 1.28	22.22 ± 0.17	2.89 ± 0.51
C264	1.41 ± 0.15	168.72 ± 1.26	22.15 ± 0.1	2.19 ± 0.14
C200	1.69 ± 0.26	169.33 ± 1.25	22.22 ± 0.14	1.44 ± 0.21
C251	1.68 ± 0.21	170.64 ± 1.3	22.2 ± 0.13	2.45 ± 0.34
C233	3.11 ± 0.28	170.64 ± 1.3	22.34 ± 0.13	1.79 ± 0.38
C214	2.54 ± 0.5	169.01 ± 1.19	22.06 ± 0.14	1.92 ± 0.31
C267	2.19 ± 0.51	169.01 ± 1.26	22.17 ± 0.14	2.41 ± 0.28
C254	2.65 ± 0.2	169.7 ± 1.2	22.09 ± 0.11	1.97 ± 0.56
C235	1.65 ± 0.28	169.26 ± 1.22	22.14 ± 0.15	2.95 ± 0.56

Table 2 The mean and standard deviation of protein backbone RMSD, solvent accessible surface area (SASA), radius of gyration (Rg), and ligand RMSD values calculated from the data of 50-ns MD trajectories of the main protease in apo form and in complex with the chalcones and the reference ligand Compound 26

C214 chalcone failed to stabilize, with the RMSD kept escalating after 20 ns and did not reach equilibrium at the time frame of 50 ns. Therefore, the RMSD mean value and its standard deviation were calculated from the last 35-ns time frame (Table 2). Seven structures with RMSD mean values lower than that of the apoprotein and the complex of reference ligand include **C250**, **C269**, **C264**, **C200**, **C251**, **C267**, and **C235**. Chalcones **C233** and **C254** had higher RMSD mean values than the apoprotein and the **Cmpd 26** complex. **C214** had RMSD mean value higher than the apoprotein but lower than the **Cmpd 26** complex. The RMSD plots of individual chalcone in 3CL^{pro} complexes were provided in Fig. S1.

The stability of the protein when folding was calculated through the solvent accessible surface areas (SASA) value. An increase of SASA value during the MDs means protein structural relaxation and instability. The mean value of protein backbone SASA is described in Table 2, while the SASA graph over time of each complex is illustrated in Fig. S2. In general, the presence of the chalcone ligands did not affect the protein stability since the whole SASA values were almost the same. Moreover, chalcones **C250** and **C264** even stabilize the protein, with a lower SASA value than the apoprotein at around 168 nm^2. The protein stability in complexes was reinforced in the result of the radius of gyration (Rg) (Table 2), in which the values of 10 chalcone complexes ranged from 22.06 to 22.34 Å. Compared to the Rg value of the apoprotein (22.09 ± 0.13 Å) and the complex of reference ligand (22.45 ± 0.17 Å), the figures for ten chalcone complexes were insignificantly different and even lower than that of the **Cmpd 26** reference complex. This result reveals that the protein in complexes was stable during the 50 ns of MD simulations.
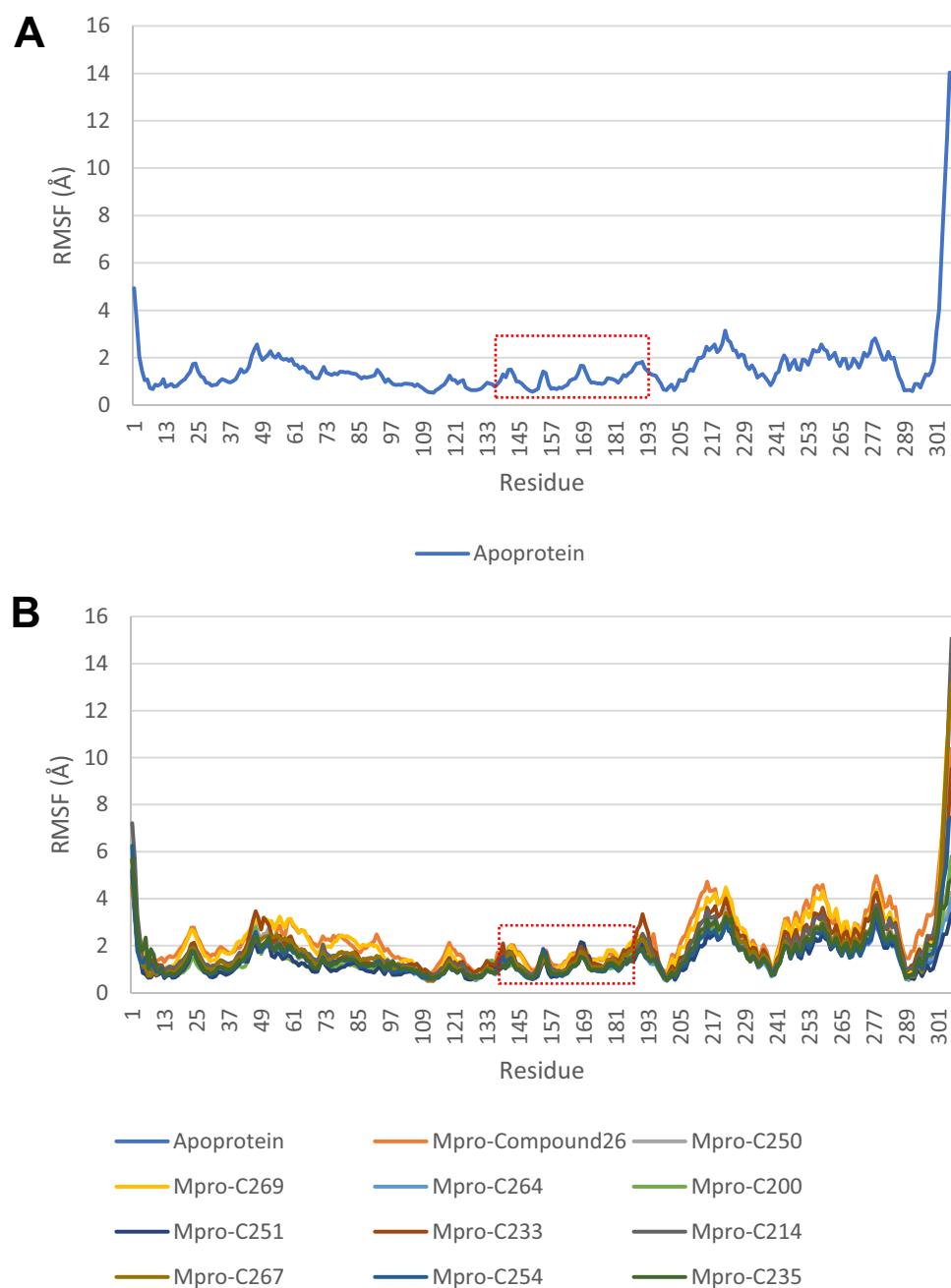
Figure 8 shows the RMSF values of residues in the apoprotein and in the complexes. The study focused on the

residues from Gly143 to Thr190 since these amino acids contributed to the binding site of 3CL^{pro}. As can be seen, these residues fluctuated around 1–2 Å in both apoprotein and chalcone–protein complexes, meaning the interaction with ligands did not affect the stability of the protein. The 10 chalcones could bind to the 3CL^{pro} in a stable during 50-ns simulation time, with a detailed RMSF plot shown in Fig. S4.

Evaluating the binding ability of top hits by long MD simulation

The MD process of the two most potential chalcones (**C264** and **C235**) was extended to 100 ns (Fig. 9). The RMSD value of both ligands started to fluctuate stably from 40 ns, with an amplitude of less than 3 Å. However, it can be seen that the ligand **C264** gained more stable fluctuation in the complex with 3CL^{pro} when compared to **C234**. Moreover, the protein's RMSD value reveals that the binding between 3CL^{pro} and **C235** becomes unstable after 65 ns, with an increase in protein instability. These results were corresponding with RMSF plotting. In the binding sites, residues from 143 to 189 were stable with the figures under 2 Å, while some residues in the 180–190 region fluctuated more unstably (> 2 Å). Regarding the radius of gyration (Rg), the figures for two complexes show downward trends with variation, especially on the **C264** complex in 70 ns. In conclusion, these two ligands **C264** and **C235** also show strong binding ability toward 3CL^{pro} of SARS-CoV-2 after a long-lasting MD process. Thus, this study suggested these ligands would be promising lead compounds for an inhibitory effect on 3CL^{pro} of SARS-CoV-2. Nevertheless, further in vitro and in vivo experiments are required to confirm the bioactivity of C264 and C235 toward 3CL^{pro} shortly.

Fig. 8 C_α RMSF values of residues of M^{pro} in apoprotein form (**A**) and in complexes with the 10 chalcones and with the reference **Compound 26** (**B**) in 50 ns of molecular dynamics simulations


Binding free energy and protein–ligand interaction

After the MDs, the 500 frames of every complex during 50 ns were used to calculate binding free energy. Figure 10 shows the MM/GBSA ΔG_{bind} values of reference and 10 chalcone complexes. Nine chalcones could bind to 3CL^{pro} with binding free energy values ranging from −20.03 to −28.24 kcal/mol (Table 3). As shown in Fig. 8, MM/GBSA ΔG_{bind} of the **C214** chalcone kept escalating during the simulation. From the 42 ns to 46 ns, the ΔG_{bind} was stable at 0 kcal/mol. This data suggested that the **C214** was not stable in the binding site and completely free from the

protease. This assumption was then confirmed by using VMD to visualize the simulation process. Regarding other substances, **C264** would be the most potent inhibitor since this chalcone had ΔG_{bind} values of −28.24 ± 3.53 kcal/mol in the complex with 3CL^{pro}, which was even lower than that of the reference ligand **Cmpd 26** (−27.64 ± 3.59 kcal/mol). Therefore, the interaction between **C264** and 3CL^{pro} would be further analyzed. The individual MM/GBSA binding free energy of 10 chalcones and **Compound 26** with the 3CL^{pro} can be found at Fig. S5.

During 50-ns simulation, chalcone **C264** bound stably to the most interactive residues of 3CL^{pro}, including His41

A Protein backbone RMSD

B Ligand RMSD

C Protein C$_\alpha$ RMSF

D Radius of gyration

**E** Solvent accessible surface
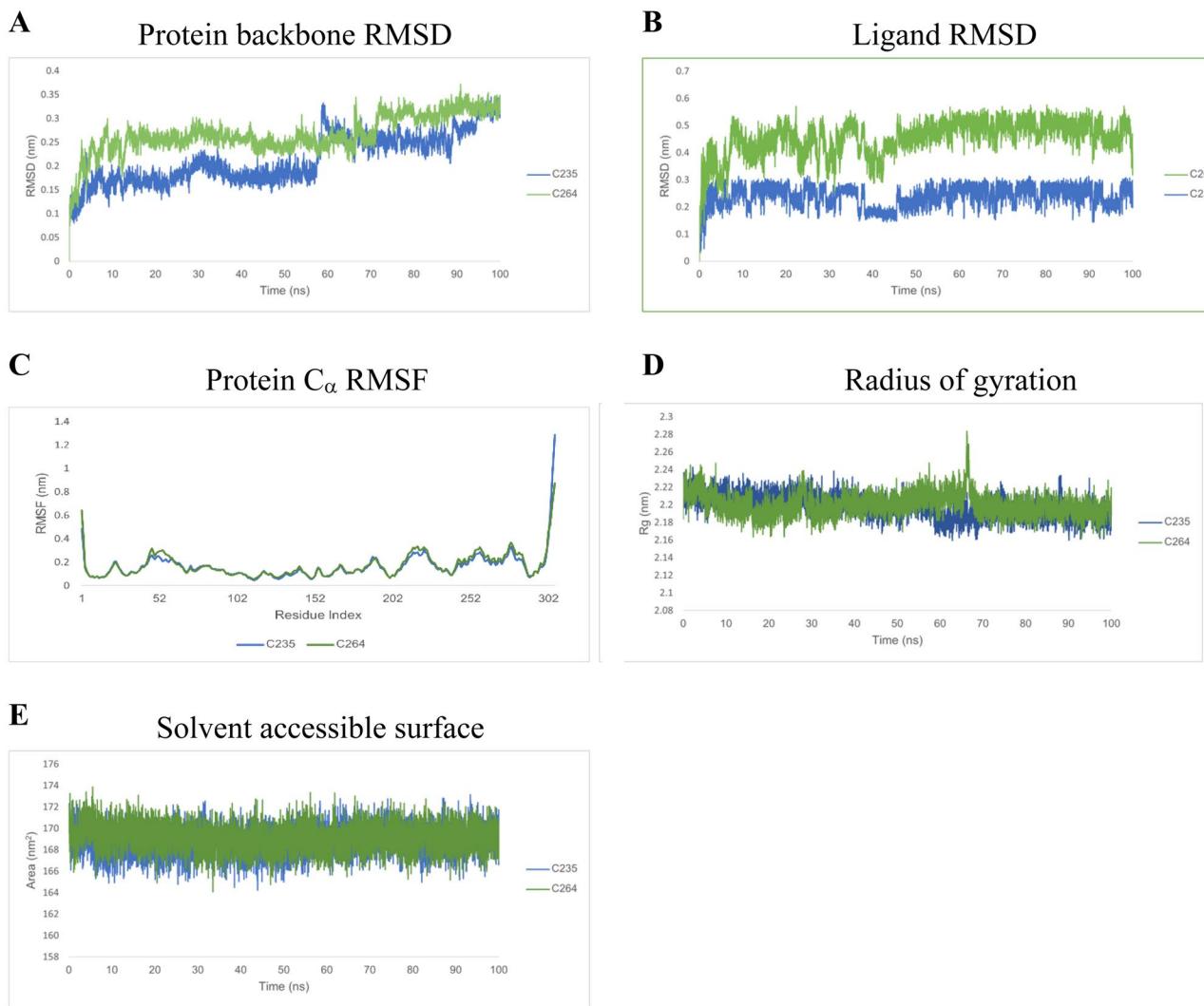
Fig. 9 A Protein carbon backbone RMSD and B ligand heavy atom RMSD in complexes of the top two hits C235 and C264 calculated using data MD simulations trajectories in 100 ns. C Carbon alpha RMSF values of the 3CL-pro in complexes with the top two hits C235 and C264 from the same MD trajectories. D Radius of gyration and E solvent accessible surface of the 3CL-pro in complexes with the top two hits C235 and C264 calculated using data of 100-ns trajectories of MD simulations

(55.97%), Met49 (74.49%), Pro168 (47.53%), and Gln189 (64.00%). The mean number of hydrogen bonds between **C264** and the 3CL^{pro} was 0.2659 ± 0.6255 per frame, nearly the same as the one of the **Cmpd 26** reference complex (0.2847 ± 0.5586 H-bond per frame). It indicates that **C264** is bound to the 3CL^{pro} mainly via hydrophobic interactions. Residue His41 and Met49 interacted with the chalcone mainly via surface and arene interactions. Residue Met49 also connect with the ligand C264 via hydrogen acceptor of NH$_4^+$ group. Gln189 was the most multifunctional residue. It served as both hydrogen donor and hydrogen acceptor and created hydrophobic interaction with the ligand. During the simulation, there were some other hydrogen bonds between the side chain NH$_4^+$ or the ketone group (C=O) of the ligand and the 3CL^{pro}. While the ketone group received hydrogen bonds from Glu166, Gln189, Thr190, and Gln192 with the frequency of 8.00%, 2.68%, 3.46%, and 0.92% in order, the side chain NH$_4^+$ group donored hydrogen to Ser46 and Gln189 with the frequency of 7.06% and 4.08%, respectively. There was another hydrogen bond between the catalytic Cys145 and the S heteroatom in the phenothiazine ring of this chalcone with a frequency of 0.34%. The interaction heat map, the MM/GBSA binding free energy of C264 with 3CL^{pro}, and the frequency of hydrogen bond can be observed in Fig. 11.

Fig. 10 MM/GBSA ΔG_{bind} energy calculated during the 50-ns MD simulation between the ten chalcones or the reference **Compound 26** with SARS-CoV-2 M^{pro}
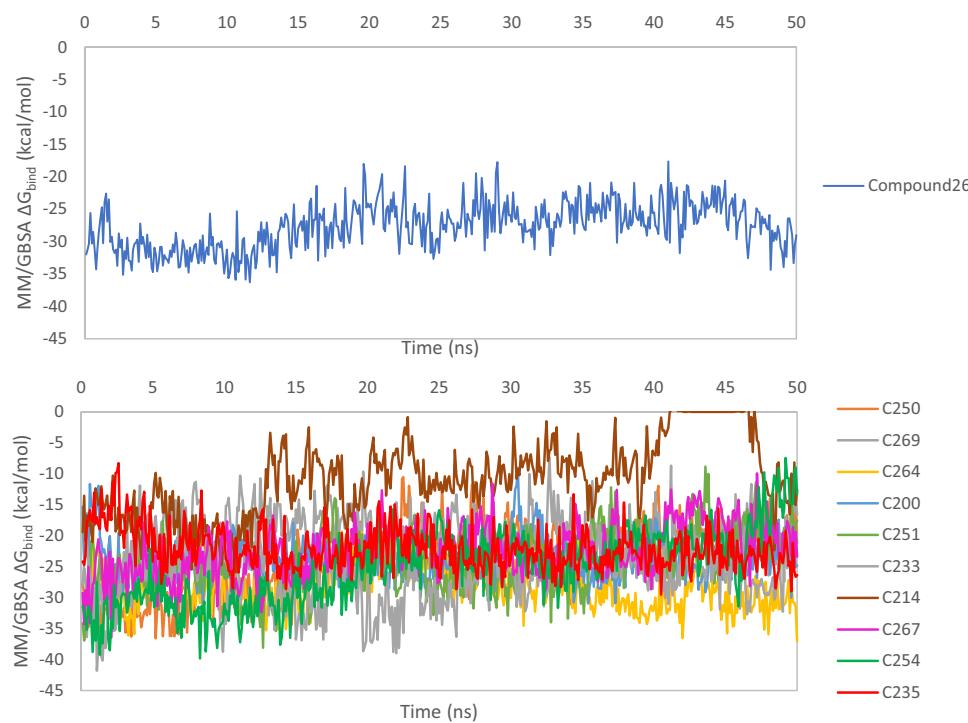
Table 3 Binding free energies of the top ten compounds and the reference **Compound 26** bound to the SARS-CoV-2 M^{pro} estimated by MM/GBSA approach from the data of the 50-ns MDs trajectories

| Ligands	ΔE_{vdW} (kcal/mol)	ΔE_{ele} (kcal/mol)	ΔG_{GB} (kcal/mol)	ΔG_{SA} (kcal/mol)	ΔG_{gas} (kcal/mol)	ΔG_{solv} (kcal/mol)	ΔG_{bind} (kcal/mol)
Compound 26	-40.82 ± 4.08	-93.30 ± 17.60	111.54 ± 17.63	-5.05 ± 0.51	-134.13 ± 19.75	16.49 ± 17.29	-27.64 ± 3.59
C250	-34.66 ± 6.23	-104.49 ± 13.56	120.00 ± 14.17	-4.53 ± 0.95	-139.15 ± 16.85	115.46 ± 13.61	-23.68 ± 5.43
C269	-35.61 ± 5.65	-89.31 ± 14.91	109.71 ± 14.09	-4.81 ± 0.71	-124.93 ± 15.57	104.89 ± 13.94	-20.03 ± 4.38
C264	-37.92 ± 3.66	-90.04 ± 14.12	104.41 ± 14.25	-4.70 ± 0.47	-127.95 ± 15.72	99.71 ± 14.01	-28.23 ± 3.53
C200	-30.62 ± 3.09	-159.67 ± 11.44	170.82 ± 10.54	-4.55 ± 0.29	-190.30 ± 11.45	166.27 ± 10.48	-24.04 ± 3.33
C251	-35.27 ± 5.37	-67.71 ± 14.49	83.68 ± 15.58	-4.91 ± 0.76	-102.99 ± 18.52	78.76 ± 14.98	-24.22 ± 4.61
C233	-36.76 ± 3.84	-16.14 ± 10.34	30.83 ± 7.38	-4.84 ± 0.46	-52.91 ± 10.59	25.99 ± 7.32	-26.91 ± 4.55
C214	-16.88 ± 9.13	-4.71 ± 4.56	13.24 ± 7.44	-2.39 ± 1.24	-21.59 ± 11.85	10.85 ± 6.39	-10.74 ± 6.23
C267	-33.08 ± 5.32	-78.87 ± 10.96	93.67 ± 11.77	-4.11 ± 0.67	-111.96 ± 13.25	89.55 ± 11.45	-22.41 ± 4.20
C254	-33.93 ± 7.20	-79.79 ± 12.94	92.69 ± 14.56	-4.18 ± 0.89	-113.73 ± 18.62	88.52 ± 13.86	-25.20 ± 5.87
C235	-26.98 ± 2.71	-24.41 ± 4.96	33.04 ± 3.77	-3.72 ± 0.29	-51.39 ± 5.98	29.33 ± 3.61	-22.07 ± 3.39
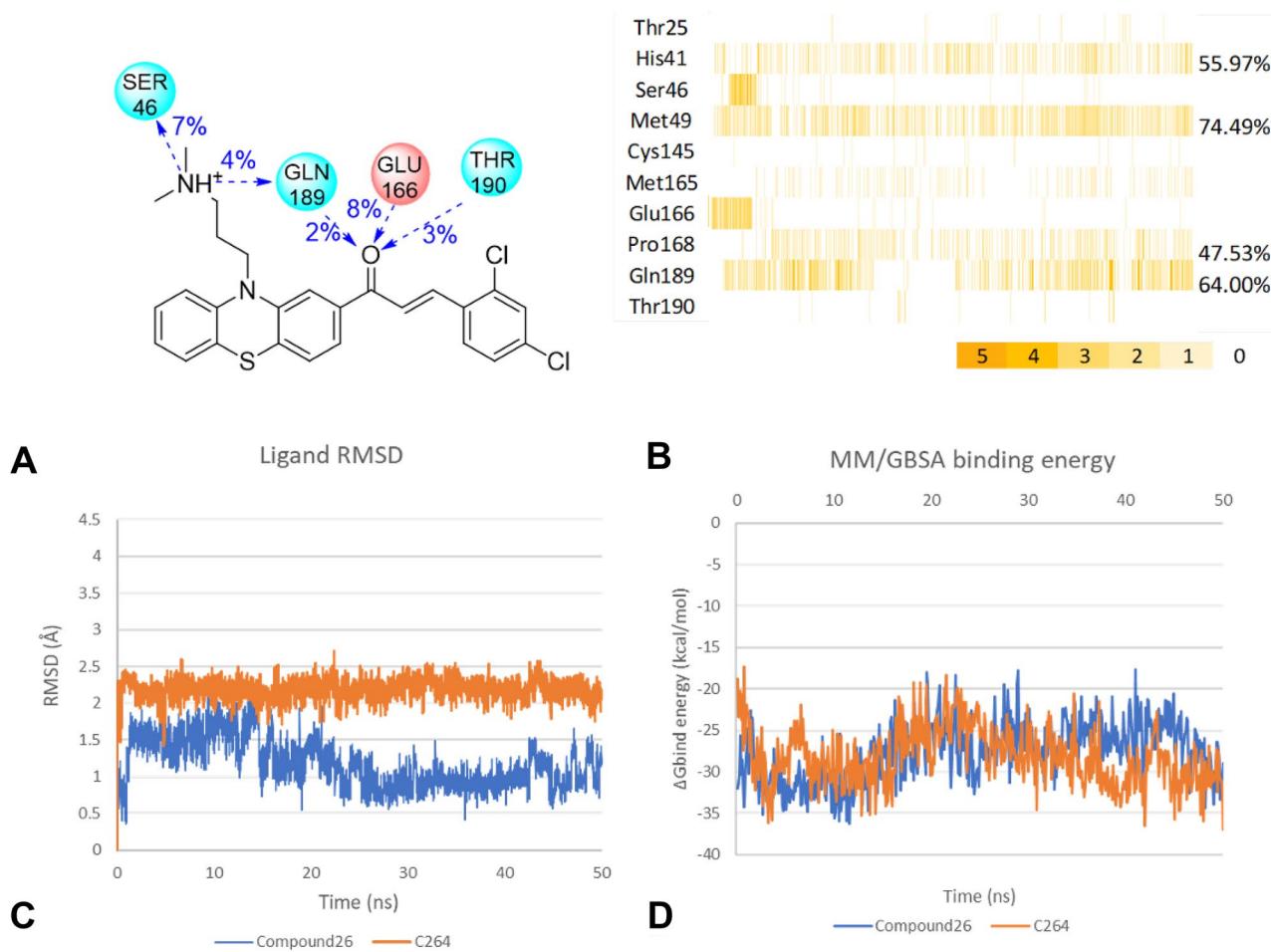
Fig. 11 Occupancy percentage of hydrogen bonds formed between M^{pro} and the top hit C264 (**A**) in 50 ns of molecular dynamics simulation; interaction heat map between M^{pro} with the C264 (**B**); ligand heavy atoms RMSD (**C**); and MM/GBSA ΔG_{bind} value (**D**) of C264 (orange) and **Compound 26** (blue) during the 50-ns MD simulation

Besides **C264**, **C235** would probably be a 3CL^{pro} inhibitor. This chalcone had the ΔG_{bind} of only -22.07 ± 3.39 kcal/mol (Table 3). This outcome could be possibly affected by the high binding free energy of the initial 10-ns time frame. In the last 10 ns of simulation, this figure tended to decrease and reached -26.36 kcal/mol, proving this chalcone stable inside the binding site. In addition, the **C235** complex had the highest number of hydrogen bonds per frame with a mean value of 1.94 ± 0.55. The O atom of the phenyl ring is formed with the His163 by the remarkably high frequency of 94.56%, while the central N atom of the ligand also received a hydrogen bond from Glu166 with the frequency of 73.35%. Besides that, Glu166 also interacted with the ligand via surface and arene attraction contacts. The interaction profile between **C235** and the protease was also analyzed using the PLIF tool. As a result, this complex witnessed similar interaction of His163, Glu166, Gln189, and Asp142 at 90.62%, 87.83%, 44.11%, and 20.96%, respectively. Although having a high contact frequency with the 3CL^{pro}, the **C235** structure was quite flexible inside the binding site. This figure can be observed in the ligand RMSD chart. All the mentioned data of the **C235** complex during the 50-ns simulation can be found in Fig. 12.
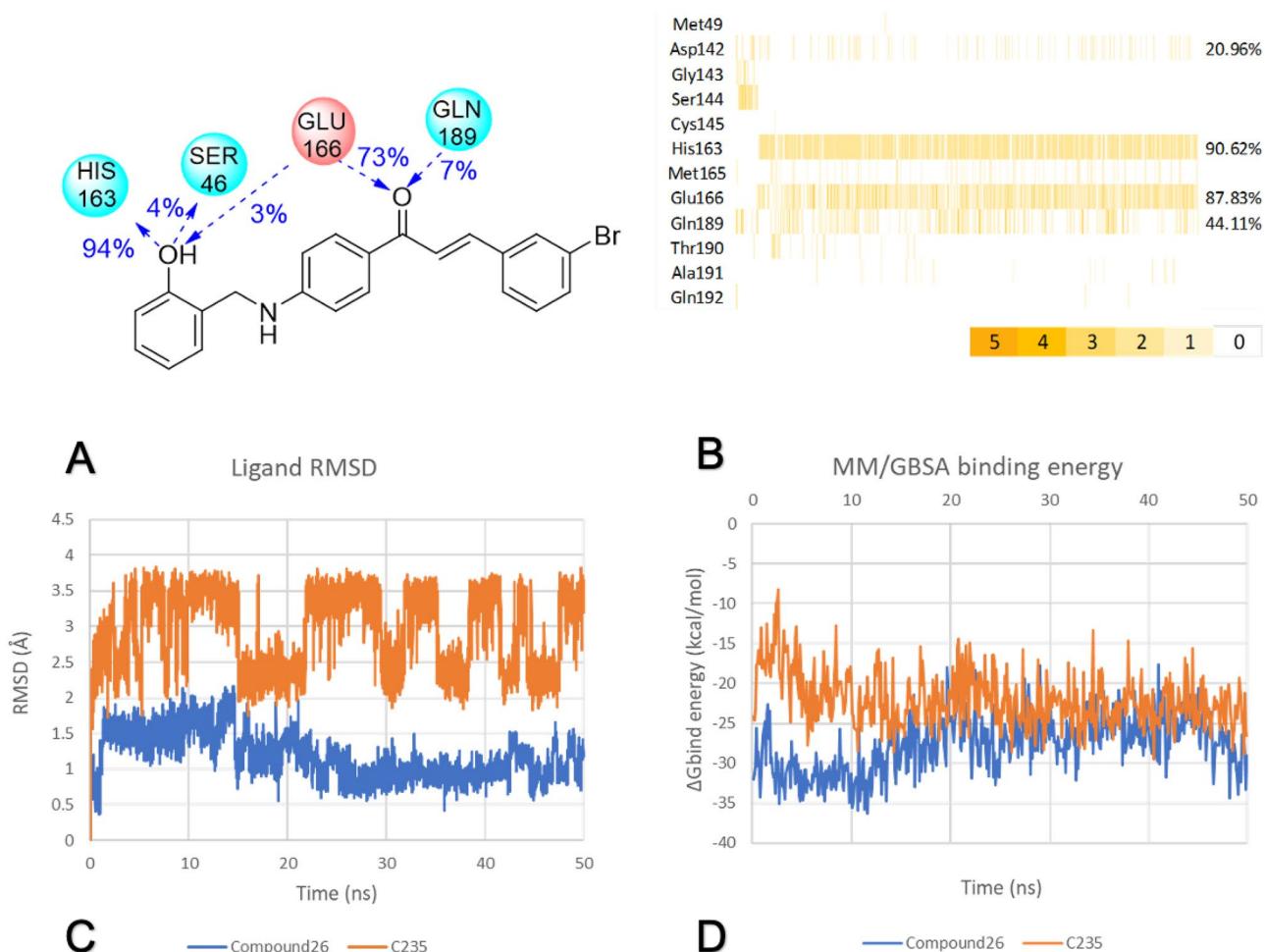
Fig. 12 Occupancy percentage of hydrogen bonds formed between M^{pro} and the hit C235 (**A**) in 50 ns of molecular dynamics simulation; interaction heat map between M^{pro} with the C235 (**B**); ligand heavy atoms RMSD (**C**); and MM/GBSA ΔG_{bind} value (**D**) of C235 (orange) and **Compound 26** (blue) during the 50-ns MD simulation

ADME evaluation

The ADME properties of the top ten chalcones were predicted by the SwissADME server (Table 4). Based on the preliminary criteria, there was no chalcone that satisfies all models provided by the SwissADME server. However, based on the most traditional rule of five proposed by Lipinski et al. [25], nine chalcones could become an oral drug with respect to bioavailability. The absorption ability of the ten hits was also evaluated using the BOILED-Egg model (Fig. S6) [31]. The two top hits **C264** and **C251** were predicted to have the ability to penetrate the blood–brain barrier to get inside the brain. The result for the high Brenk alert point observed in Table 4 because Michael acceptor in chalcone structures was 1 of 105 alert fragments proposed by Brenk et al. [30]. After the ADME evaluation, **C264** was still the most potent substance as it was predicted to inhibit only CYP2C9, and chemical optimization can be conducted to resolve this problem.

Table 4 ADME properties of the top 10 chalcones retrieved from the SwissADME server

ADME properties	PAINS	Drug-likeness					Brenk (alert)	CYP1A2	Cytochrom P450 inhibitor				
		Lipinski (Pfizer)	Ghose (Amgen)	Veber (GSK)	Egan (Pharmacia)	Muegge (Bayer)			CYP2C19	CYP2C9	CYP2D6	CYP3A4
C200	0	0	0	0	0	0	1	No	No	No	No	No
C214	0	0	0	0	0	1	1	No	Yes	Yes	Yes	Yes
C233	1	1	1	0	0	1	1	Yes	Yes	Yes	No	Yes
C235	1	1	0	0	0	1	1	Yes	Yes	Yes	Yes	Yes
C250	0	0	1	0	0	1	1	No	Yes	No	No	No
C251	0	0	1	0	0	1	1	No	No	Yes	No	No
C254	0	0	2	0	1	1	1	No	Yes	Yes	No	No
C264	0	0	3	0	1	1	1	No	No	No	No	No
C267	0	1	3	0	1	1	1	No	Yes	No	No	No
C269	0	2	4	1	1	2	2	No	No	No	No	Yes

Conclusion

This study used 269 in-house chalcone derivatives to screen for potential 3CLpro inhibitors by the in silico approach. Firstly, the study used the docking method to evaluate the binding ability of these chalcones toward the main protease of SARS-CoV-2, with the docking results optimized by JamdaScorer [18]. The top 10 structures were then subjected to molecular dynamics simulation, with the stability of the protein–ligand complex evaluated through RMSD and RMSF values. After that, the binding free energy of the top hits was calculated, and the pharmacokinetics of potential substances were also predicted. As a result, the structures **C264** and **C235** have been identified as promised candidates to inhibit the main protease of SARS-CoV-2. These selected compounds showed strong biding ability toward 3CLpro via hydrophobic and hydrogen bond interactions, with Jamda score of -2.9229 and -2.7421, respectively. The binding free energy values of **C264** and **C235** were respective at -28.23 ± 3.53 kcal/mol and -22.07 ± 3.39. These lead compounds meet the ADME requirements, but the structures must be modified in the next step. Further in vitro and in vivo experiments are also required to confirm the inhibitory ability of the two chalcone structures toward 3CLpro of SARS-CoV-2.

Code availability Not applicable.

Declarations

Conflict of interest The authors declare no competing interests.

References

1. Wu F, Zhao S, Yu B et al (2020) A new coronavirus associated with human respiratory disease in China. Nature 579:265–269
2. WHO (2021) WHO Coronavirus (COVID-19) Dashboard. https://covid19.who.int/. Accesed 13 Dec 2021
3. FDA (2021) FDA approves first treatment for COVID-19. https://www.fda.gov/news-events/press-announcements/fda-approves-first-treatment-covid-19. Accesed 13 Dec 2021
4. Mahase E (2021) Covid-19: Pfizer's paxlovid is 89% effective in patients at risk of serious illness, company reports. BMJ 375:n2713
5. Gil C, Ginex T, Maestro I et al (2020) COVID-19: drug targets and potential treatments. J Med Chem 63:12359–12386
6. Zhang L, Lin D, Sun X et al (2020) Crystal structure of SARS-CoV-2 main protease provides a basis for design of improved α-ketoamide inhibitors. Science 368:409–412
7. Abdusalam AAA, Murugaiyah V (2020) Identification of potential inhibitors of 3CL protease of SARS-CoV-2 from ZINC database by molecular docking-based virtual screening. Front Mol Biosci 7
8. Sattari A, Ramazani A, Aghahosseini H (2021) Repositioning therapeutics for COVID-19: virtual screening of the potent synthetic and natural compounds as SARS-CoV-2 3CLpro inhibitors. J Iran Chem Soc 18:2807–2827
9. Jiménez-Avalos G, Vargas-Ruiz AP, Delgado-Pease NE et al (2021) Comprehensive virtual screening of 4.8k flavonoids reveals novel insights into allosteric inhibition of SARS-CoV-2 MPRO. Sci Rep 11:15452
10. Qazi S, Das S, Khuntia BK et al (2021) In silico molecular docking and molecular dynamic simulation analysis of phytochemicals from Indian foods as potential inhibitors of SARS-CoV-2 RdRp and 3CLpro. Nat Prod Commun 16:1934578X211031707
11. Xu Z, Yang L, Zhang X et al (2020) Discovery of potential flavonoid inhibitors against COVID-19 3CL proteinase based on virtual screening strategy. Front Mol Biosci 7:556481
12. Mathpal S, Joshi T, Sharma P et al (2022) Assessment of activity of chalcone compounds as inhibitors of 3-chymotrypsin like protease (3CLPro) of SARS-CoV-2: In silico study. Struct Chem
13. Panche AN, Diwan AD, Chandra SR (2016) Flavonoids: an overview. J Nutr Sci 5:e47–e47
14. Elmer P (2021) CHEMDRAW - the gold standard for communicating chemistry research. https://perkinelmerinformatics.com/products/research/chemdraw/. Accessed 13 Dec 2021
15. Chemical Computing Group ULC (2020) Molecular operating environment (MOE), 2019.01
16. Zhang C-H, Stone EA, Deshmukh M et al (2021) Potent non-covalent inhibitors of the main protease of SARS-CoV-2 from molecular sculpting of the drug perampanel guided by free energy perturbation calculations. ACS Cent Sci 7:467–475
17. Trott O, Olson AJ (2010) AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. J Comput Chem 31:455–461
18. Flachsenberg F, Meyder A, Sommer K et al (2020) A consistent scheme for gradient-based optimization of protein-ligand poses. J Chem Inf Model 60:6502–6522
19. Daina A, Michielin O, Zoete V (2017) SwissADME: a free web tool to evaluate pharmacokinetics, drug-likeness and medicinal chemistry friendliness of small molecules. Sci Rep 7:42717
20. Abraham MJ, Murtola T, Schulz R et al (2015) GROMACS: high performance molecular simulations through multi-level parallelism from laptops to supercomputers. SoftwareX 1–2:19–25
21. Zoete V, Cuendet MA, Grosdidier A et al (2011) SwissParam: a fast force field generation tool for small organic molecules. J Comput Chem 32:2359–2368
22. Valdés-Tresanco MS, Valdés-Tresanco ME, Valiente PA et al (2021) gmx_MMPBSA: a new tool to perform end-state free energy calculations with GROMACS. J Chem Theory Comput 17:6281–6291
23. Reva BA, Finkelstein AV, Skolnick J (1998) What is the probability of a chance prediction of a protein structure with an rmsd of 6 A? Fold Des 3:141–147
24. Sundar S, Thangamani L, Manivel G et al (2019) Molecular docking, molecular dynamics and MM/PBSA studies of FDA approved drugs for protein kinase a of Mycobacterium tuberculosis; application insights of drug repurposing. Inform Med Unlocked 16:100210
25. Lipinski CA, Lombardo F, Dominy BW et al (2001) Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. Adv Drug Deliv Rev 46:3–26
26. Ghose AK, Viswanadhan VN, Wendoloski JJ (1999) A knowledge-based approach in designing combinatorial or medicinal chemistry libraries for drug discovery. 1. A qualitative and quantitative characterization of known drug databases. J Comb Chem 1:55–68
27. Veber DF, Johnson SR, Cheng HY et al (2002) Molecular properties that influence the oral bioavailability of drug candidates. J Med Chem 45:2615–2623
28. Egan WJ, Merz KM Jr, Baldwin JJ (2000) Prediction of drug absorption using multivariate statistics. J Med Chem 43:3867–3877
29. Muegge I, Heald SL, Brittelli D (2001) Simple selection criteria for drug-like chemical matter. J Med Chem 44:1841–1846
30. Brenk R, Schipani A, James D et al (2008) Lessons learnt from assembling screening libraries for drug discovery for neglected diseases. ChemMedChem 3:435–444
31. Daina A, Zoete V (2016) A BOILED-Egg to predict gastrointestinal absorption and brain penetration of small molecules. ChemMedChem 11:1117–1121