



Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.



Discovering potent inhibitors against the Mpro of the SARS-CoV-2. A medicinal chemistry approach

Aamir Mehmood^{a,b}, Sadia Nawab^c, Yanjing Wang^{d,**}, Aman Chandra Kaushik^{a,***}, Dong-Qing Wei^{a,b,*}

^a Department of Bioinformatics and Biological Statistics, School of Life Sciences and Biotechnology, Shanghai Jiao Tong University, Shanghai, 200240, PR China

^b Peng Cheng Laboratory, Shenzhen, Guangdong, 518055, China

^c State Key Laboratory of Microbial Metabolism and School of Life Sciences and Biotechnology, Shanghai Jiao Tong University, 800 Dongchuan Road, Shanghai, 200240, PR China

^d Engineering Research Center of Cell & Therapeutic Antibody, School of Pharmacy, Shanghai Jiao Tong University, Shanghai 200240, PR China

ARTICLE INFO

Keywords:

COVID-19

Mpro

MD Simulation

Respiratory syndrome

Virtual screening

ABSTRACT

The global pandemic caused by a single-stranded RNA (ssRNA) virus known as the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) is still at its peak, with new cases being reported daily. Although the vaccines have been administered on a massive scale, the frequent mutations in the viral gene and resilience of the future strains could be more problematic. Therefore, new compounds are always needed to be available for therapeutic approaches. We carried out the present study to discover potential drug compounds against the SARS-CoV-2 main protease (Mpro). A total of 16,000 drug-like small molecules from the ChemBridge database were virtually screened to obtain the top hits. As a result, 1032 hits were selected based on their docking scores. Next, these structures were prepared for molecular docking, and each small molecule was docked into the active site of the Mpro. Only compounds with solid interactions with the active site residues and the highest docking score were subjected to molecular dynamics (MD) simulation. The post-simulation analyses were carried out using the in-built GROMACS tools to gauge the stability, flexibility, and compactness. Principal component analysis (PCA) and hydrogen bonding were also calculated to observe trends and affinity of the drugs towards the target. Among the five top compounds, C1, C3, and C6 revealed strong interaction with the target's active site and remained highly stable throughout the simulation. We believe the predicted compounds in this study could be potential inhibitors in the natural system and can be utilized in designing therapeutic strategies against the SARS-CoV-2.

1. Introduction

The current outbreak of the severe acute respiratory syndrome known as coronavirus 2 (SARS-CoV-2) has been declared as a pandemic by the World Health Organization (WHO) and is still a severe threat to public health [1,2]. It is observed phylogenetically that the SARS-CoV-2 is highly associated with the previous coronaviruses SARS-CoV and MERS-CoV, infecting millions of people all-inclusive [3,4]. As of November 20, 2021, the global metrics have confirmed 257,090,259

infections and 5,158,478 deaths caused by the SARS-CoV-2, which is an alarming upsurge in the number of infections if compared with the previous data [5]. Furthermore, the novel coronavirus has a lower mortality rate, but the transmission rate is relatively high when equated with the earlier coronaviruses [6].

The SARS-CoV-2 infections could be asymptomatic, mildly symptomatic, or highly symptomatic, depending on the physiological conditions of an individual. Usually, the symptoms involve high temperature, shortness of breath/difficulty in breathing, myalgia, coughing, and

* Corresponding author. Department of Bioinformatics and Biological Statistics, School of Life Sciences and Biotechnology, Shanghai Jiao Tong University, Shanghai, 200240, PR China.

** Corresponding author. Engineering Research Center of Cell & Therapeutic Antibody, School of Pharmacy, Shanghai Jiao Tong University, Shanghai 200240, China.

*** Corresponding author.

E-mail address: dqwei@sjtu.edu.cn (D.-Q. Wei).

<https://doi.org/10.1016/j.combiomed.2022.105235>

Received 27 August 2021; Received in revised form 12 January 2022; Accepted 12 January 2022

Available online 26 January 2022

0010-4825/© 2022 Elsevier Ltd. All rights reserved.

radiological signs of ground-glass lung opacity well-matched with anomalous pneumonia in most patients [7–9]. Upon infection, the person remains asymptomatic but could be a carrier and might infect another individual. Therefore, a 14–16 days isolation is advised upon contact with the infected area or person.

The SARS-CoV-2 has a genome of ~30 kb that encodes the entire viral proteome, including structural, non-structural, and additional proteins. The structural proteins are nucleocapsid (N), matrix (M), spike (S), and small envelope (E) protein. Together with additional proteins, they are translated by 33.33% of the viral genome, while the remaining 66.66% of the genome, known as the ORF1a/b region, corresponds to the non-structural proteins [7–9].

The ORF1 (a and b) are the two non-structural proteins (polypeptides pp1a and pp1ab) encoding genes that form the RTC (replication transcription complex). The pp1a and pp1ab polypeptides, once translated, are then proteolytically cleaved by other two viral proteases, which are known as PLpro (papain-like protease) and 3CLpro (3-chymotrypsin-like protease) or main protease (Mpro). PLpro is responsible for cleaving non-structural proteins (nsp 1–3). At the same time, the 3CLpro bisects the polyprotein at 11 specific sites downstream of nsp4 to produce various non-structural proteins, which are vital for the viral cycle [10].

There have been announcements of the SARS-CoV-2 vaccines, while other researchers have reported drugs that could be proven effective. The vaccines announced could be effective in some regions and the drugs reported may or not reach the clinical trial. This is because such viral infections are pretty complicated, and they can quickly develop resistance to the drugs because of the fast mutation rate. Therefore, a new drug that could strongly bind with the virus and aid in the inhibition process is always needed.

Computer-aided drug design has proven to be quite fruitful in designing therapeutic strategies. It has helped us construct or discover molecules that could effectively cope with the desired target. In the current COVID-19 pandemic, medicinal chemists and drug designers have attempted to target several of the above structural proteins for inhibition purposes. Among these structural targets, one of the proteins is the main protease (Mpro) [11,12]. As mentioned above, it digests the pp1a and pp1b (L-Q* (S, A, G) (* is the cleave shows the cleavage spot)), which are essential polyproteins for the SARS-CoV-2 transcription and translation [13]. Therefore, the functional activity inhibition of the Mpro provides a reasonable chance of producing an effective drug against the SARS-CoV-2 infection [14].

The Mpro crystallographic structure has already been deposited in the protein databank in its apo form (PDB: 6M03). Its structural characterization reveals 96.1% of sequence similarity to the previous SARS-CoV, containing a decidedly conserved manner of the catalytic binding spot. This protease has an active site arranged in the form of a homodimer where the binding groove exists in the two domains (domain one and domain two) [15]. Besides, several crystallographic structures have been reported for the COVID-19 Mpro coupled with various inhibitors, which ultimately provided a clue of hotspots and communicating residues as part of protease catalytic role in its leading protease spot and the inhibition process.

The computer-aided drug design (CADD) approaches have shortened the timeline and lowered the expense to design drugs that could reach the clinical trial compared to those developed via traditional techniques [16–18].

Hence, we benefited from the reported crystal structure for the SARS-CoV-2 Mpro to carry out vanguard computational analysis for discovering ways of Mpro inhibition. The current work practices a multi-step *in silico* pipeline to discover drug-like compounds against the SARS-CoV-2 Mpro via virtual screening, molecular docking, and re-docking tailed by molecular dynamics simulation and free energy estimations. The outcomes obtained from these analyses could be of high importance for sure. They would be beneficial in designing an effective therapeutic strategy against the ongoing pandemic, aiming to save millions of lives.

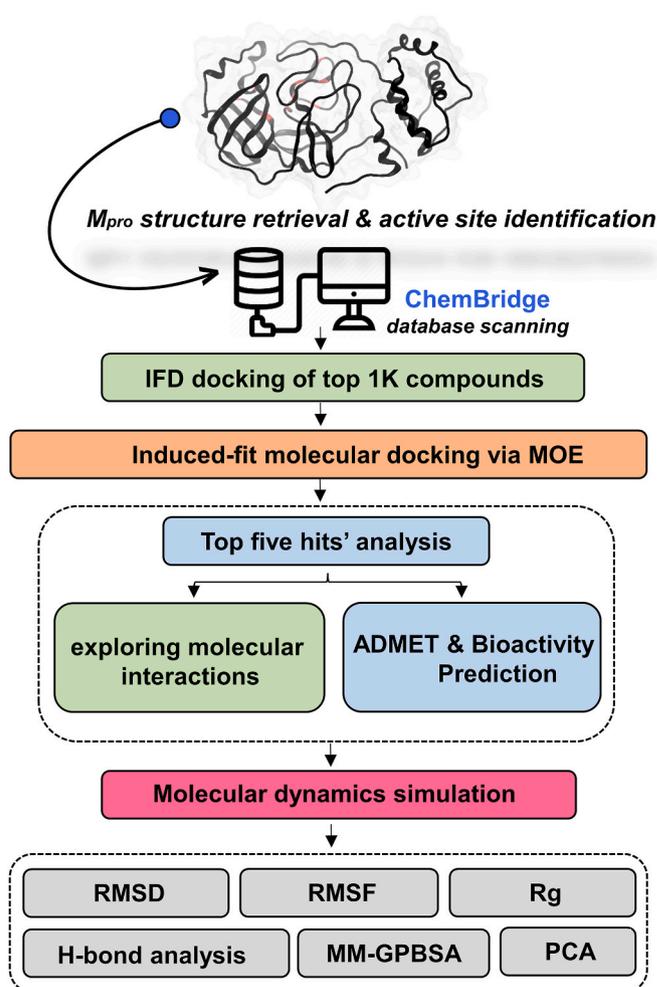


Fig. 1. This figure shows the entire workflow of our study.

2. Materials and methods

Target protein's retrieval and preparation The SARS-CoV-2 crystal structure resolved via X-Ray diffraction approach at a resolution of 2.00 Å in its apo form (PDB ID: 6M03) was retrieved from the protein databank. We used the structure preparation tool in the Molecular Operating Environment (MOE, 2018) for water stripping, charge adjustments, and 3D protonation [19–21]. The structure was minimized as well via the minimization algorithm in MOE. The overall conformation was keenly observed to assure the lack of any unintended structural defects. The overall pipeline of this study is presented as a flowchart in Fig. 1.

2.1. Molecular docking and selection criteria

There are billions of compounds, each having its features, which may or may not be similar to another compound in its physiochemical features and targets. Some might follow Lipinski's rule of five while others may not; however, every chemical compound has its own significance in medicinal chemistry. Computational molecular docking is an effective technique that enables us to see if a particular molecule will bind to the desired target or not in real-time.

In the current work, we considered scanning ~16,000 drug-like small molecules from the ChemBridge database being docked into the Mpro active site that involves Thr24, Leu27, His41, Phe140, Cys145, His163, Met165, Pro168, and His172 as reported by Wu et al., 2020 (Fig. 2) [22]. The docking parameters were slightly tweaked as the refinement was set to Induced Fit, and both the scoring algorithms were

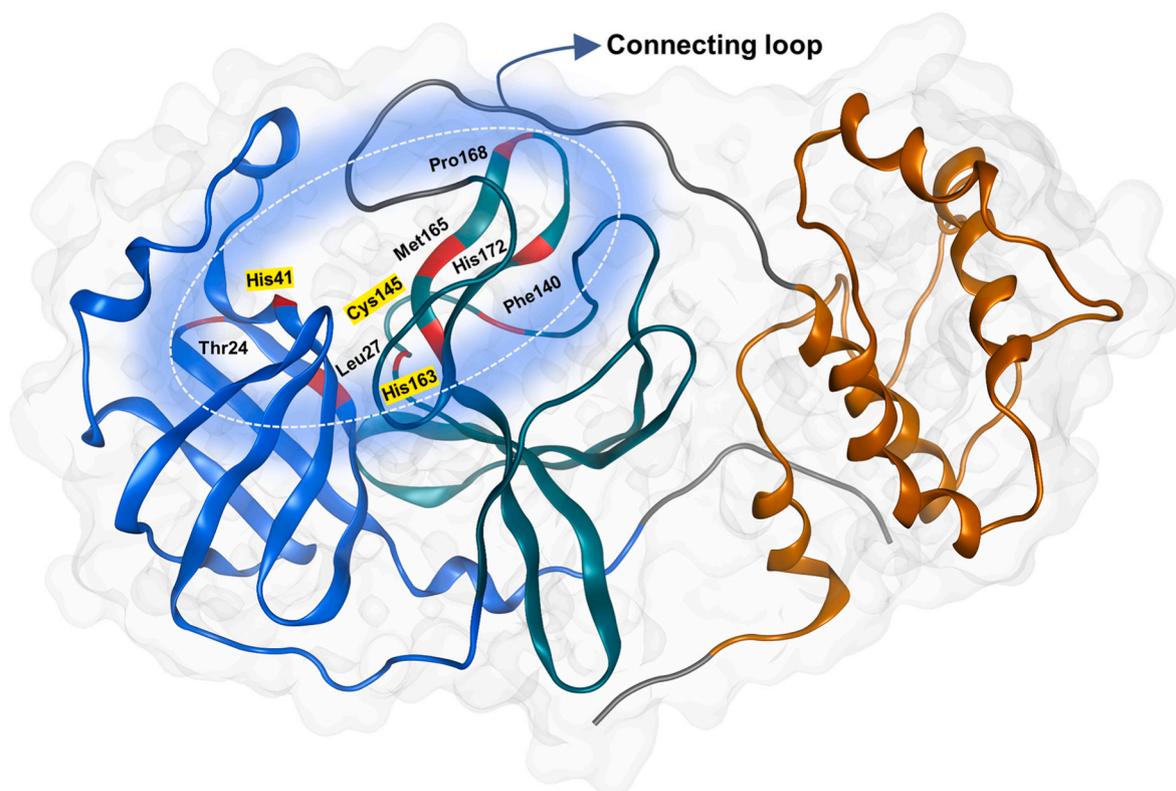


Fig. 2. Structure of the SARS-CoV-2 Mpro The structure of Mpro (3CLpro) monomer is comprised of three domains: domain I, domain II, and domain III consisting of residues 8–101, residues 102–184, and residues 201–303, respectively, as well as a long loop (encompassing residues 185–200) that connects domain II and III [22]. In the gap between domains I and II, the active site (Cys145 and His41) is located, while in the pocket, a hydrophobic surrounding is also formed by hydrophobic amino acids, i.e., T24, L27, H41, F140, C145, H163, M165, P168, and H172.

set to be London dG. Once the screening was completed, the top 1032 compounds were selected, having the docking score ranging from -11.99 to -6.30 kcal/mol.

Next, to filter out the best possible hits, a criterion was set for each molecule to be selected for further evaluation. This standard was that the molecule must have a docking score of more than -10 kcal/mol and have a minimum of three interactions with the active site residues. Any additional interaction with the neighboring residue was considered a plus for the compound. Thus, a total of ten compounds were selected, having strong interaction with the Mpro active site residues.

2.2. Induced fit redocking

Once the top-ranked ten compounds were shortlisted, they were further subjected to the induced-fit docking via MOE using the same parameters [23]. However, the conformations were set to fifty because we wanted to observe the compound, which is more likely to interact with most of its conformations. Human physiological conditions are highly dynamic, and thus, compounds that can establish interactions with the target in most of its poses are unquestionably considered to be the more effective ones.

2.3. ADMET and bioactivity prediction

The bioactivity of each top ligand was predicted by the Molinspiration Cheminformatics tool [24], whereas SwissADME [25] was used for the ADMET analysis. Approximately 4500 studies predicted their bioactivity results via Molinspiration. We further examined the drug-likeness of the selected compounds through OSIRIS Property Explorer [26].

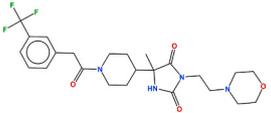
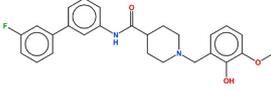
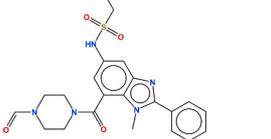
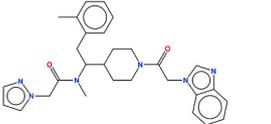
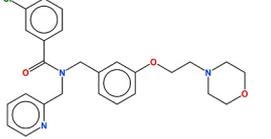
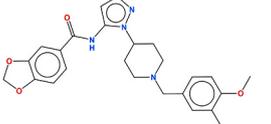
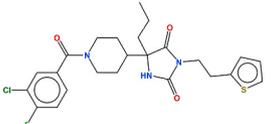
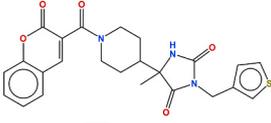
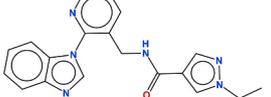
2.4. Simulation protocol

Molecular dynamics (MD) simulation is a practical approach with numerous applications. In the present work, we employed MD simulation to examine the stability of our shortlisted compounds in the active site of the Mpro at different time intervals. Hence, the shortlisted compounds individually in a complex with the Mpro were subjected to MD simulation via GROMACS v.5.1 [27]. The All-atom OPLS force field [28] was employed in this case. The ligands topology is not defined in the GROMACS forcefield; therefore, a freely accessible server LigParGen [29] was used to generate the ligands' topology. This server takes the compounds' simplified molecular-input line-entry system (SMILES) as an input and provides both the coordinates (*.gro*) and topology (*.itp*) files for the ligand. A total of five holo simulations were conducted by the addition of an explicit flexible SPC water molecule fixed in a cubic box [30], whereas the margins were placed ≥ 10 Å from all the protein atoms. The size and vectors of the box were set to be $4.256 \times 4.061 \times 4.142$ nm, and $6.7 \times 6.7 \times 6.7$ nm, respectively, whereas the box angles were equal to 90° on each side. Further we added Na⁺ to maintain a neutral system.

Next, for the neutralization of the system, we minimized the solvated structures for 50,000 steps of steepest descent minimization that dismisses when the overall maximum force is < 1000 kJ/mol/nm. A stable temperature and pressure of 300 K and 1 bar, respectively, with a timestep of 2 fs was kept to attain the equilibrium. For the heavy atoms, the LINCS (LINear Constraint SolVer) [31] constraints and non-bonded pair list were updated every ten steps under the position restraint conditions for the heavy atoms. Electrostatic interactions were calculated using the particle mesh Ewald method. The v-rescale (modified Berendsen thermostat) temperature coupling method [32] was used to maintain a constant temperature inside the box. All the simulations were carried out for a duration of 100 ns (ns).

Table 1

Chemical structures, accession numbers, docking scores, protein-ligand docking RMSDs and binding energies of all the ten compounds. The selected top five compounds are highlighted in grey color.

S. No.	Compound	ChemBridge ID	Docking score	RMSD (Å)	Binding affinity (Kcal/mol)
1		12196498	-11.64	1.22	-9.01
2		12259240	-10.75	1.53	-7.31
3		12347415	-10.04	1.28	-8.60
4		12355365	-9.80	1.23	-8.52
5		12445119	-8.20	1.20	-7.65
6		12361793	-10.32	1.22	-8.11
7		12252299	-8.40	0.84	-7.52
8		12319189	-8.46	1.6	-8.06
9		12403731	-7.98	0.54	-7.36
10		12414646	-10.21	0.71	-9.32

*S.No: Serial number.

2.5. Simulation trajectory analysis

Upon successful completion of the simulation, each system's trajectory was subjected to post-simulation analysis, the stability calculation, residual level flexibility, and overall structural compactness. All the analyses were carried out using the built-in GROMCAS functions such as *g_rms*, *g_rmsf*, and *g_gyrate*. The protein structures were rendered in PyMOL [33,34], and the graphs were constructed and visualized in Origin [35].

2.6. Estimation of free binding energy

The *g_mmpbsa* tool was used to compute the binding energy for all the protein-ligand complexes [36]. The *g_mmpbsa* tool functions on the GROMACS generated datatypes (.tpr and .xtc), and estimates mainly three components of the binding energy that involves molecular mechanical energy, polar solvation energy, and apolar solvation energy. Molecular mechanical (MM) energy consisted of electrostatic contribution (Elec) and van der Waals (EvdW) contributions. To calculate polar solvation energy, *g_mmpbsa* relies on the Assisted Poisson Boltzmann Solver (APBS) program. In case of the apolar solvation energy estimation, the Solvent Accessible Surface Area (SASA) model is used. Binding

Table 2

The list of hydrogen bonding residues along with their bond distance and energy for the selected five compounds

*The residues shaded in grey were observed to be interacting in most of the ligand's orientations.

Compound	Hydrogen bonds with:	Distance (Å)	Energy (kJ/mol)
C1	His41, Pro168	4.4, 3.6	-0.8, -0.7
C2	Leu141, Pro168, Gln189, Thr25	2.5, 3.7, 4.8, 4.0	-1.2, -0.5, -0.5, -1.1
C3	Glu166, Pro168	4.3, 4.1	-1.3, -0.5
C4	His41, Glu166, Gln189	4.4, 4.2, 4.5	-0.5, -1.0, -0.6
C5	His41, Met165, Gln189	-4.2, -3.5, 4.3	-1.7, -0.5, -0.7

free energy estimation depends on the following equation.

$$\Delta G_{\text{Binding}} = G_{\text{Complex}} - (G_{\text{Protein}} + G_{\text{Ligand}}) \quad (1)$$

where G_{Complex} indicates the total free energy of the binding complex, G_{Protein} denotes the total free energies of the individual protein and G_{Ligand} shows the total free energies of the individual ligand. In addition, for each individual entity, the free energy can be given by:

$$G_X = (E_{MM}) - TS + (G_{\text{Solvation}}) \quad (2)$$

where "X" represents the protein or ligand or protein-ligand complex. E_{MM} indicates the average MM potential energy in a vacuum. The entropic contribution to the free energy in a vacuum is jointly denoted by temperature (T) and entropy (S), and the $G_{\text{Solvation}}$ stands for the free energy of solvation. The sum of bonded as well as non-bonded interactions imply the vacuum potential energy (E_{MM}), which is calculated on the base of MM forcefield parameters.

$$E_{MM} = E_{\text{bonded}} + E_{\text{non-bonded}} = E_{\text{bonded}} + (E_{\text{vdW}} + E_{\text{elec}}) \quad (3)$$

where E_{bonded} is bonded interactions comprising of bond, angle, dihedral, as well as improper interactions. Both electrostatic (E_{elec}) and van der

Waals (E_{vdW}) interactions encompass the non-bonded interactions $E_{\text{non-bonded}}$ and are demonstrated using a Coulomb and Lennard-Jones (LJ) potential function, respectively. The free energy of solvation is well defined as the energy vital for transmitting a solute from a vacuum into a solvent. In the method of MM-GPBSA, an implicit solvent model is used for its calculation. The solvation free energy ($G_{\text{Solvation}}$) is defined as:

$$G_{\text{Solvation}} = G_{\text{polar}} + G_{\text{non-polar}} \quad (4)$$

where G_{polar} represents the electrostatic and $G_{\text{non-polar}}$ denotes the non-electrostatic contributions to the solvation-free energy. For the analysis of all the complexes' binding free energy, all frames that cover the period of 10 ns of the constant MD trajectories were utilized.

3. Results and discussion

3.1. Induced fit redocking analysis

The top ten compounds redocked into the active site of Mpro were carefully examined via the ligand interaction tool in MOE (Table 1). The objective was to search for compounds that show interactions with the active site residues in most of its poses. **Compound 1** observed stronger

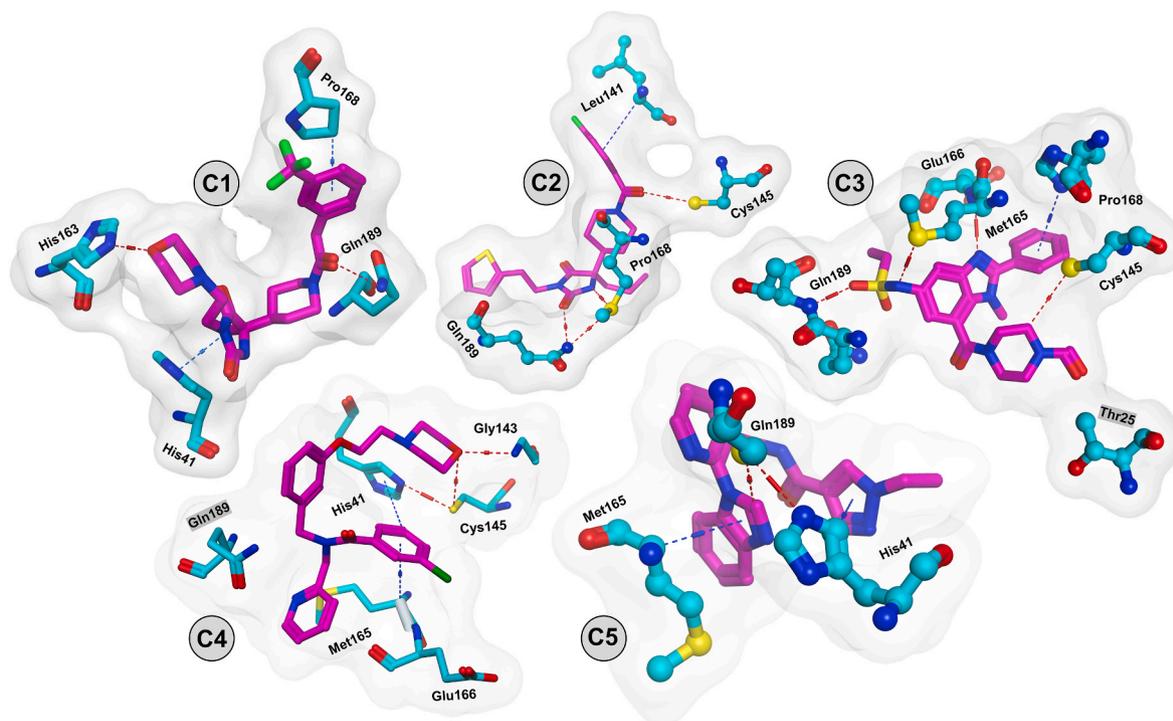


Fig. 3. Protein-drug molecular interactions The selected top five compounds are rendered in pink while the interacting residues of the Mpro are colored cyan. The active site residues are observed to be making solid interactions with the drug compound. The red dots represent hydrogen bonding, while the blue lines show additional interactions. Residues shaded grey in compounds 3 and 4 are the neighboring residues observed in most of the confirmations to be bonding.

Table 3

Bioactivity score of the top five compounds. Higher values represent the more active nature of a compound and vice versa.

Bioactivity	Compound 1	Compound 2	Compound 3	Compound 4	Compound 5
GPCR ligand	0.11	0.11	0.18	0.22	0.06
Ion channel modulator	-0.29	-0.29	-0.14	-0.06	-0.18
Kinase inhibitor	-0.38	-0.38	0.03	-0.06	0.19
Nuclear receptor ligand	-0.42	-0.42	-0.18	-0.40	-0.58
Protease inhibitor	0.10	0.10	0.01	0.07	-0.31
Enzyme inhibitor	-0.29	-0.29	0.10	-0.03	-0.05

Table 4

Additional drug-likeness prediction using the OSIRIS Property Explorer.

Compound	clogP	Solubility	TPSA	Mutagenic	Tumorigenic	Irritant	Reproductive Effective
C1	2.0	-2.51	82.1	No risk	No risk	No risk	No risk
C2	4.4	-5.3	61.8	High risk	No risk	Medium risk	No risk
C3	2.0	-4.6	112.	Medium risk	No risk	No risk	No risk
C4	3.4	-3.3	54.9	No risk	No risk	No risk	No risk
C5	2.3	-3.5	77.6	No risk	No risk	No risk	No risk

*TPSA: The Polar Surface Area Prediction.

interactions with the target, making two arene-H bonds with His41 and Pro168 and two polar side chain donors with His163 and Gln189. **Compound 2** established three arene-H bonds with Leu141, Pro168, and Gln189 while side-chain donor interactions were observed for His41 and Met165. This compound also made an arene-H bond with the threonine residue at position 25 attached to its Cyclopentane ring in several of its poses. This compound was seen to be having good interactions in every orientation. Interactions formed by **Compound 3** include two arene-H (Glu166, and Pro168), polar side chain donor

(Cys145, and Gln189), and a greasy side-chain acceptor (Met165) took part in the interaction. In the case of **Compound 4**, few poses were observed to be interacting. However, most of the conformations made zero interaction with the target, and thus this compound does not qualify our simulation standard. Interactions of **Compounds 5, 7, 8** were confined to a particular site, such as residues in the 140s or 190s or just single irregular contacts; however, we noticed stronger affinity in the case of **Compound 6** as it made five direct contacts with the key residues as well as one weaker interaction with the neighboring key

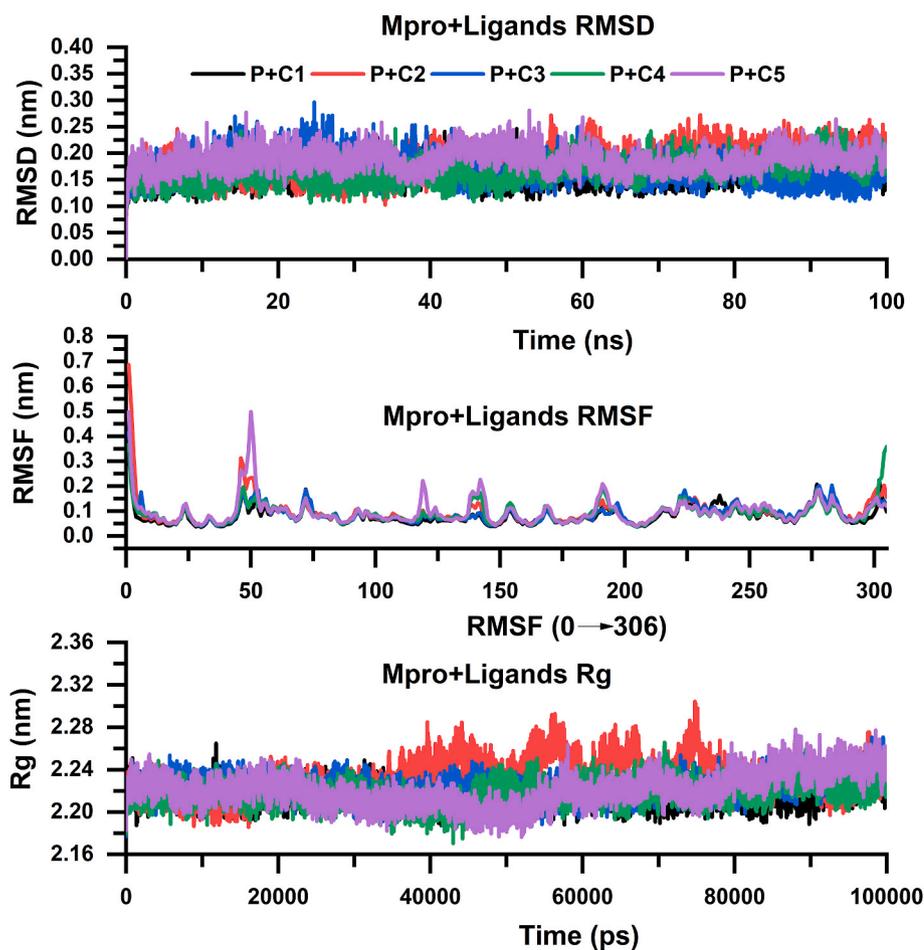


Fig. 4. RMSD and RMSF, and Rg inspection The RMSD and RMSF, and Rg for each compound are plotted as a whole for the entire simulation to analyze the stability, residual fluctuation, and compactness. The RMSD graphs are considerably stable and remain less than 0.30 nm throughout the simulation. In the RMSF analysis, few active site residues fluctuate from 25 to 70 but not those in the range of 140–175. The top selected compounds are mainly interacting with the residues from 140 to 175. Moreover, slight fluctuation occurs when there is an attachment between two chemical entities because of their affinity. The overall RMSF remains within the optimum range with no drastic changes. For the Rg, the x-axis shows the total number of frames in ps, while the y-axis represents the Rg value in nm. In the first 10,000 frames, Compound 2 has the highest Rg value and takes three major leaps but comes back in line with the rest of the compounds' Rg values. As a whole, the Rg are markedly stable, and no radical changes are experienced, proving the stable nature of our predicted compounds in the system.

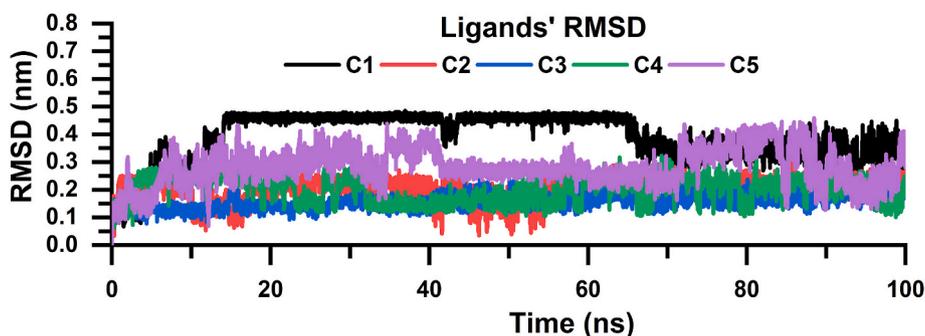


Fig. 5. Ligands' RMSD This image shows the RMSD of the top five compounds to examine if they maintain their original docking pose in the target's active site. All the compounds seem to be stable with minor swaying in the case of **Compound 5**.

amino acid Gln189. **Compounds 9** and **10** had a presentable affinity towards the target, but we were concerned with selecting those with diverse interactions. Therefore, **Compound 10** was chosen because it made three arene-H bonds with His41, Met165 and Gln189 while one greasy side-chain acceptor with Cys145.

To conclude, **Compounds 1, 2, 3, 6, and 10** (now called C1, C2, C3, C4, and C5, respectively) were finally selected as the top five compounds as a result of molecular docking analysis. The chemical structures, accession numbers and docking scores along with the RMSDs and binding energies of the top five (all the ten compounds) selected compounds are listed in [Table 1](#), while the hydrogen bond lengths and energies made by the selected top five compounds are tabulated in [Table 2](#). The interactions established by these selected compounds are given in [Fig. 3](#), while the remaining five compounds that failed to qualify the selection criteria are visually produced in the additional information ([Fig. S1](#)), showing their interactions in 2D space.

3.2. Physiological validity of the predicted compounds

For validation purposes, the selected top compounds were subjected to various biophysical and chemical properties checkups. The bioactivity values of these compounds are tabulated in [Table 3](#), while physicochemical properties, such as lipophilicity, solubility, pharmacokinetics, and drug-like features are provided in the supporting information ([Fig. S2](#)). We observed that the selected compounds are safe as they fulfill all the criteria to be an active drug, and no rules are being violated by these compounds.

Additionally, we used the OSIRIS Property Explorer further to examine the selected compounds for their drug-likeness. We see that C2 has a higher mutagenic effect, but the rest of all the compounds are in the optimal range and safe to be used, including C2 ([Table 4](#)). The substructure fragments and their drug-likeness score for the selected compounds are given in the additional information ([Fig. S3](#)).

3.3. Stability and compactness of the top hits

Molecular dynamics (MD) simulation is a practical computational approach for understanding the stability and dynamics of an entity inside a system in real-time. We analyzed the last frame simulation trajectories of all the five hits using the built-in GROMACS commands to calculate the stability (RMSD) and flexibility (RMSF) of the Mpro and our proposed hits complexes. The compactness and unstable folding were evaluated by calculating the radius of gyration (Rg).

3.4. RMSD analysis

The stability tests of the docked hits into the active site and overall systems stability were observed to be under control with acceptable deviations ([Fig. 4](#)). The average RMSD for all the systems ranged from 0.10 to 0.30 nm (1.0–3.0 Å). The RMSD for **Compound 1** starts from

0.11 nm and goes up to 0.25 nm, experiencing minor drops till 20 ns from where it keeps on fluctuating till 50 ns reaching up to 0.25 nm. Here it shows the sudden decline and rise reaching 0.20 nm and maintains this stable position till the end of the simulation. Minor fluctuations can be observed between 20 to 40 and 40–60 ns, but the overall system remains stable, and no significant escalation can be seen.

In the case of **Compound 2**, the RMSD keeps on elevating till 8.00 ns reach an altitude of 0.25 nm from where it rises and falls till 40 ns where it maintains the RMSD value greater than 0.15 nm and keeps on fluctuating with sharp peaks till 70 ns from where it remains relatively stable till the last frame. The highest and lowest RMSD value in the case of this compound is observed to be 0.27 and 0.10 nm.

The **Mpro-Compound 3** complex is observed to be fluctuating at the start of the simulation, reaching its highest RMSD value of 0.27 and 0.30 in the first 15 and 25 ns followed by a gradual decline after 25 ns and keeps on falling, reaching its lowest value of 0.12 nm. However, few significant stunts can be observed at particular time intervals, such as at 25, 42, and 60 ns. The overall stability hovers around 0.22 nm, which is believed to be a highly stable range.

Similarly, the RMSD for **Compound 4** starts from 0.12 nm and keeps on elevating till 20 ns reaching its maximum altitude of 0.24, from where it drops down back to 0.17 nm. Here it keeps on elevating till 60 ns reaching an altitude of 0.23 nm. From 60 ns, the RMSD remains stable with minor hops till 90 ns. Here it rises, reaching the altitude of 0.25 nm, and drops down to 0.16 nm. For a **complex Mpro-Compound 5**, the RMSD stays highly uniform till the last frame. No major fluctuation has been encountered, and the overall value ranges from 0.10 to 0.27 nm.

Besides the ligand-Mpro complex, we also calculated the RMSD of the ligands only ([Fig. 5](#)) to examine whether they maintain the initial docking pose or they reset to a different position along the time. We can observe that Compound 1 remains slightly unstable in the first 15 ns and then attains a stable position till 67 ns with an RMSD value of 0.49 nm, but it drops down to 0.35 nm after 70 ns and remains there with minor fluctuations as the rest of the compounds. Compound 3, 4, and 5 remains relatively stable and while Compound 5 is observed to have minor fluctuations throughout the system. Interestingly, all the five compounds stabilize after 67 ns. To conclude, Compound 1 gains the highest RMSD value, Compound 3, 4, and 5 are the most stable ones, and Compound 5 has the highest fluctuation rate.

Accordingly, these outcomes reveal the stable intrinsic motions and insignificant fluctuations during the 100 ns simulation. Stable RMSD values confirm that the ligand remains intact, while differences in the RMSD show the ligand's attachment and release at various time points. It is important to note that the small molecule attachment directly impacts the system contrarily as the ligand-binding positioning alters over the simulation period.

RMSF analysis The residual level flexibility is understood in terms of the RMSF as it depicts the flexibility at the residues level. The RMSF of all the five complexes shows a highly stable nature, excluding the loop regions that show the highest fluctuation compared to the rest of the

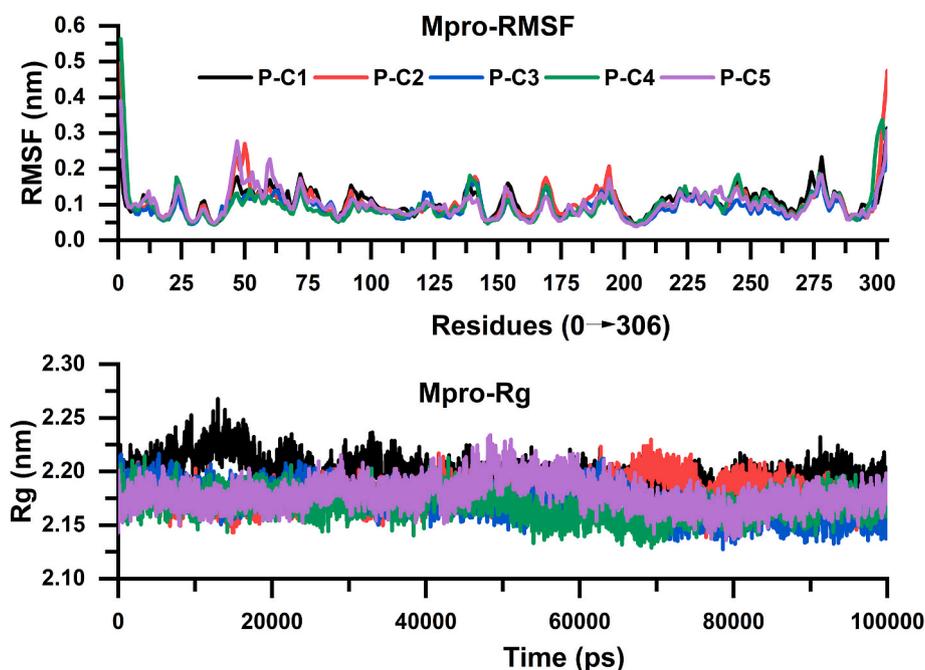


Fig. 6. RMSF and Rg for solo Mpro This image shows the RMSF and Rg calculated for the Mpro alone to see how it behaves without the ligand. This is used as a reference point for the ligands' performance to compare the target's free and bound states.

residues. The overall flexibility ranges from 0.05 to 0.25 nm, excluding a few leaps by **Compounds 3** and **5** (Fig. 4). The active site residues did not show any significant fluctuation, confirming ligand recognition in the ligand-binding domains. All the five complexes show stable behavior throughout simulation except for **Compound 3** and **Compound 5**, which were observed to cause major fluctuation from residues 30 to 60. However, the rest of the residues remain relatively low, and the average RMSF for all the five complexes did not cross the scale of 0.25 nm. We also plotted the RMSF for solo Mpro (Fig. 6) to analyze the amino acids' flexibility without the ligand, observing that the solo protein is slightly unstable compared to the Mpro-ligand complex.

This proves that our proposed ligands firmly attach to the protein throughout the system and stabilize the protein. Minor fluctuations can be observed throughout the system, which is acceptable because the ligand attachment is not a rigid phenomenon. Thus, upon its attachment, the respective bonds residues fluctuate slightly to best orient for the attachment, which results in minor crests. These analyses confirm the stability of the target protein by the attachment of all the five proposed drug compounds. Hence, the ligand binding significantly impacts the residual fluctuation because of the internal residues experiencing effects caused by various ligands' attachment.

3.5. Radius of gyration

The Rg values determine a structure's behavior under compression along an axis. A higher Rg value means less compactness (additional unfolding) and vice versa. Therefore, the Rg plots were plotted against time to examine the structural compactness of the system (Fig. 4). All the simulated complexes exhibit the Rg scores ranging from 2.20 to 2.26 nm except for **Compound 2** that exhibited significant fluctuations around 40000, 60000, and 70000 ps, reaching the highest Rg value of 2.30 nm. The Rg values in the case of the **Mpro-Compound 1** complex were recorded to be ranging from 2.20 nm to 2.20 with minor leaps reaching up to 2.25 and 2.27 nm at 30000 and 15000 ps, respectively. The average Rg values for **Compound 3** stays between 2.20 and 2.25 nm with minor ups and downs throughout the simulation. A large leap is observed at the five ns, but it drops down back to its optimal range. The complexes of **Compound 4**, and **5** range from 2.20 to 2.24 nm. In the

Table 5

Various energies calculated as a result of MM-GBSA analysis for the top five compounds.

Complexes	Electrostatic energy	Van der Wall	SASA	Binding energy
Mpro-Compound 1	-5.29	-68.45	-4.57	-60.99
Mpro-Compound 2	-3.82	-48.78	-2.16	-44.11
Mpro-Compound 3	-4.11	-58.35	-3.87	-57.20
Mpro-Compound 4	-3.06	-60.36	-4.30	-64.96
Mpro-Compound 5	-3.23	-60.27	-3.25	-55.11

case of the Rg plots for solo Mpro, no significant difference can be observed except Compound 1 that shows a slightly higher Rg value in the first quarter of the simulation, but it stabilizes as the rest of the compounds (Fig. 6). Gradual elevations and declines can be observed at different periods, but the average plots remain pretty stable. The compactness of the complexes was greatly affected by the attachment and release of ligands.

3.6. Binding free energy

For the calculation of the binding energies, the polar and apolar solvation conditions were estimated. This analysis set up the energies associated with the binding of Mpro-Compound 1/2/3/4/5 in the course of the 100 ns MD simulations. The subsequent protein-ligand energies have been calculated, such as vdW interaction energy, electrostatic energy, SASA, and average binding energy (Table 5). Compound 4 is observed to have the highest binding energy in all the complexes, while the lowest value was recorded for **Compound 2**. The obtained energy values validate the docking results as more bonds resulted in higher binding energy.

3.7. Analyzing protein-drug communication

The interaction between the top five drug-like compounds was further examined through the hydrogen bonding analysis and PCA. All five compounds show strong hydrogen bonding with the protein through

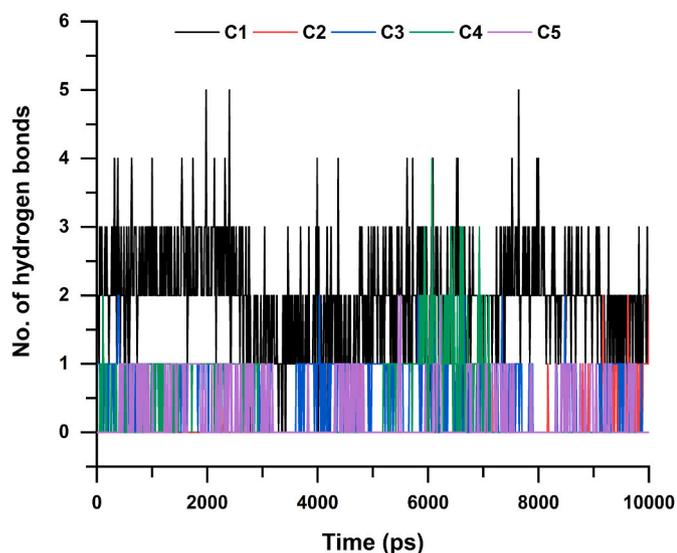


Fig. 7. H-bonding analysis All the predicted compounds established hydrogen bonds with the active site residues of the target. This figure represents the number and consistency of the hydrogen bonds formed between the protein-inhibitor through the simulation. The x-axis shows the number of frames in ps, while the number of bonds formed can be seen on the y-axis.

the course of simulation (Fig. 7 and Table S1). It can be seen that **compound 1** has the highest number of hydrogen bonds, down to which the second-highest is **compound 4** and then **compound 3**. Hydrogen bonding in the case of **compounds 3** and **5** is lower than the rest of the compounds but is more consistent throughout the simulation.

3.8. Essential dynamics analysis

In order to understand the principal structural variations revealed by each Mpro-Compound (1/2/3/4/5) complex, we plotted the PCA graphs (Fig. 8). All the compounds exhibit a clear color transition at various points that signifies the switching behavior from one conformation to another caused by the inhibitor attachment. Each dot represents an individual frame. It is observed that in most of the frames, the system remains compact throughout the 100 ns simulation with slight dispersions that could be linked to the loss of hydrogen bonds for short intervals. These analyses reveal the substantial flexibility of protein structure docked with the shortlisted compounds during primary simulation phases that ultimately reduce with the simulation interlude. Additionally, the contribution percentage in eigenmodes serially decreased too, suggesting the local instabilities in the target's structure for every complex inclined to gain stability. In the case of every protein-ligand complex, these motions are meant to be contributed by docked compounds in the viral protease active site, directing to the establishment of a strong complex. Previous studies have witnessed such observations, for instance, the G-protein coupled receptor and complexes of

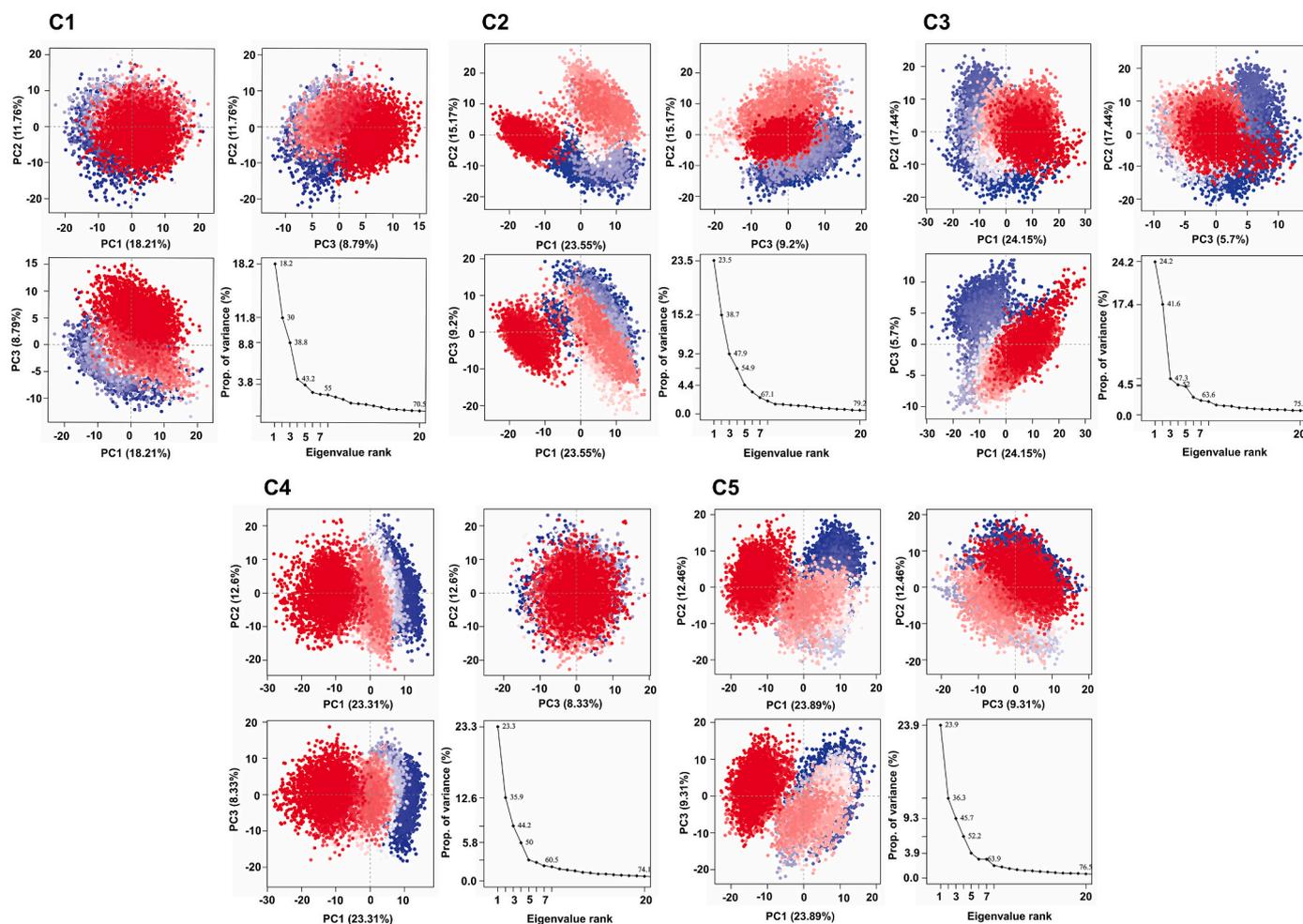


Fig. 8. Principal motions The entire simulation trajectory for each compound is used to plot the PCA for extracting the information about the conformational status. The percentage of total mean square displacement of residue positional variations recorded in each dimension is categorized by equivalent eigenvalue (PCs). Shade transitions can be observed across the frames that witness the switching from one conformation into another.

Fructose transporter GLUT5.

The given figure (Fig. 8) shows three eigenvectors or PCA for the SARS-CoV-2 main protease being docked with the top five potential ligands based on their extracted trajectories and exposed in clusters. Examining these eigenvectors supports the solid and clustered motions in the target's corresponding complexes during the simulation. The Mpro complexes formed clusters ranging from -30 to 30 on PC1, and -20 to 20 on PC2 as well as PC3. The 2D plots determine the changes in the ensemble dispersal regarding each conformation during the course of 100 ns, while the color transition from blue to red denotes the episodic hops among different conformational positions of the target's structure.

Overall, interrelated motion in SARS-CoV-2 conformation expressed the rigidity and substantial variations brought at the active site due to ligand binding during the simulation. Since these analyses suggest selected compounds' stability in the SARS-CoV-2 active site, it restricts the protein indispensable motion needed for enzymatic action and thus results in the inhibition; maintained by RMSD, RMSF, Rg, protein-ligand communication, and PCA study along with many molecular docking scores.

4. Conclusion

There is a need to discover potent compounds that could bind with the SARS-CoV-2 targets and facilitate the inhibition of this virus. Therefore, we evaluated 16,000 drug-like compounds obtained from the ChemBridge database, and the top hits were docked into the active site of the Mpro of the SARS-CoV-2. Additionally, the molecular dynamics simulation confirmed the stability of the selected compounds in the system. Considering the outcomes of molecular docking, MD simulation, and free energy calculations, we believe that the proposed compounds could act as an effective drug against the SARS-CoV-2. However, this whole study is founded on a *in silico* pipeline, and thus *in vitro* assays are recommended to confirm their validity further.

4.1. Associated content

All the interactions made by the additional five compounds are provided in Fig. S1, while the physicochemical and pharmacokinetic properties for each selected compound are provided in the supplementary information as Fig. S2. The number of hydrogen bonds formed by each selected compound is tabulated in Table S1.

Funding

Dong-Qing Wei is supported by grants from the Key Research Area Grant 2016YFA0501703 of the Ministry of Science and Technology of China, the National Science Foundation of China (Grant No. 32070662, 61832019, 32030063), the Science and Technology Commission of Shanghai Municipality (Grant No.: 19430750600), the Natural Science Foundation of Henan Province (162300410060), as well as SJTU JiRLMDS Joint Research Fund and Joint Research Funds for Medical and Engineering and Scientific Research at Shanghai Jiao Tong University (YG2017ZD14).

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

The heavy-duty computational operations were partially performed at the Peng Cheng Lab and the Center for High-Performance Computing, Shanghai Jiao Tong University.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.combiomed.2022.105235>.

References

- [1] D. Li, C. Wang, S. Wang, A. Mehmood, J. Gu, X. Cheng, P. Chen, J. Qiu, J. Zhao, J. Wang, Discovery of a Natural Product with Potent Efficacy against SARS-CoV-2 by Drug Screening, *Interdisciplinary Sciences, Computational Life Sciences*, 2021, pp. 1–9.
- [2] D. Li, P. Chen, T. Shi, A. Mehmood, J. Qiu, HD5 and LL-37 inhibit SARS-CoV and SARS-CoV-2 binding to human ACE2 by molecular simulation, *Interdiscipl. Sci. Comput. Life Sci.* 13 (2021) 766–777.
- [3] M.A. Shereen, S. Khan, A. Kazmi, N. Bashir, R. Siddique, COVID-19 infection: origin, transmission, and characteristics of human coronaviruses, *J. Adv. Res.* 24 (2020) 91–98.
- [4] J.A. Jaimes, N.M. André, J.S. Chappie, J.K. Millet, G.R. Whittaker, Phylogenetic analysis and structural modeling of SARS-CoV-2 spike protein reveals an evolutionary distinct and proteolytically sensitive activation loop, *J. Mol. Biol.* 432 (2020) 3309–3325.
- [5] P. Kalhotra, V.C. Chitpeu, G. Osorio-Revilla, T. Gallardo-Velazquez, Field-template, QSAR, ensemble molecular docking, and 3D-RISM solvation studies expose potential of FDA-approved marine drugs as SARS-CoV-2 main protease inhibitors, *Molecules* 26 (2021) 936.
- [6] E. Petersen, M. Koopmans, U. Go, D.H. Hamer, N. Petrosillo, F. Castelli, M. Storgaard, S. Al Khalili, L. Simonsen, Comparing SARS-CoV-2 with SARS-CoV and Influenza Pandemics, *The Lancet infectious diseases*, 2020.
- [7] D. Brian, R. Baric, Coronavirus Genome Structure and Replication, *Coronavirus Replication and Reverse Genetics*, 2005, pp. 1–30.
- [8] S. Adachi, T. Koma, M. Nomaguchi, A. Adachi, Commentary: origin and evolution of coronavirus, *Front. Immunol.* 11 (2020) 811.
- [9] Y.-R. Guo, Q.-D. Cao, Z.-S. Hong, Y.-Y. Tan, S.-D. Chen, H.-J. Jin, K.-S. Tan, D.-Y. Wang, Y. Yan, The origin, transmission and clinical therapies on coronavirus disease 2019 (COVID-19) outbreak—an update on the status, *Mil. Med. Res.* 7 (2020) 1–10.
- [10] A. Khan, W. Heng, Y. Wang, J. Qiu, X. Wei, S. Peng, S. Saleem, M. Khan, S.S. Ali, D.-Q. Wei, In Silico and in Vitro Evaluation of Kaempferol as a Potential Inhibitor of the SARS-CoV-2 Main Protease (3CLpro), *Phytotherapy research: PTR*.
- [11] K. Anand, G.J. Palm, J.R. Mesters, S.G. Siddell, J. Ziebuhr, R. Hilgenfeld, Structure of coronavirus main proteinase reveals combination of a chymotrypsin fold with an extra α -helical domain, *EMBO J.* 21 (2002) 3213–3224.
- [12] Z. Jin, X. Du, Y. Xu, Y. Deng, M. Liu, Y. Zhao, B. Zhang, X. Li, L. Zhang, C. Peng, Structure of M pro from SARS-CoV-2 and discovery of its inhibitors, *Nature* 582 (2020) 289–293.
- [13] S. Das, S. Sarmah, S. Lyndem, A. Singha Roy, An investigation into the identification of potential inhibitors of SARS-CoV-2 main protease using molecular docking study, *J. Biomol. Struct. Dyn.* (2020) 1–11.
- [14] W. Dai, B. Zhang, X.-M. Jiang, H. Su, J. Li, Y. Zhao, X. Xie, Z. Jin, J. Peng, F. Liu, Structure-based design of antiviral drug candidates targeting the SARS-CoV-2 main protease, *Science* 368 (2020) 1331–1335.
- [15] L. Zhang, D. Lin, X. Sun, U. Curth, C. Drosten, L. Sauerhering, S. Becker, K. Rox, R. Hilgenfeld, Crystal structure of SARS-CoV-2 main protease provides a basis for design of improved α -ketoamide inhibitors, *Science* 368 (2020) 409–412.
- [16] S.J.Y. Macalino, V. Gosu, S. Hong, S. Choi, Role of computer-aided drug design in modern drug discovery, *Arch Pharm. Res. (Seoul)* 38 (2015) 1686–1701.
- [17] L.G. Ferreira, R.N. Dos Santos, G. Oliva, A.D. Andricopulo, Molecular docking and structure-based drug design strategies, *Molecules* 20 (2015) 13384–13421.
- [18] P. Sledz, A. Cafilisch, Protein structure-based drug design: from docking to molecular dynamics, *Curr. Opin. Struct. Biol.* 48 (2018) 93–102.
- [19] A. Mehmood, A.C. Kaushik, Q. Wang, C.-D. Li, D.-Q. Wei, Bringing structural implications and deep learning-based drug identification for KRAS mutants, *J. Chem. Inf. Model.* 61 (2021) 571–586.
- [20] A. Mehmood, M.T. Khan, A.C. Kaushik, A.S. Khan, M. Irfan, D.-Q. Wei, Structural dynamics behind clinical mutants of PncA-Asp12Ala, Pro54Leu, and His57Pro of Mycobacterium tuberculosis associated with pyrazinamide resistance, *Front. Bioeng. Biotechnol.* 7 (2019) 404.
- [21] A. Wadood, A. Mehmood, H. Khan, M. Ilyas, A. Ahmad, M. Alarjah, T. Abu-Izneid, Epitopes based drug design for dengue virus envelope protein: a computational approach, *Comput. Biol. Chem.* 71 (2017) 152–160.
- [22] C. Wu, Y. Liu, Y. Yang, P. Zhang, W. Zhong, Y. Wang, Q. Wang, Y. Xu, M. Li, X. Li, Analysis of therapeutic targets for SARS-CoV-2 and discovery of potential drugs by computational methods, *Acta Pharm. Sin.* B 10 (2020) 766–788.
- [23] A. Mehmood, A.C. Kaushik, D.Q. Wei, Prediction and validation of potent peptides against herpes simplex virus type 1 via immunoinformatic and systems biology approach, *Chem. Biol. Drug Des.* 94 (2019) 1868–1883.
- [24] P. Thangarasu, S.T. Selvi, A. Manikandan, Unveiling novel 2-cyclopropyl-3-ethyl-4-(4-fluorophenyl) quinolines as GPCR ligands via PI3-kinase/PAR-1 antagonism and platelet aggregation valuations; development of a new class of anticancer drugs with thrombolytic effects, *Bioorg. Chem.* 81 (2018) 468–480.
- [25] A. Daina, O. Michielin, V. Zoete, SwissADME: a free web tool to evaluate pharmacokinetics, drug-likeness and medicinal chemistry friendliness of small molecules, *Sci. Rep.* 7 (2017) 1–13.

- [26] M.H. Satari, E. Apriyanti, H.D.A. Dharsono, D. Nurdin, M. Gartika, D. Kurnia, Effectiveness of bioactive compound as antibacterial and anti-quorum sensing agent from *myrmecodia pendans*: an in silico study, *Molecules* 26 (2021) 2465.
- [27] Q. Wang, A. Mehmood, H. Wang, Q. Xu, Y. Xiong, D.-Q. Wei, Computational screening and analysis of lung cancer related non-synonymous single nucleotide polymorphisms on the human kirsten rat sarcoma gene, *Molecules* 24 (2019) 1951.
- [28] W. Kulig, M. Pasenkiewicz-Gierula, T. Róg, Topologies, structures and parameter files for lipid simulations in GROMACS with the OPLS-aa force field: DPPC, POPC, DOPC, PEPC, and cholesterol, *Data Brief* 5 (2015) 333–336.
- [29] L.S. Dodda, I. Cabeza de Vaca, J. Tirado-Rives, W.L. Jorgensen, LigParGen web server: an automatic OPLS-AA parameter generator for organic ligands, *Nucleic Acids Res.* 45 (2017) W331–W336.
- [30] H. Bekker, E. Dijkstra, M. Renardus, H. Berendsen, An efficient, box shape independent non-bonded force and virial algorithm for molecular dynamics, *Mol. Simulat.* 14 (1995) 137–151.
- [31] B. Hess, H. Bekker, H.J. Berendsen, J.G. Fraaije, LINCS: a linear constraint solver for molecular simulations, *J. Comput. Chem.* 18 (1997) 1463–1472.
- [32] A. Oliva, J.A. Teruel, F.J. Aranda, A. Ortiz, Effect of a dirhamnolipid biosurfactant on the structure and phase behaviour of dimyristoylphosphatidylserine model membranes, *Colloids Surf. B Biointerfaces* 185 (2020) 110576.
- [33] W.L. DeLano, Pymol: an open-source molecular graphics tool, *CCP4 Newsl. Protein Crystallogr.* 40 (2002) 82–92.
- [34] A. Khan, M. Junaid, A.C. Kaushik, A. Ali, S.S. Ali, A. Mehmood, D.-Q. Wei, Computational identification, characterization and validation of potential antigenic peptide vaccines from hrHPVs E6 proteins using immunoinformatics and computational systems biology approaches, *PLoS One* 13 (2018), e0196484.
- [35] E. Seifert, *OriginPro 9.1: Scientific Data Analysis and Graphing Software* Software Review, ACS Publications, 2014.
- [36] F.I. Khan, D. Lai, R. Anwer, I. Azim, M.K.A. Khan, Identifying novel sphingosine kinase 1 inhibitors as therapeutics against breast cancer, *J. Enzym. Inhib. Med. Chem.* 35 (2020) 172–186.