ORIGINAL PAPER

In silico evaluation of Philippine Natural Products against SARS-CoV-2 Main Protease

Adrian Josiah T. Cheng[1] · Stephani Joy Y. Macalino[2] · Junie B. Billones[1] · Marilen Parungao Balolong[3] · Lyre Anni E. Murao[4] · Maria Constancia O. Carrillo[1]

Abstract

The severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), the causative agent of COVID-19, is a novel strain of coronavirus first reported in December 2019 which rapidly spread throughout the world and was subsequently declared a pandemic by the World Health Organization (WHO) in March 2020. Although vaccines, as well as treatments, have been rapidly developed and deployed, these are still spread thin, especially in the developing world. There is also a continuing threat of the emergence of mutated variants which may not be as responsive to available vaccines and drugs. Accessible and affordable sources of antiviral drugs against SARS-CoV-2 offer wider options for the clinical treatment of populations at risk for severe COVID-19. Using in silico methods, this study identified potential inhibitors against the SARS-CoV-2 main protease (Mpro), the protease directly responsible for the activation of the viral replication enzyme, from a consolidated database of 1516 Philippine natural products. Molecular docking experiments, along with in silico ADME predictions, determined top ligands from this database with the highest potential inhibitory effects against Mpro. Molecular dynamic trajectories of the apo and diosmetin-7-O-b-D-glucopyranoside (DG) in complex with the protein predicted potential mechanisms of action for the ligand—by separating the Cys145–His41 catalytic dyad and by influencing the protein network through key intra-signaling residues within the Mpro binding site. These findings show the inhibitory potential of DG against the SARS-CoV-2 Mpro, and further validation is recommended through in vitro or in vivo experimentation.

Keywords SARS-CoV-2 · Natural products · Molecular docking · ADMET · Molecular dynamics · Network analysis · Betweenness centrality

Adrian Josiah T. Cheng and Stephani Joy Y. Macalino contributed equally to this work.

✉ Stephani Joy Y. Macalino
   stephani.macalino@dlsu.edu.ph

✉ Maria Constancia O. Carrillo
   mtobrerocarrillo@up.edu.ph

1   Department of Physical Sciences and Mathematics, College of Arts and Sciences, University of the Philippines Manila, 1000 Manila, Philippines

2   Department of Chemistry, De La Salle University, 2401 Taft Avenue, 0992 Manila, Philippines

3   Department of Biology, College of Arts and Sciences, University of the Philippines Manila, 1000 Manila, Philippines

4   Department of Biological Sciences and Environmental Studies, College of Science and Mathematics, University of the Philippines Mindanao, 8000 Tugbok, Davao City, Davao del Sur, Philippines

Introduction

In December 2019, a novel coronavirus discovered in Wuhan, China, was reported to be the causative pathogen in patients exhibiting symptoms of pneumonia with unknown origins. Being highly transmissible, the virus, now named the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), spread around the world and has been declared a pandemic by the World Health Organization (WHO) last March 2020 [1]. Although treatment regimens are available, very few are specifically targeted against SARS-CoV-2. Several vaccines have also been rapidly developed and deployed. However, constraints in the supplies of both drugs and vaccines remain.

In the Philippines alone, cumulative cases amounted to 3.69 million accompanied by more than 60,000 deaths as of May 2022 [2]. And while the number of cases seems to be dwindling worldwide, there are still countries that

are severely affected by this infection, making it important to mitigate the spread of both the wild type and its highly transmissible variants to avoid further disruption of socio-economic activities.

SARS-CoV-2 shows similarities to the previous coronaviruses, such as SARS-CoV and MERS-CoV, which also belongs to the family Coronaviridae with genus *Betacoronavirus* [3]. Viruses in this taxon are an enveloped, positive-sense, single-stranded RNA virus [4]. SARS-CoV-2 has a fatality rate of 2.3%, which is lower than both of its relatives; however, it shows a higher base reproduction number (R0) of 3.1 [5–7]. The genome carried by the virus contains fourteen open reading frames (ORFs). The transcription products of these ORFs include the spike protein (S), envelope protein (E), nucleocapsid phosphoprotein (N), membrane protein (M), and the overlapping polyprotein pp1a-pp1ab. The four initial proteins are involved in forming the structure of the virus while the latter is involved in its replication [4]. The overlapping pp1a-pp1ab needs to be cleaved to function and start the replication process [8]. The fragments result in 16 non-structural proteins, among which are the RNA-dependent RNA polymerase (RdRp), the main protease (Mpro), and the papain-like protease (PLpro). Both the Mpro and the PLpro are responsible for the cleavage of this overlapping polyprotein thus releasing the RdRp, which then initiates the replication and transcription of the SARS-CoV-2 genome inside the host cell [9].

Recently, newer antiviral drugs have been reported to be effective against SARS-CoV-2. Remdesivir and Molnupiravir, both repurposed drugs originally designed against different viruses, are currently being administered to target the RdRp of SARS-CoV-2, thus inhibiting viral replication [10, 11]. Bebtelovimab, sotrovimab, and casiriviman-imdevimab, on the other hand, are developed human antibodies that target the SARS-CoV-2 spike protein, thus blocking viral transmission [10, 12]. Indirectly combatting drugs, such as tocilizumab, sarilumab, and barcinitib, the former two being IL-6 receptor blockers and the latter being a Janus kinase inhibitor limit the occurrence of a cytokine storm, thus lowering mortality rates in severely affected patients [10]. Finally, treatments not yet recommended by the WHO but have seen promising results include fluvoxamine, melatonin, atorvastatin, and ivermectin which mainly act as anti-inflammatory agents seen to reduce mortality rates among patients [10, 13, 14].

Despite the numerous available drug targets against SARS-CoV-2, the viral proteases, Mpro and PLpro, remain attractive targets, as inhibition can lead to the blocking of viral replication. Among the two enzymes, Mpro has been in greater focus as it solely cleaves polypeptide sequences after a glutamine residue with the help of the catalytic dyad His41/Cys145 [15]. No human proteases have been discovered with this specificity, therefore making it an attractive drug target compared to other viral enzymes. For example, PLpro can recognize the C-terminal sequence of ubiquitin, and therefore, inhibitors can possibly interfere with host-cell deubiquitinases [15]. Furthermore, Mpro is highly conserved among coronaviruses, with a 96% homology with the SARS-CoV [16]; thus, inhibitors against the SARS-CoV-2 Mpro may possibly be repurposed against other coronaviruses or remain effective against variants that may emerge in the future. Multiple studies dealing with drug discovery against the Mpro use an in silico approach, which allows for a cheaper and faster method of experimentation [16]. With this, nirmatrelvir-ritonavir, sold under the name Paxlovid, was developed and is currently a famous antiviral drug, as it is simply orally administered toward patients with severe symptoms through the direct inhibition of the SARS-CoV-2 Mpro [10, 11]. However, despite the availability of different antiviral drugs, the resurging cases of SARS-CoV-2 throughout the world, as well as the rapid emergence of different SARS-CoV-2 variants of concern [17], demands the continuous search for novel antiviral drugs.

Natural products have been a continued major source of drug candidates for different diseases, including viral ones, due to their notable efficacy and safety. If not the natural products themselves, drugs are often derived and synthesized from natural product structures [18, 19]. In fact, multiple herbal-derived bioactive molecules were shown to inhibit SARS-CoV-2 entry and replication. These include thymoquinone, α-hederin, and nigellidine from *Nigella sativa*, quercetin from *Ginko biloba*, ellagic acid from *Moringa oliefera*, and rosmarinic acid from *Plectranthus amboinicus* [20]. In this study, potential Mpro binders were identified from a compound library of Philippine natural products compiled by Billones et al. (unpublished) using in silico approaches, particularly docking, ADMET prediction, and molecular dynamics. Data gathered from this study can provide significant information in the design of potent leads against SARS-CoV-2 Mpro. The identified compounds can be further validated in vivo or in vitro with the hopes of developing a new and accessible drug option for COVID-19 patients. Additionally, since Mpro is highly conserved among coronaviruses, successfully identifying an inhibitor from a Philippine natural products database may lead to the discovery of a novel inhibitor which may also be effective against other coronaviruses.

Materials and methods

The simulations were conducted using the X-ray crystal structure of Mpro obtained from the RCSB PDB [21]. This structure was subjected to a series of general steps: (1) initial MD of the apo structure and clustering, (2) docking experiments, (3) MD, and (4) network analysis of the top hit. All

the MD steps were conducted using GROMACS 2020.3 [22] with the DOST-ASTI supercomputer and a GPU with AMD Ryzen 7 3800X 8-core processor, 64 GB RAM, and PNY GeForce RTX 2060 SUPER 8 GB.

Initial molecular dynamics simulation and clustering of the apo Mpro structure

The crystal structure of the SARS-CoV-2 main protease co-crystallized with an inhibitor, N3, (PDB ID:6LU7) [23] was used. Dimerization and deletion of the ligand were performed using Maestro [24]. The dimer protein was subjected to MD using the CHARMM27 all-atom force field. The system was immersed in a cubic water box and was solvated with TIP3P water. Counterions from NaCl were added to neutralize the system. The system was subjected to energy minimization using the steepest descent algorithm in a 10 ns simulation. The relaxed system was heated to 300 K in a 10 ns NVT set-up with a 1 fs time step using a Nose–Hoover temperature coupling process. The pressure of the system was subsequently equilibrated to 1 bar in an NPT set-up with a 1 fs time step using a Parrinello-Rahman process. For nonbonded interactions, the Verlet cutoff scheme was employed while for electrostatics, the PME was employed. A production run of 100 ns with a 1 fs timestep was performed, and from the clustered frames of the trajectory, a file that was used for the subsequent docking studies was retrieved. The production run was further extended to 200 ns and in triplicates, using the same parameters. These simulations were conducted using two systems: (1) 1 Tesla P40 OpenCL 1.2 CUDA GPU, 1 node with 12 tasks, 64 GB RAM, and 64 threads and (2) PNY GeForce RTX 2060 SUPER 8 GB GPU, 64 GB RAM, and 16 threads.

Virtual screening

SARS-CoV-2 main protease was used as the receptor for the docking experiment. The receptor was prepared for docking using AutoDockTools 1.5.7 [25]. Kollman charges were used for the protein. The grid box was defined as the area covering the co-crystallized inhibitor, N3, and was centered only on chain A ($x = 10.102$, $y = 2.905$, and $z = 30.5$) with an approximate size ratio of 8:22:24 (x:y:z) and a spacing of 1.

The ligands that were used in this study were retrieved from a compilation of 1516 Philippine natural products (Supplementary Table 1). This database was created by Billones et al. (unpublished) through compiling published natural product compounds that can be retrieved from the 10 Philippine medicinal plants, approved by the Philippine Department of Health. These comprise the following: *Allium sativum* (garlic/bawang), *Blumea balsamifera* (nagal camphor/sambong), *Cassia alata* (ringworm bush/akapulko), *Clinopodium douglasii* (mint/yerba buena),

Ehretia microphylla (scorpion bush/tsaang Gubat), *Momordica charantia* (bitter melon/ampalaya), *Peperomia pellucida* (silver bush/ulasimang bato), *Psidium Guajava* (guava/bayabas), *Quisqualis indica* (rangoon creeper/niyug-niyugan), and *Vitex negundo* (five-leaved chaste tree/lagundi). Common moieties shared by the compounds in the database include substituted hydrocarbons, alcohols, esters, ketones, aldehydes, carboxylic acids, and common plant secondary metabolites such as flavonoids, terpenoids, quinones, sterols, carotenes, and aromatics. Another major group of compounds includes pyranosides and other carbohydrate-linked structures.

Prior to docking, most structures were minimized using MMFF94 while structures containing Selenium were minimized using UFF, and the minimized files were then converted to PDBQT format. All minimization and file conversion steps were performed through OpenBabel 2.3.1 [26] using Avogadro 1.2.0 [27]. AutoDock Vina [28] was used for the docking procedures with default exhaustiveness, 30 modes, and energy range of 4. The co-crystallized ligand was docked to use as a measure on whether the docking method is suitable for the protein target. The backbone of the peptide co-crystallized ligand was made rigid to minimally constrain its high flexibility while keeping the rest of its structure flexible. After docking, the top 5% ligands that produced the lowest binding affinities were then subjected to clustering using Canvas [29, 30] to determine the major binding conformation. Ligands with major clusters proceeded to absorption, distribution, metabolism, and excretion (ADME) screening using SwissADME [31]. Those that showed favorable ADME characteristics were then analyzed for specific ligand-active site interactions using Maestro to determine the top potential inhibitor. Prior to any MD procedures, the selected ligand was also docked into the active site of chain B ($x = -11.250$, $y = 2.305$, $z = -23.135$) using the same docking parameters and procedure. The selected inhibitor in complex with the Mpro was outputted as a single PDB file using PyMol [32].

Molecular dynamics simulation of complexes

The docked complex of the top ligand, along with the Mpro-N3 complex structure, was then subjected to molecular dynamics. Each protein–ligand complex was subjected to the same set of steps. The topology of the protein structure was generated using the CHARMM27 all-atom force field while the topology of the ligand was generated using SwissParam [31]. The ligands were converted to the compatible mol2 format using Avogadro [27]. The same parameters from the initial MD simulations on the apo structure were introduced to this system for energy minimization, NVT, and NPT processes. In these simulations, however, the protein and the ligand were coupled as one group. Once these steps were

completed, the system was subjected to 200 ns of production run in triplicates with a timestep of 1 fs. Trajectory analysis was performed using GROMACS 2020.3 [22], VMD 1.9.3 [33], and gRINN 1.1.0.hf1 [34] to determine complex stability through root-mean-square deviation (RMSD) and root-mean-square fluctuation (RMSF) calculations, protein–ligand interactions, H-bonding network, and centrality measures. These simulations were conducted using 2 Intel(R) Xeon(R) CPUs E5-2699 v4 @ 2.20 GHz, 2 nodes with 32 tasks per node, 64-bit memory, and 64 threads.

Results and Discussion

MPro apo structure MD simulation

RMSD calculations on the initial 100 ns trajectory showed that the system stabilized at around 43 ns (Fig. 1A). For both chains, it was seen that the active site residues His41 and Cys145 showed relatively low fluctuations (< 1 Å fluctuations in both chains, Fig. 1B). It is noteworthy, however, that the neighboring residues of His41, particularly Tyr45 to Arg60, showed a region of flexibility.

Clustering of the trajectory frames was performed to retrieve a representative structure that can be used for the subsequent docking study. Only 1 cluster was retrieved from the clustering procedure despite the notable fluctuations that were previously mentioned. The clustered MD structure and the crystal structure of Mpro were superimposed and resulted in an RMSD of 1.20 Å indicating high similarity to the published structure. A difference in the dyad side chain orientation is observed between the clustered and the crystal structure, but may be accounted for by incorporating dynamics after docking. The His41 imidazole ring of the crystal structure is positioned such that the His41@Nε atom is seen to point directly toward the Cys145@SγH atom, and the His41@Nδ atom points directly toward Asp187 (Fig. 2). This orientation is induced by the presence of a co-crystal ligand, a peptidomimetic inhibitor, such that the catalytic dyad is positioned appropriately for proteolysis thereby mimicking its native function. In the clustered structure, the imidazole ring of His41 was oriented almost perpendicularly compared to that of the crystal structure, distancing the imidazole from Cys145 and Asp187.

Virtual screening of the Philippine natural products

The co-crystallized peptidomimetic inhibitor and 1516 compounds in the Philippine natural products library were docked in the clustered Mpro apo structure using Auto-Dock Vina. Redocking of N3 was performed to obtain a cut-off docking score and control. This resulted in a score of −5.4 kcal/mol and a 3.18 Å RMSD from the co-crystal pose. The slight deviations seen in the clustered trajectory with respect to the crystal structure could have caused the high RMSD for N3 binding. Moreover, N3 is a peptide mimetic; thus, it is highly flexible. Furthermore, AutoDock Vina does not account for possible covalent bond interactions which is seen in the 6LU7 crystal structure with a N3–Cys145 covalent bond. This may have resulted in a conformation that is not accessible by docking through Auto-Dock Vina. Despite this, the redocked N3 was still oriented in a similar manner as the co-crystal pose, indicating that the grid box parameters can suitably encapsulate the residues along the active site.

After docking all the library compounds (Supplementary Table 2), the ligands that produced the most negative docking score were selected as the top 5% hits. However, the best scoring conformation is not necessarily an accurate prediction of the bioactive conformation. In order to acquire a more reliable docking pose, clustering of the docking results was performed. Conformations that are 2.5 Å from each other were defined to belong to one cluster, and clusters with the highest number of conformers were labeled as the major cluster. The best scoring conformer from the major cluster was used as the representative docking pose for a particular ligand. Ligands that did not produce a major cluster were excluded from the sample as these ligands are predicted to not give stable conformers. From this clustering, 38 ligands from the initial top 5% were selected (Supplementary Table 3). The 38 docking hits were subjected to ADME screening by filtering them based on their lipophilicity and structural properties. Compounds were filtered using a logP value of 4 and below as the threshold to provide room for possible future synthetic modifications. Pan-assay interference (PAINS) [35] filter was also utilized to remove compounds that contain certain moieties which can react non-specifically to several biological targets, leading to false positives. The remaining ligands were assessed for their synthetic accessibility wherein lower values signify that the query structure is easier to synthesize. Given that natural products are finite resources, considering the structures' synthetic feasibility can help select compounds that can be synthetically manufactured should they be developed into drugs. An arbitrary cut-off value of 7 was used to further reduce hits to 13. To prioritize the hits to only the top 10, gastrointestinal (GI) absorption, cytochrome P450 inhibition, and Pgp substrate characteristics were additionally used (Supplementary Table 4).

Protein-ligand interactions and selection of the top hit

To determine the top ligand that will be subjected to molecular dynamics simulations, the remaining ten natural products underwent ligand interaction analysis. Since the active

Fig. 1 A Protein RMSD plot of the 100 ns Mpro apo production run. **B** RMSF of chain **A** and **B**. Important residues are color-coded: active site (red), residues interacting with the dyad (brown), oxyanion loop (peach), substrate binding sites (green), N-finger region (blue), Domain II/III linker (violet), protomer-protomer interaction (gold)
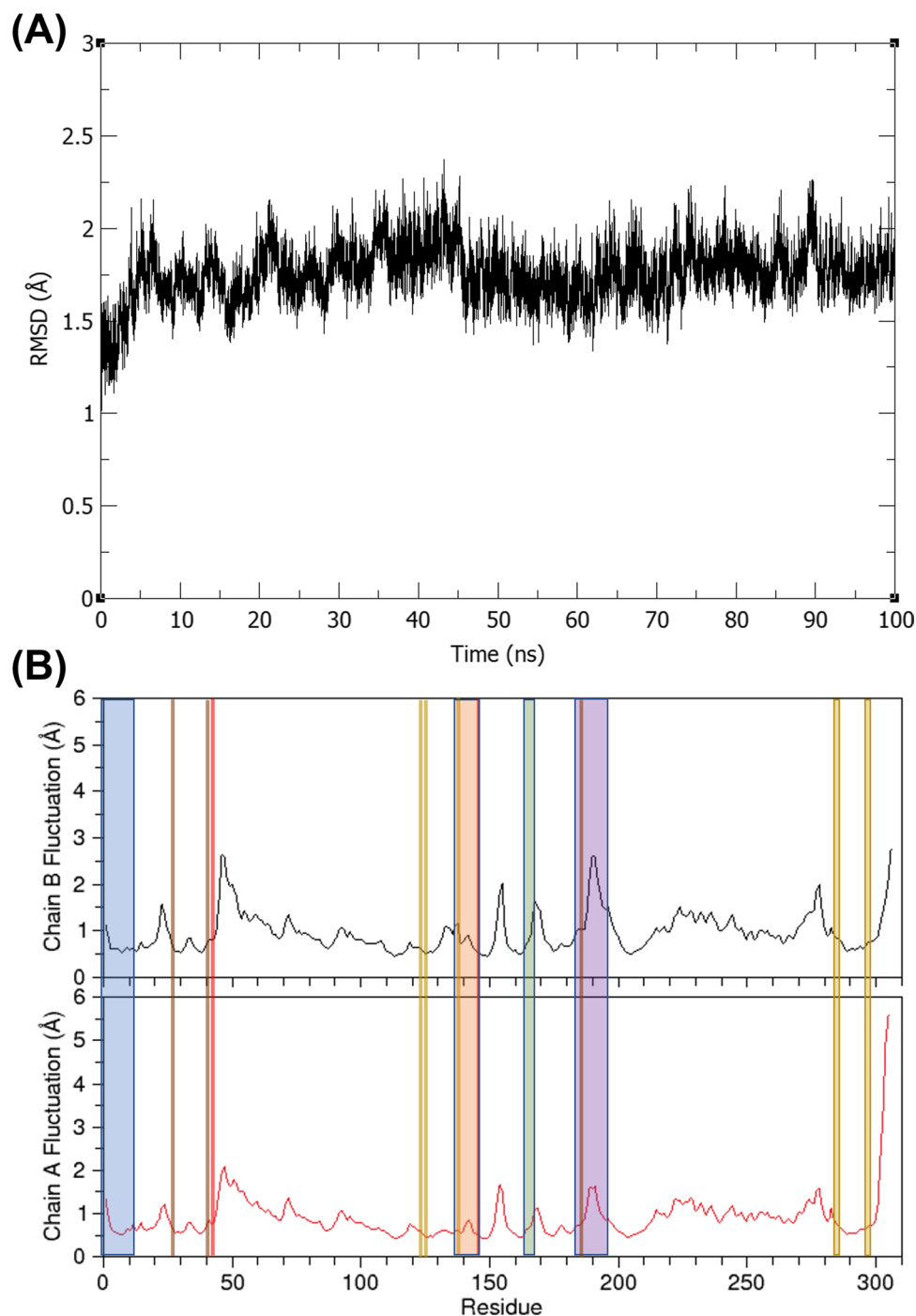
site of the Mpro involves a Cys145–His41 catalytic dyad, interactions with these residues served as the basis of the potential of these ligands. In particular, hindering the deprotonation of Cys145@SγH by His41@Nε can repress the proteolytic activity of the Mpro [36]. Thus, H-bond interactions formed or the proximity of a ligand H-bond donor with the His41@Nε atom (H-bond acceptor) were used to select the final hit that will be subjected to MD simulation (Supplementary Table 5).

Diosmetin-7-O-b-D-glucopyranoside (DG), a polyhydroxylated flavonoid conjugated with a glucose moiety, was seen to contain a hydroxyl H-bond donor group in close proximity with the His41@Nε atom with a distance of 3.05 Å (Fig. 3A and C). It was posited that the proximity of the H-bond donor from the His41@Nε atom may allow the favorable positioning of the atoms leading to the formation of H-bonds once the static system is allowed to progress dynamically in MD simulations. Aside from this,

Fig. 2 Orientation of His41 of the cluster representative structure (cyan) and the crystal structure (green), with respect to Cys145 and Asp187. The respective nitrogen atoms are labeled according to color: blue text for the cluster representative structure and green text for the crystal structure
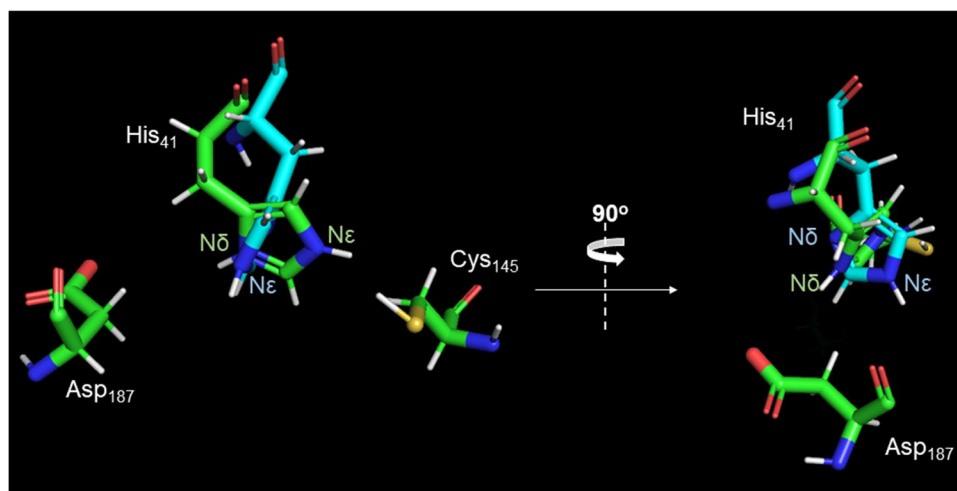

Fig. 3 **A** 3D binding orientation of DG in SARS-CoV-2 Mpro. The ligand is shown as sticks with green carbon atoms while the protein is in white cartoon with the interacting residues shown as sticks with white carbon. **B** 2D protein–ligand interaction of DG in the Mpro binding site. Polar interactions are shown in cyan, hydrophobic interactions are shown in light green, and negative and positive charged interactions are shown in red and blue, respectively. Hydrogen bonding between interacting partners is shown as magenta arrows. **C** Distance between DG hydroxyl group and His41@Nε atom exhibiting proximity for potential formation of polar interaction
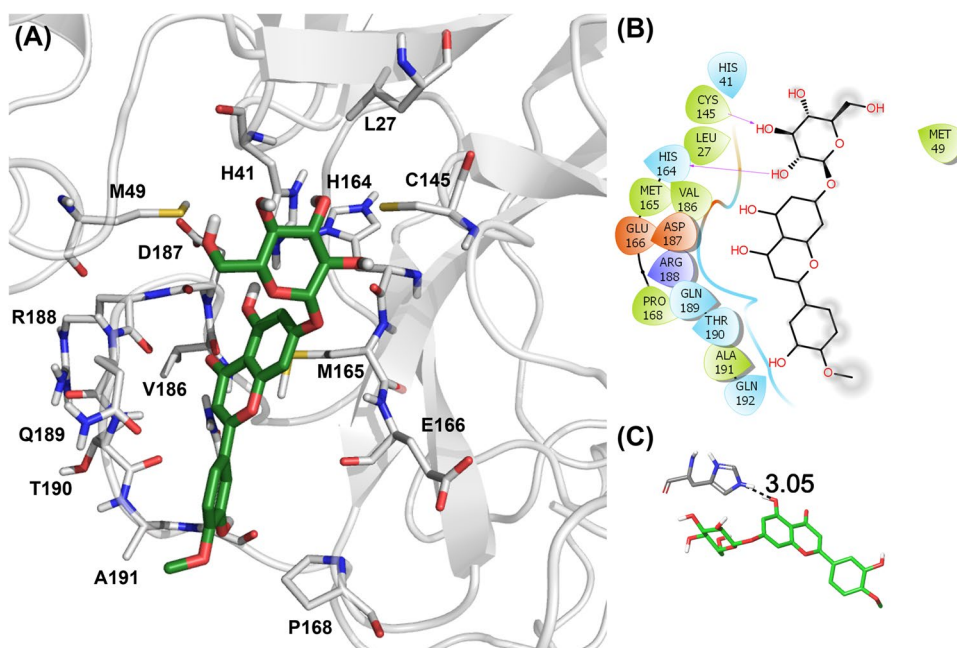
due to the polyhydroxyl substituents of DG, it was seen to form other hydrogen bond interactions with Cys145 and His164. Other interactions include charged contacts with Glu166, Asp187, and Arg188; polar interactions with His41, Gln189, Thr190, and Gln192; and hydrophobic interactions with Leu27, Met49, Met165, Pro168, Val186, and Ala191 (Fig. 3B). Important interactions with Mpro were made by DG, such as the interactions made to the S1/S2 subsites (Glu166, Leu27, Met49, Met165, and Pro168) and to the Domain II/III linker (Val186, Asp187, Arg188, Gln189, Thr190, Ala191, and Gln192). DG also formed an interaction with Asp187, which could possibly hinder the H-bond interactions between His41 and Asp187, thus lowering the pKa of the His41@Nε atom, decreasing its ability to deprotonate Cys145@SγH. DG is a natural product isolated from the flowers of *Chrysanthemum morifolium* and is recognized

for its antioxidant activity [37]. As such, different studies have already presented protocols for the extraction of this compound [38]. In fact, DG, in its molecular form, is already distributed by various sources for research purposes. With this, repurposing of DG as a viral protease inhibitor is highly feasible. Hence, because of the favorable positioning of DG in the Mpro active site and its commercial availability, DG was chosen as the top hit for further dynamics studies.

Molecular dynamics simulations of protein-ligand complexes

The initial 100 ns production run of the Mpro-DG complex was performed using the same parameters as the apo production run. The trajectory of DG in complex with the SARS-CoV-2 Mpro was analyzed to better understand its dynamic
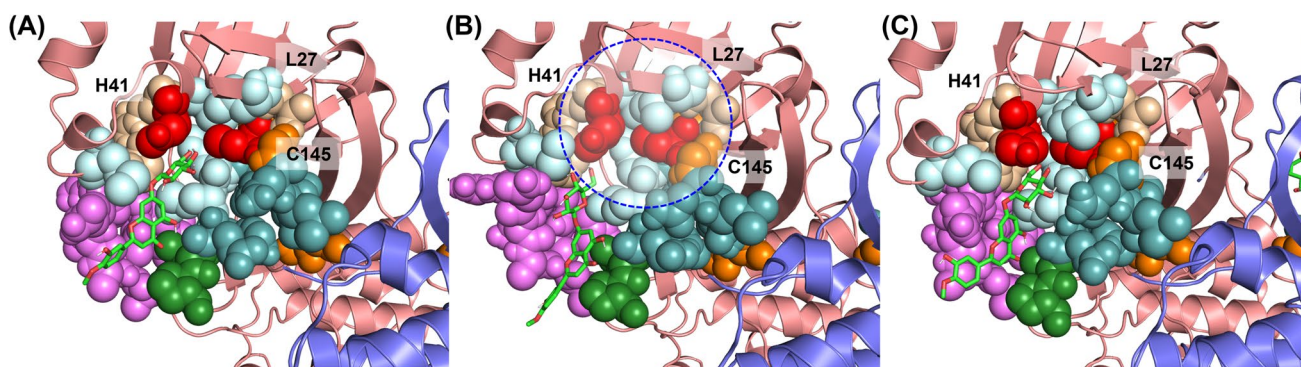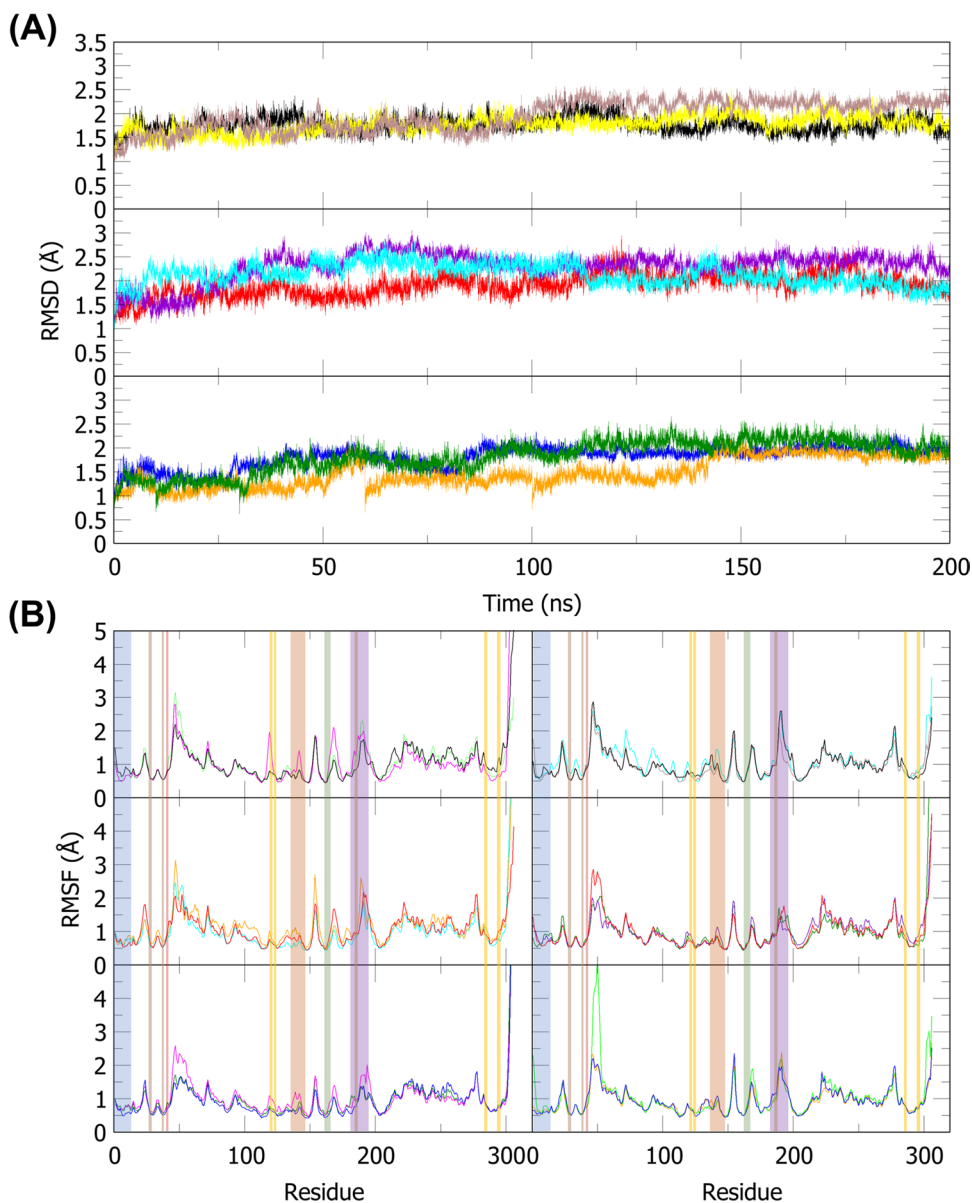
Fig. 4 Trajectory frame at **A** 30 ns, **B** 70 ns, and **C** 100 ns (first replicate) showing the changes in the side chain conformation of Leu27, His41, and Cys145 in the Mpro active site. The ligand is represented as sticks with green carbon atoms. Important active site residues are shown as spheres: catalytic dyad (red), residues interacting with the dyad (wheat), oxyanion loop (orange), substrate binding site (light teal, S1; pale cyan, S2; forest green, ß binding sheet), and the Domain II/III linker (violet)

Fig. 5 **A** Protein RMSD plots for 200 ns triplicates of Mpro apo (top), Mpro-N3 (middle), and Mpro-DG (bottom) production runs. **B** RMSF for trajectory triplicates of chain A (left panel) and B (right panel) for Mpro apo (top), Mpro-N3 (middle), and Mpro-DG (bottom)
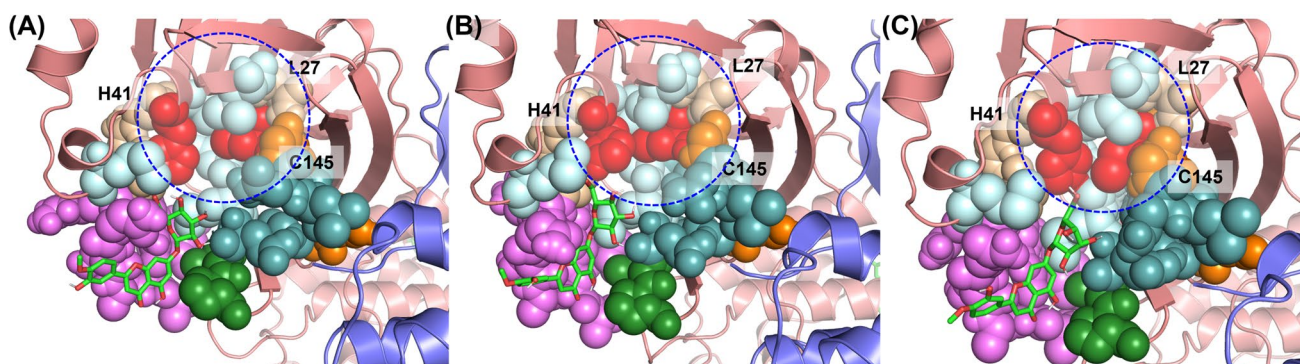
Fig. 6 Trajectory frame at **A** 130 ns, **B** 170 ns, and **C** 200 ns (first replicate) showing the changes in the side chain conformation of Leu27, His41, and Cys145 in the Mpro active site. The ligand is represented as sticks with green carbon atoms. Important active site resi-dues are shown as spheres: catalytic dyad (red), residues interacting with the dyad (wheat), oxyanion loop (orange), substrate binding site (light teal, S1; pale cyan, S2; forest green, ß binding sheet), and the Domain II/III linker (violet)

interactions with Mpro. To aid in the analysis of the potential of DG as a hit, the trajectory of the elucidated Mpro-N3 complex (PDB ID: 6LU7) [23] was used as the positive control due to its known inhibitory effects to Mpro in vitro. The protein RMSD of the Mpro-DG trajectory remained stable throughout the initial 100 ns simulation (Supplementary Fig. 1A). Interestingly, contrary to the positive control, the binding of DG to the Mpro was seen to slightly stabilize the enzyme structure as indicated by the lower and more stable RMSD, as well as lower RMSF values (Supplementary Fig. 1B). Noteworthy residues that were stabilized are Thr190 to Ala191 in the Domain II/III linker, Asp48 to Met49 in the S2 subsite, and the N-finger residues. These observations suggest that throughout the 100 ns run, DG can interact with these residues such that residues in the Domain II/III linker and Asp48 to Met49, which have high fluctuations in the apo structure, are stabilized. This contrasts with the positive control wherein the mentioned residues saw a greater fluctuation in the N3-bound protein. On the other hand, there were also some residues that were destabilized by the binding of DG, such as Thr23 to Thr24 which is the linker directly connected to Leu27 in the S2 subsite.

In order to more clearly understand the effect of the ligand on the active site residues, inspection of 10 ns frames of the MD trajectory was conducted (Supplementary Fig. 2). It was observed that from 0 to 20 ns, DG was positioned near the S1 subsite and the β-binding sheet. At 30 ns, however, the DG "head" was seen to slowly position closer to the catalytic dyad (Fig. 4A). This transition causes the ligand tail to start interacting with the Domain II/III linker and the S2 subsites as indicated by the changes in the conformations of those residues. From this point onwards, the ligand maintained a close proximity with the catalytic dyad. This proximity allows the imidazole ring of His41 to interact with the hydroxyl groups of DG. This interaction is indicated by the periodic shifts in orientation of the imidazole ring from

being parallel to the ligand at approximately 30 ns, 50 ns, and 70 ns, to being pointed outwards directly toward the ligand at approximately 40 ns, 60 ns, and 80 ns onwards (Supplementary Fig. 2). Interestingly, the outward point-ing of the His41 imidazole ring is only seen when DG approaches this moiety. Moreover, this results in the flat-tening of the imidazole ring, causing the distancing of the His41@Nε atom from Cys145. Consequently, this allows the bulky tert-butyl group of Leu27 in the S2 subsite to occupy the initial space of the imidazole ring (Fig. 4B). This block-ing of Leu27 is also supported by the increase in fluctuations of Thr23 and Thr24 from the RMSF calculations (Supple-mentary Fig. 1B), as they are the linker residues directly in contact with Leu27. Furthermore, another transition was seen between the 90 and 100 ns frame where DG switched its conformation such that its heterocyclic ring directly faced the His41 imidazole ring. This conformation allows more polar groups from the ligand to form interactions with the His41 imidazole, causing this ring to completely point out-wards at 100 ns. As a result, a fuller occupancy from the Leu27 tert-butyl group in the interaction space of His41 and Cys145 was seen (Fig. 4C). The repositioning of these residues leading to the repositioning of the His41 side chain away from Cys145 was not observed in both the apo protein and the positive control. Due to the observed transition near the end of the production run, the simulation was extended another 100 ns to determine if there are any more major changes observed.

The protein RMSD of the Mpro-DG trajectory remained stable throughout the extended simulation (Fig. 5A). The binding of DG continued to stabilize the Mpro structure as seen from the decrease in residue fluctuation in the RMSF plot (Fig. 5B). Residues in the S2 subsite, the N-finger region, and the Domain II/III linker stabilized as compared to the apo structure indicating the continued influence of DG in these regions. On the other hand, Thr23 and Thr24

Fig. 7 Betweenness central-
ity plot for the **A** covalent and
non-covalent interactions and **B**
non-covalent interactions only
of the Mpro apo (upper panel,
left: chain A, right: chain B),
Mpro-N3 (middle panel, left:
chain A, right: chain B), and
Mpro-DG (lower panel, left:
chain A, right: chain B) produc-
tion runs in triplicates
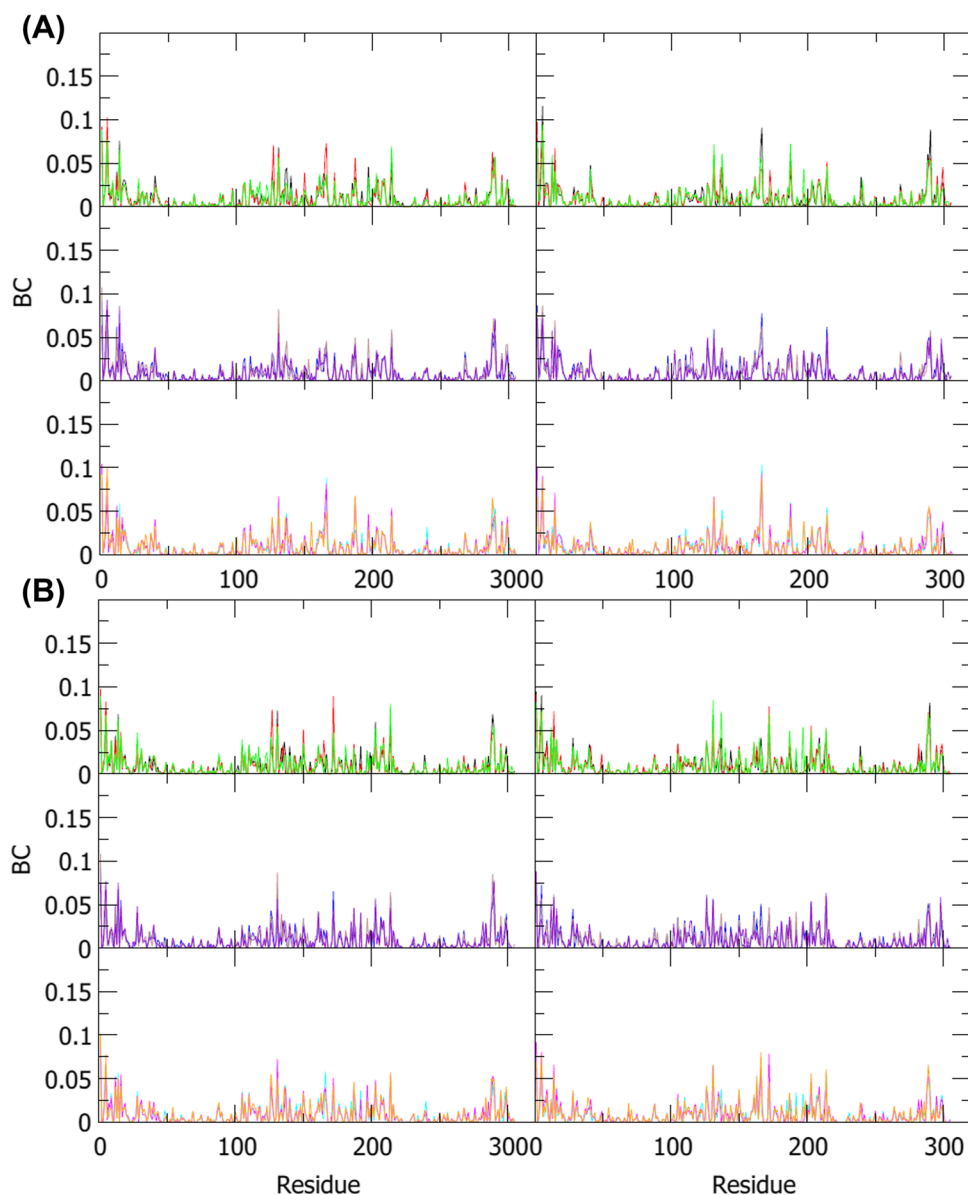
are still destabilized indicating the continued fluctuation of
Leu27 in the S2 subsite throughout the 200 ns simulation.

In the extended simulation of the first replicate, DG bind-
ing conformation initially allowed the continued distanc-
ing of the His41 side chain away from Cys145 by the same
mechanism as explained earlier (Supplementary Fig. 3).
However, at the 130 ns time frame, the ligand seemed to
have been pushed out from this pocket, forming interactions
with the S1 and S2 subsites and the Domain II/III linker
residues (Fig. 6A). This allowed the His41 side chain to
reorient itself pointing toward Cys145 from 130 to 160 ns,
while Leu27 shifted away from occupying the space between
the catalytic dyad residues, albeit still closer to the dyad than
in the apoprotein. Finally, from 170 to 200 ns, the hydroxyl
ends of DG regained interaction with the His41 side chain as
seen by the recurrence of the pointing away conformation of

the His41 imidazole at the 170 ns and the 200 ns time frames
(Fig. 6B and C). This caused the Leu27 tert-butyl group to
reoccupy the space between the dyad as well. These obser-
vations are in line with the DG RMSD showing increasing
fluctuation from when it shifted in conformations at 96 ns
before stabilizing again in a new binding pocket at 175 ns.
While a second MD run of the Mpro-DG complex showed
DG moving away from the original binding site and situating
itself in the Domain II/III linker (Supplementary Fig. 4 and
5), a third trial showed fluctuations in the S2 subsite, similar
to the first run, with Leu27 positioned between the dyad
(Supplementary Fig. 6 and 7), which can potentially block
the enzymatic activity of Mpro. In one trajectory replicate,
high fluctuation is observed for protomer B residues 47–51,
near the S2 subsite, which may be due to the DG "head" and
"tail" orienting toward the solvent (Supplementary Fig. 3

Table 1 Betweenness centrality value for covalent and non-covalent and non-covalent only interactions of notable residues for Mpro apo, Mpro-N3 complex, and Mpro-DG complex measured in triplicates

Chain-residue	Apo			Mpro-N3 complex			Mpro-DG complex		
	1	2	3	1	2	3	1	2	3
Covalent and non-covalent interactions									
A-His164	0.038	0.010	0.027	0.011	0.030	0.020	0.031	0.016	0.009
A-Glu166	0.031	0.073	0.038	0.041	0.046	0.034	0.089	0.082	0.051
A-Asp187	0.034	0.056	0.031	0.033	0.049	0.042	0.028	0.065	0.068
B-His164	0.010	0.017	0.013	0.035	0.030	0.028	0.009	0.013	0.026
B-Glu166	0.091	0.043	0.050	0.078	0.062	0.038	0.104	0.093	0.089
B-Asp187	0.061	0.045	0.072	0.035	0.031	0.042	0.060	0.058	0.042
Non-covalent interactions									
A-Glu166	0.0001	0.014	0.021	0.010	0	6.98E-05	0.058	0.040	0.026
A-His172	0.036	0.089	0.047	0.065	0.042	0.045	0.024	0.051	0.042
B-Glu166	0.042	0.007	0	0.051	4.29E-05	0.023	0.056	0.061	0.080
B-His172	0.049	0.078	0.068	0.038	0.047	0.014	0.038	0.079	0.046

and 4). The glucopyranoside head is a highly polar functional group, and thus, would not be able to interact well with the S2 subsite which is hydrophobic in nature, making the glucopyranoside group a good optimization point for further studies intended to improve the potency of DG. All in all, despite the fluctuations in the conformations accessed by DG in the Mpro binding pocket, DG was able to retain its interactions with the Mpro binding site and influence S2 subsite movement, thereby distancing the catalytic dyad from each other.

Network and H-bond analysis

To further understand the effect of DG binding on Mpro, network analysis was performed. Figure 7A shows that the betweenness centrality (BC) measure of the Mpro structure based on covalent and non-covalent interactions changed after DG binding, particularly for His164, Glu166, and Asp187 (Table 1). The higher covalent and non-covalent BC for these residues in chain A of the DG-bound complex suggest that Mpro relies on these residues for protein communication when DG is in the binding site. Given that Glu166 is part of the S2 subsite while Asp187 is found in the Domain II/III linker and can participate in the catalytic activity by

increasing the pKa of the His41@Nε atom, any changes in the intra-signals passing through these residues can affect the overall protein network. On the other hand, His164 may influence the movement and function of adjacent residues, Met165 and Glu166, or other nearby residues within the S1/S2 subsite of chain A, thereby exerting its influence over the flow of information within the protein. Interestingly, shortest paths analysis of the apo structure showed that communications of different residue partners pass through His164, such as the signal from His41 to Cys145 (His41-His164-Cys145) or from an S1 (e.g. Leu27) to an S2 (e.g. Glu166) subsite residue (Leu27-Cys145-His164-Met165-Glu166). This suggests that changes in BC for this residue can change the signal flow within the protein, and thus, influence the action of several residues in the binding site. The non-covalent BC analysis (Fig. 7B and Table 1) also shows that Glu166 is a critical residue for non-covalent communication in the DG-bound complex. In contrast, His164 and Asp187 did not show any large differences between the apo and DG-bound BC, suggesting that these residues are primarily used for communication via covalent interactions. Another residue with a significant change in non-covalent BC is His172. Shortest paths analysis showed that it can send signals from the S1 (e.g., Leu27) to the S2 (e.g., Glu166) subsite through

Table 2 Hydrogen bond donor and acceptor occupancy between Mpro and DG (left) and Mpro and N3 (right)

Donor	Acceptor	Average occupancy	Donor	Acceptor	Average occupancy
GLU166-Main	DG	18.60%	N3	GLU166-Main	109.30%
DG	GLU166-Main	11.12%	N3	GLN189-Side	46.77%
DG	HIS41-Side	8.14%	N3	SER144-Side	37.67%
			GLU166-Main	N3	22.81%

Main signifies that the main chain or the backbone participates in hydrogen bonding, while Side signifies that the residue side chain participates in hydrogen bonding

the path Leu27-Cys145-His163-His172-Ser139-Glu166. In the DG-bound protein, this path is shortened to Leu27-Cys145-His163-Glu166. These results indicate that DG can potentially influence the protein network by affecting signals sent through the binding site residues, and thus, potentially inhibiting normal enzymatic process.

Hydrogen bonding analysis was also performed to identify key H-bond interactions that were formed between the target and ligands during the production run (Table 2 and Supplementary Table 6 and 7). Glu166 acted as a donor and acceptor with 22.81% and 109.30% average occupancy, respectively, when it interacted with N3. The high occupancy percentage can be attributed to the presence of more than one H-bond that can be formed between N3 and the Glu166 amide. Moreover, N3 formed H-bonds with Gln189, which is found in the Domain II/III linker, and can therefore help stabilize the Mpro structure upon N3 binding. In total, N3 formed 32 to 35 H-bonds based on the analysis of the trajectory triplicates. Notably, the backbone of Glu166 was found to have 2 H-bond interactions with DG, wherein the Glu166 backbone acts as a donor and as an acceptor with 18.60% and 11.12% average occupancy, respectively. The observed interaction between the His41 side chain and DG showed an average occupancy of 8.14%. While these occupancy values may be lower compared to N3, DG had a more extensive H-bonding network with Mpro, forming 61 to 77 H-bonds based on the analysis of the trajectory triplicates. These results, when combined with those from the trajectory and network analysis, indicate that the binding of DG potentially plays a role in influencing the functional integrity of Mpro by altering the centrality of Glu166 from that of the native protein energy network.

Summary and conclusions

This study identified a potential inhibitor, DG, from a Philippine natural products database through molecular docking, ADME screening, and molecular dynamics simulations. DG displayed suitable binding and interactions with Mpro. It must be noted, however, that DG has a low GI absorption, is a PGP substrate, and is a CYP 450 inhibitor. Thus, further structural optimization must be conducted to improve its pharmacokinetics prior to in vivo testing.

From the trajectory analyses, it was seen that DG was able to cause conformational changes in the Mpro active site, such that the catalytic dyad is potentially distanced and blocked from each other. Furthermore, network and H-bond analyses showed that DG binding can potentially influence the native protein energy network of Mpro by changing the BC of important binding site residues, thereby disturbing communications between the dyad and the protein as a whole. Overall, these results indicate that DG has a high potential of becoming an inhibitor for SARS-CoV-2 Mpro protein, though in vitro or in vivo validation and further optimization are needed to confirm its activity and improve its properties, respectively.

Data availability Not applicable.

Code availability Not applicable.

Declarations

Conflict of interest The authors declare no competing interests.

References

1. Shereen MA, Khan S, Kazmi A, Bashir N, Siddique R (2020) COVID-19 infection: origin, transmission, and characteristics of human coronaviruses. J Adv Res 24:91–98. https://doi.org/10.1016/j.jare.2020.03.005
2. Worldometer (2022). Philippine Covid-19 data. https://www.worldometers.info/coronavirus/country/philippines/. Accessed 10 May 2022
3. Petrosillo N, Viceconte G, Ergonul O, Ippolito G, Petersen E (2020) COVID-19, SARS and MERS: are they closely related? Clin Microbiol Infect 26(6):729–734. https://doi.org/10.1016/j.cmi.2020.03.026
4. Chen Y, Liu Q, Guo D (2020) Emerging coronaviruses: genome structure, replication, and pathogenesis. J Med Virol 92(10):2249. https://doi.org/10.1002/jmv.26234
5. Hu T, Liu Y, Zhao M, Zhuang Q, Xu L, He Q (2020) A comparison of COVID-19. SARS and MERS PeerJ 8:e9725. https://doi.org/10.7717/peerj.9725
6. Lai A, Bergna A, Acciarri C, Galli M, Zehender G (2020) Early phylogenetic estimate of the effective reproduction number of SARS-CoV-2. J Med Virol 92(6):675–679. https://doi.org/10.1002/jmv.25723
7. Liu Y and Rocklov J (2021) The reproductive number of the Delta variant of SARS-CoV-2 is far higher compared to the ancestral SARS-CoV-2 virus. J Travel Med. 28(7). https://doi.org/10.1093/jtm/taab124

8. Zhou P, Yang XL, Wang XG, Hu B, Zhang L, Zhang W, Si HR, Zhu Y, Li B, Huang CL, Chen HD, Chen J, Luo Y, Guo H, Jiang RD, Liu MQ, Chen Y, Shen XR, Wang X, Zheng XS, Zhao K, Chen QJ, Deng F, Liu LL, Yan B, Zhan FX, Wang YY, Xiao GF, Shi ZL (2020) A pneumonia outbreak associated with a new coronavirus of probable bat origin. Nature 579(7798):270–273. https://doi.org/10.1038/s41586-020-2012-7

9. Harrison AG, Lin T, Wang P (2020) Mechanisms of SARS-CoV-2 transmission and pathogenesis. Trends Immunol 41(12):1100–1115. https://doi.org/10.1016/j.it.2020.10.004

10. World Health Organization (2022) Therapeutics and COVID-19: living guideline. World Health Organization: Geneva. https://www.who.int/publications/i/item/WHO-2019-nCoV-therapeutics-2022.4. Accessed 10 May 2022

11. Cully M (2022) A tale of two antiviral targets - and the COVID-19 drugs that bind them. Nat Rev Drug Discov 21(1):3–5. https://doi.org/10.1038/d41573-021-00202-8

12. US Food and Drug Adminstration (2022) Coronavirus (COVID-19) update: FDA authorizes new monoclonal antibody for treatment of COVID-19 that retains activity against Omicron variant. US FDA. https://www.fda.gov/news-events/press-announcements/coronavirus-covid-19-update-fda-authorizes-new-monoclonal-antibody-treatment-covid-19-retains. Accessed 10 May 2022

13. Reiter RJ, Sharma R, Simko F, Dominguez-Rodriguez A, Tesarik J, Neel RL, Slominski AT, Kleszczynski K, Martin-Gimenez VM, Manucha W, Cardinali DP (2022) Melatonin: highlighting its use as a potential treatment for SARS-CoV-2 infection. Cell Mol Life Sci 79(3):143. https://doi.org/10.1007/s00018-021-04102-3

14. Kouhpeikar H, KhosaravizadeTabasi H, Khazir Z, Naghipour A, MohammadiMoghadam H, Forouzanfar H, Abbasifard M, Kirichenko TV, Reiner Z, Banach M, Sahebkar A (2022) Statin use in COVID-19 hospitalized patients and outcomes: a retrospective study. Front Cardiovasc Med 9:820260. https://doi.org/10.3389/fcvm.2022.820260

15. Ullrich S, Nitsche C (2020) The SARS-CoV-2 main protease as drug target. Bioorg Med Chem Lett 30(17):127377. https://doi.org/10.1016/j.bmcl.2020.127377

16. Zhang L, Lin D, Sun X, Curth U, Drosten C, Sauerhering L, Becker S, Rox K, Hilgenfeld R (2020) Crystal structure of SARS-CoV-2 main protease provides a basis for design of improved alpha-ketoamide inhibitors. Science 368(6489):409–412. https://doi.org/10.1126/science.abb3405

17. Miller JK, Elenberg K, Dubrawski A (2022) Forecasting emergence of COVID-19 variants of concern. PLoS ONE 17(2):e0264198. https://doi.org/10.1371/journal.pone.0264198

18. Lin LT, Hsu WC, Lin CC (2014) Antiviral natural products and herbal medicines. J Tradit Complement Med 4(1):24–35. https://doi.org/10.4103/2225-4110.124335

19. El Sayed KA (2000) Natural products as antiviral agents. In: Atta ur R (ed) Studies in Natural Products Chemistry. Elsevier, pp 473–572

20. Isidoro C, Chiung-Fang Chang A, Sheen LY (2022) Natural products as a source of novel drugs for treating SARS-CoV2 infection. J Tradit Complement Med 12(1):1–5. https://doi.org/10.1016/j.jtcme.2022.02.001

21. Bernstein FC, Koetzle TF, Williams GJ, Meyer EF Jr, Brice MD, Rodgers JR, Kennard O, Shimanouchi T, Tasumi M (1977) The Protein Data Bank: a computer-based archival file for macromolecular structures. J Mol Biol 112(3):535–542. https://doi.org/10.1016/s0022-2836(77)80200-3

22. Van Der Spoel D, Lindahl E, Hess B, Groenhof G, Mark AE, Berendsen HJ (2005) GROMACS: fast, flexible, and free. J Comput Chem 26(16):1701–1718. https://doi.org/10.1002/jcc.20291

23. Jin Z, Du X, Xu Y, Deng Y, Liu M, Zhao Y, Zhang B, Li X, Zhang L, Peng C, Duan Y, Yu J, Wang L, Yang K, Liu F, Jiang R, Yang X, You T, Liu X, Yang X, Bai F, Liu H, Liu X, Guddat LW, Xu W, Xiao G, Qin C, Shi Z, Jiang H, Rao Z, Yang H (2020) Structure of M(pro) from SARS-CoV-2 and discovery of its inhibitors. Nature 582(7811):289–293. https://doi.org/10.1038/s41586-020-2223-y

24. Schrödinger (2021) Maestro, Schrödinger, LLC: New York, NY.

25. Goodsell DS, Morris GM, Olson AJ (1996) Automated docking of flexible ligands: applications of AutoDock. J Mol Recognit 9(1):1–5. https://doi.org/10.1002/(sici)1099-1352(199601)9:1%3c1::aid-jmr241%3e3.0.co;2-6

26. O'Boyle NM, Banck M, James CA, Morley C, Vandermeersch T, Hutchison GR (2011) Open Babel: an open chemical toolbox. J Cheminform 3:33. https://doi.org/10.1186/1758-2946-3-33

27. Hanwell MD, Curtis DE, Lonie DC, Vandermeersch T, Zurek E, Hutchison GR (2012) Avogadro: an advanced semantic chemical editor, visualization, and analysis platform. J Cheminform 4(1):17. https://doi.org/10.1186/1758-2946-4-17

28. 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(2):455–461. https://doi.org/10.1002/jcc.21334

29. Duan J, Dixon SL, Lowrie JF, Sherman W (2010) Analysis and comparison of 2D fingerprints: insights into database screening performance using eight fingerprint methods. J Mol Graph Model 29(2):157–170. https://doi.org/10.1016/j.jmgm.2010.05.008

30. Schrödinger Canvas [Computer Software] (2021) Schrödinger, LLC: New York, NY

31. 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. https://doi.org/10.1038/srep42717

32. DeLano S (2020) PyMol, Schrödinger, LLC: New York, NY.

33. Humphrey W, Dalke A, Schulten K (1996) VMD: visual molecular dynamics. J Mol Graph. 14(1):33–38. https://doi.org/10.1016/0263-7855(96)00018-5

34. Sercinoglu O, Ozbek P (2018) gRINN: a tool for calculation of residue interaction energies and protein energy network analysis of molecular dynamics simulations. Nucleic Acids Res 46(W1):W554–W562. https://doi.org/10.1093/nar/gky381

35. Baell JB, Holloway GA (2010) New substructure filters for removal of pan assay interference compounds (PAINS) from screening libraries and for their exclusion in bioassays. J Med Chem 53(7):2719–2740. https://doi.org/10.1021/jm901137j

36. Swiderek K, Moliner V (2020) Revealing the molecular mechanisms of proteolysis of SARS-CoV-2 M(pro) by QM/MM computational methods. Chem Sci 11(39):10626–10630. https://doi.org/10.1039/d0sc02823a

37. Matsuda H, Morikawa T, Toguchida I, Yoshikawa M (2002) Structural requirements of flavonoids and related compounds for aldose reductase inhibitory activity. Chem Pharm Bull (Tokyo) 50(6):788–795. https://doi.org/10.1248/cpb.50.788

38. Wang YJ, Yang XW, Guo QS (2008) Studies on chemical constituents in Huangjuhua (flowers of *Chrysanthemum morifolium*). Zhongguo Zhong Yao Za Zhi 33(5):526–530