



# Allosteric Modulation of the Main Protease ( $M^{Pro}$ ) of SARS-CoV-2 by Casticin—Insights from Molecular Dynamics Simulations

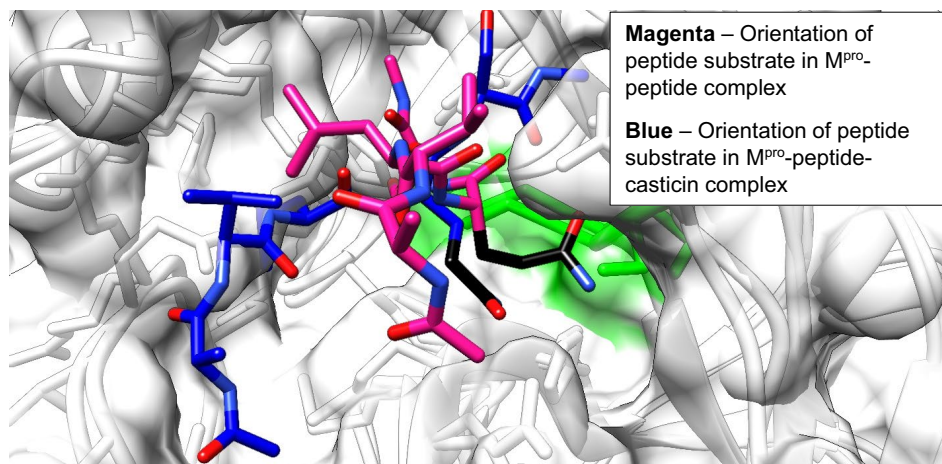
Jehoshaphat Oppong Mensah<sup>1</sup> · Gilbert Boadu Ampomah<sup>2</sup> · Edward Ntim Gasu<sup>1,2</sup> · Abigail Kusiwaa Adomako<sup>1</sup> · Elliott Sarpong Menkah<sup>1</sup> · Lawrence Sheringham Borquaye<sup>1,2</sup>

Received: 1 April 2022 / Accepted: 15 June 2022 / Published online: 7 July 2022  
© The Tunisian Chemical Society and Springer Nature Switzerland AG 2022

## Abstract

Inhibition of the main protease ( $M^{Pro}$ ) of SARS-CoV-2 has been suggested to be vital in shutting down viral replication in a host. Most efforts aimed at inhibiting  $M^{Pro}$  activity have been channeled into competitive inhibition at the active site, but this strategy will require a high inhibitor concentration and impressive inhibitor- $M^{Pro}$  binding affinity. Allosteric inhibition can potentially serve as an effective strategy for alleviating these limitations. In this study, the ability of antiviral natural products to inhibit  $M^{Pro}$  in an allosteric fashion was explored with *in silico* techniques. Molecular docking revealed a strong interaction between casticin, an antiviral flavonoid, and  $M^{Pro}$  at a site distant from the active site. This site, characterized as a distal site, has been shown to have an interdependent dynamic effect with the active site region.  $M^{Pro}$  only,  $M^{Pro}$ -peptide (binary) and  $M^{Pro}$ -peptide-casticin (ternary) complexes were subjected to molecular dynamics simulations for 50 ns to investigate the modulatory activity of casticin binding on  $M^{Pro}$ . Molecular dynamic simulations revealed that binding of casticin at the distal site interferes with the proper orientation of the peptide substrate in the oxyanion hole of the active site, and this could lead to a halt or decrease in catalytic activity. This study therefore highlights casticin as a potential allosteric modulator of the SARS-CoV-2 main protease, which could be optimized and developed into a potential lead compound for anti-SARS-CoV-2 chemotherapy.

## Graphical Abstract



**Keywords** Covid-19 chemotherapy · Natural products · Molecular docking · Distal site · Oxyanion hole · Proteolysis

✉ Lawrence Sheringham Borquaye  
slborquaye@yahoo.com; lsborquaye.sci@knust.edu.gh

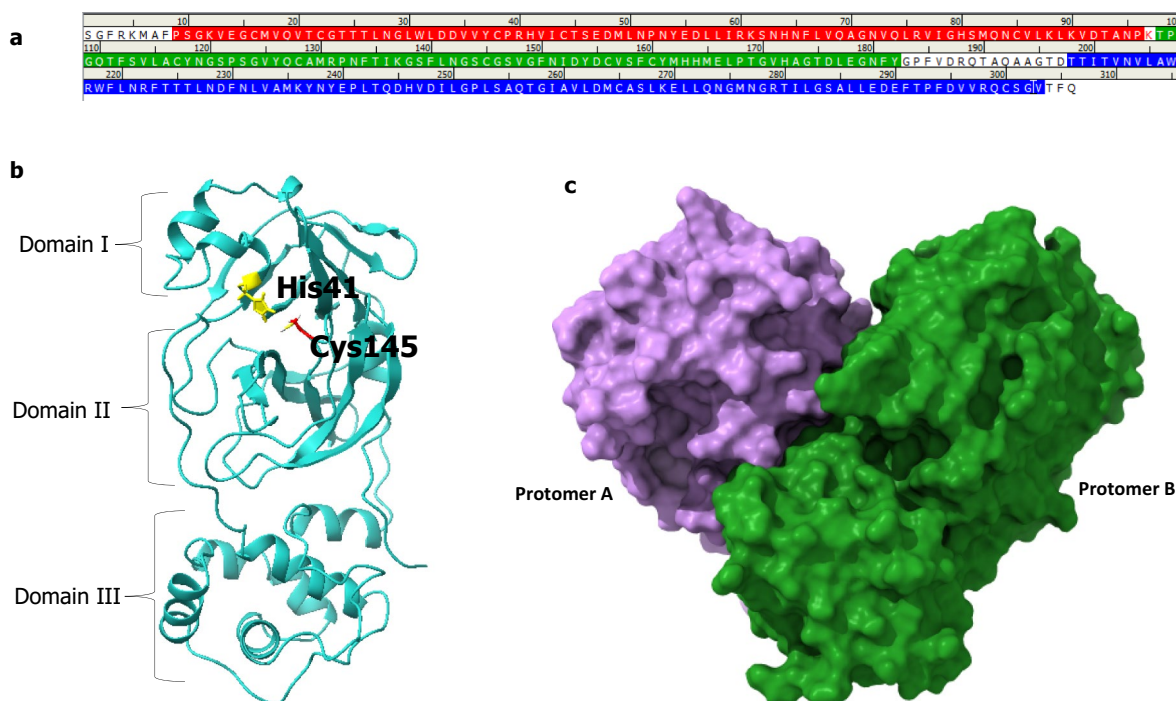
Extended author information available on the last page of the article

## 1 Introduction

The novel coronavirus was declared a global pandemic after its devastating effect was felt in most countries across the world [1]. Its adverse impact on the health, social, and economic sectors has pressed the need for rapid solutions. With intense ongoing scientific research resulting in the development of vaccines, the rapid mutation of the virus poses another major concern. As a result, efficacious therapeutics with lower side effects are required to combat present and future viral strains. Coronaviruses can be classified into four main subgroups—alpha, beta, gamma, and delta—based on their genomic structure [2]. These viruses are well noted for their structural [spike(S), nucleocapsid (N), matrix (M), envelope (E)] and non-structural proteins (nsp1 to nsp16) [3]. Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) belongs to the family Coronaviridae and the genus *Betacoronavirus*. It is characterized as a positive-sense, single-stranded RNA virus with a genome sequence length of about 32 kb [4]. The proteins encoded by SARS-CoV-2 function together to aid in transcription, replication, transmission and virulence. SARS-CoV-2 possesses over 50% genome sequence similarity with both SARS-CoV and MERS-CoV [5].

The main protease ( $M^{pro}$ ) of SARS-CoV-2 is an important enzyme for cleaving translated polyproteins from the viral RNA. Proteolysis occurs at 11 distinct cleavage sites on the translated polyprotein, with Leu-Gln↓ (Ala, Gly, Ser) specified as the recognition sequence [6].  $M^{pro}$  has two protomers, with each protomer possessing three domains, domains I, II, and III. Domains I (residues 10–99) and II (residues 100–182) are made up of six-stranded anti-parallel barrels with the substrate-binding site positioned in the middle of the barrel. Domain III (residues 198–303), which is characterized by a globular cluster of five alpha-helices, is responsible for regulating protease dimerization [7] (Fig. 1).

$M^{pro}$  induces its catalytic activity in a dimeric state, with His41 and Cys145—a catalytic dyad situated in the gap between domains I and II—playing a crucial role in proteolysis [6]. Some highly conserved binding sites—S1', S1, S2, S4—are also present in the catalytic machinery, with the S1 site specifying the enzyme specificity for glutamine at the P1 position of the peptide substrate. During the first step of the hydrolysis reaction catalyzed by  $M^{pro}$ , Cys145 serves as a nucleophile with His41 acting as a base catalyst. An oxyanion hole is generated by the backbone amido groups of Gly143, Ser144, and Cys145, and these residues are arranged in the so-called oxyanion loop (residues 138–146) to stabilize the partial negative charge developed at the P1



**Fig. 1** **a** Sequence map of  $M^{pro}$  showing the various domains. Domain I (residues 10–99) is colored red, Domain II (residues 100–182) blue, and Domain III (residues 198–303) green. **b** 3D structure of SARS-CoV-2  $M^{pro}$  marking specified domains. His41 is indicated

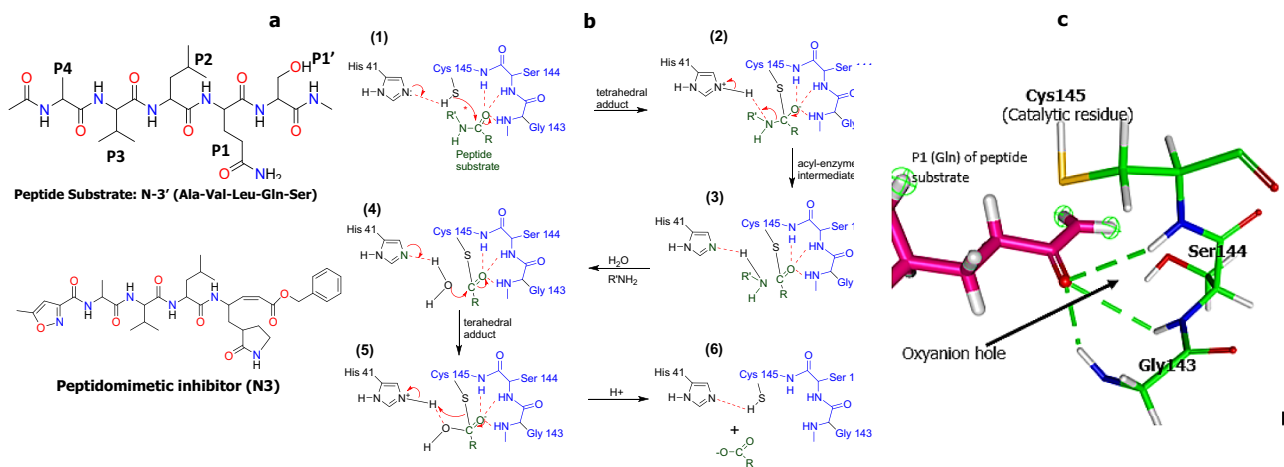
as yellow and Cys145 as red. **c** Surface representation of a dimer SARS-CoV-2  $M^{pro}$  showing protomer A (purple) and protomer B (green)

carbonyl group of the peptide substrate during the hydrolysis of the P1–P1' bond (Fig. 2) [8]. Because there are no known human proteases with a similar cleavage specificity, the M<sup>Pro</sup> has become an attractive target for developing potential SARS-CoV-2 inhibitors [9].

The viability of M<sup>Pro</sup> as a potential drug target has seen efforts channeled towards its inhibition by experimental [10] and computational strategies [11]. These efforts have focused on the development of inhibitors that compete with the native substrate at the active site. The effectiveness of these inhibitors could however be limited by competition with the natural substrate or issues of affinity [12]. Exclusion of electrophiles from potential lead compounds for reasons of safety and other negative effects such as carcinogenesis has also limited the pool of compounds available for development [13]. Interrogation of compounds capable of allosteric modulation of M<sup>Pro</sup> has, thus, become an attractive option that is being earnestly pursued. El-Baba and his colleagues, for example, employed fragment molecules that bind to distant areas from the active site to inhibit M<sup>Pro</sup> [14]. Sztain and coworkers also used Gaussian Accelerated Molecular Dynamics to identify and characterize three areas of M<sup>Pro</sup>, namely the active site, dimer interface, and distal site, which could serve as druggable sites for inhibition [15]. The distal site, found in the C-terminal region, was

shown to be highly druggable while having an interdependent dynamic effect on the active site, which is located in the N-terminal region [15, 16]. Allosteric inhibition, via the proposed distal site of M<sup>Pro</sup>, could serve as a new strategy for targeting SARS-CoV-2.

Herbal remedies have existed in folkloric medicine for a long time [17–19]. Some of the plants used in such herbal remedies provide elegant structural scaffolds that can serve as lead compounds in drug discovery efforts. Natural products have served well in this regard, and various successes in other therapeutic areas—artemisinin and quinine for malaria, taxol for cancer—have encouraged a search for antiviral agents from natural products. A major strategy utilized in the identification of potential therapeutic agents against the covid-19 disease is repurposing, where either old drugs or existing compounds with antiviral activities are screened against the virus. Several natural products have been reported to possess significant antiviral activities and could therefore serve as a starting point for such screens. The alkaloids of *Cryptolepis sanguinolenta* have been shown to potentially inhibit some of the viral proteins of SARS-CoV-2 in an in silico study [20]. Jo and coworkers also showed that some natural flavonoids possess significant activity against SARS-CoV-2 in an experimental study that utilized a Fluorescence



**Fig. 2** **a** Structures of peptide substrate with labels of the SARS-CoV-2 M<sup>Pro</sup> proteolytic cleavage sites, and a peptidomimetic inhibitor (N3). **b** Molecular details SARS-CoV-2 M<sup>Pro</sup> polypeptide hydrolysis in the oxyanion hole (blue structures, Gly143, Ser144, Cys145), (1) Substrate (green) binds to the active site such that the thiol group of Cys145 is positioned to attack the substrate carbonyl carbon. Simultaneously, the carbonyl oxygen hydrogen bonds to three -NH groups of the protein and the histidine removes a proton from the cysteine -SH, thus activating it. The three -NH groups are part of the peptide backbone forming a loop called the oxyanion hole (colored in blue). (2) In the transition state for the acyl-enzyme intermediate, negative charge has developed on the carbonyl oxygen; this is also stabilized by hydrogen bonding in the oxyanion hole. The first product is beginning to be released in this step and the bond between

the carbonyl carbon of the substrate and the cysteine sulphur is beginning to form. (3) The acyl-intermediate. In this intermediate, the cysteine side chain is covalently bound to the carbonyl carbon of the substrate. The histidine gives up the proton it acquired from the cysteine to the nitrogen of the substrate. (4) In the transition state for hydrolysis of the acyl-enzyme intermediate, a water molecule attacks the carbonyl carbon of the acyl-enzyme ester as the histidine accepts a proton from it, analogous to the acylation step (2). (5) The transient negative charge on the oxygen of the acyl-enzyme is again stabilized by hydrogen bonding in the oxyanion hole. (6) Formation of product. The transient state collapses to release the other portion of the substrate and regenerate the cysteine -SH. **c** Figure showing the oxyanion hole of M<sup>Pro</sup>. Glutamine of substrate shown in magenta. Backbone NH groups of the oxyanion hole residues are colored blue

Resonance Energy Transfer (FRET) protease assay. This was further corroborated by an *in silico* study, which further highlighted baicalin, herbacetin, and pectolarin as potential inhibitors of SARS-CoV-2 M<sup>Pro</sup> [21]. The use of computational tools to facilitate lead identification and optimization, and also to tease out plausible mechanisms of actions of lead compounds exist in the literature [22]. In particular, molecular docking and molecular dynamics simulations have proved very useful in this regard [23]. Molecular docking was used to identify potent inhibitors of Aurora Kinase A (AKA) and this resulted in the identification of the (4-methoxy-pyrimidin-2-yl)-phenyl-amine scaffold as a promising warhead. Using a computer-aided drug design (CADD) protocol involving structure-based virtual screening, *de novo* design and free energy perturbations, the study served as a starting point for the discovery of anticancer therapeutics [24]. Molecular docking and MD simulations have also been used to tease out new targets for HIV-1 therapeutics [25].

The *Hibiscus* plant is widely used in ethnomedicine to treat a wide range of diseases. Pharmacological evidence has shown that several species from the genus *Hibiscus* possess anti-diarrhoeic, antihypertensive, anti-inflammatory, antipyretic, hepaprotective, anti-spermatogenic, antihelminthic, anti-tumor, antidiabetic, anticonvulsant, immunomodulatory, antioxidant, antimutagenic and antiviral properties. This study therefore aimed to evaluate the potential of selected natural product-derived compounds to interfere with the normal functioning of the M<sup>Pro</sup> of SARS-CoV-2 as a means of identifying structural scaffolds that could serve as lead compounds using *in silico* tools. Fifteen isolated compounds (gossypin, canthin-6-one, hibiscanal, myriceric acid, syriancusin A, syriancusin B, hibiscatin, gossypetin, cyanin, hibiscatin, rutin) from various plants of the genus *Hibiscus* were selected for this study based on their reported antiviral action [26–28]. Isolated compounds with biological activities from other plants were also considered. Cynaratriol and 1,4-benzoquinone, which are active compounds of *Cynara scolymus* and *Embelia ribes* respectively, are known inhibitors of angiotensin converting enzymes (ACE) [29]. Casticin, present in most plants from the genus *Vitex*, is a known immunomodulatory compound and an inhibitor of the Lassa virus [30] and was therefore also included in the study. Casticin was observed to interact with M<sup>Pro</sup> primarily at its distal site and hence was hypothesized to be capable of modulating the activity of the enzyme through a possible allosteric mechanism. Molecular dynamics simulation was used to investigate the effect of casticin binding on events at the active site in the enzyme. We herein report that casticin binding at the distal site distorts the orientation of the peptide substrate, and this could markedly decrease or potentially, abolish enzyme catalytic activity.

## 2 Methods

### 2.1 Pharmacokinetic Properties

The structures of the compounds used are shown in Fig. 3.

The SWISS ADME webserver was used to analyze the pharmacokinetic properties of the compounds [31]. Lipinski's rule of five was employed in assessing the drug-likeness of the selected compounds [32]. Drug-likeness properties investigated included absorption, distribution, metabolism, and excretion. To be drug-like, a compound was supposed to have a molecular weight < 500, logP < 5, hydrogen-bond donors < 5, and hydrogen-bond acceptors < 10. Compounds that had 0 violation of these requirements were predicted to be most drug-like.

#### 2.1.1 Ligand Preparation and Substrate Peptide Modeling

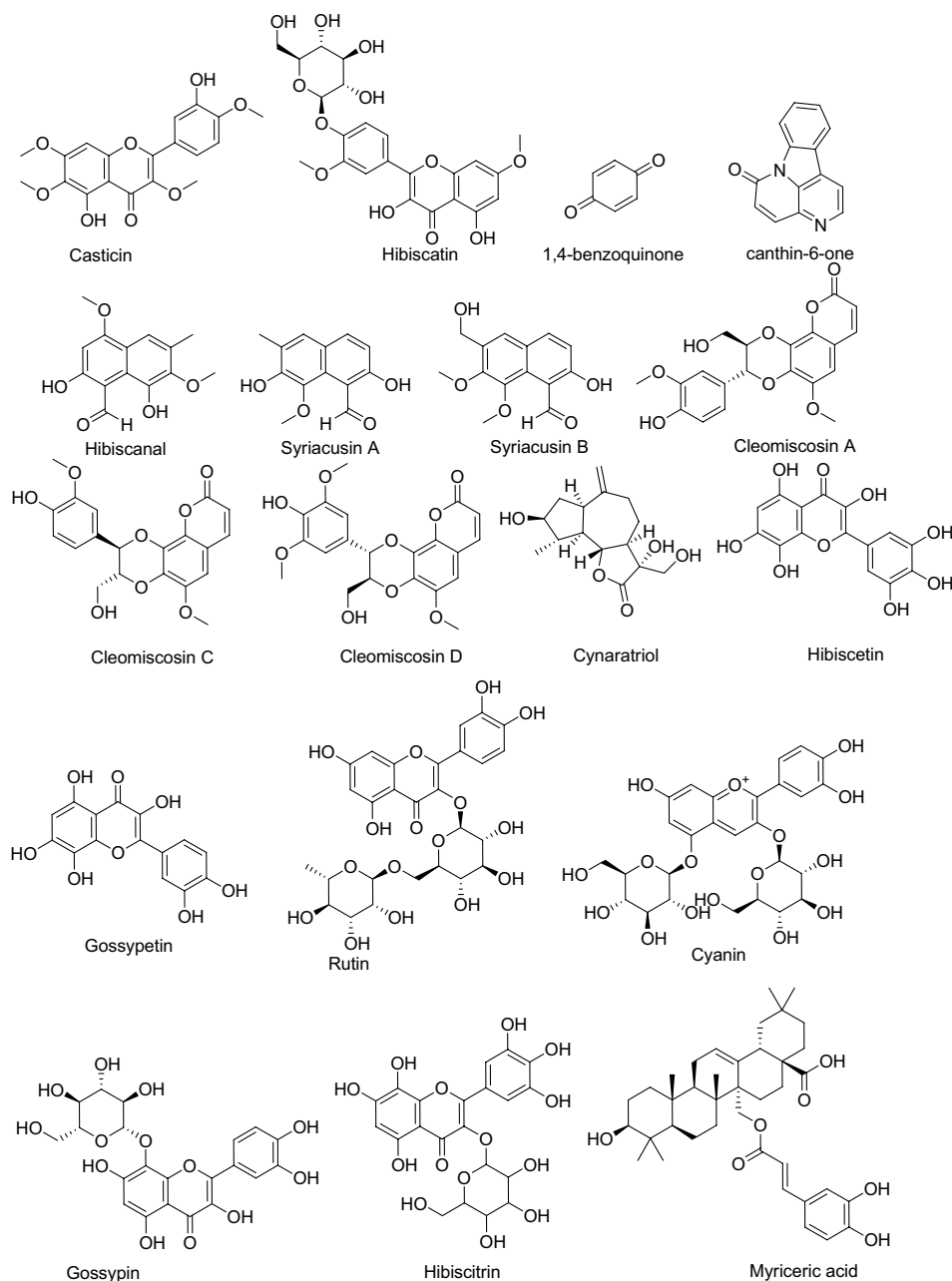
The three-dimensional structures of the library of selected “drug-like” compounds were modeled and optimized using quantum mechanical methods as reported elsewhere [20, 33]. To monitor events at the substrate-binding site of M<sup>Pro</sup>, a peptide substrate was required. A peptide that mimics the natural substrate of M<sup>Pro</sup> and bears the recognition sequence needed for catalytic activity Ala-Val-Leu-Gln-Ser was designed. The residues of the peptide were represented as P4, P3, P2, P1 and P1' respectively. This peptide (Fig. 2) was labeled N-3' (due to its similarity with N3, a known peptidic inhibitor of M<sup>Pro</sup>). The energy of N-3' was minimized in Discovery Studio 2017 R2 client (BIOVIA, Discovery Studio Visualizer, [2017 R2 Client], San Diego: Dassault Systèmes, (2017)) with the CHARMM force field. All ligands were prepared for docking by adding polar hydrogens, merging non-polar hydrogens, and calculating Gasteiger charges for all atoms [20].

#### 2.1.2 Protein Selection and Preparation

6WNP (X-ray Structure of SARS-CoV-2 main protease bound to boceprevir at 1.45 Å) was obtained from the Protein Data Bank (PDB). To prepare the protein for docking, all crystallographic water molecules and co-crystallized ligands were removed, polar hydrogens added, and Gasteiger charges calculated for all receptor atoms. All these were achieved with AutoDockTools-1.5.7rc1 [34].

#### 2.1.3 Molecular Docking

Receptors used for molecular docking simulations included the SARS-CoV-2 main protease under the PDB accession codes 6LU7 and 6WMP. The *apo* protein was obtained by

**Fig. 3** Chemical structures of compounds used in this study

deleting all co-crystallized ligands and solvents. Global docking was performed with Autodock Vina embedded in PyRx to ascertain the preferential binding modes and their respective binding affinities of the ligands. Ligand candidates, yielding a binary complex were chosen based on the highest binding affinity at the binding site which contained majority of the predicted ligand conformations. Following these considerations, casticin was ranked best, and molecular docking with AutoDock Vina in UCSF Chimera was used to model ternary complexes. The peptide substrate from 6LU7 was initially docked into the substrate-binding site of 6WMP. To generate the ternary complex, casticin was subsequently docked to the 6WMP-peptide complex using the following

grid dimensions: center:  $x = -13.4736$ ,  $y = 10.7715$ ,  $z = -0583865$  and size:  $x = 23.4867$ ,  $y = 17.6548$ ,  $z = 20.9149$ . The selected casticin pose for further MD analysis was based on interactions with at least one of the residues, namely Ser284, Ala285 and Leu286.

## 2.2 Molecular Dynamics Simulation

### 2.2.1 System Preparation and MD Production

Complexes considered for MD simulations were built in UCSF Chimera. The peptide substrate was placed into the active site by superimposing the 6LU7-peptide complex



obtained from Suarez et al. work [35] with 6WNP. The `gmx pdb2gmx` tool was used to construct protein topologies using the CHARMM 36 all-atom force field in GROMACS v. 2018.2 on a local workstation. The CgenFF webserver was used to produce ligand topologies [36]. A high performance and parallel GROMACS v.2018.6 [37] module were used for all MD production runs. This was pre-compiled on the Lengau cluster (Center For High-Performance Computing, Cape Town, South Africa). Complexes were solvated with the TIP3P water model in a dodecahedron box set. Sodium and chloride ions were added for system-specific neutralization. The steepest descent algorithm was used to minimize energy without constraints over 50,000 steps until a tolerance of 10 kJ/mol was reached. The systems were then pressure-equilibrated at 1 atm for 100 ps using the Berendsen barostat after temperature equilibration at 300 K using V-rescale (modified Berendsen thermostat) for 100 ps under constant number and volume. For positional constraints, the linear constraints solver (LINCS) algorithm was utilized, while for long-range electrostatics, the particle-mesh Ewald technique was used. Coulombs and van der Waal's cut-offs were set to 1.2 nm. The time step for the production run was set at 2 fs, writing trajectories every 10 ps. Finally, periodic boundary conditions (pbc) were set before every production run of 50 ns [38].

**2.2.1.1 Trajectory Analysis, Visualization, and Plotting** The total pressure, temperature, potential and kinetic energies of all systems were monitored before the production run. The `gmx trjconv` tool was used to correct all output trajectory files prior to further analyses. Root mean square deviation (RMSD), radius of gyration (Rg), and root mean square fluctuation were computed using `gmx rms`, `gmx gyrate` and `gmx rmsf` tools, respectively [39]. Visualization of trajectories was performed with VMD 1.9.3 [40] using the generated structural (.gro) and pbc-corrected compressed trajectory (.xtc) files of all complexes. Molecular graphic images were generated with UCSF Chimera and DS where necessary. Datasets were plotted with XMGRACE.

**2.2.1.2 Binding Free Energy Calculations** The average binding energies were estimated with `g_mmpbsa` implemented in GROMACS. The estimated binding energies comprise molecular mechanical ( $E_{MM}$ ) as well as polar ( $G_{pol}$ ) and apolar ( $G_{apol}$ ) solvation terms.  $E_{MM}$  calculates the molecular mechanics contribution of each system in vacuum and this consists of bonding (dihedrals, bond stretching, angle bending) and non-bonding (electrostatics and van der Waal's) interactions. For  $G_{pol}$ , `g_mmpbsa` relies on solving for the Poisson–Boltzman Surface Area (PBSA) while the solvent accessible surface area (SASA) model was used for  $G_{apol}$  estimations. Available python scripts (MmPbSaDecomp.py and MmPSaStat.py) were used for these calculations [41].

## 3 Results

### 3.1 Pharmacokinetic Predictions for Ligands

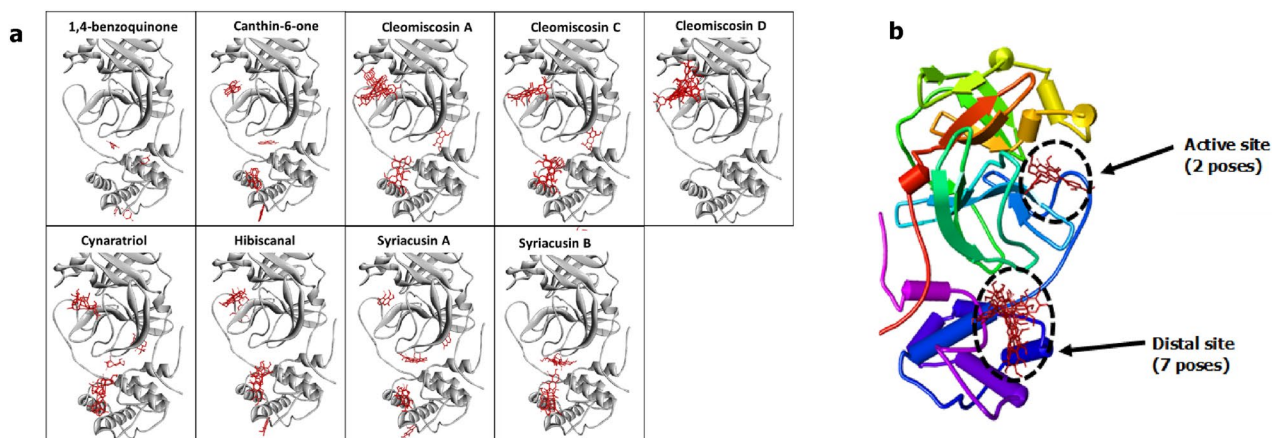
To be considered drug-like, a compound should not violate more than one of the following requirements set by Lipinski; molecular weight < 500, logP < 5, hydrogen-bond donors < 5, and hydrogen-bond acceptors < 10 [32]. A summary of the pharmacokinetic properties of the compounds is presented in Table 1. For this work, a strict criterion of zero (0) violation was used for ligand candidate selection. The logP values of most of the compounds were less than 5, ranging from  $-3.89$  to  $2.41$ . Myriceric acid was however, an exception, with a logP value of  $5.07$ . Two glycosides, rutin and cyanin, as well as myriceric acid (with 6 rings), had molecular weights greater than 500. The rest of the compounds had molecular weights of less than 500. Based on the hydrogen-bond donor requirement, gossypetin, hibiscatin, hibiscetin, rutin, cyanin and gossypin failed the set criteria, since they possessed more than 5 hydrogen bond donor groups. Hibiscatin, rutin, cyanin, gossypin and hibiscitrin had more than 10 hydrogen acceptors, violating the hydrogen bond acceptor rule. In all, 8 compounds—myriceric acid, rutin, cyanin, gossypetin, hibiscatin, hibiscetin, rutin, cyanin and gossypin—violated at least one of Lipinski's criteria. The compounds that did not violate any of the set rules were 1,4-benzoquinone, canthin-6-one, casticin, cleomiscosin A, C and D, cynaratriol, hibiscanal, syriacusin A and B. These compounds were therefore selected for molecular docking studies.

### 3.2 Molecular Docking Studies

The results of the blind docking of ligands with  $M^{Pro}$  are summarized in Fig. 4 and Table 2. The ligands had binding preference for three notable sites on the  $M^{Pro}$  – the active site, the active site-distal site interface, and the distal site (Fig. 4). Canthin-6-one, casticin and hibiscanal had majority of their poses in the distal site. Cleomiscosin A, C and D were clustered mainly at the active site of the protease. The other ligands had poses scattered around the aforementioned binding pockets. From the clusters obtained, docking scores obtained for the best pose of each compound were  $-6.4$  kcal/mol for canthin-6-one,  $-7.1$  kcal/mol for casticin and  $-5.9$  kcal/mol for hibiscanal. Due to the high affinity of casticin to the distal site, it had the highest number of poses (78%) within the distal site hence, formed the binary complex of choice hinting at the allosteric effect. Casticin was subjected to precision docking at the  $M^{Pro}$  distal site. The best pose considered for MD simulations was selected based on high binding affinity

**Table 1** Pharmacokinetic properties of compounds

Name	Pharmacokinetic properties				
	LogP	Molecular weight (g/mol)	No. of hydrogen bond donors	No. of hydrogen bond acceptors	No. of violations
1,4-benzoquinone	-0.23	108.09	0	2	0
Canthin-6-one	2.41	220.23	0	2	0
Casticin	-0.12	374.34	2	8	0
Cleomiscocin A	0.66	386.35	2	8	0
Cleomiscocin C	0.37	416.38	2	9	0
Cleomiscocin D	0.37	416.38	2	9	0
Cynaratriol	0.88	282.33	3	5	0
Gossypetin	-1.67	336.25	7	9	1
Hibiscanal	0.88	262.26	2	5	0
Hibiscatin	-2.24	494.45	6	12	2
Hibiscetin	-1.59	334.23	7	9	1
Syriacusin A	1.18	232.23	2	4	0
Syriacusin B	0.62	262.26	2	5	0
Rutin	-3.89	610.52	10	16	3
Cyanin	-3.82	611.53	11	16	3
Gossypin	-3.07	480.38	9	13	2
Hibiscitrin	-3.55	496.38	4	14	2
Myriceric acid B	5.07	634.84	4	7	2

**Fig. 4** **a** Collection of docked poses of compounds after blind docking, depicting preferential ligand binding regions. Protein is shown in grey while binding modes of ligands are shown in red. **b** Figure showing the low energy clusters of casticin at the active and distal sites

and interaction with regulatory residues Ser284, Ala285, and Leu286. N-3' was placed in the active site with the help of UCSF Chimera to generate an M<sup>PRO</sup>-N-3' (binary) complex and an M<sup>PRO</sup>-N-3'-casticin (ternary) complex. These complexes were then subjected to MD simulations.

### 3.3 MD Simulations

To explore the allosteric potential of casticin, MD simulations were performed in explicit solvent. Since the

SARS-CoV-2 main protease performs its catalysis by recognizing conserved glutamine at the P1 site (i.e., cleavage occurs at the P1–P1' peptide bond), M<sup>PRO</sup>-N-3' (binary complex) was considered to provide insights into the orientation of the peptide substrate during the catalytic event across the simulation period. Another complex, with N-3' bound at the active site and casticin at the distal site, was generated to explore the effect casticin may have on the orientation of N-3' across time. This casticin—M<sup>PRO</sup>—N-3' system was the ternary complex used. An unbound protein (*apo* form)

**Table 2** Binding locations of the different poses of the selected compounds in blind docking against M<sup>Pro</sup>

Compound	Binding sites		
	Active site	Active site-distal site interface	Distal site
1,4-benzoquinone	0	5	4
Canthin-6-one	2	1	6
Casticin	0	2	7
Cleomiscocin A	6	1	2
Cleomiscocin C	6	1	2
Cleomiscocin D	9	0	0
Cynaratriol	3	2	4
Hibiscanal	4	0	5
Syriacusin A	1	4	4
Syriacusin B	1	4	4

was used to explore the protein dynamics in the bound and unbound states. The binary and ternary complexes were sampled at various time frames to explore the changes occurring across time. This was to unravel the effect of casticin binding on the orientation of the peptide substrate at the catalytic site. The 3D structures of the complexes were extracted from trajectory times at 10 ns, 20 ns, 30 ns, 40 ns, and 50 ns to sample the trajectories across the entire simulation period. The sampled structures were compared to their minimized structures (0 ns) prior to simulation to explore the change in dynamic events. 3D and 2D interactions were used to elucidate detailed observations made across the simulation. Comparative studies were made for both complexes at the selected time frames.

### 3.3.1 Protein-Peptide (binary) Complex

From the initial structure for the binary complex, N-3' interacted with M<sup>Pro</sup> via hydrogen bonds with Asn142, Gly143, Ser144, Cys145, His163, Glu166, Pro168, Gln189, Gln192 and hydrophobic interactions with His41, Met49, Met165. Particularly, the carbonyl group of the glutamine residue of N-3' formed hydrogen bond interactions with the backbone NH of Gly143, Ser144 and Cys145, with bond distances of 2.04, 2.97 and 2.76 Å respectively. Gly143, Ser144, and Cys145 form the all-important oxyanion hole in the S1 pocket of the active site of M<sup>Pro</sup> during proteolysis. Other hydrogen bonds and non-polar interactions were observed between the remaining residues of M<sup>Pro</sup> and N-3'. These residues aid in the binding of N-3' in the active site. At 10 ns, a slight change in the orientation, and hence bonding interactions of N-3' was observed. While interactions with residues such as Glu166, Met49 and Met165 persisted, some contacts were lost and a few other new ones were formed. The carbonyl group of the glutamine residue of the peptide

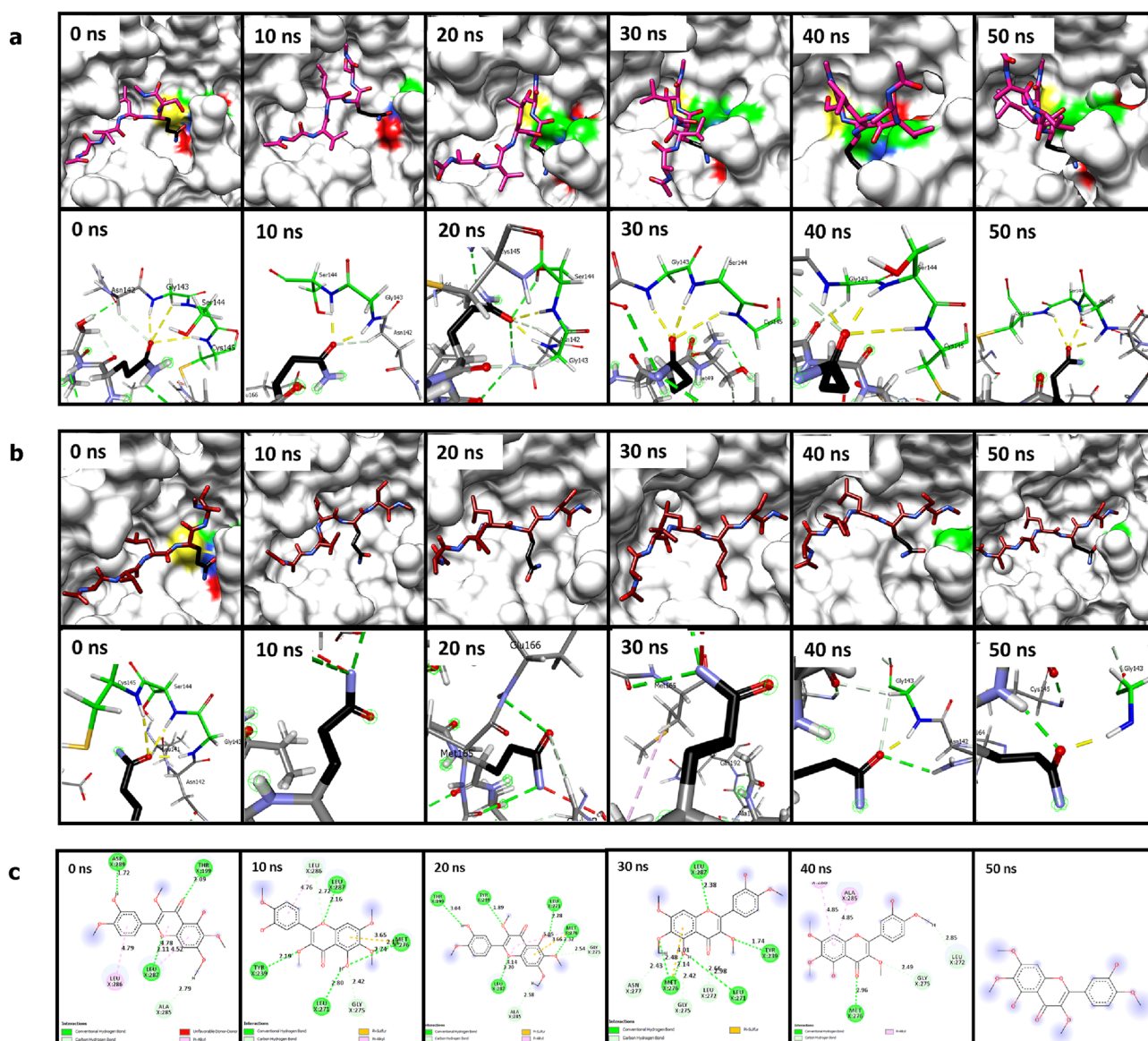
substrate (N-3') maintained its hydrogen bond interactions with Gly143 (1.86 Å) and Ser144 (1.94 Å) of the oxyanion hole but lost contact with Cys145 (> 5 Å). Similar observations were made at 20 ns—the glutamine residue of N-3' maintained interactions with Gly143 (1.79 Å) and Ser144 (3.7 Å) at shorter distances but none with Cys145. Interestingly, the hydrogen bond interactions of N-3' glutamine carbonyl with the oxyanion hole residues (Gly143, Ser144 and Cys145) were restored at 30 ns. Even though interactions with other residues differed across the rest of the trajectory, the hydrogen bond contact between the glutamine carbonyl with the oxyanion hole formed was ever-present throughout the rest of the simulation at distances ranging from 1.91 to 3.0 Å. (Figs. 5a and S1).

### 3.3.2 Peptide-protein-Casticin (ternary) Complex

**3.3.2.1 Events at the Active Site** The initial structure of the ternary complex featured similar interactions between N-3' and M<sup>Pro</sup> as was seen in the binary complex. Hydrogen bond interactions between M<sup>Pro</sup> with N-3' were through Thr25, Leu141, Asn142, Gly143, Ser144, Cys145, Glu166, Gln189, while hydrophobic interactions with His41, Met49, Met165 were observed. Similar to the binary complex, the glutamine carbonyl made hydrogen bond contacts with the oxyanion hole residues—Gly143, Ser144, and Cys145 at 1.79, 2.06 and 2.53 Å respectively. A change in residue contacts was observed at 10 ns. Interestingly, the peptide substrate lost interactions with all 3 oxyanion hole residues, resulting in large bond distances (> 5 Å) as seen in Fig. 5b. Even though N-3' had a strong hydrogen bond interaction with Gly143 (2.43 Å) and Cys145 (2.45 Å) at 20 ns, these interactions, however, were with the Ser residue of N-3' and not the glutamine residue as seen in the binary complex. The glutamine carbonyl, instead, interacted with Glu166 via hydrogen bond at a distance of 2.80 Å. No interaction was observed with the glutamine residue of the peptide substrate and any of the oxyanion hole residues at 30 ns, leading to higher bond distances (> 5 Å). At 40 ns, the glutamine residue interacted with only Gly143 (1.84 Å) and a similar scenario prevailed at 50 ns. Details of the interactions have been shown in Figs. 5b and S2.

**3.3.2.2 Events at the Distal Site** In the starting structure, casticin interacted with M<sup>Pro</sup> via hydrogen bonds to Thr199, Ala285, Leu286 and Asp289, and hydrophobic interactions with Leu286. The residues, Ala285 and Leu286, have been reported in previous works to be important for modulating activity at the active site of the protein. Much focus was therefore placed on these residues during the analysis. At 10 ns, hydrogen bond and hydrophobic interactions involving the oxygen in casticin's pyran ring (2.72 Å) and pi system of ring A with Leu286 (4.76 Å) respectively were





**Fig. 5** Influence of allosteric binding of casticin on the interactions of the oxyanion hole residues with peptide substrate by 3D and 2D comparative analysis of binary and ternary complexes (represented by 3D and 2D interactions). 3D diagram showing the peptide substrate poses in the active site. The oxyanion hole is indicated by red-yellow-green surface patches. Hydrogen bond interactions (green dashes) of P1 carbonyl group of the peptide substrate (black) with the backbone

amido groups of oxyanion hole, are shown with 3D interaction diagrams (bottom panels of **a**, **b**). Residue is indicated as green stick model while backbone amido group is labeled as blue a. Trajectory based time-specific analysis of the residence times of the conformations of N-3' and presence of oxyanion hole residues in the binary (**a**) and ternary (**b**) complex at 0, 10, 20, 30, 40 and 50 ns. **c** 2D view of interactions of casticin at the distal site in the ternary complex

observed. This was accompanied by hydrogen interactions with Leu271 (2.80 Å), Tyr239 (2.19 Å), Leu287 (2.16 Å), Met276 (2.74 Å) and Gly275 (2.42). Interaction with Ala285 was however lost. The snapshot at 20 ns revealed hydrogen bond contact with Ala285 but not with Leu286, while hydrogen bonds with other distal site residues were observed. Although interactions with Ala285 and Leu286 were lost at 30 ns, they were however restored at 40 ns via hydrophobic interactions (4.85 Å). However, binding of cas-

ticin at the distal site was short-lived as it exited the site after 40 ns. Summary of the interactions has been shown in Fig. 5c.

### 3.3.3 Comparison of Binary and Ternary Complexes

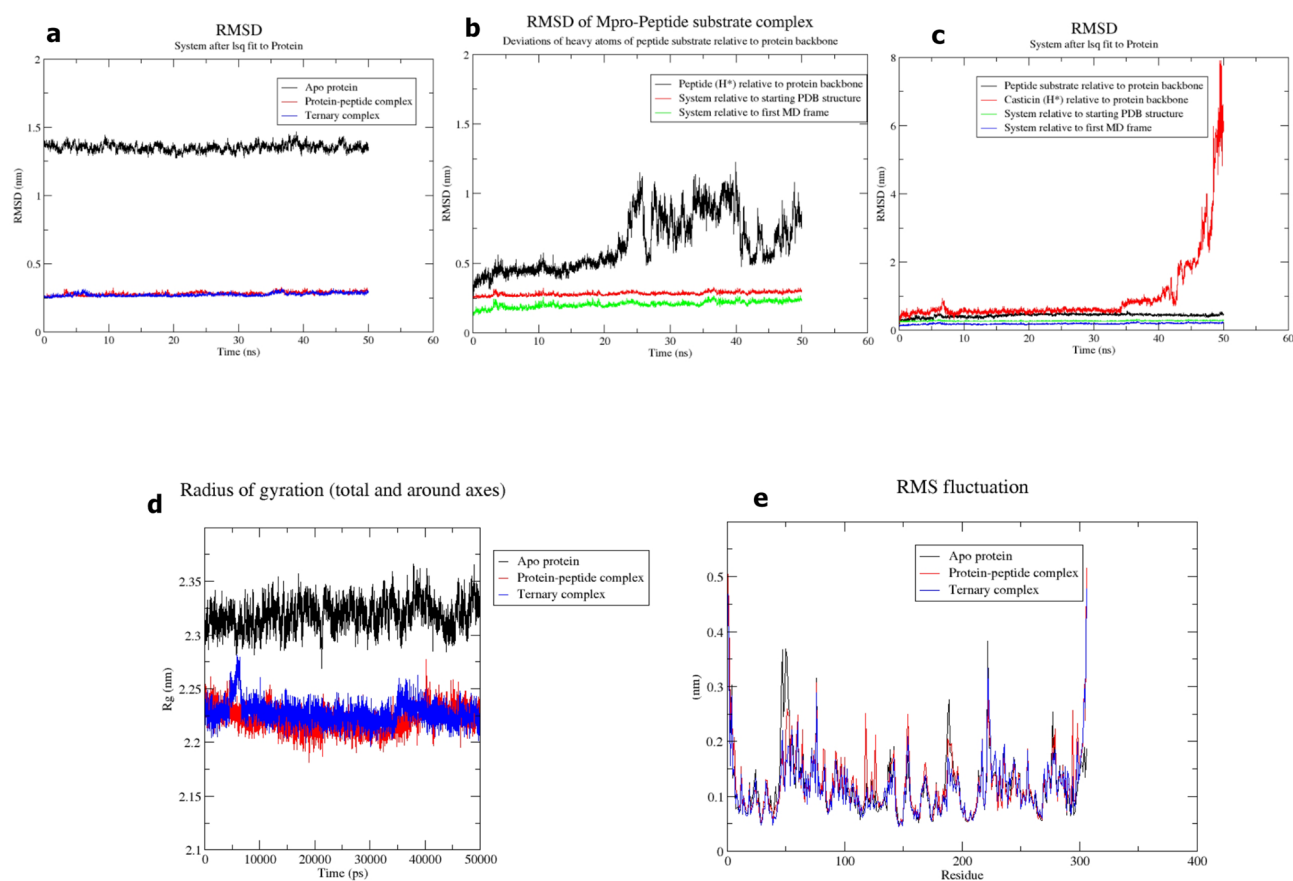
Both complexes exhibited similar interactions between N-3' and the active site residues at the start of the MD run. Interactions in the binary complex involved His41, Met49,

Asn142, Gly143, Ser144, Cys145, His163, Met165, Glu166, Pro168, Gln189, and Gln192 while Thr25, His41, Met49, Leu141, Asn142, Gly143, Ser144, Cys145, Met165, Glu166, Gln189, and Thr190 was observed in the ternary complex. Notably, both complexes exhibited strong hydrogen bond interactions between the glutamine carbonyl of N-3' and the backbone NH groups of Gly143, Ser144, and Cys145. At 10 ns and 20 ns, Gly143 and Ser144 hydrogen bond interactions with the N-3' glutamine's carbonyl were still present in the binary complex. However, hydrogen bond interactions involving the glutamine carbonyl and the three oxyanion hole residues were displaced in the ternary complex. An interesting observation was however made at 30 ns. In the binary complex, the glutamine carbonyl of N-3' established hydrogen bond interactions with the backbone NH groups of Gly143, Ser144 and Cys145. This is in contrast to the ternary complex where none of the aforementioned hydrogen

bond interactions was observed. At 40 ns, the hydrogen bond interactions in question were retained in the binary complex. In the ternary complex, only hydrogen bond interaction with Gly143 was observed. These observations made in both complexes at 40 ns persisted to the end of the simulation.

### 3.4 Stability of Complexes

The RMSD of the apo protein, the binary complex, and the ternary complex were analyzed to determine the deviation of the protein backbone atoms in reference to their initial positions in the trajectory. This was used to examine the complexes' stability across the simulated time frames. As can be seen in Fig. 6a, an RMSD value of  $\sim 0.3$  nm was recorded for both binary and ternary complexes, while a much higher  $\sim 1.3$  nm was observed for the apo protein. The RMSD of N-3' and casticin was used



**Fig. 6** Summary of system stability analysis from RMSD, RMSF, and Rg plots of apo protein and bound complexes. **a** RMSD of the C $\alpha$  atoms of the protein. Bound complexes were relatively stable compared to the apo protein. **b** RMSD of binary system relative to first MD frame. The peptide substrate showed deviation from 25 to 50 ns, which was attributed to torsions in the ligand. Complex was however stable throughout the simulation **c**. RMSD of the ternary complex relative to first MD frame. Even though peptide substrate had a lower

RMSD value, however, its conformation was disoriented, compared to the binary complex. Casticin binding was generally stable for 40 ns ( $\text{RMSD} \leq 1.5$ ), before exiting the distal site ( $\text{RMSD} \geq 1.5$  nm and  $\geq 8$  nm). **d** Radius of gyration of apo protein and bound complexes. Protein in bound complexes was more compact, compared to the apo protein. **e** RMSF of apo and bound complexes. Fluctuations were observed mostly in the N-terminal and the C-domain compared to the other regions of the protein

to determine the deviation of their heavy atoms to the protein backbone. This was used to measure their stability in their respective pockets in the dynamic event. In the binary complex, the peptide substrate had an RMSD value of ~0.5 nm for the first 25 ns, which later increased to ~1 nm for the rest of the simulation (Fig. 6b). In the ternary complex, an RMSD value of ~0.4 nm was recorded for the peptide substrate from the start of the simulation and this was stable throughout the rest of the simulation period. RMSD for casticin was measured as ~0.4 nm for the initial 40 ns of the trajectory. However, there was a huge rise in the RMSD value (~8 nm) after 40 ns to the end of the simulation (Fig. 6c). This corresponded to casticin exiting its binding pocket after 40 ns.

The root mean square fluctuation (rmsf) was used to identify the regions of the protein that undergo residue fluctuations during the simulation period. This provides insight into the effect of ligand binding on the flexibility of the protein. Unsurprisingly, the graphs show that the most flexible regions occurred at the N and C terminus of the protein, with rmsf values less than 0.5 nm. Although similar observations were made in most parts of the various protein structures under consideration (apo, binary, and ternary complexes), residues 50–80 and 190–220 of the apo protein exhibited higher fluctuations compared to the bound states. The protein in its bound state, therefore, undergoes fewer fluctuations as compared to its unbound state, depicting its stability upon substrate only, or peptide and casticin binding (Fig. 6e). To determine the compactness of the protein structures, the radius of gyration (rg) was determined. The radius of gyration of the apo protein was ~2.33 nm. In the bound complexes (binary and ternary complexes), an rg value of ~2.23 nm was also observed. Comparing the apo structure to the bound complexes, the protein in its bound states was slightly more compact (Fig. 6d).

### 3.5 Binding Free Energy Calculations

End-point binding free energy was computed to determine the binding affinity and stability of casticin at the distal site. From the results, van der Waals, electrostatic, and non-polar solvation energy contributed significantly to the total binding energy of casticin at –100.156 kJ/mol, –36.768 kJ/mol, and –11.485 kJ/mol respectively. The polar solvation energy however had an unfavorable contribution with a binding energy of 71.257 kJ/mol. The overall binding free energy of casticin was reported as –74.760 kJ/mol. A summary of the results can be found in Table 3. The lower binding energy observed indicates a stable interaction between casticin and M<sup>Pro</sup>.

**Table 3** MM-PBSA calculations of casticin in the ternary complex (kJ/mol)

Ternary Complex	$E_{vdW}$	$E_{elec}$	$E_{polar}$	$E_{non-polar}$	$\Delta G_{bind}$
Casticin	–100.156	–36.768	71.257	–11.485	–77.151

## 4 Discussion

Computer-aided drug design (CADD) has become one of the most important methodologies in modern drug development because it may greatly reduce the cost and effort involved in the drug discovery process. CADD speeds up lead development by cutting down experimental efforts. In particular, molecular docking, molecular dynamics, and ADME technologies have become crucial aspects of CADD [42, 43]. Drugs with good pharmacokinetic features are needed to meet efficacy and safety requirements, and these features must therefore be optimized to pass standard clinical trials [44]. Out of 19 compounds considered for this study, 11 compounds showed zero violation of the 5 Lipinski rules, and this is an indication of their high drug-likeness. Compounds with MW < 500 are usually drug-like because they could rapidly diffuse across cell membranes to reach target site. Also, a compound with the right hydrophilicity/lipophilicity ensures the permeability of the drugs to reach their targets in the body. The number of hydrogen bond donors (< 5) and acceptors (< 10) plays a major role in oral bioavailability by impacting diffusion across cell membranes [32]. The pharmacokinetic properties of these 11 selected compounds, therefore, merited further investigations as potential anti-SARS-CoV-2 agents. Given the important role of M<sup>Pro</sup> in the viral cycle of the coronavirus, (cleavage of the translated polypeptide into functional subunits) there have been numerous efforts to identify potent inhibitors of this protease. Many natural products have been explored as potential inhibitors of the major SARS-COV-2 viral proteins [45–47]. Some of these natural products were suggested to function by acting as active site inhibitors of M<sup>Pro</sup>. While no natural product or natural product-derived compound has yet to be approved as a drug for covid-19, the surge in research highlights the significance of natural products in the development of potential anti-SARS-COV-2 drugs. Blind docking studies displayed various binding modes of the compounds to M<sup>Pro</sup>. Notably, 3 binding sites were observed: the active site, the active site-distal site interface, and the distal site (Table 2/Fig. 4). Several promising strategies for inhibiting M<sup>Pro</sup> have been investigated and this includes covalent inhibition of the active site by peptidomimetic inhibitors such as N3 and boceprevir [10, 48]. Non-covalent modulators, such as compounds that alter the catalytic activity

of the protease at distant sites from the active site have not been studied extensively. Recently, Sztain and coworkers used Gaussian Accelerated Molecular Dynamics to identify and characterize two areas of  $M^{\text{Pro}}$ , namely the dimer interface, and distal site, which could serve as druggable sites for allosteric inhibition. The distal site, found in the C-terminal region, was proven to be highly druggable using the PockDrug19 webserver, which predicts druggability probability based on a variety of factors including geometry, hydrophobicity, and aromaticity [15]. Mass spectrometric analyses by El-Baba identified small molecules that inhibit  $M^{\text{Pro}}$  through a potential allosteric mechanism by binding at the distal site. Some allosteric inhibitors were identified to bind to the proposed distal site. Interactions of these allosteric inhibitors (such as 2-[[[(1H-benzimidazol-2-yl)amino]methyl]phenol and (2R,3R)-1-benzyl-2-methylpiperidin-3-ol) via hydrogen bonding with Thr196, Thr198 and Glu240 of  $M^{\text{Pro}}$  was observed [14]. Of the molecules screened in this study, casticin exhibited drug-like and allosteric features that caught our eye. In addition to zero violation of the Lipinski rules, 78% of its poses were in the distal site and the binding affinity was the strongest of all compounds that were docked to  $M^{\text{Pro}}$  in the distal site. Casticin, a natural product identified in *Achillea millefolium*, *Artemisia annua* and most *Vitex* species (eg. *Vitex agnus-castus*) has been reported to have immunomodulatory and anti-inflammatory effects on the lungs [49]. This flavonoid has also been reported for its antiviral effect against Lassa virus by blocking low-pH-induced membrane fusion [30] and possesses anticancer, antifibrotic, and anti-helminthic effects [50, 51].

The coronavirus main protease invariably recognizes peptide cleavage sites within polyproteins with a strictly conserved glutamine residue at the P1 position [52]. Given that the hydrolysis event mediated by the enzyme takes place at the P1–P1' amide bond (Fig. 2a), the glutamine residue should necessarily be rightly positioned in the active site [53]. The insertion of the glutamine carbonyl group into the protease's oxyanion hole, formed by Gly143, Ser144, and Cys145, is a critical pre-requisite for catalysis (Fig. 2b). Oxyanion holes are formed from two or three hydrogen bonds oriented towards a central atom in the substrate. In cysteine proteases, the oxyanion hole contains a cysteine residue in which the thiol group reacts with the carbonyl group of the substrate, resulting in the tetrahedral intermediate [54]. The optimum binding arrangement for anions also improves binding for the reaction's starting materials [55]. As the reaction proceeds towards the transition state, insertion of the carbonyl group of the substrate into the oxyanion hole is the most effective arrangement that best stabilizes the transition state. Also, since the oxyanion hole plays a major role in the proper orientation of the substrate for hydrolysis,

any substrate that does not occupy the hole would not meet the structural requirements of the catalytic protease [56, 57]. The presence of the oxyanion hole in cysteine proteases is therefore vital for proteolysis [6, 58].

In the binary complex, the carbonyl group of the glutamine residue was well placed within the oxyanion hole defined by hydrogen bond interactions with the backbone amido groups of Gly143, Ser144, and Cys145 in the initial structure (Fig. 5a). Within the first 20 ns of the trajectory, the hydrogen bonding network that established the oxyanion hole was unstable, probably due to the peptide searching for a stable position within the active site to establish the necessary interaction required for the formation of the oxyanion hole. By the 30th ns, the glutamine-oxyanion hole network was re-established, and this conformation was stable enough to persist throughout the entire simulation. The stable hydrogen bond contacts between the glutamine carbonyl and the oxyanion hole result in a stable transition state, which lowers the activation energy required for the reaction and so promotes catalysis. The alignment of the substrate chain, together with the accommodation of glutamine of the peptide substrate in the S1 pocket, effectively locks the P1–P1' amide bond into the Gly143/Ser144/Cys145 oxyanion hole. This observation made in the protein-peptide substrate suggests a stable orientation of the cleavage site for catalytic activity.

Analysis of the ternary complex sheds light on the effect of casticin binding on the orientation of the peptide substrate at the active site. In the starting structure, the glutamine carbonyl of the peptide substrate was inserted right into the oxyanion hole, forming hydrogen bond contacts with the backbone NH groups of the oxyanion hole residues. For the first 20 ns, the peptide substrate searched the active site for a stable conformation with the pocket residues. This resulted in the distortion of the glutamine carbonyl position in the oxyanion hole pocket. However, in contrast to stabilization of the hydrogen bond network in the binary complex at 30 ns, the position of the glutamine of the peptide substrate was never restored to the initial state, leading to the observed disorientation of the glutamine–oxyanion hole hydrogen bond network in the ternary complex. This was confirmed by large bond distances ( $> 5 \text{ \AA}$ ) between the glutamine carbonyl and the backbone NH groups of the oxyanion hole residues (Fig. 5b). At 40 and 50 ns, the glutamine of the peptide substrate was observed to move closer to the S1 pocket oxyanion hole. However, the glutamine carbonyl was stabilized by a hydrogen bond with only Gly143. The loss of hydrogen bond contacts with the other oxyanion hole residues (Ser144 and Cys145) could destabilize the peptide substrate orientation and thus lead to a reduction in the rate of hydrolysis, or altogether abolish catalysis. Disruption of the hydrogen bonding network that forms the oxyanion hole in subtilisin via site-directed mutagenesis revealed that the



catalytic contribution of the oxyanion hole to the overall stability of the complex was about 3.2–3.4 kcal/mol [54]. It has been reported that double mutations removed two hydrogen bond donors in the oxyanion hole of 3-Oxo- $\Delta$ 5-Steroid Isomerase in *Pseudomonas putida* and *Comamonas testosteroni* resulted in a  $10^5$ – $10^8$  fold decrease in proteolytic activity [59, 60]. Furthermore, strong interactions exhibited by N-3' with the active site residues of the ternary complex might prevent rotation that will lead to N-3' achieving the right orientation in the active site.

Analysis of the interactions at the distal site revealed hydrogen bond contacts with Thr199, Ala285, Leu286 and Asp289 and hydrophobic interaction with Leu286 in the starting structure (Fig. 5c). Ala285 and Leu286 have shown to be critically important residues responsible for regulating motions in the active site of  $M^{Pro}$  [15, 16]. The first 20 ns showed interactions between casticin and at least one of the two aforementioned residues. At 30 ns, there was deviation in the position of casticin in the pocket such that interactions with Ala285 and Leu286 were not observed. However, casticin quickly attained hydrophobic interactions with these residues as observed at 40 ns. This observation was short-lived though as casticin exited the pocket after 40 ns. However, the residence time of casticin coupled with its interaction with Ala285 and Leu286 was sufficient to distort the orientation of the peptide substrate in the active site of  $M^{Pro}$ . The importance of Ala285 and Leu286 has been highlighted in some reports. It has been shown that the replacement of Ser284, Thr285, and Ile286 by alanine residues in SARS-CoV  $M^{Pro}$  led to enhanced catalytic activity of the protease. This also led to changes in dynamic events in the protease that transmit the effect of the mutation to the catalytic site [61]. Another study also revealed that Ser284, Ala285 and Leu286 influenced catalytic efficiency, with Ala285 shown to make SARS-CoV-2  $M^{Pro}$  much more efficient than SARS-CoV  $M^{Pro}$  [62]. A molecular docking study also suggested potential allosteric inhibition of SARS-CoV-2  $M^{Pro}$  after the compounds interacted with the regulatory residues, Ala285 and Leu286 [63]. Another work by Dubanevics et al. also confirmed dynamic allostery by the SARS-CoV-2  $M^{Pro}$  via Ala285 and Leu286, where a dynamic motion correlation between the active site and the distal site was observed [16]. We, hereby, propose that interaction between casticin and the regulatory residues Ala285 and Leu286 caused a change in the internal motion of the active site, thereby causing disorientation of the position of the peptide substrate. This study, therefore, highlights casticin as a potential allosteric modulator that could be considered for further studies. Another careful observation throughout the trajectory shows hydrogen bond interaction between the oxygen in ring C of casticin and Leu271. This observation was consistent from the start of the simulation to 30 ns. At 40 ns, this interaction was however lost, which led to casticin's exit from the distal

site. It therefore suggests that the hydrogen bond interaction with Leu271 is responsible for anchoring casticin at the distal site, while casticin interacts with Ala285 and Leu286 to regulate motion at the active site.

Analysis of the RMSD, RMSF, and Rg helps to provide insights into the structural dynamics of the proteins and the ligands. RMSD analysis of the protein in the apo structure, binary and complexes showed that the protein gets stabilized after peptide substrate and casticin binding. Results from RMSD analysis of the peptide substrate in the binary and ternary complexes showed less deviation in the heavy atoms, depicting stability in the active site (Fig. 6). In the ternary complex, casticin was stable in the distal site for 40 ns, before it exited the site. Analysis of the protein in the apo structure, the binary and ternary complexes revealed less deviation in the side chain residues of the bound complexes (binary and ternary) compared to the apo protein. This shows that binding of the ligands induced less fluctuations in the protein, making it more stable. The radius of gyration helps to determine the compactness of the protein. Comparing the apo protein to the binary and ternary complexes, the bound systems were more compact compared to the apo protein (Fig. 6). Overall, the binary and ternary complexes were stable throughout the trajectory. This correlates with previous studies where it was reported that ligand binding stabilized  $M^{Pro}$  after trajectory analysis [44, 64]. The binding free energy of casticin during its residence time at the distal site (40 ns) was 77.151 kJ/mol, with van der Waals, electrostatic and non-polar solvation energy contributing significantly to binding. The estimated binding energy was sufficient enough to distort the orientation of the peptide substrate at the active site.

## 5 Conclusion

The  $M^{Pro}$  protein has been widely considered one of the most druggable targets for anti-SARS-COV-2 drug discovery. The protease activity of the protein is required to process viral protein in the host and thus aid in viral replication. Our study revealed that casticin could serve as a potential allosteric modulator of the SARS-COV-2 main protease. Interaction of casticin with reported regulatory residues Ala285 and Leu286 at the distal site could probably lead to a distortion in the conformation of the peptide substrate (N-3') in the active site of the protein. This distortion disrupts the key oxyanion hole-peptide substrate interactions needed for proteolysis. RMSD, RMSF, and Rg analysis suggest a stable conformation of the protein upon casticin binding in its pocket for 40 ns. Binding energy calculation reveals a favorable interaction between  $M^{Pro}$  and casticin. The results of this study reveal that casticin possesses potential allosteric inhibitory against the SARS-COV-2 main protease. Further

in vitro and in vivo assays could be considered to explore the inhibitory potential of casticin against the main protease.

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1007/s42250-022-00411-7>.

**Acknowledgements** The authors are grateful to the Center for High Performance Computing, Cape Town, South Africa who granted the authors generous access to the Lengau cluster for the MD simulations.

**Author contributions** LSB, ENG and ESM conceived the study. All experiments were designed by LSB, ESM, GBA, JOM and ENG. Computations were made by JOM, GBA, AKA and ENG. Data analysis was by ENG, JOM, AKA, GBA, ESM and LSB. Manuscript was prepared by ENG, JOM, ESM, GBA and LSB. All authors read and approved the final manuscript.

**Funding** Authors received no direct funding for this study.

**Data availability** All data generated or analyzed during this study are included in this published article.

## Declarations

**Conflict of interest** All authors declare that there were no financial, professional, or personal competing interests that might have influenced the performance or presentation of this study.

## References

- Phan T (2020) Genetic diversity and evolution of SARS-CoV-2. *Infect Genet Evol* 81:104260
- Chu DKW et al (2020) Molecular diagnosis of a Novel Coronavirus (2019-nCoV) causing an outbreak of pneumonia. *Clin Chem* 66(4):549–555. <https://doi.org/10.1093/clinchem/hvaa029>
- Wu C et al (2020) Analysis of therapeutic targets for SARS-CoV-2 and discovery of potential drugs by computational methods. *Acta Pharm Sin B* 10(5):766–788. <https://doi.org/10.1016/j.apsb.2020.02.008>
- Subissi L et al (2014) One severe acute respiratory syndrome coronavirus protein complex integrates processive RNA polymerase and exonuclease activities. *Proc Natl Acad Sci* 111(37):E3900–E3909
- Tang X et al (2020) On the origin and continuing evolution of SARS-CoV-2. *Natl Sci Rev* 7(6):1012–1023
- Zhang L et al (2020) Crystal structure of SARS-CoV-2 main protease provides a basis for design of improved  $\alpha$ -ketoamide inhibitors. *Science* 368(6489):409–412
- Shi J, Song J (2006) The catalysis of the SARS 3C-like protease is under extensive regulation by its extra domain. *FEBS J* 273(5):1035–1045
- Hilgenfeld R (2014) From SARS to MERS: crystallographic studies on coronaviral proteases enable antiviral drug design. *FEBS J* 281(18):4085–4096
- Kim Y et al (2016) Correction: reversal of the progression of fatal coronavirus infection in cats by a broad-spectrum coronavirus protease inhibitor. *Plos Pathog* 12(5):e1005650
- Jin Z et al (2020) Structure of Mpro from SARS-CoV-2 and discovery of its inhibitors. *Nature* 582(7811):7811. <https://doi.org/10.1038/s41586-020-2223-y>
- Mahanta S et al (2021) Potential anti-viral activity of approved repurposed drug against main protease of SARS-CoV-2: an in silico based approach. *J Biomol Struct Dyn* 39(10):3802–3811
- Verma S, Pandey AK (2021) Factual insights of the allosteric inhibition mechanism of SARS-CoV-2 main protease by quercetin: an in silico analysis. *3 Biotech* 11(2):1–10
- Macchiagodena M, Pagliai M, Procacci P (2020) Identification of potential binders of the main protease 3CLpro of the COVID-19 via structure-based ligand design and molecular modeling. *Chem Phys Lett* 750:137489. <https://doi.org/10.1016/j.cplett.2020.137489>
- El-Baba TJ et al (2020) Allosteric inhibition of the SARS-CoV-2 main protease: insights from mass spectrometry based assays. *Angew Chem Int Ed* 59(52):23544–23548
- Sztain T, Amaro R, McCammon JA (2021) Elucidation of cryptic and allosteric pockets within the SARS-CoV-2 main protease. *J Chem Inf Model* 61(7):3495–3501
- Dubanevics I, McLeish TC (2021) Computational analysis of dynamic allostery and control in the SARS-CoV-2 main protease. *J R Soc Interface* 18(174):20200591
- Ayisi F, Mensah CN, Borquaye LS (2021) Antiplasmodial potential and safety evaluation of the ethanolic stem bark extract of *Distemonanthus benthamianus* Baill. (Leguminosae). *Sci Afr* 12:e00809. <https://doi.org/10.1016/j.sciaf.2021.e00809>
- Ayisi F, Mensah CN, Borquaye LS (2021) In vivo antiplasmodial activity and toxicological analyses of the ethanolic leaf and twig extract of *Faurea speciosa* Welw. (Proteaceae). *J Parasitol Res*. <https://doi.org/10.1155/2021/7347532>
- Laryea MK, Sheringham Borquaye L (2021) Antimalarial, antioxidant, and toxicological evaluation of extracts of *Celtis africana*, *Grosseria vignei*, *Physalis micrantha*, and *Stachytarpheta angustifolia*. *Biochem Res Int*. <https://doi.org/10.1155/2021/9971857>
- Borquaye LS et al (2020) Alkaloids from *Cryptolepis sanguinolenta* as potential inhibitors of SARS-CoV-2 viral proteins: an in silico study. *BioMed Res Int* 2020:5324560. <https://doi.org/10.1155/2020/5324560>
- Jo S, Kim S, Kim DY, Kim M-S, Shin DH (2020) Flavonoids with inhibitory activity against SARS-CoV-2 3CLpro. *J Enzyme Inhib Med Chem* 35(1):1539–1544
- Ramachandran B, Muthupandian S, Jeyaraman J (2021) Prospects of covalent approaches in drug discovery: an overview. In: *Computation in bioinformatics*. John Wiley & Sons, Ltd, pp 295–319. <https://doi.org/10.1002/9781119654803.ch15>
- Sankar M et al (2021) In silico Screening of natural phytochemicals towards identification of potential lead compounds to treat COVID-19. *Front Mol Biosci* 8:637122. <https://doi.org/10.3389/fmolb.2021.637122>
- Park H, Jung H-Y, Mah S, Hong S (2018) Systematic computational design and identification of low Picomolar inhibitors of Aurora kinase a. *J Chem Inf Model* 58(3):700–709
- Schames JR, Henchman RH, Siegel JS, Sotriffer CA, Ni H, McCammon JA (2004) Discovery of a novel binding trench in HIV integrase. *J Med Chem* 47(8):1879–1881
- Maganha EG, da Costa Halmenschlager R, Rosa RM, Henriques JAP, de Paula Ramos ALL, Saffi J (2020) Pharmacological evidences for the extracts and secondary metabolites from plants of the genus *Hibiscus*. *Food Chem* 118(1):1–10
- Vasudeva N, Sharma SK (2008) Biologically active compounds from the genus *Hibiscus*. *Pharm Biol* 46(3):145–153
- Salem MZ, Olivares-Pérez J, Salem AZM (2014) Studies on biological activities and phytochemicals composition of *Hibiscus* species—a review. *Life Sci J* 11(5):1–8
- Gomathi M, Padmapriya S, Balachandar V (2020) Drug studies on Rett syndrome: from bench to bedside. *J Autism Dev Disord* 50(8):2740–2764

30. Liu Y et al (2021) Screening of botanical drugs against Lassa virus entry. *J Virol* 95(8):e02429–e2520
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. Lipinski CA, Lombardo F, Dominy BW, Feeney PJ (2001) Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv Drug Deliv Rev* 46(1–3):3–26. [https://doi.org/10.1016/s0169-409x\(00\)00129-0](https://doi.org/10.1016/s0169-409x(00)00129-0)
33. Kyei LK, Gasu EN, Ampomah GB, Mensah JO, Borquaye LS (2022) An in silico study of the interactions of alkaloids from *Cryptolepis sanguinolenta* with *Plasmodium falciparum* dihydrofolate reductase and dihydroorotate dehydrogenase. *J Chem.* <https://doi.org/10.1155/2022/5314179>
34. Forli S, Huey R, Pique ME, Sanner MF, Goodsell DS, Olson AJ (2016) Computational protein-ligand docking and virtual drug screening with the AutoDock suite. *Nat Protoc* 11(5):905–919. <https://doi.org/10.1038/nprot.2016.051>
35. Suárez D, Díaz N (2020) SARS-CoV-2 main protease: a molecular dynamics study. *J Chem Inf Model* 60(12):5815–5831. <https://doi.org/10.1021/acs.jcim.0c00575>
36. Vanommeslaeghe K et al (2010) CHARMM.pdf. *J Comput Chem* 31(4):671–690. <https://doi.org/10.1002/jcc.21367>. CHARMM
37. Kutzner C, Páll S, Fechner M, Esztermann A, de Groot BL, Grubmüller H (2019) More bang for your buck: Improved use of GPU nodes for GROMACS 2018. *J Comput Chem* 40(27):2418–2431. <https://doi.org/10.1002/jcc.26011>
38. Che Omar MT (2020) Data analysis of molecular dynamics simulation trajectories of  $\beta$ -sitosterol, sonidegib and cholesterol in smoothed protein with the CHARMM36 force field. *Data Brief* 33:106350. <https://doi.org/10.1016/j.dib.2020.106350>
39. Amusengeri A, Tata RB, Tasthan Bishop Ö (2020) Understanding the pyrimethamine drug resistance mechanism via combined molecular dynamics and dynamic residue network analysis. *Molecules* 25(4):4. <https://doi.org/10.3390/molecules25040904>
40. 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](https://doi.org/10.1016/0263-7855(96)00018-5) (27–28)
41. Kumari R, Kumar R, Lynn A (2014) g\_mmpbsa—a GROMACS tool for high-throughput MM-PBSA calculations. *J Chem Inf Model* 54(7):1951–1962. <https://doi.org/10.1021/ci500020m>
42. Jain A (2017) Computer aided drug design. *J Phys Conf Ser* 884:012072. <https://doi.org/10.1088/1742-6596/884/1/012072>
43. Talele TT, Khedkar SA, Rigby AC (2010) Successful applications of computer aided drug discovery: moving drugs from concept to the clinic. *Curr Top Med Chem* 10(1):127–141. <https://doi.org/10.2174/156802610790232251>
44. Islam R et al (2021) A molecular modeling approach to identify effective antiviral phytochemicals against the main protease of SARS-CoV-2. *J Biomol Struct Dyn* 39(9):3213–3224. <https://doi.org/10.1080/07391102.2020.1761883>
45. Gyebi GA, Ogunro OB, Adegunloye AP, Ogunyemi OM, Afolabi SO (2021) Potential inhibitors of coronavirus 3-chymotrypsin-like protease (3CLpro): an in silico screening of alkaloids and terpenoids from African medicinal plants. *J Biomol Struct Dyn* 39(9):3396–3408. <https://doi.org/10.1080/07391102.2020.1764868>
46. Joshi RS et al (2021) Discovery of potential multi-target-directed ligands by targeting host-specific SARS-CoV-2 structurally conserved main protease. *J Biomol Struct Dyn* 39(9):3099–3114. <https://doi.org/10.1080/07391102.2020.1760137>
47. Majumder R, Mandal M (2022) Screening of plant-based natural compounds as a potential COVID-19 main protease inhibitor: an in silico docking and molecular dynamics simulation approach. *J Biomol Struct Dyn* 40(2):696–711. <https://doi.org/10.1080/07391102.2020.1817787>
48. Ma C et al (2020) Boceprevir, GC-376, and calpain inhibitors II, XII inhibit SARS-CoV-2 viral replication by targeting the viral main protease. *Cell Res* 30(8):8. <https://doi.org/10.1038/s41422-020-0356-z>
49. Liou C-J, Cheng C-Y, Yeh K-W, Wu Y-H, Huang W-C (2018) Protective effects of casticin from *Vitex trifolia* alleviate eosinophilic airway inflammation and oxidative stress in a murine asthma model. *Front Pharmacol.* <https://doi.org/10.3389/fphar.2018.00635> (Accessed: Feb. 16, 2022. [Online])
50. Lam HYP, Liang T-R, Lan Y-C, Chang K-C, Cheng P-C, Peng S-Y (2021) Antifibrotic and anthelmintic effect of casticin on *Schistosoma mansoni*-infected BALB/c mice. *J Microbiol Immunol Infect Wei Mian Yu Gan Ran Za Zhi.* <https://doi.org/10.1016/j.jmii.2021.03.017>
51. Rasul A et al (2014) Molecular mechanisms of casticin action: an update on its antitumor functions. *Asian Pac J Cancer Prev* 15(21):9049–9058. <https://doi.org/10.7314/APJCP.2014.15.21.9049>
52. Ramos-Guzmán CA, Ruiz-Pernía JJ, Tuñón I (2021) A microscopic description of SARS-CoV-2 main protease inhibition with Michael acceptors. Strategies for improving inhibitor design. *Chem Sci* 12(10):3489–3496. <https://doi.org/10.1039/d0sc04978f>
53. Fernandes HS, Sousa SF, Cerqueira NMFS (2021) New insights into the catalytic mechanism of the SARS-CoV-2 main protease: an ONIOM QM/MM approach. *Mol Divers.* <https://doi.org/10.1007/s11030-021-10259-7>
54. Ménard R, Storer AC (1992) Oxyanion hole interactions in serine and cysteine proteases. *Biol Chem Hoppe-Seyler.* 373(2):393–400. <https://doi.org/10.1515/bchm3.1992.373.2.393>
55. Simón L, Goodman JM (2010) Enzyme catalysis by hydrogen bonds: the balance between transition state binding and substrate binding in oxyanion holes. *J Org Chem* 75(6):1831–1840. <https://doi.org/10.1021/jo901503d>
56. Kursula P, Ojala J, Lambeir A-M, Wierenga RK (2002) The catalytic cycle of biosynthetic thiolase: a conformational journey of an acetyl group through four binding modes and two oxyanion holes. *Biochemistry* 41(52):15543–15556. <https://doi.org/10.1021/bi026623z>
57. Yang H et al (2003) The crystal structures of severe acute respiratory syndrome virus main protease and its complex with an inhibitor. *Proc Natl Acad Sci* 100(23):13190–13195. <https://doi.org/10.1073/pnas.1835675100>
58. Zhang L et al (2020)  $\alpha$ -Ketoamides as broad-spectrum inhibitors of coronavirus and enterovirus replication: structure-based design, synthesis, and activity assessment. *J Med Chem* 63(9):4562–4578. <https://doi.org/10.1021/acs.jmedchem.9b01828>
59. Choi G et al (2000) Asp-99 donates a hydrogen bond not to Tyr-14 but to the steroid directly in the catalytic mechanism of  $\Delta^5$ -3-ketosteroid isomerase from *Pseudomonas putida* biotype B. *Biochemistry* 39(5):903–909. <https://doi.org/10.1021/bi991579k>
60. Thornburg LD, Goldfeder YR, Wilde TC, Pollack RM (2001) Selective catalysis of elementary steps by Asp-99 and Tyr-14 of 3-Oxo- $\Delta^5$ -steroid isomerase. *J Am Chem Soc* 123(40):9912–9913. <https://doi.org/10.1021/ja016683f>
61. Lim L, Shi J, Mu Y, Song J (2014) Dynamically-driven enhancement of the catalytic machinery of the SARS 3C-like protease by the S284–T285–I286/A mutations on the extra domain. *PLoS ONE* 9(7):e101941. <https://doi.org/10.1371/journal.pone.0101941>
62. Wei P et al (2006) The N-terminal octapeptide acts as a dimerization inhibitor of SARS coronavirus 3C-like proteinase. *Biochem Biophys Res Commun* 339(3):865–872. <https://doi.org/10.1016/j.bbrc.2005.11.102>

63. Vijayakumar BG, Ramesh D, Joji A, Jayachandra Prakasan J, Kannan T (2020) In silico pharmacokinetic and molecular docking studies of natural flavonoids and synthetic indole chalcones against essential proteins of SARS-CoV-2. *Eur J Pharmacol* 886:1173448. <https://doi.org/10.1016/j.ejphar.2020.173448>
64. Ghosh R, Chakraborty A, Biswas A, Chowdhuri S (2021) Evaluation of green tea polyphenols as novel corona virus (SARS CoV-2) main protease (Mpro) inhibitors—an in silico docking and molecular dynamics simulation study. *J Biomol Struct Dyn* 39(12):4362–4374. <https://doi.org/10.1080/07391102.2020.1779818>

## Authors and Affiliations

Jehoshaphat Oppong Mensah<sup>1</sup>  · Gilbert Boadu Ampomah<sup>2</sup>  · Edward Ntim Gasu<sup>1,2</sup>  · Abigail Kusiwaa Adomako<sup>1</sup> · Elliott Sarpong Menkah<sup>1</sup> · Lawrence Sheringham Borquaye<sup>1,2</sup> 

Jehoshaphat Oppong Mensah  
jehoshaphatmensah@gmail.com

Gilbert Boadu Ampomah  
gilbertboadu@outlook.com

Edward Ntim Gasu  
engasu24@gmail.com

Abigail Kusiwaa Adomako  
aadomako80@gmail.com

Elliott Sarpong Menkah  
elliottsmenkah@gmail.com

<sup>1</sup> Department of Chemistry, Kwame Nkrumah University of Science and Technology, Kumasi, Ghana

<sup>2</sup> Central Laboratory, Kwame Nkrumah University of Science and Technology, Kumasi, Ghana