



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

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



Unravelling lead antiviral phytochemicals for the inhibition of SARS-CoV-2 M^{pro} enzyme through in silico approach

Arun Bahadur Gurung^{a,*}, Mohammad Ajmal Ali^b, Joongku Lee^c, Mohammad Abul Farah^d, Khalid Mashay Al-Anazi^d



^a Department of Basic Sciences and Social Sciences, North-Eastern Hill University, Shillong 793022, Meghalaya, India

^b Department of Botany and Microbiology, College of Science, King Saud University, Riyadh 11451, Saudi Arabia

^c Department of Environment and Forest Resources, Chungnam National University, 99 Daehak-ro, Yuseong-gu, Daejeon 34134, Republic of Korea

^d Genetics Laboratory, Department of Zoology, College of Science, King Saud University, Riyadh 11451, Saudi Arabia

ARTICLE INFO

Keywords:

Molecular docking
Antiviral properties
Phytochemicals
Medicinal plants
SARS-CoV-2
COVID-19
Binding affinity
SARS-CoV-2 M^{pro}

ABSTRACT

A new SARS coronavirus (SARS-CoV-2) belonging to the genus *Betacoronavirus* has caused a pandemic known as COVID-19. Among coronaviruses, the main protease (M^{pro}) is an essential drug target which, along with papain-like proteases catalyzes the processing of polyproteins translated from viral RNA and recognizes specific cleavage sites. There are no human proteases with similar cleavage specificity and therefore, inhibitors are highly likely to be nontoxic. Therefore, targeting the SARS-CoV-2 M^{pro} enzyme with small molecules can block viral replication. The present study is aimed at the identification of promising lead molecules for SARS-CoV-2 M^{pro} enzyme through virtual screening of antiviral compounds from plants. The binding affinity of selected small drug-like molecules to SARS-CoV-2 M^{pro}, SARS-CoV M^{pro} and MERS-CoV M^{pro} were studied using molecular docking. Bonducellpin D was identified as the best lead molecule which shows higher binding affinity (-9.28 kcal/mol) as compared to the control (-8.24 kcal/mol). The molecular binding was stabilized through four hydrogen bonds with Glu166 and Thr190 as well as hydrophobic interactions via eight residues. The SARS-CoV-2 M^{pro} shows identities of 96.08% and 50.65% to that of SARS-CoV M^{pro} and MERS-CoV M^{pro} respectively at the sequence level. At the structural level, the root mean square deviation (RMSD) between SARS-CoV-2 M^{pro} and SARS-CoV M^{pro} was found to be 0.517 Å and 0.817 Å between SARS-CoV-2 M^{pro} and MERS-CoV M^{pro}. Bonducellpin D exhibited broad-spectrum inhibition potential against SARS-CoV M^{pro} and MERS-CoV M^{pro} and therefore is a promising drug candidate, which needs further validations through *in vitro* and *in vivo* studies.

1. Introduction

Coronaviruses (CoVs) are positive-sense RNA enveloped viruses which derive their name from the crown-like spikes on their surface and they belong to Coronaviridae family. They are classified into four main subgroups- alpha, beta, gamma, and delta depending on their genomic structure [1]. Alpha- and beta coronaviruses cause respiratory infections in humans and gastroenteritis in other mammals [2,3]. Until the last year, only six different coronaviruses causing infection in humans were known. Four of this (HCoV-NL63, HCoV-229E, HCoV-OC43 and HKU1) cause mild common cold-type symptoms in healthy individuals and the other two which caused epidemics in the last two decades. The

severe acute respiratory syndrome coronavirus (SARS-CoV) caused a SARS epidemic in 2002–2003 with a 10% mortality rate. Likewise, the Middle East respiratory syndrome coronavirus (MERS-CoV) caused a disastrous pandemic in 2012 leading to 37% mortality [1]. All coronaviruses infecting humans usually known to have intermediate hosts such as bats or rodents [4]. Previous outbreaks of SARS-CoV and MERS-CoV involved civet cats and dromedary camels for their direct transmission to humans [1]. A new coronavirus caused an outbreak of the pulmonary disease in Wuhan (the capital of Hubei province in China) in December 2019 and has since spread across different parts of the world [5,6]. Since its RNA genome shows about 82% identity to that of the SARS coronavirus (SARS-CoV), the new virus has been termed as SARS-

Abbreviations: ACE-2, Angiotensin Converting Enzyme 2; BLAST, Basic Local Alignment Search; cLogP, partition coefficient between N-octanol and water; HBA, hydrogen bond acceptor; HBD, hydrogen bond donor; MERS-CoV, Middle East respiratory syndrome coronavirus; Mpro, main protease; MW, molecular weight; PDB, Protein Data Bank; RMSD, root mean square deviation; ROF, rule of five; SARS-CoV, severe acute respiratory syndrome coronavirus

* Corresponding author.

E-mail address: arunbgurung@gmail.com (A.B. Gurung).

<https://doi.org/10.1016/j.lfs.2020.117831>

Received 28 March 2020; Received in revised form 11 May 2020; Accepted 19 May 2020

Available online 22 May 2020

0024-3205/© 2020 Elsevier Inc. All rights reserved.

Table 1

The grid box dimensions selected for molecular docking studies.

Target enzyme	Grid box dimension		
	Number of grid points (npts)	Center (xyz coordinates)	Grid point spacing (Å)
SARS-CoV-2 M ^{pro}	60 × 60 × 60	11.476, -1.396, 21.127	0.375
SARS-CoV M ^{pro}	65 × 65 × 65	22.803, 40.132, -10.620	0.375
MERS-CoV M ^{pro}	65 × 65 × 65	-16.001, 25.308, 10.579	0.375

CoV-2 [7]. However, both these viruses belong to the same clade of the genus *Betacoronavirus* [5,6]. The SARS-CoV-2 caused a disease known as COVID-19. At the initial outbreak, cases were linked to the Huanan seafood and animal market in Wuhan but active human-to-human transmission caused exponential growth in the number of reported cases. The World Health Organization (WHO) confirmed the outbreak a pandemic on March 11, 2020. There have been > 170,000 cumulative cases worldwide accounting for approximately 3.7% case-fatality rate as of March 15, 2020 [8].

Due to the close similarity to SARS-CoV, the biochemical interactions and the pathogenesis of SARS-CoV-2 are highly likely to be similar [1]. The virus entry into the host cell is mainly mediated through the binding of the SARS spike (S) protein to the angiotensin-converting enzyme 2 (ACE-2) receptor on the cell surface [9]. Among coronaviruses, the main protease (M^{pro}, also called 3CL^{pro}) has emerged as the best-described drug target [10]. The polyproteins that are translated from the viral RNA are processed by this enzyme together with the papain-like protease(s) [11]. The M^{pro} recognizes and acts remarkably on eleven cleavage sites typically Leu-Gln↓(Ser,Ala,Gly) on the large polyprotein 1ab (replicase 1ab) of approximately 790 kDa. Blocking the activity of this enzyme would help in inhibiting viral replication. There are no reported human proteases with a similar cleavage specificity and therefore, inhibitors against this enzyme are less probable to be toxic [8]. The three dimensional X-ray crystal structure of this enzyme in complex with α-ketoamide inhibitor 13b (O6K) was recently solved by Zhang et al. [8] (PDB ID: 6Y2F) which offers an opportunity for structure-based drug design against the enzyme target. Understanding the relevance of the steady rise in the number of infected and death cases in recent time from COVID-19 and lack of effective therapeutic interventions such as drugs and vaccines, computer-aided drug design is an important strategy to be sought after. This rational based drug design will reduce the cost and time incurred in the drug discovery process. Structure-based drug design primarily relies on molecular docking to identify lead molecules against the target proteins from chemical libraries [12,13]. Compared to the synthetic inhibitors plant based-drugs have less toxicity and much safer to use. The natural products such as traditional medicines and plant-derived compounds (phytochemicals) are the rich sources of promising antiviral drugs [14]. Around 44% of the approved antiviral drugs between 1981 and 2006 were derived from natural products [15]. The plant extracts have been extensively used and screened for drug molecules to evaluate theirs in vitro antiviral activities. Few examples of medicinal plants with proven antiviral activities include *Phyllanthus amarus* Schum. and Thonn which blocks human immunodeficiency virus (HIV) replication both in vitro and in vivo [16]; *Azadirachta indica* Juss. (Neem) shows in vitro and in vivo inhibition properties against Dengue virus type-2 (DENV-2) [17]; *Geranium sanguineum* L. significantly inhibits the replication of Herpes simplex virus type-1 and 2 (HSV-1 and HSV-2) in vitro [18]; *Acacia nilotica* L. possesses activity against Hepatitis C virus (HCV) in vitro etc. [19].

In the present study, we have screened small drug-like molecules from a dataset of phytochemicals possessing antiviral activities using

Table 2

List of phytochemicals selected for virtual screening.

Class	Phytochemicals	Name
Flavonoids	Cmpd1	(+)-Catechin
	Cmpd2	(-)-Epigallocatechin gallate (EGCG)
	Cmpd3	(-)-Epicatechin gallate (ECG)
	Cmpd4	(-)-Epigallocatechin (EGC)
	Cmpd5	Rutin
	Cmpd6	5,7-dimethoxyflavaN-4'-O-β-d-glucopyranoside
	Cmpd7	5,7,3'-trihydroxy-flavaN-4'-O-β-d-glucopyranoside
	Cmpd8	Genistein
	Cmpd9	Baicalein
	Cmpd10	Orientin
	Cmpd11	Vitexin
	Cmpd12	Chrysosplenol C
	Cmpd13	Quercetin 3-rhamnoside (Q3R)
	Cmpd14	Luteoforol
Organic acids	Cmpd15	3,4-di-O-caffeoylequinic acid
	Cmpd16	3,5-di-O-caffeoylequinic acid
	Cmpd17	Raoulic acid
	Cmpd18	Caffeic acid
	Cmpd19	Chlorogenic acid
	Cmpd20	Naphthoindozidine alkaloid
Alkaloids	Cmpd21	7-demethoxytylophorine
	Cmpd22	7-demethoxytylophorine N-oxide
	Cmpd23	Sophoridine
	Cmpd24	Lycorine
Essential oils	Cmpd25	Himachalol
	Cmpd26	α-himachalene
	Cmpd27	β-himachalene
Stilbenes	Cmpd28	Resveratrol
	Cmpd29	Balanocarpol
	Cmpd30	Shegansu B
Others	Cmpd31	Gnetupendin B
	Cmpd32	Gnetin D
	Cmpd33	Isorhapontigenin
Others	Cmpd34	Caesalmin B
	Cmpd35	Bonducellpin D
	Cmpd36	Chrysophanic acid (1,8-dihydroxy-3-methylanthraquinone)
	Cmpd37	Emodin
	Cmpd38	Hippomanin A

drug-like filters and toxicity studies. The selected molecules were evaluated for their binding affinity to SARS-CoV-2 M^{pro} enzyme using molecular docking. The sequences and structures of SARS-CoV-2 M^{pro} were compared with SARS-CoV M^{pro} and MERS-CoV M^{pro}. The identified lead molecules were further examined for their broad-spectrum inhibition potential against SARS-CoV M^{pro} and MERS-CoV M^{pro}.

2. Material and methods

2.1. Ligand preparation

The information about a set of thirty-eight phytochemicals from

Table 3

Physicochemical properties of the compounds (asterisk indicates the compounds which passed the toxicity studies). a: Molecular weight, b: Partition coefficient between n-octanol and water, c: Hydrogen bond acceptor, d: Hydrogen bond donor, N: None, H: High, L: Low.

Compounds	MW ^a (Da)	clogP ^b	HBA ^c	HBD ^d	Mutagenic	Tumourigenic	Reproductive effective	Irritant
Cmpd1*	290.270	1.5087	6	5	N	N	N	N
Cmpd2	458.374	2.0543	11	8	N	N	N	N
Cmpd3	442.375	2.4	10	7	N	N	N	N
Cmpd4	306.269	1.163	7	6	N	N	N	N
Cmpd5	610.519	-1.2573	16	10	N	N	N	N
Cmpd6*	462.449	0.7178	10	4	N	N	N	N
Cmpd7	450.395	-0.1793	11	7	N	N	N	N
Cmpd8	270.239	1.6272	5	3	H	H	H	N
Cmpd9*	270.239	2.3357	5	3	N	N	N	N
Cmpd10	448.379	-0.4237	11	8	H	N	N	N
Cmpd11	432.380	-0.078	10	7	H	N	N	N
Cmpd12	360.317	2.1238	8	3	H	N	N	N
Cmpd13	448.379	0.5798	11	7	N	N	N	N
Cmpd14*	290.270	1.6348	6	5	N	N	N	N
Cmpd15	516.454	0.7977	12	7	N	N	N	N
Cmpd16	516.454	0.7977	12	7	N	N	N	N
Cmpd17	370.575	8.2146	2	1	N	N	N	N
Cmpd18	180.159	0.7285	4	3	H	H	H	N
Cmpd19	354.310	-0.7685	9	6	N	N	N	N
Cmpd20	353.460	3.5777	4	0	N	H	N	H
Cmpd21	363.456	4.0868	4	0	L	H	N	N
Cmpd22	379.455	3.1134	5	0	L	H	N	N
Cmpd23*	248.369	1.7074	3	0	N	N	N	N
Cmpd24*	287.314	1.2078	5	2	N	N	N	N
Cmpd25*	222.370	3.5349	1	1	N	N	N	N
Cmpd26	204.356	4.3321	0	0	N	N	N	L
Cmpd27*	204.356	4.4247	0	0	N	N	N	N
Cmpd28	228.246	2.8295	3	3	H	N	H	N
Cmpd29	470.476	3.3075	7	6	N	N	N	N
Cmpd30	514.528	4.6839	8	5	N	N	N	N
Cmpd31	380.395	3.884	6	5	N	N	H	N
Cmpd32	470.476	4.4782	7	6	N	N	H	N
Cmpd33	258.272	2.7595	4	3	N	N	H	N
Cmpd34*	388.458	2.5856	6	1	N	N	N	N
Cmpd35*	404.457	1.7335	7	2	N	N	N	N
Cmpd36	254.240	2.6859	4	2	H	N	N	H
Cmpd37	270.239	2.3402	5	3	H	H	H	H
Cmpd38	634.454	0.0527	18	11	N	N	N	N
Control	594.687	1.7119	12	4	N	N	N	N

Table 4

Standard Protein BLAST result using SARS-CoV-2 M^{pro} as a query sequence.

Subject sequence	Maximum score	Total score	Query coverage	E value	Percent identity
SARS-CoV M ^{pro}	623	623	100%	0.0	96.08%
MERS-CoV M ^{pro}	322	322	100%	1e-114	50.65%

medicinal plants with antiviral activities was retrieved through literature search [14]. The summary of the selected phytochemicals (class, plant source and antiviral activity) is provided in Suppl. Table 1. The two-dimensional structures of the phytochemicals were generated using Chemsketch and modelled into 3D using open Babel and the structure optimization was carried out with an MMFF94 force field [20] using our previously defined protocol [21]. These optimized structures were used for virtual screening. A set of twelve FDA approved antiviral drugs such as Raltegravir (CID54671008), Dolutegravir (CID54726191), Nelfinavir (CID64143), Lopinavir (CID92727), Zidovudine (CID35370), Nevirapine (CID4463), Simeprevir (24873435), Boceprevir (CID10324367), Zanamivir (CID60855), Oseltamivir (CID65028), Acyclovir (CID135398513) and Ganciclovir (CID135398740) were retrieved from PubChem database [22].

2.2. Screening of small drug-like molecules

The phytochemicals were screened based on physicochemical properties obeying Lipinski's rule of five (ROF) filters [23] and further tested for in silico toxicity studies such as mutagenicity, tumorigenicity, reproductive effects and irritancy. The physicochemical properties of the phytochemicals were determined using DataWarrior program version 4.6.1 [24].

2.3. Receptor preparation

The three-dimensional structures of SARS-CoV-2 M^{pro}, SARS-CoV M^{pro} and MERS-CoV M^{pro} were retrieved from Protein Data Bank (PDB) (<http://www.rcsb.org/>) using PDB IDs: 6Y2F, 3TNT and 5WKK at a

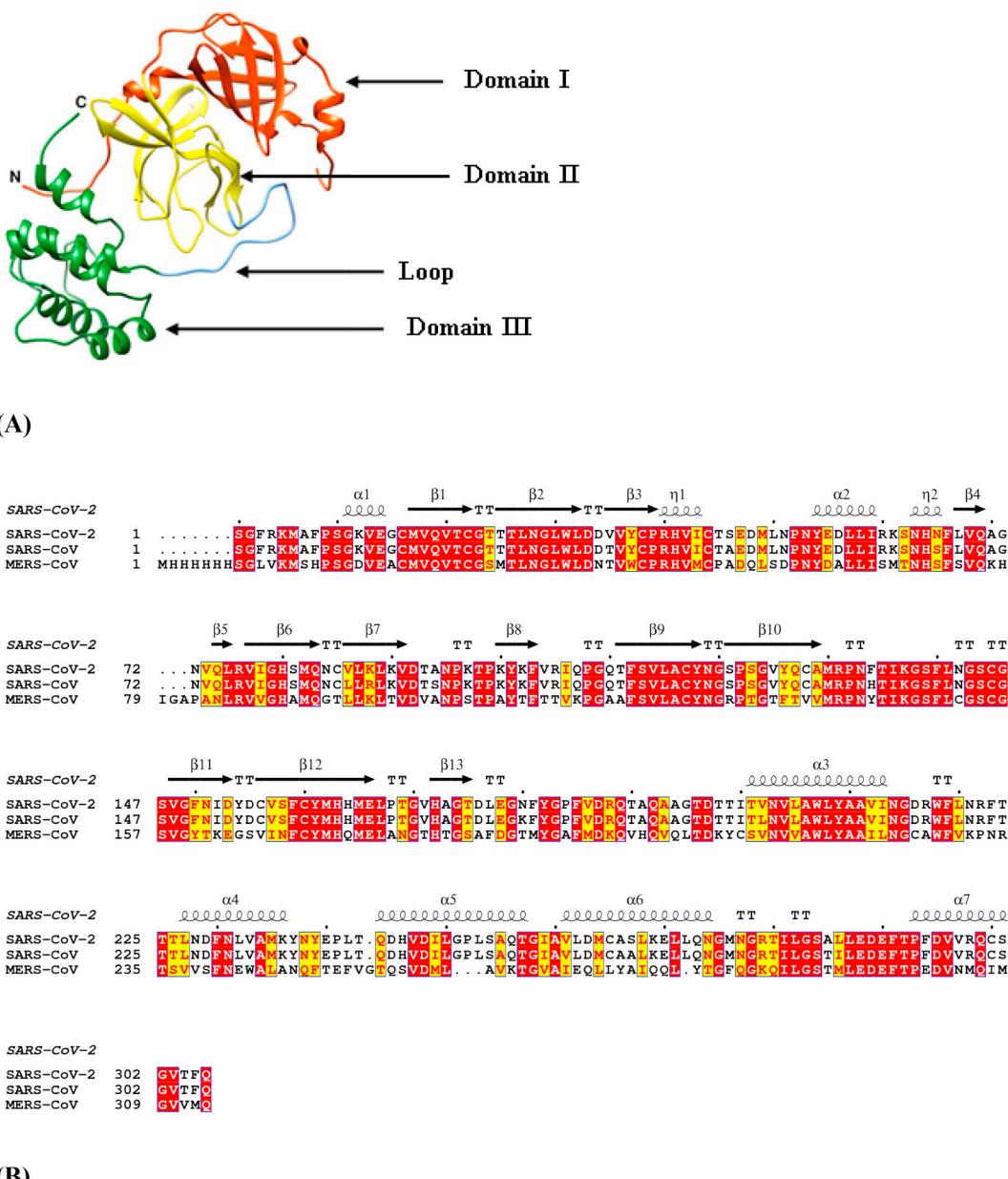


Fig. 1. Sequence and structural analysis of SARS-CoV-2 M^{pro}. (A) three-dimensional structure of SARS-CoV-2 M^{pro} consists of three domains-I (orange-red), II (yellow) and III (forest green) where domain II and III are connected by a loop (cornflower blue), (B) Multiple Sequence alignment (MSA) of SARS-CoV-2 M^{pro}, SARS-CoV M^{pro} and MERS-CoV M^{pro}. The red box with white characters indicates strict identity whereas yellow box with red characters corresponds to semi-conserved residues. The secondary structure elements such as α -helices and 310-helices (η) are displayed as squiggles and β -strands and strict β -turns are rendered as arrows and TT letters respectively. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

resolution of 1.95 Å, 1.59 Å and 1.55 Å respectively. The receptors were prepared for molecular docking using our previously defined protocol [25]. The binding sites were defined by choosing grid boxes of suitable dimensions around the bound co-crystal ligand. The grid box dimensions for SARS-CoV-2 M^{pro}, SARS-CoV M^{pro} and MERS-CoV M^{pro} are represented in Table 1.

2.4. Determination of sequence percentage identity, multiple sequence alignment and pairwise structural clustering

The sequence percentage identity of the SARS-CoV-2 M^{pro} to SARS-CoV M^{pro} and MERS-CoV M^{pro} was determined using a standard protein

Basic Local Alignment Search (BLASTp) tool (<https://blast.ncbi.nlm.nih.gov/Blast.cgi?PAGE=Proteins>). Multiple sequence alignment of the three sequences were performed using CLUSTAL W algorithm [26]. Pairwise structural clustering of SARS-CoV-2 M^{pro}, SARS-CoV M^{pro} and MERS-CoV M^{pro} was analyzed using UCSF Chimera tool [27].

2.5. Validation of docking method

To check the suitability of molecular docking parameters and algorithm to reproduce the native binding poses, a redocking experiment was performed using the co-crystal compound.

2.6. Molecular docking

Molecular docking was executed using AutoDock4.2 software [28]. Lamarckian Genetic Algorithm was used for performing molecular docking studies and the docking parameters were chosen from our previously defined protocol [21]. A total number of fifty independent docking runs were performed for each compound. The conformational clustering of the poses was performed by considering a root mean square deviation (RMSD) of $< 2.0 \text{ \AA}$. The most favourable binding poses of the compounds were analyzed by choosing the lowest free energy of binding (ΔG) and the lowest inhibition constant (K_i) which is calculated using the following equation (Eq. (1)):

$$K_i = \text{exponential}(\Delta G/RT) \quad (1)$$

where ΔG is the binding energy in kcal/mol, R is the universal gas constant ($1.987 \text{ cal K}^{-1} \text{ mol}^{-1}$) and T is the temperature (298.15 K).

A stable complex is formed between a protein and ligand which exhibits more negative free energy of binding and low K_i indicates high potency of an inhibitor [29,30]. The hydrogen bonds and hydrophobic interactions between the compounds and the target enzymes were studied using LigPlot+ v 1.4.5 [31].

3. Results and discussion

3.1. Selection of phytochemicals for the study

A total of 38 bioactive phytochemicals possessing antiviral activities were selected for the study. These compounds were chosen based on the previous reports of their potent antiviral effects against various pathogenic viruses such as adenovirus, influenza virus, respiratory syncytial virus, human cytomegalovirus, herpes simplex virus, poliovirus, varicella-zoster virus etc. (Suppl. Table 1). The compound set consists of different classes of phytochemicals including active flavonoids ($N = 14$), active organic acids ($N = 5$), active alkaloids ($N = 5$), active essential oils ($N = 3$), active stilbenes ($N = 6$) and other phytoconstituents ($N = 5$). The three-dimensional structures the compounds were modelled and optimized. A list of these phytochemicals is enumerated in Table 2. These optimized structures were used further for

virtual screening and molecular docking studies.

3.2. Screening of small drug-like molecules

From a total of 38 phytochemicals, 10 compounds (four active flavonoids, two active alkaloids, two active essential oils and two other other phytoconstituents) were found to be orally bioactive with respect to ROF criterion (Molecular weight (MW) ≤ 500 , cLogP (partition coefficient between n-octanol and water) ≤ 5 , number of hydrogen bond donors (HBD) ≤ 5 and number of hydrogen bond acceptors (HBA) ≤ 10 [23]) without any significant toxicity issues such as being non-mutagenic, non-tumourigenic, non-irritant and no adverse effects on reproductive health (Table 3). These drug-like compounds were further taken for molecular docking studies. The drug-attrition rate in pre-clinical and clinical trials is quite high due to the poor pharmacokinetic studies [25] and therefore initial screening of these drug-like molecules can increase the chances of passing through the clinics.

3.3. Comparative analysis of SARS-CoV-2 M^{pro}, SARS-CoV M^{pro} and MERS-CoV M^{pro} at the sequence and structural level

The primary sequences of SARS-CoV-2 M^{pro} was compared with that of SARS-CoV M^{pro} and MERS-CoV M^{pro} and it was found that SARS-CoV-2 M^{pro} shares identity percentages of 96.08 and 50.65 with that of SARS-CoV M^{pro} and MERS-CoV M^{pro} respectively (Table 4). The MSA results show the conservation of residues in the three domains-I (residues 1–99), II (residues 100–182) and III (residues 199–306) of the SARS-CoV-2 M^{pro} enzyme which includes catalytic dyad residues His41-Cys145 located in a cleft between domains I and II along with canonical oxyanion hole residues Gly143, Ser144 and Cys145 (Fig. 1). Domains I and II have a typical chymotrypsin-like two-β-barrel fold. Domain III possesses a globular structure consisting of five α-helices which is connected to Domain II by a loop of residues 183–198 [32]. The pairwise structural imposition of SARS-CoV-2 M^{pro} and SARS-CoV M^{pro} yields an RMSD value of 0.517 \AA while an RMSD value of 0.817 \AA was obtained when SARS-CoV-2 M^{pro} was superimposed on MERS-CoV M^{pro} (Fig. 2). This indicates that SARS-CoV-2 M^{pro} is structurally more close to SARS-CoV M^{pro} as compared to MERS-CoV M^{pro}. This comparative

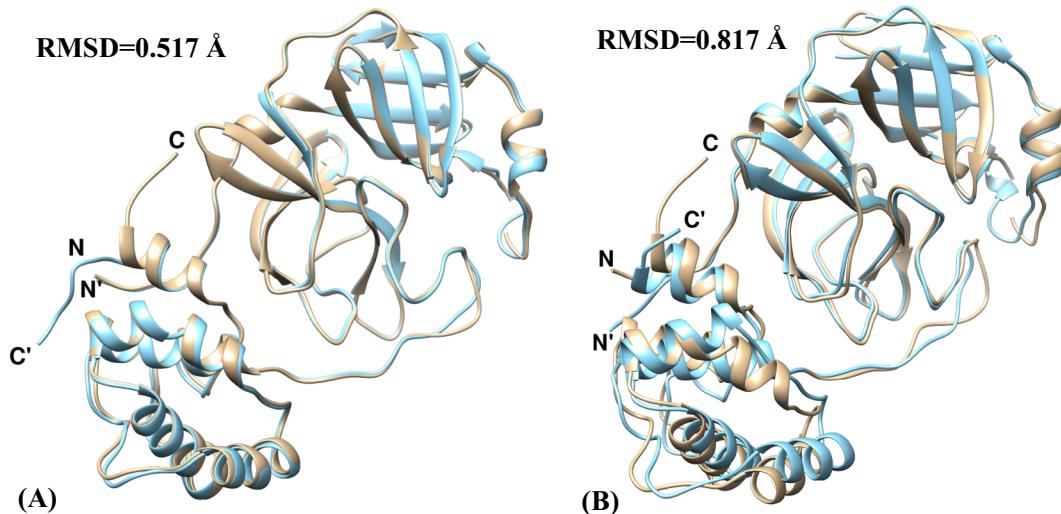


Fig. 2. Structural clustering- (A) SARS-CoV-2 M^{pro} (tan) and SARS-CoV M^{pro} (cyan), (B) SARS-CoV-2 M^{pro} (tan) and MERS-CoV M^{pro} (cyan). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 5

Molecular docking result for selected compounds and control against SARS-CoV-2 M^{pro}.

Compounds	Name	Structure	Binding energy (kcal/mol)	Inhibition constant (nM)
Cmpd1	(+)-Catechin		-7.86	1720
Cmpd6	5,7-dimethoxyflavanone-4'-O-β-d-glucopyranoside		-9.23	170.31
Cmpd9	Baicalin		-6.51	16,950
Cmpd14	Luteoforol		-7.02	7100
Cmpd23	Sophoridine		-6.76	11,120
Cmpd24	Lycorine		-7.67	2380
Cmpd25	Himachalol		-7.00	7390
Cmpd27	β-himachalene		-5.80	55,970
Cmpd34	Caesalmin B		-8.82	342.79
Cmpd35	Bondurellipin D		-9.28	156.75
Control	α-ketoamide 13b		-8.24	905.88

study at the sequence and structural level can be exploited for the design of broad-spectrum inhibitors which can block viral infection in host cells.

3.4. Binding interaction studies of drug-like molecules with SARS-CoV-2 M^{pro}

The molecular docking protocol and algorithm was validated through the redocking experiment. The redocking result of co-crystal compound (6OK) shows an RMSD value of 3.603 Å between the docked and native co-crystal position. This small deviation indicates that the docking protocols and parameters used in the study can reliably estimate the biological conformations of the compounds [33].

The ten selected drug-like small molecules were docked into the active site pocket of SARS-CoV-2 M^{pro}. The molecular docking results are shown in Table 5. The docking scores were bench-marked using the standard inhibitor, α-ketoamide 13 b (6OK). The α-ketoamide 13 b is a broad-spectrum inhibitor with an IC₅₀ value of 0.67 ± 0.18 μM, 0.90 ± 0.29 μM and 0.58 ± 0.22 μM respectively against purified recombinant SARS-CoV-2 M^{pro}, SARS-CoV M^{pro} and MERS-CoV M^{pro} [8]. On comparing the binding energies and inhibition constants with that of the standard inhibitor, we have shortlisted three lead molecules viz., Bondurellipin D, 5,7-dimethoxyflavanone-4'-O-β-d-glucopyranoside and Caesalmin B. Bondurellipin D and Caesalmin B are terpenoids isolated from the seeds of traditional Chinese medicinal plant *Cae-salpinia minax* and exhibit in vitro inhibitory activities against Parainfluenza-3 (PI-3) virus. [34] The flavonoid 5,7-dimethoxyflavanone-4'-O-β-d-glucopyranoside has previously been reported to have antiviral activity against PI-3 [35]. The molecular interactions of LigPlot results are represented in Table 6. The best-ranked lead molecule i.e., Bondurellipin D interacts with the enzyme target with a binding energy of -9.28 kcal/mol and inhibition constant of 156.75 nM. Bondurellipin D establishes four hydrogen bonds with residues Glu166 and Thr190 and hydrophobic interactions via eight residues (Met49, His164, Met165, Pro168, Asp187, Arg188, Gln189 and Gln192) (Fig. 3A). The second lead molecule i.e., 5,7-dimethoxyflavanone-4'-O-β-d-glucopyranoside interacts with the target enzyme with a binding energy of -9.23 kcal/mol and inhibition constant of 170.31 nM. This interaction is strengthened by the establishment of five hydrogen bonds with residues Glu166, Thr190 and Gln192 and with the involvement of thirteen residues (Phe140, Leu141, Asn142, Gly143 Ser144, Cys145, His164, Met165, Leu167, Pro168, His172, Arg188 and Gln189) contributing to

Table 6

LigPlot analysis for top three lead compounds along with the control against SARS-CoV-2 M^{pro}.

Compounds	Name	Hydrogen bonds	Hydrophobic interactions
Cmpd35	Bondurellipin D	O2....O(Glu166) (2.83 Å) O2....N(Glu166) (2.91 Å) O4....O(Glu166) (2.61 Å) O6....N(Thr190) (2.92 Å)	Met49, His164, Met165, Pro168, Asp187, Arg188, Gln189 and Gln192 (N = 8)
Cmpd6	5,7-dimethoxyflavanone-4'-O-β-d-glucopyranoside	O8....O(Glu166) (2.76 Å) O1....N(Glu166) (3.22 Å) O7....N(Thr190) (3.03 Å) O7....N(Thr190) (2.44 Å) O9....O(Gln192) (2.91 Å)	Phe140, Leu141, Asn142, Gly143 Ser144, Cys145, His164, Met165, Leu167, Pro168, His172, Arg188 and Gln189 (N = 13)
Cmpd34	Caesalmin B	O2....N(Glu166) (3.32 Å) O2....O(Glu166) (2.46 Å) O6....N(Thr190) (3.01 Å) O6....NE2(Gln192) (2.87 Å)	Met49, His41, Tyr54, Cys145, His164, Met165, Leu167, Asp187, Arg188 and Glu189 (N = 10)
Control	α-ketoamide 13b	N23....OD1(Asn142) (2.83 Å) O37....SG(Cys145) (3.31 Å) O41....N(Glu166) (3.12 Å) O40....N(Glu166) (2.98 Å)	Thr26, His41, Gly143, Phe140, Leu141, Ser144, His163, His164, Met165, Pro168, Asp187, Gln189, Thr190 and Gln192 (N = 14)

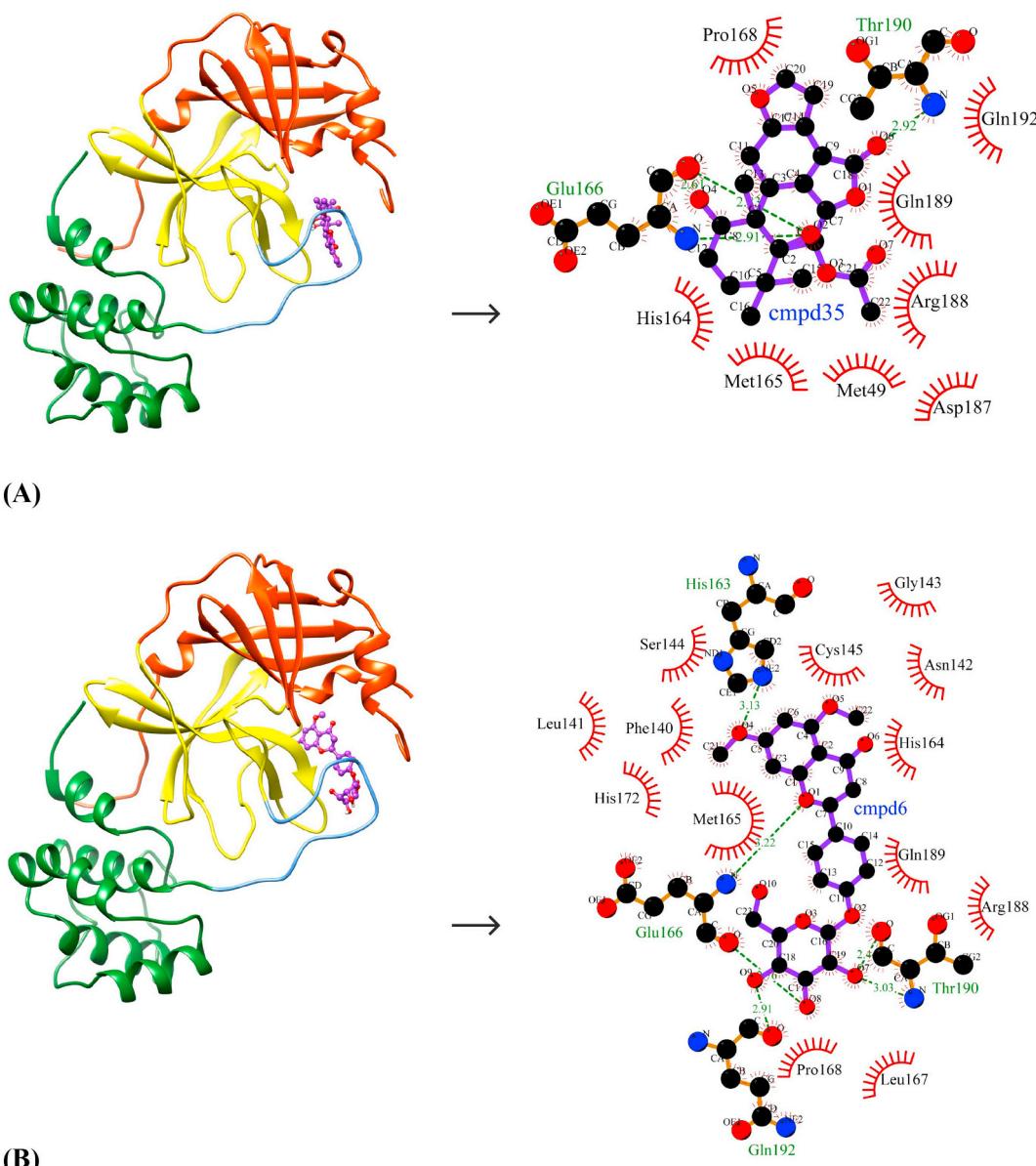


Fig. 3. Binding poses and molecular interactions between lead compounds and SARS-CoV-2 M^{pro}. (A) Bonducelpin D, (B) 5,7-dimethoxyflavanone-4'-O-β-d-glucopyranoside, (C) Caesalmin B, (D) α-ketoamide 13b (Control). The binding poses display the target enzyme in ribbon form with structural domains-I (orange-red), II (yellow) and III (forest green) where domain II and III are connected by a loop (cornflower blue) and the bound compounds are rendered as ball-and-stick (purple). The molecular interactions show hydrogen bonds as green dashed lines and hydrophobic interactions as semi-arcs with red eyelashes. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

hydrophobic interactions (Fig. 3B). The third lead molecule- Caesalmin B interacts with SARS-CoV-2 M^{pro} with a binding energy of -8.82 kcal/mol and inhibition constant of 342.79 nM. The molecular interaction is facilitated through four hydrogen bonds with residues Glu166, Thr190 and Gln192 and hydrophobic interactions via ten residues (His41, Met49, Tyr54, Cys145, His164, Met165, Leu167, Asp187, Arg188 and Glu189) (Fig. 3C). The standard inhibitor interacts with SARS-CoV-2 M^{pro} with a binding energy of -8.24 kcal/mol and inhibition constant of 905.88 nM. The binding interaction is mediated through four hydrogen bonds with residues Asn142, Cys145 and Glu166, and hydrophobic interactions via fourteen residues (Thr26, His41, Phe140, Leu141, Gly143, Ser144, His163, His164, Met165, Pro168, Asp187,

Gln189, Thr190 and Gln192) (Fig. 3D). Interestingly, the residues His41 and Cys145 which form catalytic dyad residues are also found interacting with the inhibitor. Thus all the three lead molecules have better binding affinity to SARS-CoV-2 M^{pro} compared to the standard inhibitor.

3.5. Binding interaction studies of the lead compounds with SARS-CoV M^{pro} and MERS-CoV M^{pro}

Due to the close similarity of SARS-CoV-2 M^{pro} to SARS-CoV M^{pro} and MERS-CoV M^{pro} at the structural level, the identified lead molecules may have broad-spectrum inhibitory activity and therefore, the

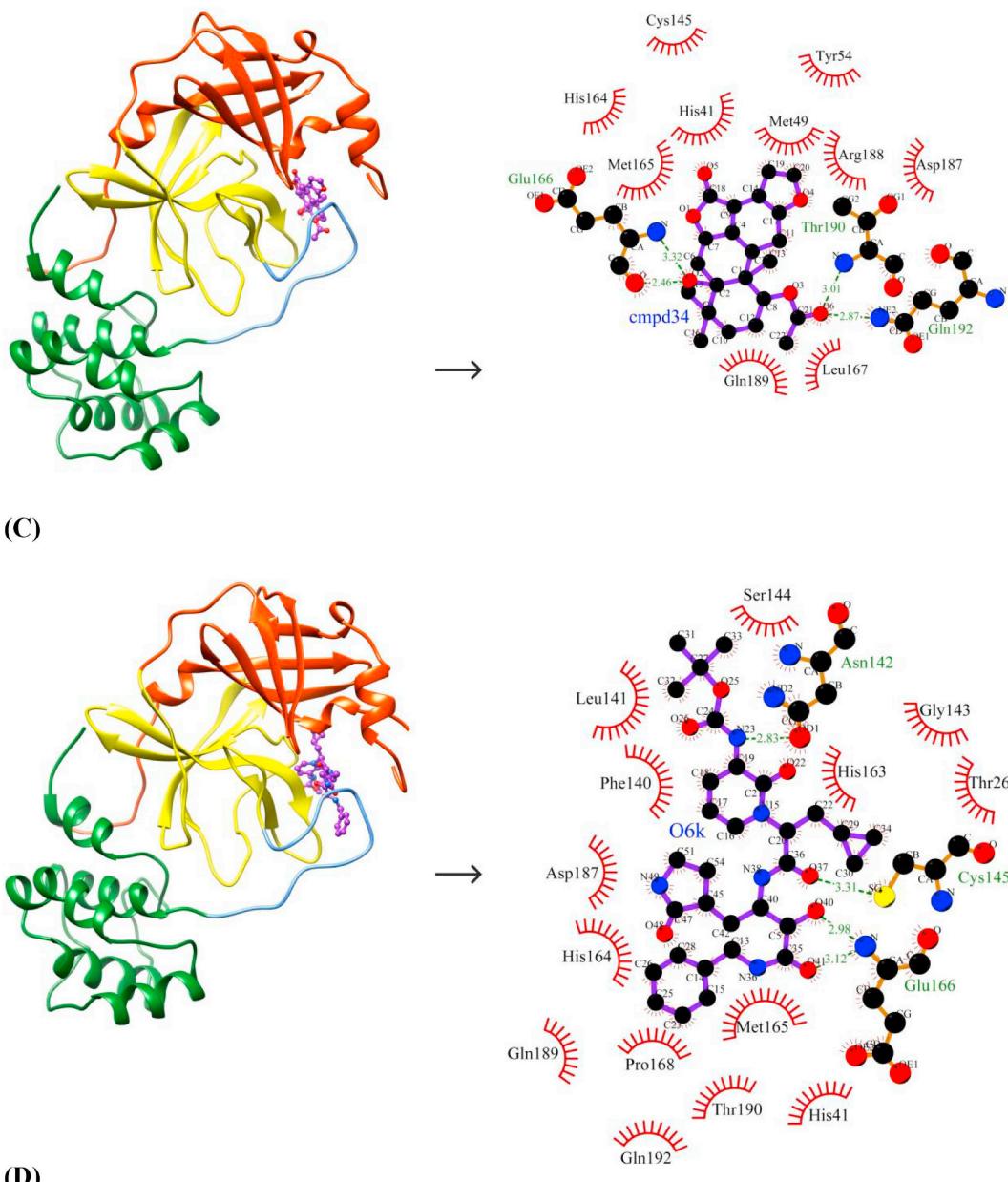


Fig. 3. (continued)

Table 7Molecular docking results for top 3 leads and control against SARS-CoV M^{pro} and MERS-CoV M^{pro}.

Compounds	Name	SARS-CoV M ^{pro}		MERS-CoV M ^{pro}	
		Binding energy (kcal/mol)	Inhibition constant (nM)	Binding energy (kcal/mol)	Inhibition constant (nM)
Cmpd35	Bonducellpin D	-8.66	467.11	-8.93	284.86
Cmpd6	5,7-dimethoxyflavan-4'-O- β -d-glucopyranoside	-8.94	282.15	-8.55	539.13
Cmpd34	Caesalmin B	-8.71	412.24	-9.49	111.50
control	α -ketoamide 13b	-8.62	482.81	-11.36	4.75

binding affinities of the lead molecules against the other two target enzymes were studied. The binding energies and inhibition constants of the compounds against SARS-CoV M^{pro} and MERS-CoV M^{pro} are shown in Table 7. The best-ranked lead molecule i.e., Bonducellpin D exhibits binding energies and inhibition constants of -8.64 kcal/mol and 467.11 nM; -8.93 kcal/mol and 284.86 nM against SARS-CoV M^{pro} and MERS-CoV M^{pro} respectively. It binds to SARS-CoV M^{pro} by establishing two hydrogen bonds with Gln189 and hydrophobic interacting

residues include His41, Cys44, Met49, Tyr54, Asn142, Cys145, His164, Met165, Glu166, Asp187 and Arg188 ($N = 11$) (Fig. 4A). Bonducellpin D also shows good molecular interaction with MERS-CoV M^{pro} using four hydrogen bonds with residues His41, Ser147, Cys148 and Glu169 and residues such as Met25, Leu49, Phe143, Leu144, Cys145, Gly146, His166, Gln167, Met168 and Gln192 ($N = 10$) participate in hydrophobic interactions (Fig. 5A). With binding energies and inhibition constants of -8.94 kcal/mol and 282.15 nM; -8.55 kcal/mol and

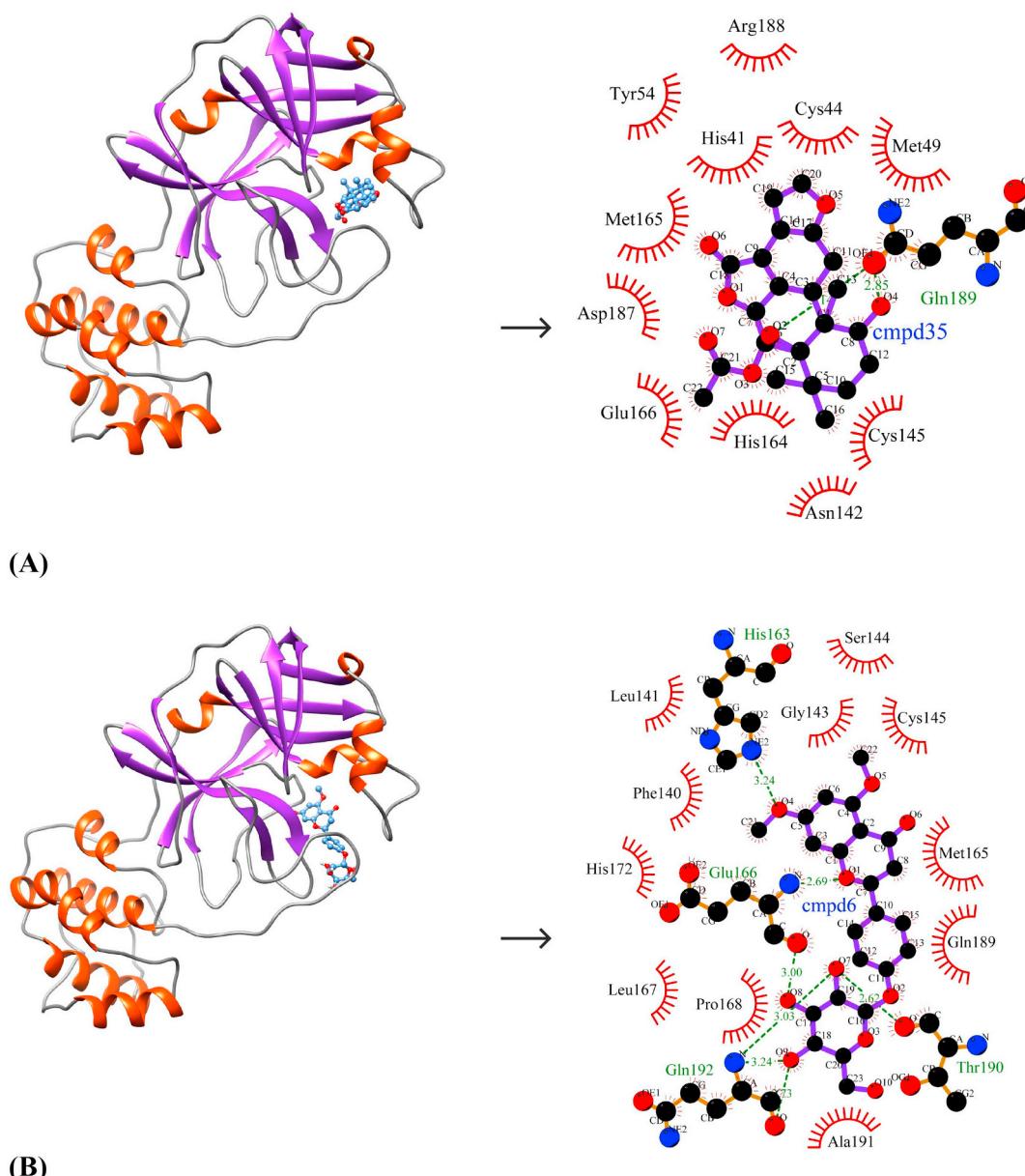


Fig. 4. Binding pose and molecular interaction between lead compounds and SARS-CoV M^{pro}. (A) Bonducellpin D, (B) 5,7-dimethoxyflavanone-4'-O-β-d-glucopyranoside, (C) Caesalmin B, (D) α-ketoamide 13b (Control). The binding poses represent the target enzyme as ribbon where helices, sheets and loops are indicated by orange-red, purple and grey respectively and the bound compounds as ball-and-stick (cornflower blue). The molecular interactions are represented by the green dashed lines for hydrogen bonds and semi-arcs with red eyelashes for hydrophobic interactions. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

539.13 nM against SARS-CoV M^{pro} and MERS-CoV M^{pro} respectively, the second lead molecule i.e., 5,7-dimethoxyflavanone-4'-O-β-d-glucopyranoside also shows broad-spectrum antiviral activities. 5,7-dimethoxyflavanone-4'-O-β-d-glucopyranoside shows favourable interactions with SARS-CoV M^{pro} through seven hydrogen bonds with His163, Glu166, Thr190 and Gln192 and hydrophobic interactions via residues Phe140, Leu141, Gly143, Ser144, Cys145, Met165, Leu167, Pro168, His172, Gln189 and Ala191 ($N = 11$) (Fig. 4B). It also exhibits good binding to MERS-CoV M^{pro} through four hydrogen bonds via

residues Gly146, Gln192 and Gln195 and the interaction is further strengthened through hydrophobic interactions via Met25, His41, Leu49, Cys148, Gln167, Met168, Glu169, Leu170, Ala171, Lys191, Val193 and His194 ($N = 12$) (Fig. 5B). The third lead compound Caesalmin B shows binding energies and inhibition constants of -8.71 kcal/mol and 412.24 nM; -9.49 kcal/mol and 111.50 nM against SARS-CoV M^{pro} and MERS-CoV M^{pro} respectively. Caesalmin B shows optimum binding to SARS-CoV M^{pro} by forming two hydrogen bonds with residues Gly143 and Glu166 and hydrophobic interactions

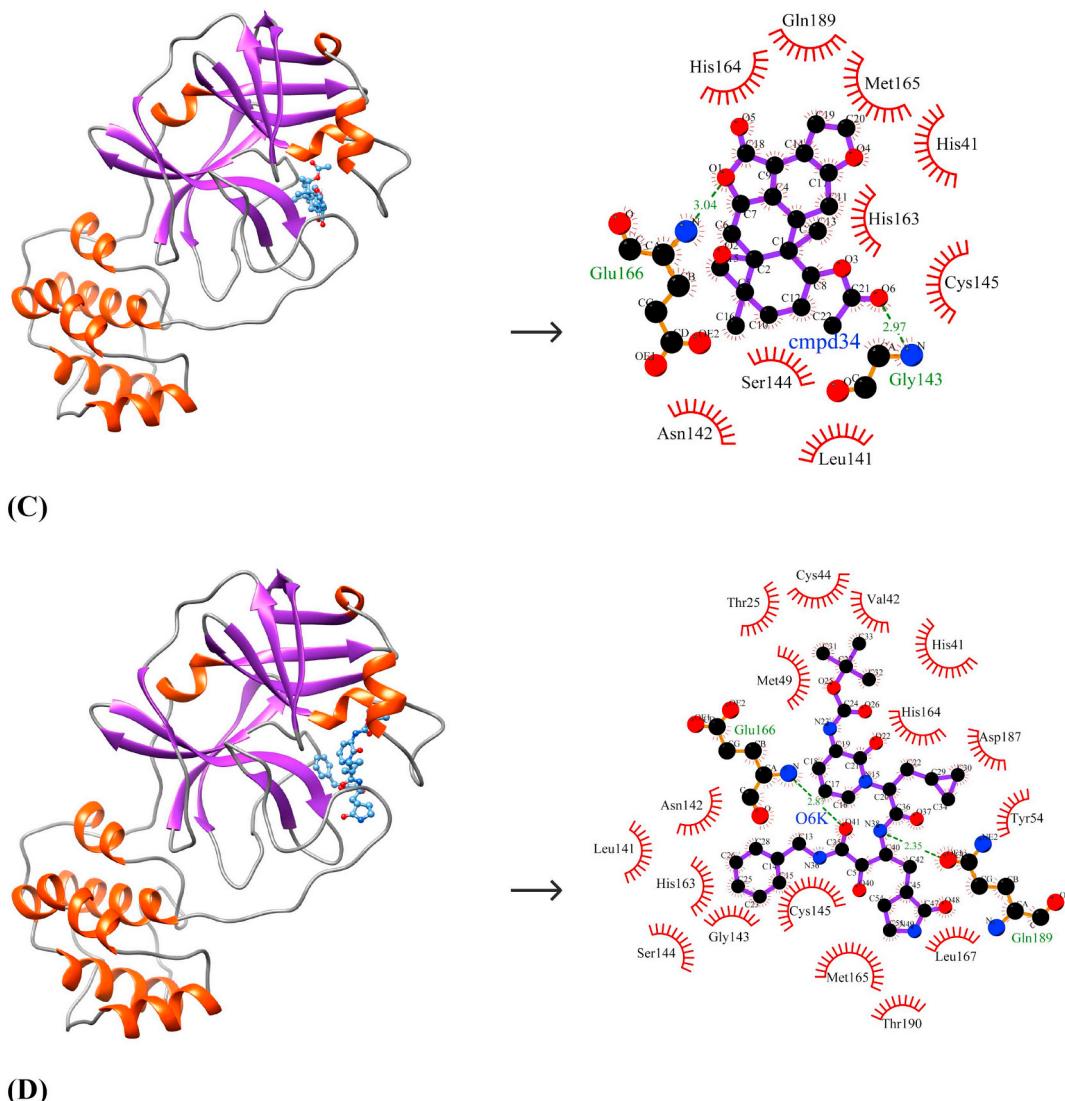


Fig. 4. (continued)

via residues His41, Leu141, Asn142, Ser144, Cys145, His163, His164, Met165 and Gln189 ($N = 9$) (Fig. 4C). It also exhibits good binding to MERS-CoV M^{pro} by establishing three hydrogen bonds with residues Cys145, Cys148 and His166 and hydrophobic interactions via residues Met25, Thr26, Leu27, Phe143, Leu144, Gly146, Ser147, Gln167, Met168, Glu169 and Gln192 ($N = 11$) (Fig. 5C). The control shows binding energies and inhibition constants of -8.62 kcal/mol and 482.81 nM; -11.36 kcal/mol and 4.75 nM against SARS-CoV M^{pro} and MERS-CoV M^{pro} respectively. The interaction between the control and SARS-CoV M^{pro} is mediated through two hydrogen bonds with Glu166 and Gln189 and hydrophobic interactions through residues-Thr25, His41, Val42, Cys44, Met49, Tyr54, Leu141, Asn142, Gly143, Ser144, Cys145, His163, His164, Met165, Leu167, Asp187 and Thr190 ($N = 17$) (Fig. 4D). It also shows good binding to MERS-CoV M^{pro} which involves seven hydrogen bonds with Cys145, Ser147, Cys148, Gln167 and Glu169 and hydrophobic interactions via residues Met25, Thr26, Leu27, His41, Phe143, Leu144, Gly146, His166, Met168,

Leu170, Ala171, Gln192, Val193, His194 and Gln195 ($N = 15$) (Fig. 5D).

3.6. Comparison of binding affinity of the lead molecules with FDA-approved antiviral drugs

The binding energies and inhibition constants of the phytochemicals with the SARS-CoV-2 M^{pro} enzyme were compared with that of a set of twelve FDA approved antiviral drugs-a) Viral integrase inhibitors (Raltegravir and Dolutegravir) b) HIV-1 protease inhibitors (Nelfinavir and Lopinavir) c) HIV-1 reverse transcriptase inhibitors (Zidovudine and Nevirapine) d) HCV NS3/NS4B protease inhibitors (Simeprevir and Boceprevir) e) Neuraminidase inhibitors (Zanamivir and Oseltamivir) and f) HSV-1 thymidine kinase inhibitors (Acyclovir and Ganciclovir) (Suppl. Table 2). The results show that the binding energies and inhibition constant values of the phytochemicals were lower as compared to the majority of the approved antiviral drugs except for Nelfinavir

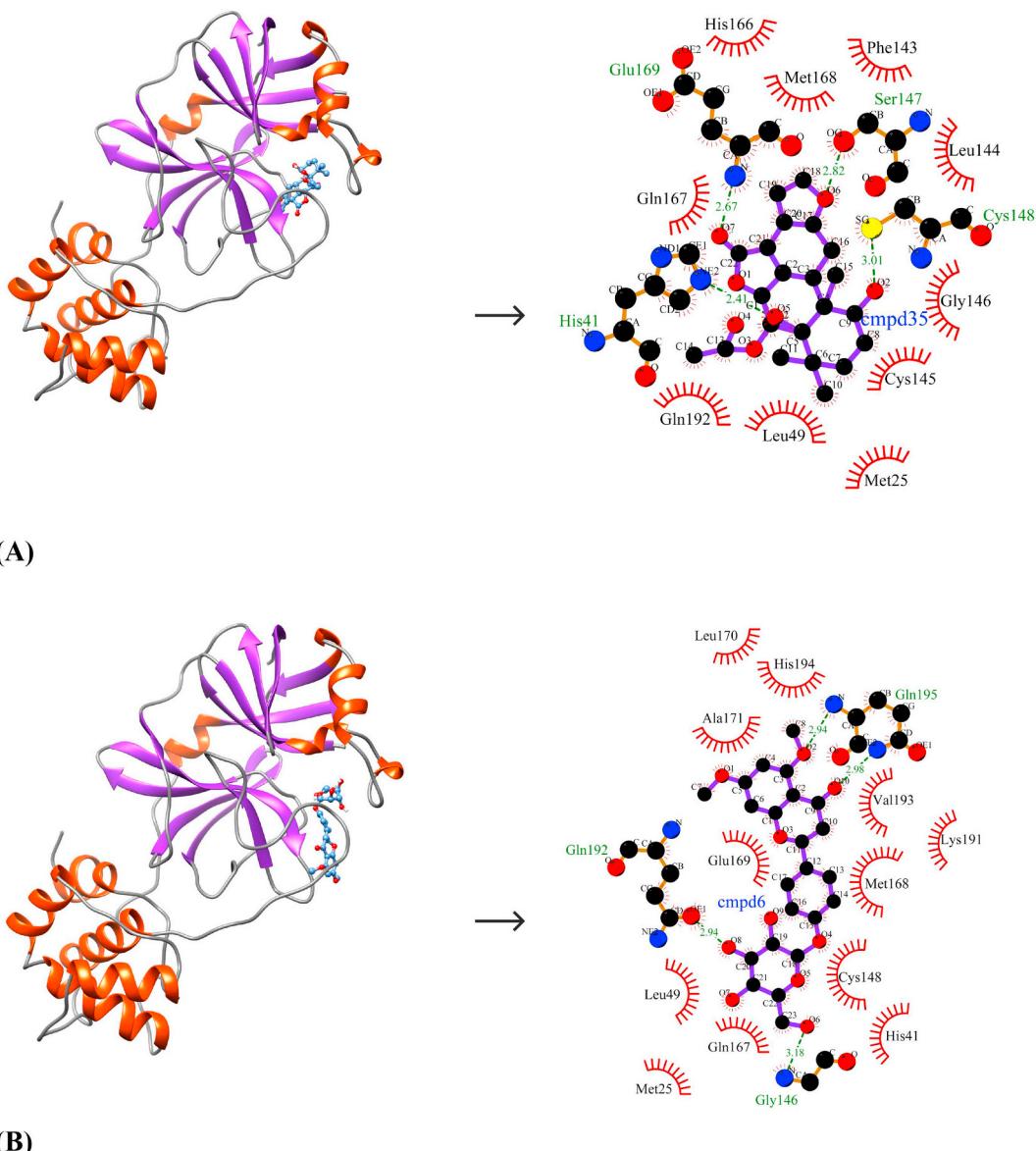


Fig. 5. Molecular interaction between lead compounds and MERS-CoV M^{pro}. (A) Bonducellpin D, (B) 5,7-dimethoxyflavanone-4'-O-β-d-glucopyranoside, (C) Caesalmin B, (D) α-ketoamide 13b (Control). The binding poses represent the target enzyme as ribbon where helices, sheets and loops are indicated by orange-red, purple and grey respectively and the bound compounds as ball-and-stick (cornflower blue). The molecular interactions are represented by the green dashed lines for hydrogen bonds and semi-arcs with red eyelashes for hydrophobic interactions. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

(−9.54 kcal/mol), Boceprevir (−9.3 kcal/mol) and Simeprevir (−11.1 kcal/mol) which indicates that these phytochemicals have higher binding affinity to the target enzyme as compared to the anti-viral drugs.

4. Conclusion

With the exponential increase in the mortality rate and lack of therapeutic interventions for the treatment of SARS-CoV-2 infection in humans, the discovery of novel drug molecules is crucial. SARS-CoV-2 M^{pro} is a well-characterized drug target and its recent structural elucidation through X-ray crystallography has opened an avenue for structure-based drug design. Herein, we have explored a small library of

phytochemicals with previously reported antiviral properties for the identification of small molecular inhibitors against SARS-CoV-2 M^{pro} using a computational approach. We identified three lead molecules—Bonducellpin D, 5,7-dimethoxyflavanone-4'-O-β-d-glucopyranoside and Caesalmin B which exhibit higher binding affinities as compared to the control. These lead molecules also demonstrate broad-spectrum anti-viral activities against SARS-CoV M^{pro} and MERS-CoV M^{pro}. The current findings need further validations through in vitro and in vivo studies for developing into drug candidate molecules.

Declaration of competing interest

The authors report no conflicts of interest in this work.

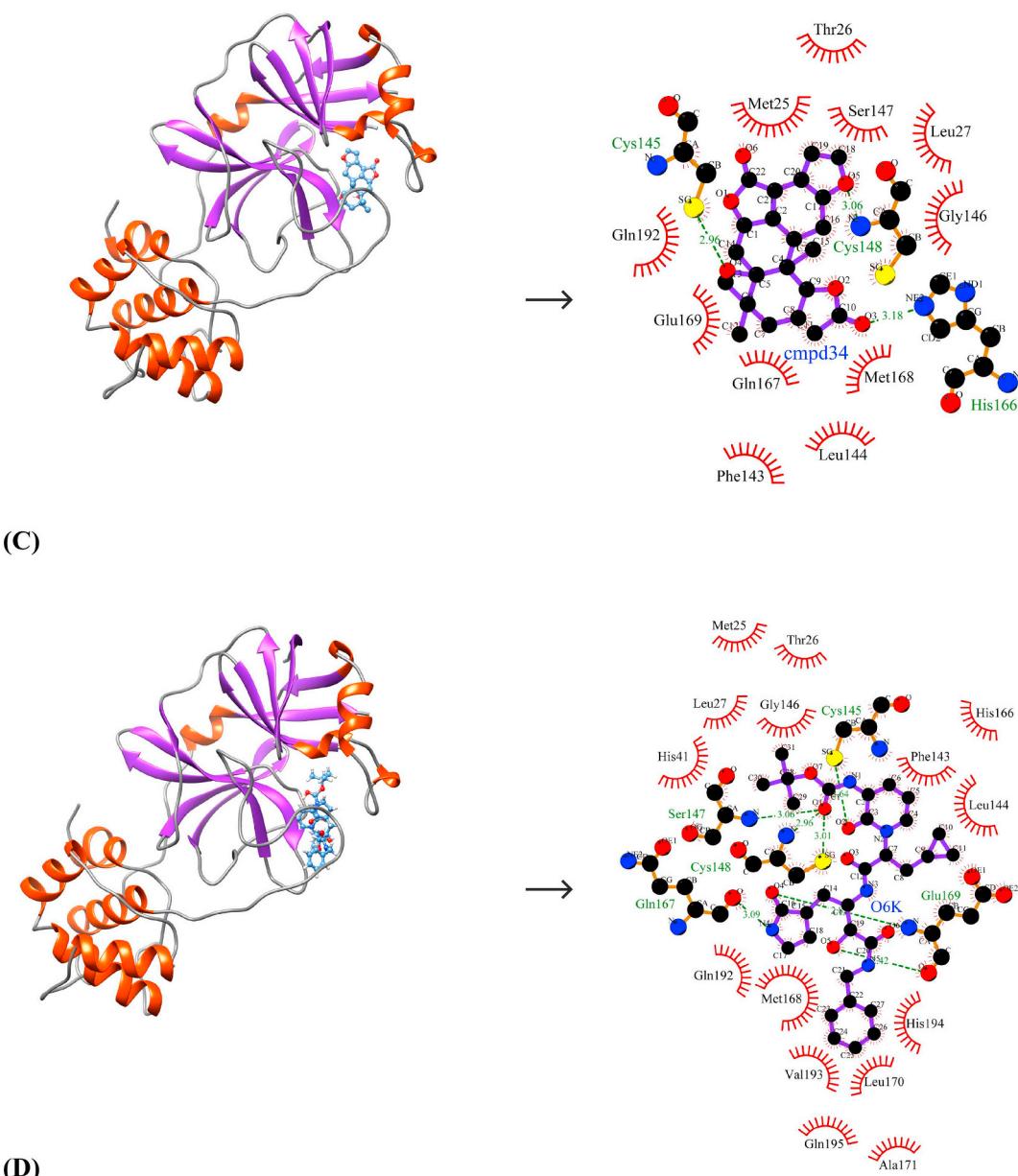


Fig. 5. (continued)

Acknowledgement

The author ABG would like to thank the Department of Basic Sciences and Social Sciences for providing the necessary facilities for the research work. The authors would like to extend their sincere appreciation to the Deanship of Scientific Research at King Saud University for its funding of the research through the research group project #RG-1438-015.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.lfs.2020.117831>.

References

- [1] F.A. Rabi, M.S. Al Zoubi, G.A. Kasasbeh, D.M. Salameh, A.D. Al-Nasser, SARS-CoV-2 and coronavirus disease 2019: what we know so far, *Pathogens* 9 (2020) 231.
- [2] J. Cui, F. Li, Z.-L. Shi, Origin and evolution of pathogenic coronaviruses, *Nat. Rev. Microbiol.* 17 (2019) 181–192.
- [3] P. Zhou, H. Fan, T. Lan, X.-L. Yang, W.-F. Shi, W. Zhang, Y. Zhu, Y.-W. Zhang, Q.-M. Xie, S. Mani, et al., Fatal swine acute diarrhoea syndrome caused by an HKU2-related coronavirus of bat origin, *Nature* 556 (2018) 255–258.
- [4] Y. Fan, K. Zhao, Z.-L. Shi, P. Zhou, Bat coronaviruses in China, *Viruses* 11 (2019) 210, <https://doi.org/10.3390/v11030210>.
- [5] P. Zhou, X.-L. Yang, X.-G. Wang, B. Hu, L. Zhang, W. Zhang, H.-R. Si, Y. Zhu, B. Li, C.-L. Huang, et al., A pneumonia outbreak associated with a new coronavirus of probable bat origin, *Nature* (2020) 1–4.
- [6] F. Wu, S. Zhao, B. Yu, Y.-M. Chen, W. Wang, Z.-G. Song, Y. Hu, Z.-W. Tao, J.-H. Tian, Y.-Y. Pei, et al., A new coronavirus associated with human respiratory disease in China, *Nature* 579 (2020) 265–269.
- [7] A.E. Gorbalenya, S.C. Baker, R.S. Baric, R.J. de Groot, C. Drosten, A.A. Gulyaeva, B.L. Haagmans, C. Lauber, A.M. Leontovich, B.W. Neuman, D. Penzar, S. Perlman, L.L.M. Poon, D.V. Samborskiy, I.A. Sidorov, I. Sola, J. Ziebuhr, C.S.G. of the I.C. on T. of Viruses, The species Severe acute respiratory syndrome-related coronavirus: classifying 2019-nCoV and naming it SARS-CoV-2, *Nat. Microbiol.* 5 (2020) 536–544, <https://doi.org/10.1038/s41564-020-0695-z>.
- [8] L. Zhang, D. Lin, X. Sun, U. Curth, C. Drosten, L. Sauerhering, S. Becker, K. Rox, R. Hilgenfeld, Crystal structure of SARS-CoV-2 main protease provides a basis for design of improved α -ketoamide inhibitors, *Science* (80-.) (2020), <https://doi.org/10.1126/science.abb3405>.
- [9] K. Kuba, Y. Imai, S. Rao, H. Gao, F. Guo, B. Guan, Y. Huan, P. Yang, Y. Zhang, W. Deng, et al., A crucial role of angiotensin converting enzyme 2 (ACE2) in SARS coronavirus-induced lung injury, *Nat. Med.* 11 (2005) 875–879.
- [10] K. Anand, J. Ziebuhr, P. Wadhwani, J.R. Mesters, R. Hilgenfeld, Coronavirus main

- proteinase (3CLpro) structure: basis for design of anti-SARS drugs, *Science* (80-) 300 (2003) 1763–1767.
- [11] R. Hilgenfeld, From SARS to MERS: crystallographic studies on coronaviral proteases enable antiviral drug design, *FEBS J.* 281 (2014) 4085–4096.
- [12] X.-Y. Meng, H.-X. Zhang, M. Mezei, M. Cui, Molecular docking: a powerful approach for structure-based drug discovery, *Curr. Comput. Aided. Drug Des.* 7 (2011) 146–157.
- [13] A. Sethi, K. Joshi, K. Sasikala, M. Alvala, Molecular docking in modern drug discovery: principles and recent applications, *Drug Discov. Dev. Adv.*, IntechOpen, 2019.
- [14] A.-L. Liu, G.-H. Du, Antiviral properties of phytochemicals, *Diet. Phytochem. Microbes*, Springer, 2012, pp. 93–126.
- [15] D.J. Newman, G.M. Cragg, Natural products as sources of new drugs over the last 25 years, *J. Nat. Prod.* 70 (2007) 461–477.
- [16] F. Notka, G. Meier, R. Wagner, Concerted inhibitory activities of Phyllanthus amarus on HIV replication in vitro and ex vivo, *Antivir. Res.* 64 (2004) 93–102.
- [17] M.M. Parida, C. Upadhyay, G. Pandya, A.M. Jana, Inhibitory potential of neem (*Azadirachta indica* Juss) leaves on dengue virus type-2 replication, *J. Ethnopharmacol.* 79 (2002) 273–278.
- [18] J. Serkedjieva, S. Ivancheva, Antitherpes virus activity of extracts from the medicinal plant Geranium sanguineum L., *J. Ethnopharmacol.* 64 (1998) 59–68.
- [19] S. Rehman, U.A. Ashfaq, S. Riaz, T. Javed, S. Riazuddin, Antiviral activity of Acacia nilotica against hepatitis C virus in liver infected cells, *Virol. J.* 8 (2011) 220.
- [20] T.A. Halgren, Merck molecular force field. I. Basis, form, scope, parameterization, and performance of MMFF94, *J. Comput. Chem.* 17 (1996) 490–519, [https://doi.org/10.1002/\(SICI\)1096-987X\(199604\)17:5/6<490::AID-JCC1>3.0.CO;2-P](https://doi.org/10.1002/(SICI)1096-987X(199604)17:5/6<490::AID-JCC1>3.0.CO;2-P).
- [21] A.B. Gurung, A. Bhattacharjee, M.A. Ali, Exploring the physicochemical profile and the binding patterns of selected novel anticancer Himalayan plant derived active compounds with macromolecular targets, *Informatics Med. Unlocked.* 5 (2016) 1–14.
- [22] S. Kim, P.A. Thiessen, E.E. Bolton, J. Chen, G. Fu, A. Gindulyte, L. Han, J. He, S. He, B.A. Shoemaker, J. Wang, B. Yu, J. Zhang, S.H. Bryant, PubChem substance and compound databases, *Nucleic Acids Res.* 44 (2016) D1202–D1213, <https://doi.org/10.1093/nar/gkv951>.
- [23] C.A. Lipinski, Lead- and drug-like compounds: the rule-of-five revolution, *Drug Discov. Today Technol.* 1 (2004) 337–341, <https://doi.org/10.1016/j.ddtec.2004.11.007>.
- [24] T. Sander, J. Freyss, M. von Korff, C. Rufener, DataWarrior: an open-source program for chemistry aware data visualization and analysis, *J. Chem. Inf. Model.* 55 (2015) 460–473, <https://doi.org/10.1021/ci500588j>.
- [25] A.B. Gurung, M.A. Ali, A. Bhattacharjee, M. AbulFarah, F. Al-Hemaid, F.M. Abou-Tarboush, K.M. Al-Anazi, F.S.M. Al-Anazi, J. Lee, Molecular docking of the anticancer bioactive compound proceraside with macromolecules involved in the cell cycle and DNA replication, *Genet. Mol. Res.* 15 (2016), <https://doi.org/10.4238/gmr.15027829>.
- [26] J.D. Thompson, D.G. Higgins, T.J. Gibson, CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice, *Nucleic Acids Res.* 22 (1994) 4673–4680.
- [27] E.F. Pettersen, T.D. Goddard, C.C. Huang, G.S. Couch, D.M. Greenblatt, E.C. Meng, T.E. Ferrin, UCSF chimera—a visualization system for exploratory research and analysis, *J. Comput. Chem.* 25 (2004) 1605–1612, <https://doi.org/10.1002/jcc.20084>.
- [28] G.M. Morris, R. Huey, W. Lindstrom, M.F. Sanner, R.K. Belew, D.S. Goodsell, A.J. Olson, AutoDock4 and AutoDockTools4: automated docking with selective receptor flexibility, *J. Comput. Chem.* 30 (2009) 2785–2791, <https://doi.org/10.1002/jcc.21256>.
- [29] X. Du, Y. Li, Y.-L. Xia, S.-M. Ai, J. Liang, P. Sang, X.-L. Ji, S.-Q. Liu, Insights into protein-ligand interactions: mechanisms, models, and methods, *Int. J. Mol. Sci.* 17 (2016) 144, <https://doi.org/10.3390/ijms17020144>.
- [30] I.A. Adejoro, S.O. Waheed, O.O. Adeboye, Molecular docking studies of *Lonchocarpus cyanescens* triterpenoids as inhibitors for malaria, *J. Phys. Chem. Biophys.* 6 (2016) 398–2161.
- [31] R.A. Laskowski, M.B. Swindells, LigPlot+: multiple ligand-protein interaction diagrams for drug discovery, *J. Chem. Inf. Model.* 51 (2011) 2778–2786, <https://doi.org/10.1021/ci200227u>.
- [32] X. Xue, H. Yu, H. Yang, F. Xue, Z. Wu, W. Shen, J. Li, Z. Zhou, Y. Ding, Q. Zhao, et al., Structures of two coronavirus main proteases: implications for substrate binding and antiviral drug design, *J. Virol.* 82 (2008) 2515–2527.
- [33] H.I. Januar, A.S. Dewi, E. Marraskuranto, T. Wikanta, In silico study of fucoxanthin as a tumor cytotoxic agent, *J. Pharm. Bioallied Sci.* 4 (2012) 56.
- [34] R.-W. Jiang, S.-C. Ma, Z.-D. He, X.-S. Huang, P.P.-H. But, H. Wang, S.-P. Chan, V.E.-C. Ooi, H.-X. Xu, T.C.W. Mak, Molecular structures and antiviral activities of naturally occurring and modified cassane furanoditerpenoids and friedelane triterpenoids from *Caesalpinia minax*, *Bioorg. Med. Chem.* 10 (2002) 2161–2170.
- [35] D.D. Orhan, B. Özçelik, S. Özgen, F. Ergun, Antibacterial, antifungal, and antiviral activities of some flavonoids, *Microbiol. Res.* 165 (2010) 496–504.