



# Repositioning therapeutics for COVID-19: virtual screening of the potent synthetic and natural compounds as SARS-CoV-2 3CLpro inhibitors

Ahmad Sattari<sup>1</sup> · Ali Ramazani<sup>1,2</sup> · Hamideh Aghahosseini<sup>1,2</sup>

Received: 31 October 2020 / Accepted: 1 March 2021 / Published online: 26 March 2021  
© Iranian Chemical Society 2021

## Abstract

The widespread transmission of *SARS-CoV-2* has sparked alarm worldwide. Attaining the best drugs to treat *COVID-19* at the shortest possible time is one of the most critical issues in this urgent situation. Molecular docking investigation of the therapeutic potential of marketed drugs is a fast and cost-effective approach to provide a solution to this problem. The recent research efforts have led to the resolving of the *3CLpro* structure as a key protease in the lifecycle of coronavirus, which could facilitate *in silico* evaluation of drug candidates. Herein, the similarity between the SARS-CoV-2-3CL main protease and the other SARS-CoV receptors was evaluated via multiple sequence alignment and phylogenetic tree. The reported structure of the *3CLpro* was considered as a target to identify potential inhibitors for treating *COVID-19* using molecular docking based virtual screening protocol. Accordingly, a database of 50 synthetic compounds with various pharmacological usage such as antiviral, anti-inflammatory, anti-human immunodeficiency viruses, antimalarial, antibacterial, anticancer, and antioxidant including approved drugs and those undergoing clinical trials, and 40 natural compounds particularly those employed in traditional Iranian medicine was constructed. The output of multiple sequence alignment analysis showed that SARS-CoV-2 main protease shares a similarity of 96% with SARS-CoV. Also, the docking results indicated that the *licofelone acyl glucuronide* as an anti-inflammatory drug and *delta-bilirubin* as an antioxidant and anti-inflammatory agent as well as *kappa-carrageenan conformer*, *beta-D-galactopyranosyl* and *calycosin 7-O-glucoside* as natural compounds with minimal side-effects, according to *in vitro* studies, are good candidates to block the enzymatic activity of *SARS-CoV-2 3CLpro*. Moreover, the *compound 1* with the highest negative binding energy is a chemical compound that due to its favorable interactions with the *3CLpro* can be identified as a representative potential drug candidate for *COVID-19*.

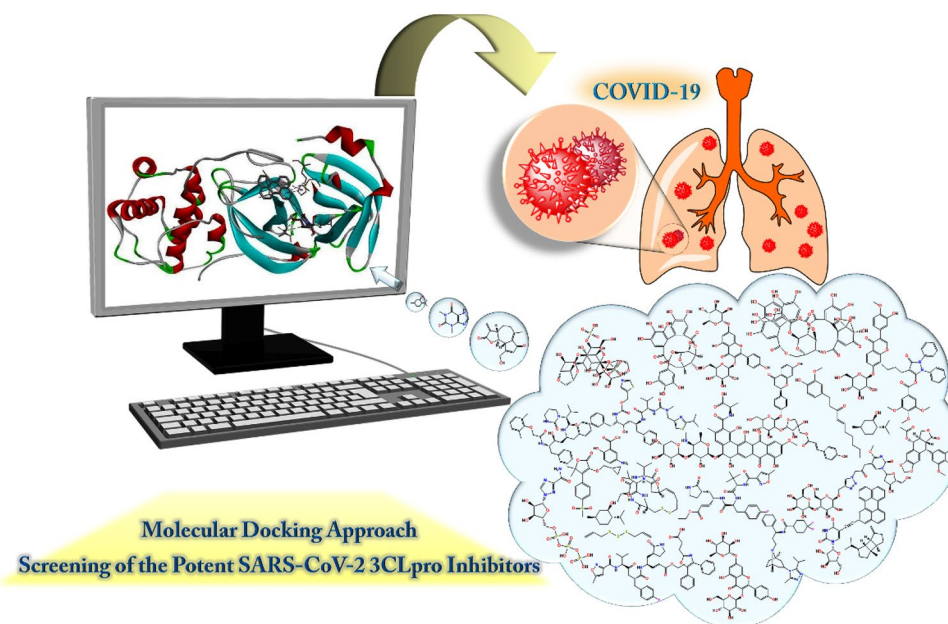
---

✉ Ali Ramazani  
aliramazani@gmail.com; aliramazani@znu.ac.ir

<sup>1</sup> Department of Chemistry, Faculty of Science, University of Zanjan, 45371-38791 Zanjan, Iran

<sup>2</sup> Department of Biotechnology, Research Institute of Modern Biological Techniques (RIMBT), University of Zanjan, 45371-38791 Zanjan, Iran

## Graphic abstract



**Keywords** COVID-19 · SARS-CoV-2 · 3CLpro · Virtual docking · Synthetic and natural ligands

### Abbreviations

ACE2	Angiotensin-converting enzyme 2
MERS	Middle East respiratory syndrome
SARS	Severe acute respiratory syndrome
SARS-CoV	Severe acute respiratory syndrome coronavirus
MERS-CoV	Middle East respiratory syndrome coronavirus
SARS-CoV-2	Severe acute respiratory syndrome coronavirus 2
COVID-19	Coronavirus disease
HIV	Human immunodeficiency viruses
AIDS	Acquired immunodeficiency syndrome
HRV	Human rhinoviruses
HCV	Hepatitis C virus
HBV	Hepatitis B virus
NSAID	Non-steroidal anti-inflammatory drug
Anti-TB	Antituberculosis
MS	Multiple sclerosis
HSV	Herpes simplex viruses
ADV	Adenoviruses
MNV	Murine norovirus
RSV	Respiratory syncytial virus
SDF	Structure-data file
PDB	Protein Data Bank
ADT	Auto-Dock Tools

PDBQT	Protein Data Bank, Partial Charge (Q), & Atom Type (T)
HETATOM	Heteroatom
MSA	Multiple sequence alignment
T-Coffee	Tree-based Consistency Objective Function for alignment Evaluation
ESPrpt3	Easy Sequencing in PostScript
BLASTp	Basic Local Alignment Search Tool
CADD	Computer-assisted drug design
RMSD	Root-mean-square deviation

### Introduction

Coronaviruses are enveloped, positive-sense, single-stranded RNA viruses, which in humans range from the mild respiratory tract infections such as common cold to lethal infections such as *Middle East respiratory syndrome coronavirus (MERS-CoV)*, *severe acute respiratory syndrome coronavirus (SARS-CoV)* and *severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2)*. Human coronaviruses date back to the late 1960s [1, 2]. The viral spike peplomers created a crown-like morphology on the surface of the virus, which is the basis for naming the coronaviruses [3, 4]. Coronaviruses particles have enveloped and pleomorphic structure [5] with the diameter of around 120 nm [6] and a distinct pair of electron dense

shells formed their envelope [7]. The coronaviruses are protected outside the host cells by their lipid bilayer envelope and nucleocapsids inside them, as well as membrane proteins [8]. The coronaviruses subfamily is divided into the four genera called *alpha*-, *beta*-, *gamma*- and *delta*-coronavirus [9]. The *SARS-CoV-2* belongs to the genus *Beta*-coronavirus from group 2B, which represented close to 79% sequence similarity to the *SARS-CoV* according to the next-generation sequencing technology [10, 11]. Only earlier reports indicated that Covid-19 has been originated from bat, but there are serious doubts regarding such claims and its origin has not been identified now. Since humans and bats have very limited close contact with each other, it is thought that an intermediate host such as domestic animal, a wild animal, or a domesticated wild animal could be a cause of transmission of *SARS-CoV-2* to humans [12, 13]. The rapid spread of the *SARS-CoV-2* has sparked alarm worldwide. The outbreak is believed to have begun in Wuhan, China, in late December 2019 [10, 14] although today, the epicenter of the outbreak is Europe. This pathogen was named as *2019 novel coronavirus (2019-nCoV)* by the World Health Organization (WHO) [15] and later renamed as *SARS-CoV-2* by the International Committee on Taxonomy of Viruses and the causing disease named as *coronavirus disease 2019 (COVID-19)* [16]. The virus seems to spread from person-to-person very easily, which makes containment efforts difficult. As of October 26, 2020, a total of more than 43,187,134 people have been infected by the *COVID-19* and the total number of deaths reached 1,155,653 across the world [17]. So far no antiviral agent has been proven for treating human coronavirus infections, and preventive vaccines are still being explored. This is why, the outbreak caused massive disruptions to the nations' health and economy. Therefore, the dire need to find potential therapeutic agents is strongly felt. In this regard, many research teams have focused their researches on finding an effective way for the treatment of *COVID-19* as one of the most critical issues of our time. For the first time, Zhu et al. determined whole-genome sequence of *SARS-CoV-2* which can help to quickly detect the virus in patients [18]. Then several laboratories have been submitted this whole-genome sequences to *global initiative on sharing all influenza data* [19]. Four major structural proteins have been encoded in coronaviruses: Spike (S) protein, envelope(E) protein, membrane (M) protein, and nucleocapsid (N) protein [20]. The study of biological structures of these proteins in *SARS-CoV-2* is still at a preliminary stage and heretofore only the crystal structure of *SARS-CoV-2 3CLpro (3-chymotrypsin-like proteinase, 3CLpro)* was solved and released (Protein Data Bank code: 6LU7) [21]. According to the target types, the potential anticoronavirus therapies are subdivided into human cells- and virus-based therapeutics subdivisions. If

the human cells were considered as a target, the anticoronavirus effect could be induced via blocking of the human cells signaling pathways which are essential for virus replication [20]. Moreover, the blocking of the entry receptor proteins on the surface of human cells could prevent from virus attachment to the target cells. As instances, the *angiotensin-converting enzyme 2 (ACE2)* was identified as a *SARS-CoV* receptor [22–24] and the *dipeptidyl peptidase-4 (DPP4)* was identified as a *MERS-CoV* receptor [25, 26]. If the coronavirus was considered as a target, the antiviral effect could be induced by blocking the receptor-binding domain of virus, hampering viral self-assembly process, preventing the virus RNA synthesis and inhibiting viral replication. The *3CLpro* has a vital role in coronaviruses replication [27]; hence, it could be a promising target to develop *anti-SARS-CoV-2* drugs [28–30]. It is worthy of note that a ~800 kDa polypeptide is produced upon transcription of the genome of beta-coronaviruses. This polypeptide is proteolytically cleaved by *3CL<sup>pro</sup>* and *PL<sup>pro</sup>* to produce various proteins. The *3CLpro* role is production of important non-structural proteins for viral replication via cleavage of the mentioned polyprotein at 11 distinct sites [27, 31]. Some potential inhibitors were identified against *SARS-CoV 3CLPro* and *MERS-CoV 3CLPro* according to the structure–activity analysis [32–34]. Given the vital role of *3CLpro* in the life cycle of the coronaviruses, studying this protein to find therapeutics against the *SARS-CoV-2* could be very important. Considering the rapidly spreading *COVID-19* pandemic and the utmost importance of rapid access to the safe and effective medicines, molecular modeling investigation of the therapeutic potential of marketed drugs could be a fast and cost-effective way to help solve this problem. Herein, the molecular docking studies were performed on a broad range of reported synthetic drugs and natural compounds employing AutoDock Vina program [35], with the aim of rapid investigating their inhibition potential against *SARS-CoV-2 3CLpro* and ultimately repurposing them as a possible treatment for *COVID-19*. In this regard, we used *3CLpro* as a target to screen 90 compounds including synthetic compounds (50 compounds) with various pharmacological usages (such as antiviral, anti-inflammatory, anti-human immunodeficiency viruses (HIV) & acquired immunodeficiency syndrome (AIDS), antimalarial, antibacterial, anticancer, antioxidant, etc.) and natural compounds particularly those employed in traditional Iranian medicine, with its great history of medicine and pharmacy (40 compounds) by virtual screening protocol [36–40]. The prediction of the inhibition potential of these compounds against *SARS-CoV-2 3CLpro* could allow researchers to increase the likelihood of success for compounds selected for clinical trials after validating their antiviral effects in vitro and in vivo.

## Computational methods

### Structures of inhibitors and targets

The information and structure-data file (SDF) files of different synthetic and natural *Covid-19* inhibitors were achieved from PubChem and Zinc15 databases and are recorded in Table 1 and Table 2, respectively. The 2D chemical structure of suggested inhibitors are illustrated in Fig. S1 and Fig. S2 followed by ChemDraw Professional V15.0 drawing and analysis. The library converted subsequently to Protein Data Bank (PDB) files by using Open Babel. The PDB files state the 3D coordination of constituent atoms and chemical bonding. The particular programs within Open Babel enable the software to minimize the input files and select the conformer with the lowest energy by systematic determination of conformations and calculation of their in vacuo free energy [41]. The structural file of target molecule (*3CLpro*, PDB ID: 6LU7 [21]) was fetched out from RCSB PDB ([www.rcsb.org/pdb](http://www.rcsb.org/pdb)) with resolution of 2.16 Å. It is edited by removing the hetero atoms like water and ligand molecules followed by adding polar hydrogens. From here, Auto-Dock Tools 1.5.7 (ADT) was used to do all the pre-processing steps according to the more reports [42].

### Preparation of inhibitors and targets

ADT converts PDB files of the ligands and receptors to the AutoDock Vina program [35]. We use Vina in this study, inputs in Protein Data Bank, Partial Charge ( $Q$ ), and Atom Type ( $T$ ) format (PDBQT) during the process naming the preparation of inhibitor and target structures. In this way, the PDB format extended to PDBQT via addition of partial charge and atom type to ATOM and heteroatom (HETATM) records and recording the information of molecule rigid blocks. For the rigid docking running in this study, the rotatable bonds of ligand explicitly changed to non-rotatable bonds.

### Multiple sequence alignment and phylogenetic tree

Multiple sequence alignment (MSA) of all nucleotide sequences was performed using T-Coffee (Tree-based Consistency Objective Function for alignment Evaluation) [43], and the alignment figure was generated using ESPrpt3 (Easy Sequencing in PostScript) [44]. The phylogenetic tree analyses and similarity/identity values were carried out by using BLAST (Basic Local Alignment Search Tool) [45] and Discovery Studio, respectively.

Physicochemical parameters of proteins were investigated using the ProtParam tool of ExPASy [46].

### Molecular docking study

Molecular docking was performed using Vina program, version 1.1.2, on Windows 8.1 platform (64-bit) with Asus X450C machine (Intel Pentium ULV 1.8 GHz, 4 GB memory). After preparing the PDBQT files, it is required to adjust the size and center point of a 3D box for ligand docking. In the set of ligands docked to the receptor, the grid center was selected as the middle point between extreme value of  $x$ ,  $y$ , and  $z$  coordinates. The grid dimensions were chosen so as to include all atoms of the ligand set and then augmented by 10 Å in  $\pm x$ ,  $\pm y$ , and  $\pm z$  directions [47]. The Num\_modes was 50 for each ligand also. The options employed for other parameters were default. In especial, the grid spacing was 1.0 Å.

### Analysis of docking outcome

Vina results, including multiple modes in PDBQT format, describe the docked ligand position, orientation and conformation. However, many visualization programs are not capable of reading these files with nonstandard format, and AutoDock Tools, discovery studio and LigPlot are freely available options used to visualize and analyze the protein–ligand interactions in this project [48].

## Results and discussion

### Multiple sequence alignment and phylogenetic tree

Sequence alignment is a computational program to find the relationship or similarity between biological sequences. Unlike Pairwise sequence alignment that attempts to search the optimal alignment in parts or whole of sequences, multiple sequence alignment identifies common regions within different ones. In other words, the investigation of distantly related genes or proteins by this tool is a good way to find out the evolutionary relationships between genes and identify shared patterns among functionally or structurally related genes. This tool is used not only for construction of phylogenetic trees but also creation of sequence profiles to identify closely related sequences in the database [49].

The physicochemical parameters of SARS-CoV-2 3CLpro and its closest homologs are summarized in Table 3. For example, the results revealed that SARS-CoV-2 main protease categorized as a stable and hydrophilic protein contains 306 amino acids long. The Fig. 1a illustrates the phylogenetic tree of all protein sequences. It is inferred from the image that SARS-CoV-2 M<sup>pro</sup> protein sequence

**Table 1** The physical information and pharmaceutical activities of several drugs and a number of chemical compounds from different resources as *3CLpro* inhibitors

No	Name/compound code	No. <sup>a</sup> HBD	No. <sup>b</sup> HBA	HAC	Pharmaceutical function <sup>c</sup>
1	Compound 1 CID: 134,816,013	4	6	39	–
2	Licofelone acyl glucuronide CID: 71,749,786	4	8	39	Anti-inflammatory, relevant inhibitor of CYP2C8
3	Ritonavir impurity H [EP] CID: 66,832,842	1	8	41	–
4	delta-Bilirubin CID: 129,320,333	4	6	43	Antioxidant
5	Raltegravir CID: 54,671,008	3	9	32	Anti-HIV
6	Nigericin ZINC000085552063	3	11	51	Antimicrobial and antibacterial
7	Pradimicin A ZINC000169346835	11	19	60	Anti-HIV and antifungal
8	Rupintrivir ZINC000003919807	3	9	43	Anti-human rhinoviruses (HRV)
9	Lopinavir ZINC000003951740	4	5	46	Anti-HIV and AIDS
10	Adenylyl-(3'-5') ribavirin CID: 196,553	7	17	39	Anti-influenza A virus
11	Novobiocin ZINC000076945632	5	11	44	Antibacterial
12	Megazone CID: 27,624	0	7	37	Anti-inflammatory
13	CHEMBL21082 ZINC000028231984	4	9	44	Anti-HRV
14	Simeprevir ZINC000085540268	2	10	52	Anti-hepatitis C virus (HCV)
15	CGP 75,136 ZINC000004394015	5	10	50	Anti-HIV
16	Amenavir CID: 11,397,521	1	7	34	Antiviral
17	Conivaptan CID: 151,171	2	3	38	Inhibitor of antidiuretic hormone
18	CHEMBL140521 CID: 6,479,024	3	8	43	Anti-HRV
19	Indinavir CID: 5,362,440	4	7	45	Anti-HIV and AIDS
20	Compound 20 CID: 134,814,833	3	11	48	Anti-HIV and antituberculosis (TB) activity
21	Oxaprozin CID: 4614	1	4	22	Anti-inflammatory
22	Telmisartan CID: 65,999	1	4	39	Anti-hypertensives
23	Boceprevir CID: 10,324,367	4	5	37	Anti-HCV
24	Elvitegravir CID: 5,277,135	2	7	31	Anti-HIV
25	Telaprevir CID: 3,010,818	4	8	49	Anti-HCV
26	Saquinavir ZINC000003914596	5	7	49	Anti-HIV and AIDS
27	Atazanavir ZINC000003941496	5	9	51	Anti-HIV

**Table 1** (continued)

No	Name/compound code	No. <sup>a</sup> HBD	No. <sup>b</sup> HBA	HAC	Pharmaceutical function <sup>c</sup>
28	Maraviroc CID: 3,002,977	1	6	37	Anti-HIV & AIDS
29	Compound 29 CID: 134,815,433	1	10	46	Anti-HIV and Anti-TB activity
30	CHEMBL289920 CID: 6,477,669	3	9	43	Anti-HRV
31	CHEMBL345023 CID: 6,478,681	3	8	42	Anti-HRV
32	Geneticin CID: 134,688,573	10	14	34	Antibacterial
33	Ritonavir CID: 392,622	4	9	50	Anti-HIV
34	Vancomycin Mimic ZINC000150553684	7	11	75	Antimicrobial
35	Ribavirin 5'- Triphosphate CID: 122,108	7	16	29	Antiviral
36	Oseltamivir CID: 65,028	2	5	22	Anti-influenza viruses
37	Phenylbutazone CID: 4781	0	2	23	Anti-inflammatory, antipyretic, and analgesic
38	Licofelone CID: 133,021	1	2	27	Anti-inflammatory, anti-analgesic
39	Parecoxib CID: 119,828	1	5	26	Anti-inflammatory, antipyretic activities
40	Sofosbuvir CID: 45,375,808	3	11	36	Anti-HCV
41	Lopinavir free Amine CID: 17,755,107	4	4	34	–
42	Isoxicam CID: 54,677,972	2	7	23	Anti-inflammatory and antirheumatic
43	(1R,2R)-2-azido-1,2-dihydro oseltamivir CID: 76,968,516	2	7	25	–
44	Etoricoxib CID: 123,619	0	4	24	Non-steroidal anti-inflammatory drug (NSAID), antipyretic, analgesic
45	Fingolimod CID: 107,970	3	3	22	Immunomodulatory drug, used to treat Multiple sclerosis (MS)
46	Chloroquine CID: 2719	1	3	22	Anti-inflammatory, antimalarial
47	Firocoxib CID: 208,910	0	5	23	Anti-inflammatory and antineoplastic
48	Romidepsin CID: 123,135,747	4	8	36	Antineoplastic activity
49	Mesalazine CID: 4075	3	4	11	Anti-inflammatory
50	Favipiravir CID: 492,405	2	4	11	Antiviral

<sup>a</sup> The number of hydrogen bond donors

<sup>b</sup> The number of hydrogen bond acceptors

<sup>c</sup> All of the pharmaceutical function information are recorded from PubChem except the ligands containing references which are mentioned in supplementary information

**Table 2** The physical information and pharmaceutical activities of SARS-CoV-2 inhibitors from natural sources

No	Name/compound code	No. HBD	No. HBA	HAC	Pharmaceutical function <sup>a</sup>	Source <sup>a</sup>
1	kappa-Carrageenan conformer ZINC96061851	10	25	51	Antiviral activity against myxoviridae, and coronaviridae	Red Algae
2	beta-D- Galactopyranosyl CID: 23,656,242	9	19	50	Anti-inflammatory	<i>Rosa canina</i>
3	Calycosin 7-O-glucoside CID: 5,318,267	5	10	32	Antiviral	Astragalus
4	Gallic acid 3-cholesteryl ester CID: 101,021,751	3	5	39	Antimicrobial activity	<i>Ficus carica</i>
5	Spicoside A CID: 44,258,517	0	7	30	Neuroprotective potency	<i>Cichorium intybus</i>
6	Corilagin CID: 73,568	11	18	45	NSAID, anti-hypertensive agent	Euphorbiaceae
7	Astragalin CID: 5,282,102	7	11	32	Anti-inflammatory	<i>Rosa canina</i>
8	Podophyllotoxin acetate CID: 164,791	0	9	33	Anticancer, antiviral	<i>Euphorbia</i>
9	Rhamnetin 3-O-beta-glucopyranoside CID: 14,704,554	7	12	34	–	<i>Syzygium aromaticum</i>
10	Astragalus polyphenol CID: 5,321,884	7	9	29	Anti-inflammatory antioxidant	<i>Astragalus</i>
11	Geraniin ZINC000169289506	14	27	68	antiviral	<i>Nepheleium lappaceum</i>
12	Gallic Acid Tribenzyl Ether CID: 11,133,969	1	5	33	Antimicrobial	<i>Ficus carica</i>
13	Linalool-3-Rutinoside CID: 21,630,850	6	10	32	Antibacterial, antifungal and antiviral	<i>Myrtus communis</i>
14	Rhamnopyranoside CID: 21,606,527	9	15	42	Inhibitor of topoisomerase I and II	<i>Cichorium intybus</i>
15	Catechin gallate CID: 5,276,454	7	10	32	Anti-inflammatory	<i>Rosa canina</i>
16	Chicoric acid CID: 5,281,764	6	12	34	Anti-HIV	Echinacea
17	Tiliroside CID: 5,320,686	7	13	43	Anti-inflammatory	<i>Rosa canina</i>
18	Glycyrrhizic acid CID: 14,982	8	16	58	Anti-allergic, antiviral and anti-inflammatory	Licorice
19	Hyperin CID: 133,568,467	8	12	33	Anti-inflammatory	<i>Rosa canina</i>
20	Licorice glycoside E CID: 101,938,909	7	14	50	–	Licorice
21	Rosmarinic acid CID: 5,281,792	5	8	26	Anti-inflammatory, antiviral	Pepermint
22	Syringin CID: 5,316,860	5	9	26	Immunopotentiating, immunostimulatory, radioprotective	Ginseng
23	Carboxymethyl inulin CID: 446,984	3	2	32	–	<i>Cichorium intybus</i>
24	Rutin CID: 5,280,805	10	16	43	Anti-inflammatory	<i>Rosa canina</i>
25	Quercetin CID: 5,280,343	5	7	22	Anti-inflammatory	<i>Rosa canina</i>
26	Rhamnetin CID: 5,281,691	4	7	23	Anti-inflammatory, antioxidant	<i>Syzygium aromaticum</i>
27	beta-Sitosterol CID: 222,284	1	1	30	Anticarcinogenic, anti-atherogenic	<i>Cichorium intybus</i>

**Table 2** (continued)

No	Name/compound code	No. HBD	No. HBA	HAC	Pharmaceutical function <sup>a</sup>	Source <sup>a</sup>
28	Sageone CID: 6,481,824	2	3	22	Anti-HIV-1	Sage
29	Apigenin CID: 5,280,443	3	5	20	Anti-herpes simplex viruses (HSV), anti-adenoviruses (ADV), anti-hepatitis B virus (HBV)	Basill
30	Oleanolic acid CID: 10,494	2	3	33	Antiviral, anti-HIV, anti-influenza	Rosemary
31	Deoxylactucin CID: 442,196	1	4	19	Antifungal,	<i>Cichorium intybus</i>
32	Sambucus nigra Degraded cyanogenic glycosides (2'-Epimer) CID: 131,751,786	2	9	25	Anti-influenza virus, reducer of upper respiratory symptoms	<i>Black elderberry (Sambucus nigra)</i>
33	Luteine (Xanthophyll) CID: 5,368,396	2	2	42	anti-inflammatory	<i>Rosa canina</i>
34	Shogaol CID: 5,281,794	1	3	20	anti-inflammatory, antimicrobial	Ginger
35	Theophylline CID: 2153	1	3	13	Bronchodilator & Vasodilator Agents	<i>Amellia sinensis</i> and <i>Coffea arabica</i>
36	Carvacrol CID: 10,364	1	1	11	Anti-murine norovirus (MNV), anti-respiratory syncytial virus (RSV) anti-HSV-1,	Oregano
37	Menthol ZINC000001482164	1	1	11	Anti-inflammatory, antiviral	Peperminte
38	Germacrene D CID: 5,373,727	0	0	15	Antiviral, antifungal, antibacterial,	<i>Myrtus communis</i>
39	Trans-Pinocarveol CID 88,302	1	1	11	antimicrobial activity	<i>Erodium cicutarium</i>
40	Diallyl trisulfide ZINC000001531082	0	2	8	Anticancer and antiviral	Garlic

<sup>a</sup> All of the pharmaceutical function and sources information are recorded from PubCheme except the ligands containing references which are mentioned in supplementary information

clusters with SARS coronaviruses. The multiple sequence alignment of this protein sequence with its closest homologs revealed that SARS-CoV-2 3CLpro shares 96.1, 48.7, 40.8 and 44.4% sequence identity with SARS-CoV, MERS-CoV, Human-CoV and TGEV-CoV homologs, respectively (Fig. 1b). These findings are very close to the initial reports that SARS-CoV-2 is very similar to SARS-CoV compared to MERS-CoV [13, 18]. It is noticed that the sequence similarity of SARS-CoV main protease with SARS-CoV-2-3CLpro is remarkably higher than all other human coronaviruses studied here. The main protease of SARS-Cov-2 and SARS-CoV categorizing three distinct domains have nine  $\alpha$ -helices and 13  $\beta$ -strands. The target consists of I, II and III domain identified by residues 1–101, residues 102–184 and residues 201–306, respectively. A long loop between domains II and III and the active site between domains I and II are the other characteristics of 3CLpro target [50]. However, despite the high similarity between this two proteins, 12-point mutations may affect 3CLpro structure and function. The mutations

are include: Val35Thr, Ser46Ala, Asn65Ser, Val86Leu, Lys88Arg, Ala94Ser, Phe134His, Asn180Lys, Val202Leu, Ser267Ala, Ala285Tr and Leu286Ile. Regardless of the mutation in amino acid at 286 position, Lucien, it appears that the mutation effects on SARS-CoV-2 3CLpro structure are conservation of polarity and hydrophobicity. Finally, it is worthy of note that mutations could not affect the critical proteolytic activity of the SARS-CoV-2 M<sup>Pro</sup>, because they are not present in catalytic and substrate-binding regions [51].

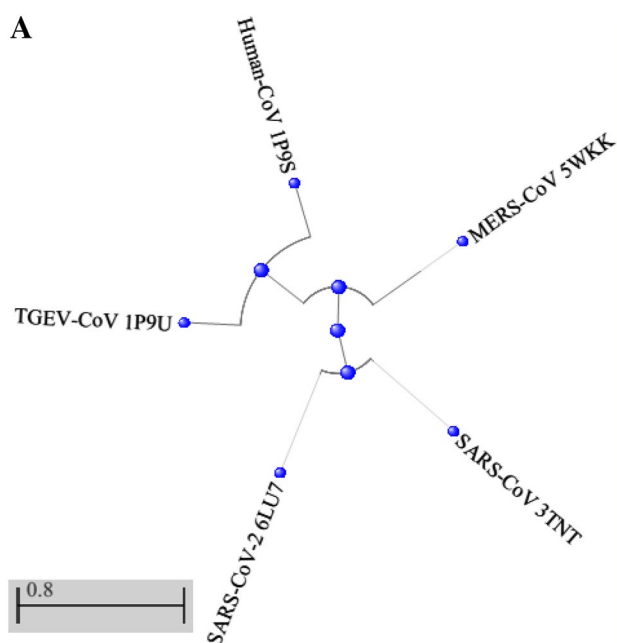
### Molecular docking study

Protein–ligand docking is a process in which protein-related binding mode, and affinity of ligand is predicted. Docking programs, as a key tool in computer-assisted drug design (CADD) and structural molecular biology, generally used for estimating the modeled system free energy and sampling its positional space by using a scoring function and an



**Table 3** 3CLpro physicochemical parameters of SARS-CoV-2, SARS-CoV, MERS-CoV, Human-CoV and TGEV-CoV

Parameters	SARS-CoV-2 3CLpro	
Mol. Weight	33,796.64 Dalton	
No. of amino acids	306	
Theoretical pI	5.95	
Instability index (II)	27.65 (stable)	
No. of negatively charged residues (Asp/Glu)	26	
No. of positively charged residues (Arg/Lys)	22	
Aliphatic index	82.12	
Grand average of hydropathicity (GRAVY)	− 0.019	
Atomic composition	Carbon-1499; Hydrogen-2318; Nitrogen-402; Oxygen-445; Sulfur-22	
Amino acid composition	Ala-17 (5.6%); Arg-11 (3.6%); Asn-21 (6.9%); Asp-17 (5.6%); Cys-12 (3.9%); Gln-14 (4.6%); Glu-9 (2.9%); Gly-26 (8.5%); His-7 (2.3%); Ile-11 (3.6%); Leu-29 (9.5%); Lys-11 (3.6%); Met-10 (3.3%); Phe-17 (5.6%); Pro-13 (4.2%); Ser-16 (5.2%); Thr-24 (7.8%); Trp-3 (1.0%); Tyr-11 (3.6%); Val-27 (8.8%); Pyl-0 (0.0%); Sec-0 (0.0%)	
Parameters	SARS-CoV 3CLpro	MERS-CoV 3CLpro
Mol. Weight	33,845.72	34,284.33
No. of amino acids	306	313
Theoretical pI	6.22	6.24
Instability index (II)	29.67 (stable)	28.77 (stable)
No. of negatively charged residues(Asp/Glu)	26	21
No. of positively charged residues(Arg/Lys)	23	16
Aliphatic index	81.83	78.12
Grand average of hydropathicity (GRAVY)	− 0.049	0.071
Atomic composition	Carbon-1499; Hydrogen-2325; Nitrogen-405; Oxygen-445; Sulfur-22	Carbon-1520; Hydrogen-2335; Nitrogen-405; Oxygen-446; Sulfur-27
Amino acid composition	Ala-17 (5.6%); Arg-12 (3.9%); Asn-19 (6.2%); Asp-17 (5.6%); Cys-12 (3.9%); Gln-14 (4.6%); Glu-9 (2.9%); Gly-26 (8.5%); His-8 (2.6%); Ile-12 (3.9%); Leu-30 (9.8%); Lys-11 (3.6%); Met-10 (3.3%); Phe-16 (5.2%); Pro-13 (4.2%); Ser-16 (5.2%); Thr-26 (8.5%); Trp-3 (1.0%); Tyr-11 (3.6%); Val-24 (7.8%); Pyl-0 (0.0%); Sec-0 (0.0%)	Ala-23 (7.3%); Arg-5 (1.6%); Asn-16 (5.1%); Asp-12 (3.8%); Cys-10 (3.2%); Gln-17 (5.4%); Glu-9 (2.9%); Gly-26 (8.3%); His-14 (4.5%); Ile-9 (2.9%); Leu-24 (7.7%); Lys-11 (3.5%); Met-17 (5.4%); Phe-14 (4.5%); Pro-12 (3.8%); Ser-20 (6.4%); Thr-26 (8.3%); Trp-5 (1.6%); Tyr-11 (3.5%); Val-32 (10.2%); Pyl-0 (0.0%); Sec-0 (0.0%)
Parameters	Human-CoV 3CLpro	TGEV-CoV 3CLpro
Mol. weight	32,812.38	33,094.49
No. of amino acids	300	302
Theoretical pI	6.38	6.09
Instability index (II)	29.47 (stable)	31.57 (stable)
No. of negatively charged residues (Asp/Glu)	23	26
No. of positively charged residues (Arg/Lys)	20	23
Aliphatic index	84.43	83.48
Grand average of hydropathicity (GRAVY)	0.058	− 0.089
Atomic composition	Carbon-1464; Hydrogen-2241; Nitrogen-397; Oxygen-427; Sulfur-18	Carbon-1463; Hydrogen-2275; Nitrogen-395; Oxygen-449; Sulfur-16
Amino acid composition	Ala-20 (6.7%); Arg-9 (3.0%); Asn-23 (7.7%); Asp-8 (2.7%); Cys-8 (2.7%); Gln-11 (3.7%); Glu-15 (5.0%); Gly-35 (11.7%); His-10 (3.3%); Ile-16 (5.3%); Leu-23 (7.7%); Lys-11 (3.7%); Met-10 (3.3%); Phe-17 (5.7%); Pro-5 (1.7%); Ser-20 (6.7%); Thr-15 (5.0%); Trp-5 (1.7%); Tyr-11 (3.7%); Val-28 (9.3%); Pyl-0 (0.0%); Sec-0 (0.0%)	Ala-13 (4.3%); Arg-11 (3.6%); Asn-25 (8.3%); Asp-7 (2.3%); Cys-6 (2.0%); Gln-8 (2.6%); Glu-19 (6.3%); Gly-31 (10.3%); His-6 (2.0%); Ile-13 (4.3%); Leu-26 (8.6%); Lys-12 (4.0%); Met-10 (3.3%); Phe-13 (4.3%); Pro-8 (2.6%); Ser-31 (10.3%); Thr-16 (5.3%); Trp-3 (1.0%); Tyr-14 (4.6%); Val-30 (9.9%); Pyl-0 (0.0%); Sec-0 (0.0%)



**Fig. 1** **a** Phylogenetic tree inferred from closest homologs of SARS-CoV-2 3CLpro and **b** the multiple alignment sequence between main protease of SARS-CoV-2 (PDB ID: 6LU7) and SARS-CoV (PDB ID: 3TNT), MERS-CoV (PDB ID: 5WKK), Human-CoV (PDB ID: 1P9S) and TGEV-CoV (PDB ID: 1P9U)

exploration method, respectively. AutoDock is one of the well-established and open source protein–ligand docking software available. Moreover, it is an advantageous program in teaching, research and designing bioactive compounds. To improve the performance and accuracy of docking process, Vina is published under a free software license by the same group as AutoDock in 2010 [35] which was used in this project.

### Internal validation of molecular docking

In order to substantiate the validation of docking method, the co-crystal ligand (Fig. 2a) extracted from crystal structure of CoV-2019 main protease (6LU7) and re-docked. The root-mean-square deviation (RMSD) value of redocking process calculated by ADT1.5.7 was 0.46 Å. The comparison of RMSD values between docked and X-ray pose conforms accuracy for pose prediction [52] and the ability of docking simulation for this prediction [53]. In other words, RMSD was used to evaluate the difference of obtained docking orientation with the corresponding co-crystallized pose of the same ligand molecule [54]. In terms of docking accuracy, the RMSD values between the docked and X-ray pose below 3 Å could be considered as adoptable [55], below 2 Å is fairly good [53] and with a threshold of 1.0–3.0 Å has been generally considered to be a successful [56]. Moreover, the

other different RMSD classifications for docking simulations could be include:

RMSD  $\leq$  2.0 Å (good), 2.0 Å < RMSD < 3.0 Å (acceptable), and RMSD  $\geq$  3.0 Å (bad) [54].

RMSD < 1.0 Å (excellent), 1.0 Å < RMSD < 3.0 Å (good), 3.0 Å < RMSD < 5.0 Å (fair) and RMSD  $\geq$  5.0 Å (poor or unacceptable) [57].

The binding poses of docked and crystallographic ligands are compared as illustrated in Fig. 2b. It can be deduced from Fig. 2b that the docking process is valid because not only the RMSD value is less than 1.0 Å but also the cognate ligand docked in the active site of target has very little difference with crystallographic ligand in 3-methyl-2-pyrrolidinone ring and benzene motif. The parameters of protein internal validation are summarized in Table 4.

### Molecular docking results

Blocking of the SARS-CoV-2 main protease, 3CLpro, to prevent the synthesis of virus RNA and its replication is one of the current suggested therapies for *Covide-19* diseases [58]. Based on the relevant target fetched from PDB, 6LU7, we screened potential bio-active synthetic and natural chemical compounds from PubChem and Zinc database using Vina. The ranking of AutoDock results is based on the lowest binding free energies and RMSD values of determined binding site. On the other hand, in Vina the RMSD value is related to the top ranked pose which presents that the highest negative binding energy is 0. Therefore, Vina ranks docking results based on the top ranked binding free energy not the relevant RMSD value [59]. The other binding affinity indicator, ligand efficiency (LE), is the size-dependent binding energy and calculated by Eq. (1) [60].

$$LE = -\Delta G_b / HAC \quad (1)$$

where  $\Delta G_b$  stands for calculated binding energy and HAC is heavy atom counts of a ligand, a number of non-hydrogen atoms that express ligand size. Based on this parameter, the larger ligand provides more interactions with target and shows greater binding energy. However, the ligand efficiency of large ligands is reduced because these compounds interact with other regions beside ‘hot spots’ and may not necessarily be the most efficient binders [61]. Hence, Vina results to clarify the ligands with the highest binding affinity to 3CLpro are charted for this research based on the lowest binding energy not subsequent RMSD values. In the same amount of binding energy, the ligands with higher ligand efficiency are preferred. Results are summarized in Table 5 and Table 6. As shown in Table 5, chemical compound 1,

## B

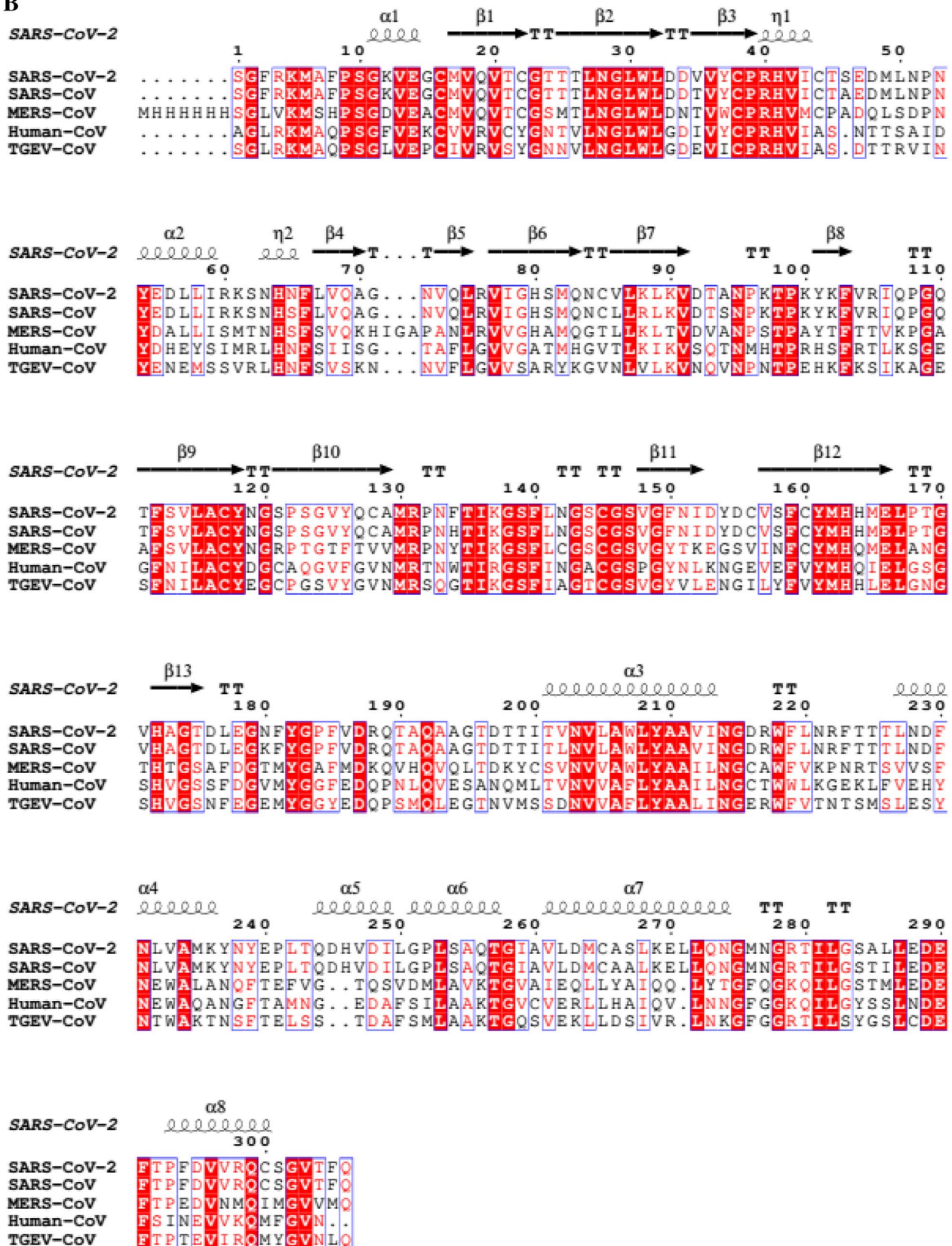
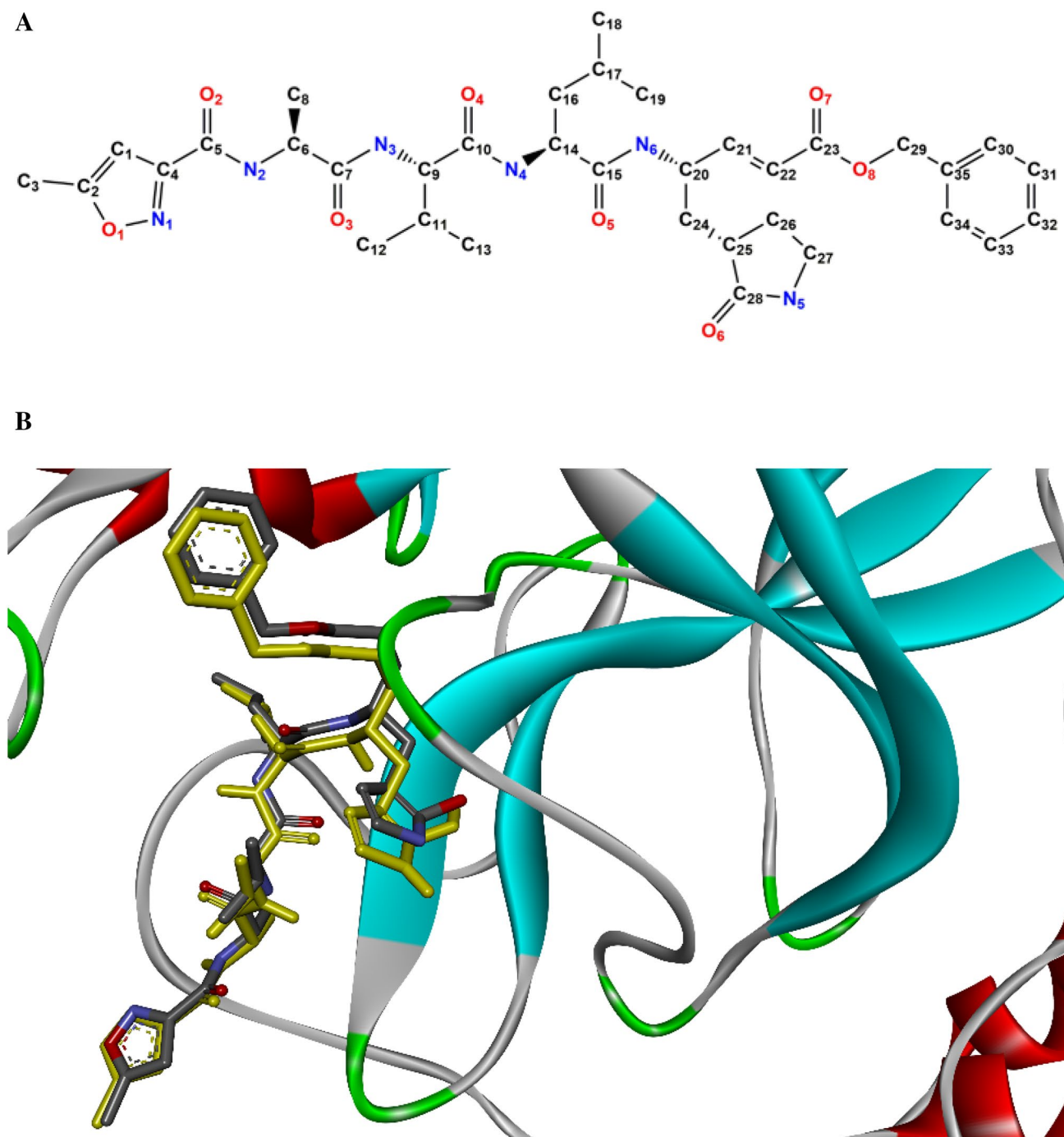


Fig. 1 (continued)



**Fig. 2** **a** 2D structure of cognate ligand (N3 inhibitor) and **b** The Superposition of docked (●) and crystallographic (●) conformation of cognate ligand in the binding pose of relevant receptor

**Table 4** The internal validation parameters of cognate ligand

Ligand	Binding energy (kcal/mol)	Ligand efficiency	RMSD (Å)
Cognate ligand	-13.1	0.27	0.46

licofelone acyl glucuronide (anti-inflammatory drug), des(isopropylthiazolyl) hydantoin-oxazolidinone ritonavir, delta-bilirubin (antioxidant), raltegravir (anti-HIV agent), nigericin (antimicrobial and antibacterial agent) and Pradimicin A (anti-HIV and antifungal agents) had the lowest binding energy. A number of other marketed drugs such as rupintrivir (anti-HRV), novobiocin (antibacterial

**Table 5** The binding energy and ligand efficiency (LE) indices of synthetic ligands docked to 3CLpro receptor. The binding energy of cognate ligand and ligand efficiency are -13.1 kcal/mol and 0.27, respectively

No	$\Delta G$	LE	No	$\Delta G$	LE	No	$\Delta G$	LE	No	$\Delta G$	LE
1	-10.9	0.23	14	-8.7	0.17	27	-8.3	0.16	40	-7.4	0.20
2	-10.3	0.26	15	-8.7	0.17	28	-8.2	0.22	41	-7.3	0.21
3	-10.0	0.24	16	-8.6	0.25	29	-8.2	0.18	42	-7.3	0.32
4	-9.9	0.23	17	-8.6	0.23	30	-8.2	0.19	43	-7.2	0.29
5	-9.8	0.31	18	-8.5	0.20	31	-8.1	0.19	44	-7.2	0.30
6	-9.8	0.19	19	-8.5	0.19	32	-8.1	0.24	45	-7.1	0.32
7	-9.8	0.16	20	-8.4	0.17	33	-8.0	0.16	46	-6.9	0.31
8	-9.6	0.22	21	-8.4	0.38	34	-7.9	0.10	47	-6.7	0.29
9	-9.5	0.21	22	-8.4	0.21	35	-7.9	0.27	48	-6.7	0.19
10	-9.4	0.21	23	-8.4	0.23	36	-7.7	0.35	49	-5.6	0.51
11	-9.3	0.21	24	-8.3	0.27	37	-7.5	0.33	50	-5.3	0.48
12	-9.0	0.24	25	-8.3	0.17	38	-7.4	0.27			
13	-8.9	0.20	26	-8.3	0.17	39	-7.4	0.28			

**Table 6** The binding energy and ligand efficiency (LE) indices of natural ligands docked to 3CLpro receptor

No	$\Delta G$	LE	No	$\Delta G$	LE	No	$\Delta G$	LE	No	$\Delta G$	LE
1	-11.5	0.23	11	-10.0	0.15	21	-9.3	0.36	31	-7.3	0.38
2	-11.2	0.22	12	-9.7	0.29	22	-8.9	0.34	32	-6.5	0.15
3	-10.5	0.32	13	-9.7	0.30	23	-8.8	0.27	33	-6.5	0.15
4	-10.2	0.26	14	-9.7	0.23	24	-8.8	0.20	34	-6.0	0.30
5	-10.2	0.34	15	-9.7	0.30	25	-8.7	0.39	35	-5.6	0.43
6	-10.2	0.23	16	-9.6	0.28	26	-8.4	0.36	36	-5.2	0.4
7	-10.2	0.34	17	-9.5	0.22	27	-8.4	0.28	37	-5.1	0.4
8	-10.1	0.31	18	-9.4	0.16	28	-8.0	0.36	38	-5.1	0.34
9	-10.1	0.30	19	-9.4	0.28	29	-7.9	0.39	39	-4.6	0.42
10	-10.0	0.34	20	-9.3	0.19	30	-7.9	0.24	40	-3.5	0.44

agents), megazone (anti-inflammatory), simeprevir (anti-HCV) and amenamevir (anti-varicella-zoster virus and anti-herpes simplex virus types I & II) showed relatively low binding energy which are worth studying more. However, some others like sofosbuvir (anti-HCV), isoxicam (antirheumatic), fingolimod (immunomodulatory drug used to treat relapsing multiple sclerosis), romidepsin (antineoplastic activity), and several anti-inflammatory drugs (etoricoxib, firocoxib, and mesalazine) showed the highest binding energy. For natural ligands extracted from various sources, Table 6, kappa-Carrageenan conformer, beta-D-galactopyranosyl, calycosin 7-O-glucoside, gallic acid 3-cholesteryl ester, spicoside A, corilagin, astragal- in, podophyllotoxin acetate, rhamnetin and 3-O-beta-glu- copyranoside showed the lowest binding energy, respec- tively. Like synthetic ligands, natural ligands with the highest estimated ligand efficiency do not have lower bind- ing energy. According to the mentioned results, the small- molecule compounds containing the lowest binding energy could probably have the inhibitory potential of 3CLpro target and used to treat the SARS-CoV-2. Based on the different theoretical and clinical studies, several research

groups claimed that the telmisartan (anti-hypertensives agent) [58], conivaptan (treatment of hyponatremia) [58], chloroquine (antimalarial) [62], favipiravir (antiviral) [63] and several anti-HIV agents such as lopinavir [64], indi- navir [63], saquinavir [65], ritonavir [66] and atazanavir [67] could be the best 3CLpro inhibitors. Moreover, we docked these marketed drugs to compare them with other research group studies. The data from Table 5 showed that most of these drugs may have relatively acceptable bind- ing affinity to 3CLpro target, except the chloroquine and favipiravir. However, the applied receptor structure and scoring function are the same, the predicted binding con- stants are non-similar for different research groups. These differences are related to the not only different ligand and receptor preparation parameters but also to the differ- ent search procedure. For example, in preparation step, the different assigned charge, relaxation and flexibility of receptor besides the different applied united atoms, added charge type and number of bond torsions for ligand could not provide the same results. In docking step, the exhaustiveness and randomness of the search procedure in addition to the size and centering of the grid box could

increase these differences, as well [68]. If we want to look on the bright side, the various binding energy estimated by different groups provide valuable information for further computational and experimental studies.

### Analysis of molecular docking results

The *3CLpro* or *Nsp5*, the *COVID-19* main protease, which has important role in virus RNA synthesis and replication is one of the most important targets for the introduction of efficient small-molecule inhibitors. The interactions of *3CLpro* target with reference ligand, N3, and suggested bioactive inhibitors are discussed in the coming sections. For comparison, the 2D images of crystallographic and redocked ligands interaction with active site of receptor are illustrated in Fig. S3. The further hydrogen bond between N<sub>5</sub> of methyl-2-Pyrrolidinone motif in cognate ligand and Glu166 is a cause of distinct difference between crystallographic and cognate one which is identified by dashed red cycle in Fig. 2. The other indistinct differences include hydrophobic interactions between C<sub>8</sub> and C<sub>17</sub> with Pro168 and Met165, respectively. This means the physiological conditions, especially various solvents, may influence on the ligand–protein interactions in crystal structure.

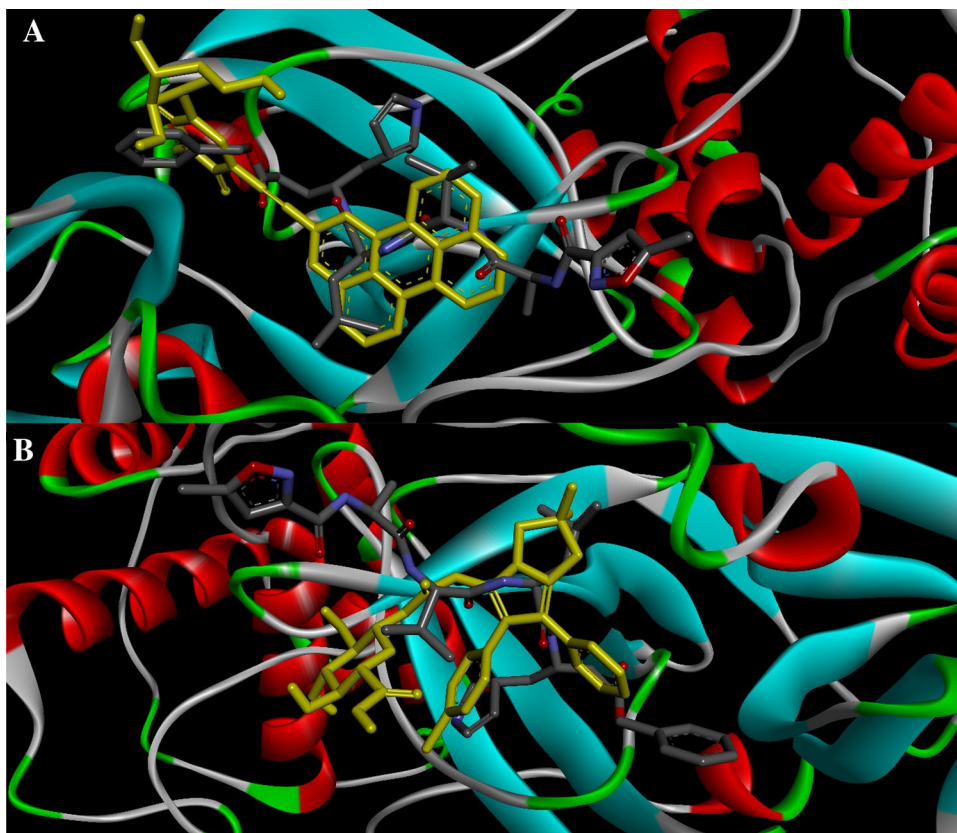
### Synthetic compound analysis

The similarity of binding mode for potentially more effective inhibitors containing lower binding energy was further investigated. The chemical compound 1 with the lowest binding energy (− 10.9 kcal/mol) showed a relatively similar binding mode, but licofelone acyl glucuronide with binding energy of − 10.5 kcal/mol showed less similar binding mode in comparison with reference ligand (Table 4, Fig. 3). The superposition images of ritonavir impurity H [EP] and delta-bilirubin illustrated their lower binding similarity modes, as well (Fig. S4).

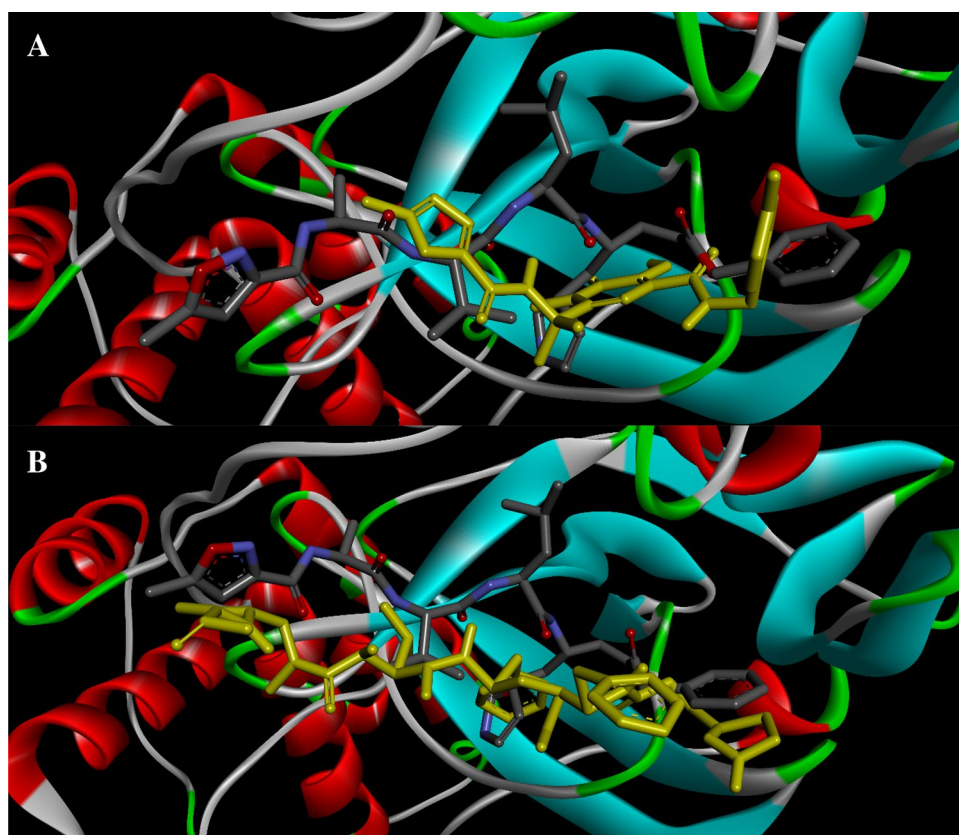
It is surprising that raltegravir (anti-HIV agent) with the highest binding energy (− 9.8 kcal/mol) toward the above-mentioned compounds has more binding mode similarity and could be one of the best candidate drugs for *SARS-CoV-2* (Table 5, Fig. 4a).

In other research studies, the marketed drugs such as lopinavir, indinavir, and ritonavir have been reported as potential inhibitors to block *3CLpro* of *SARS-CoV-2*. The results of this study containing molecular docking and binding mode similarity based on the X-ray crystallographic structure of *Mpro* are compatible with the other predictions [66] (Table 5, Fig. S5). However, ritonavir with an estimated binding energy of − 8.0 kcal/mol could be the best candidate drug due to high similarity of binding mode.

**Fig. 3** The overlap images of crystallographic binding mode of ligand N3 (●) and predicted binding mode of potential inhibitors (●). **a** Compound 1, **b** licofelone acyl glucuronide



**Fig. 4** The overlap images of crystallographic binding mode of ligand N3 (●) and predicted binding mode of potential inhibitors (●). A) Raltegravir and B) Ritonavir



The detailed investigation of ligand and receptor interactions uncover the affinity of suggested inhibitors and facilitate the chance of introducing potential drug candidates for Mpro blocking. As shown in Fig. 5a and Fig. S6A, the compound 1 fitness with active pocket of receptor is well. A number of  $\pi$ - $\pi$  and  $\pi$ -alkyl hydrophobic interactions between ligand and amino acids such as Gln189, Gln166, His41, Cys145, His164, Met165, Met49, Arg188 and Asp167 conform the compound in the pocket of receptor. The predicted hydrogen bonds of Asn142 with oxygen atoms and Thr26 with hydrophilic hydrogen atom of the compound guarantee the conformer stabilization, also. The presence of 4 hydrogen bond donor and 6 hydrogen bond acceptor atoms in the ligand structure and hydrophilic amino acids provides these hydrophilic environment (Table 1 and Fig. 5a).

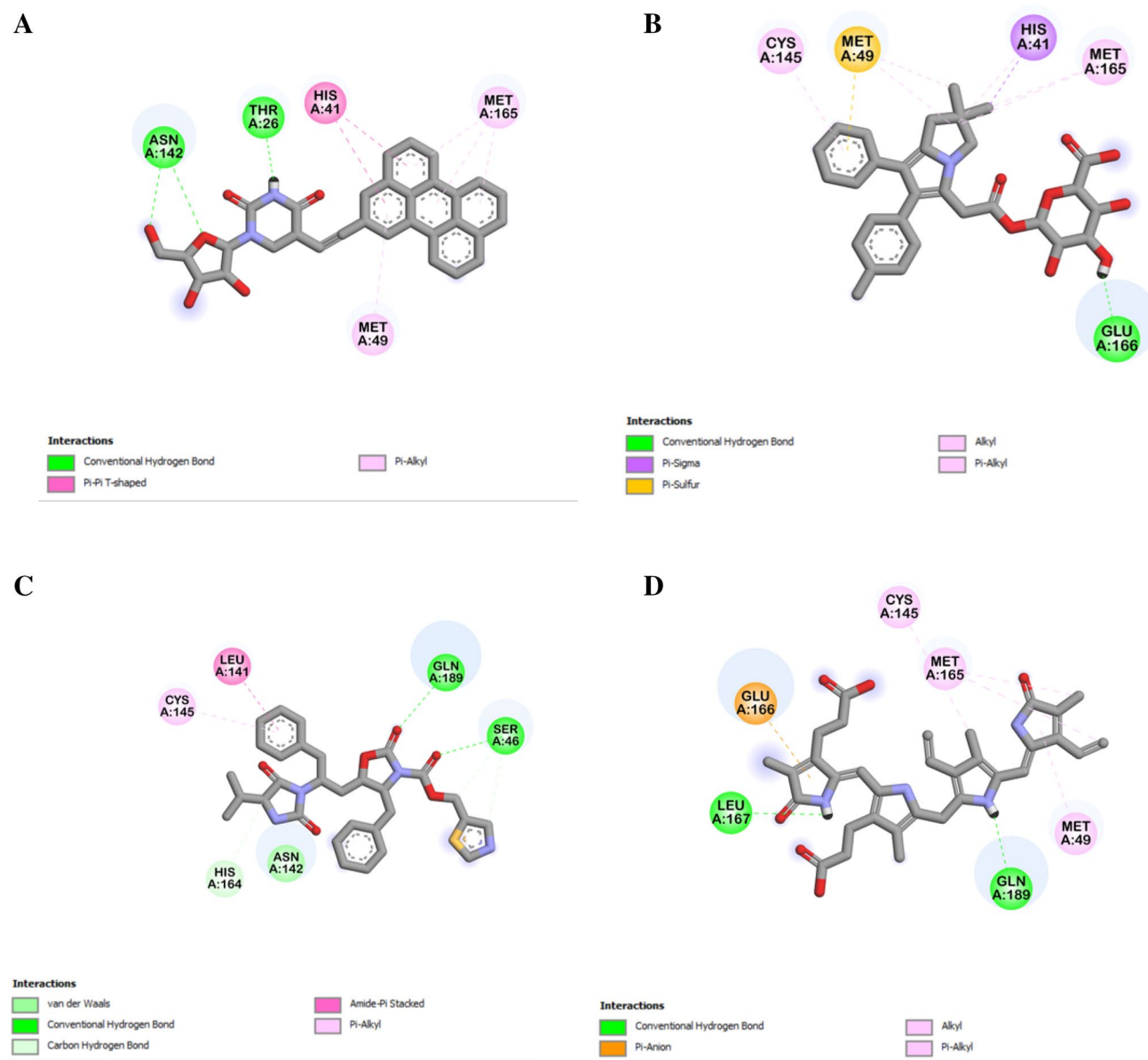
Anti-inflammatory drug licofelone acyl glucuronide which is the relevant inhibitor of CYP2C8 [69] was predicted to bind to *3CLpro* with low binding energy (Scores = -10.3 kcal/mol). The generated docking model shows that the drug conjunction with the active site of the enzyme is created by hydrogen bond between hydroxyl group of drug and Glu166 (Fig. 5b, and Fig. S6B). Moreover, lots of interactions between drug and hydrophobic amino acids, like His41 ( $\pi$ -sigma), Met49 ( $\pi$ -sulfur, and Cys145 and Met165 (Alkyl and  $\pi$ -alkyl) imply that it may

be a potent *3CLpro* inhibitor. Figure 6b) shows the 3D image of provided hydrophobic environment.

The other compound, ritonavir impurity H [EP], with docking scoring of (-10.0 kcal/mol) was well fitted into the active pocket of *3CLpro*, also. The hydrogen bonds between Gln189 and Ser46 with the carbonyl group of the compound and the hydrophobic bonds between ligand atoms and Leu141 (Amide- $\pi$  stacked), Cys145 ( $\pi$ -alkyl) and His164 (carbon hydrogen bond) stabilize the ligand conformation and introduce it as a good inhibitor for target (Fig. 5c and Fig. S6C).

Moreover, the results of delta-bilirubin docking in the active site of *3CLpro* were analyzed and the output is illustrated in Fig. S8D and Fig. S6D. The images show two hydrogen bond between ligand N-H groups and Gln189 and Leu167, a  $\pi$ -anion bond between 5-member ring of ligand and Glu166 and several alkyl and  $\pi$ -alkyl bonds related to for example, Cys145, Met165 and Met49. This hydrophobic environment besides two hydrogen bond could ensure the stability of ligand and receptor complex. Our findings revealed that all of the analyzed compounds possess docking sites that strongly overlap with the protein pockets, and could be potential therapeutic agents.

Moreover, the marketed drugs like lopinavir and indinavir provide more hydrophilic and hydrophobic interactions



**Fig. 5** Docking result analysis for synthetic inhibitors: The 2D image of **a** Compound 1, **b** licoferone acyl glucuronide, **c** ritonavir impurity H [EP] and **d** delta-bilirubin

with different amino acids, and these figures are not shown in this study and conform their stabilization in the pocket. According to Table 1, ritonavir with 4 hydrogen bond donor and 9 hydrogen bond acceptor atoms provides no more hydrogen bonds with amino acids and has less hydrophobic interactions, and consequently it could not be stabilized in the pocket of target. Because licoferone acyl glucuronide, delta-bilirubin, lopinavir and indinavir have been used in clinical practices with limited toxicity, we recommend them to treat *COVID-19*.

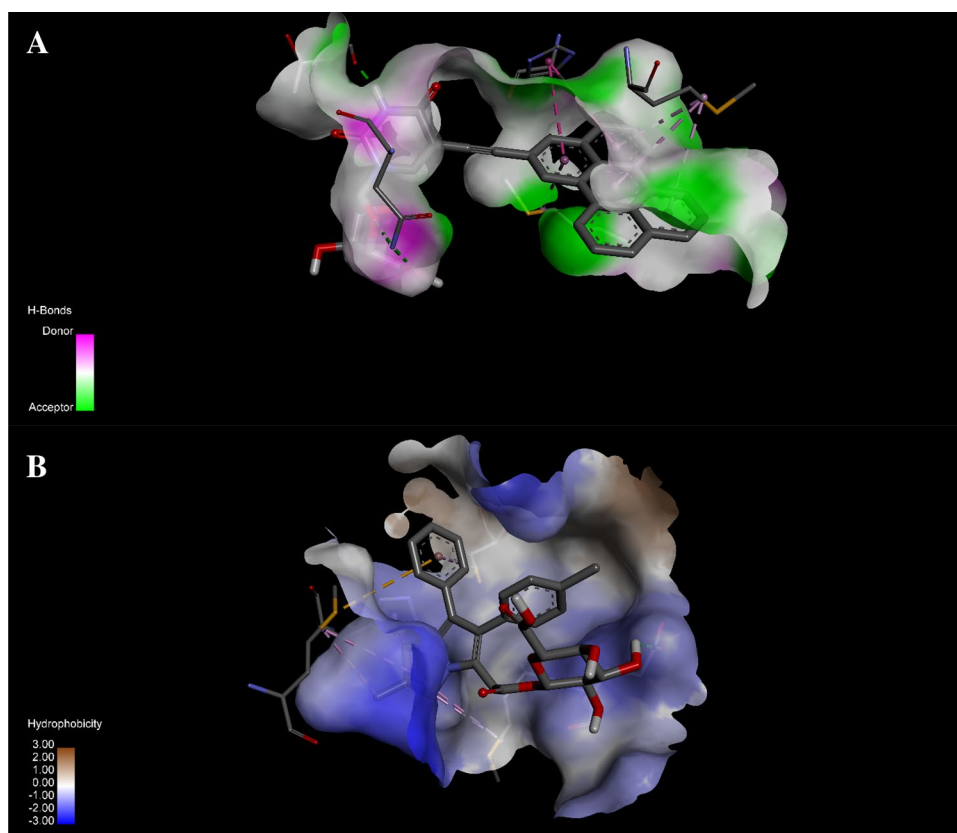
### Analysis of natural ligands

According to the docking results, lots of natural compounds from different sources were predicted to be *3CLpro* inhibitors with high binding affinity (Table 5) through virtual ligand screening. The binding similarity mode and docking result analysis of a number of these compounds containing the highest negative binding energy were studied in detail.

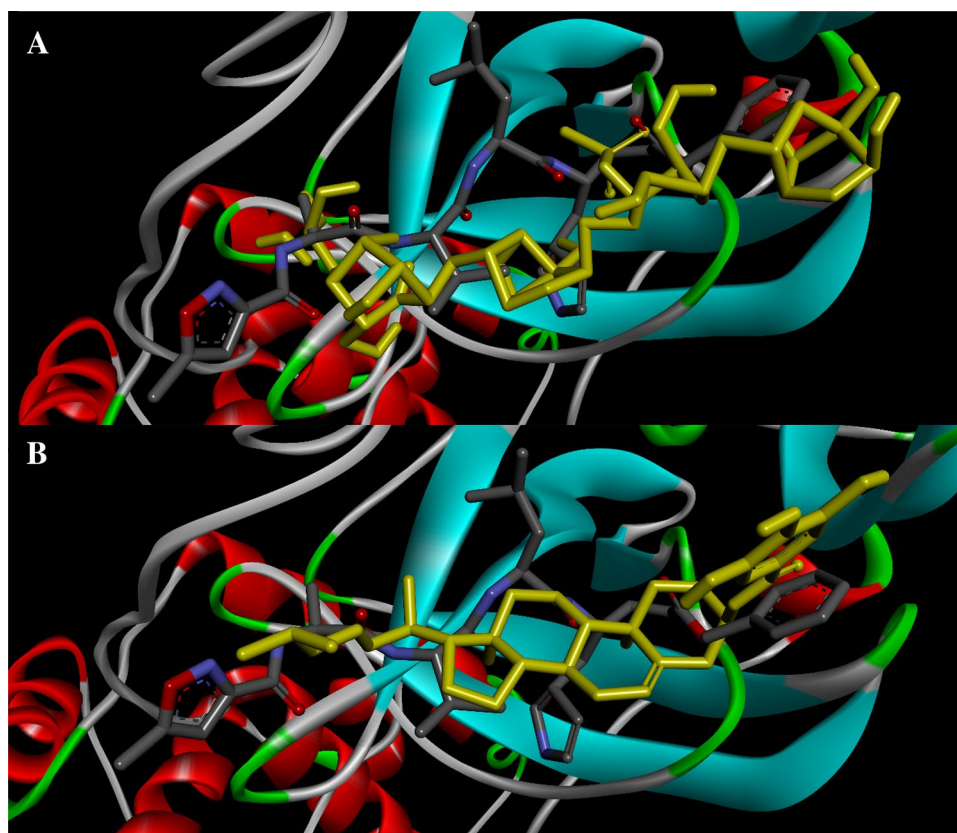
The antiviral activity of kappa-carrageenan extracted from *Red Algae* against myxoviridae, paramyxoviridae, adenoviridae and coronaviridae increases the chance of this



**Fig. 6** The 3D image of **a** the hydrophilic environment created by Compound1, and **b** the hydrophobic environment created by licofelone acyl glucuronide



**Fig. 7** The overlap images of crystallographic binding mode of ligand N3 (●) and predicted binding mode of potential inhibitors (●). **a** kappa-Carrageenan and **b** Gallic acid 3-cholesteryl ester

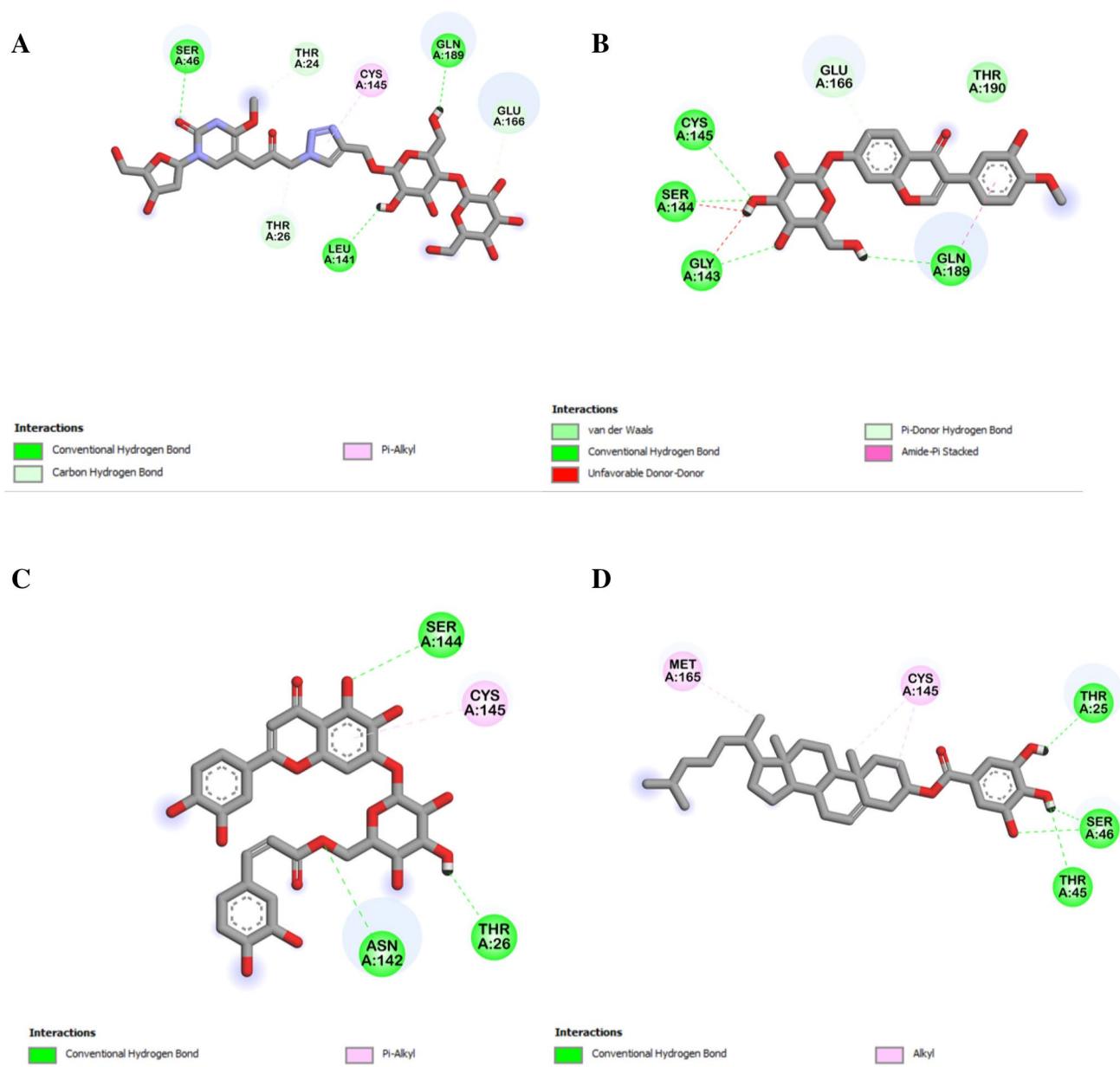


ligand to inhibit the SARS-Cov-2 main protease [70]. One of the conformers of this compound, ZINC96061851, with the lowest binding energy ( $-11.5$  kcal/mol) showed similar binding mode when overlapped with reference ligand (Fig. 7a).

For beta-D-galactopyranosyl [71] ( $-11.2$  kcal/mol) with anti-inflammatory effect, which is extracted from *Rosa canina L.*, the binding similarity mode is relatively good (Fig. S7A). *Rosa canina L.*, which is called *Nasrarane vahshi*, is *Rosaceae* family plant and grows Kordestan Province, Iran [72].

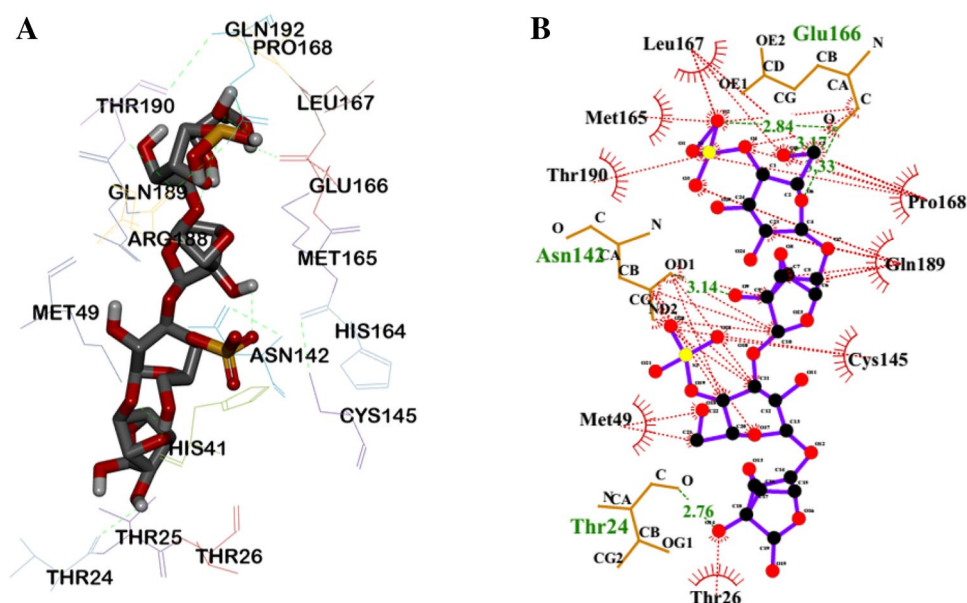
The extracted compound from the Astragalus plant, calycosin 7-O-glucoside [73], which is proved to have antiviral activity might be a candidate for inhibiting target showed relatively similar binding mode, as well (Fig. S7B).

*Cichorium intybus L.* is the scientific name of *Asteraceae* family plant, locally called Sechertghi, and find in the north of Iran, Turkmen Sahra [72]. Spicoside A [74], plant extract, which has the Neuroprotective potency docked into the relevant target and gained ( $-10.2$  kcal/mol) binding energy. Unfortunately, the similarity of binding mode for this ligand



**Fig. 8** Docking result analysis for natural inhibitors: 2D images of **a** beta-D-galactopyranosyl, **b** Calycosin 7-O-glucoside, **c** Spicoside A, and **(D)** Gallic acid 3-cholesteryl ester

**Fig. 9** kappa-Carrageenan conformer-receptor interactions: a 3D and b 2D images



was lower. The superposition figure of binding mode similarity for this ligand is not shown in this study.

*Ficus carica L.* is another Iranian medicinal plant that grows in Golestan, Fars and Khuzestan Province and locally called *Anjeir* [72]. The plant extract, gallic acid 3-cholesteryl ester [75], with proven antimicrobial activity showed well binding affinity and higher similar binding mode when docked to *3CLpro* receptor (Fig. 7b).

The 2D images of docking result analysis for beta-D-galactopyranosyl are illustrated in Fig. 8b and Fig. S8A. It can be inferred from the images that the hydroxyl and carboxyl groups of ligand provide hydrogen bonds with Ser46, Leu141, Gln189 and Glu166. The presence of 9 hydrogen bond donor and 19 hydrogen bond acceptor atoms in the ligand structure conforms the existence of more hydrophilic bonds in the pocket. The interactions of carbon atoms on the ligand structure with amino acids such as Thr24, Thr26, and Glu166 created carbon hydrogen bonds, also. These interactions and other hydrophobic interactions such as  $\pi$ -alkyl one between 5-member ring of ligand and Cys145 cause beta-D-galactopyranosyl to be a good inhibitor for target blocking.

The different types of bonds between ligand and receptor based on Fig. 8b and Fig. S8B for Calycosin 7-O-glucoside include:

- Hydrogen bonds from the interaction of Cys145, Ser144, Gly143 and Gln189 with different hydroxyl groups of ligand.
- Unfavorable donor-donor bond for Ser144 and Gly143 and hydroxyl groups of ligand.
- Amid- $\pi$  stacked of ligand aromatic ring and Gln189.
- Hydrophobic bonds of ligand atoms with Asn142, Met165, Glu166, Thr190 and Aln191.

The further analysis of docking results for spicoside A illustrated in Fig. 8b and Fig. S8C that showed more hydrogen bonds between hydroxyl and oxygen groups of ligand and different amino acids including Ser144, Asn142, Thr26, Gln189, Gly143, Lue141 and Cys145. In addition, hydrophobic interactions with Phe140, Thr26, Thr25, Glu166, Asn142 and Cys145 may further direct the favorite conformation of this inhibitor.

The data from Fig. 8b and Fig. S8D showed three hydrogen bonds for Ser46 and one hydrogen bond for Thr24, Thr25 and Thr45. The alkyl bonds between carbon atoms of ligand and Cys145 and Met165 are the other characteristics of analysis of docking results.

It's worth mentioning that, as shown in Fig. 9, Thr24, Glu166 and Asn142 formed five hydrogen bonds with the oxygen, hydroxyl and sulfate groups of the kappa-Carrageenan conformer. The hydrophobic interactions between the compound atoms and Leu167, Met165, Thr190, Pro168, Gln189, Cys145, Met49 and Thr26 may further stabilize its conformation (Fig. 8b).

The results of analysis indicated that all of the above-mentioned compounds could be connected to the active site of target via desirable and strong hydrophilic and hydrophobic bonds. These strong interactions are related to the affinity of compound atoms to various amino acids presence in the conserved region which are the key factors in enzymatic catalysis. The compounds could be suitable and potent substitutes for synthetic drugs to treat new coronavirus infections due to their natural origin and fewer side effects.

## Conclusions

The emergence of *COVID-19* as a potential global health threat caused massive disruptions to the nations' health and economy. The employment of effective and time-efficient protein–ligand docking process to discover potent anti-*COVID-19* compounds at the shortest possible time is critical. The aim of this study was the construction of 50 synthetic compounds with various pharmacological usage including approved drugs and those undergoing clinical trials, and 40 natural compounds database, molecular docking of selected compounds, and evaluation of their binding interaction against the *SARS-CoV-2 3CLpro* [76]. Accordingly, the compound 1, licofelone acyl glucuronide (anti-inflammatory drug), ritonavir impurity H [EP], *delta*-bilirubin (antioxidant), raltegravir (anti-HIV agent), nigericin (antimicrobial and antibacterial agent) and pradimicin A (anti-HIV and antifungal agents) had the lowest binding energy. For natural ligands, kappa-Carrageenan conformer, *beta*-D-galactopyranosyl, calycosin 7-O-glucoside, gallic acid 3-cholesteryl ester, spicoside A, corilagin, astragaloside, podophyllotoxin acetate, rhamnetin and 3-O-*beta*-glucopyranoside showed the lowest binding energy, respectively. Moreover, the results showed that among investigated marketed drugs, telmisartan, conivaptan, lopinavir, indinavir, saquinavir, ritonavir and atazanavir may have relatively low binding energy. The similarity of binding mode and ligand–receptor interactions were investigated for potential inhibitors, optionally. Compound 1, raltegravir, kappa-carrageenan conformer and gallic acid 3-cholesteryl ester showed higher similarity binding mode, as well. The analysis of ligand–receptor interactions revealed that most studied compounds have the ability to bind to the target pocket. Overall, the compound 1, CID:134,816,013, was identified as the best inhibitor of *3CLpro* due to the lowest binding energy, the highest similarity mode and more ligand–receptor interactions and introduced to further *in vitro* and *in vivo* studies. Moreover, the small-molecules like licofelone acyl glucuronide and *delta*-bilirubin in addition to some of natural compounds with the highest negative binding energy could probably have the inhibitory potential of *3CLpro* target and they have the potential to become an anti-*COVID-19* clinical drug.

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1007/s13738-021-02235-7>.

**Funding** This work supported by the “Iran National Science Foundation: INSF” and the “University of Zanjan. Also, the present work has been done in line with Ahmad Sattari PhD thesis.

## Declarations

**Conflicts of interest** Authors declare that they have no conflict of interest.

**Availability of data and material** The authors confirm that the data supporting the findings of this study are available within the article and/or its supplementary materials.

## References

- J.S. Kahn, K. McIntosh, *Pediatr. Infect. Dis. J.* **24**, S223 (2005)
- C. Geller, M. Varbanov, R.E. Duval, *Viruses* **4**, 3044 (2012)
- J. Almeida, D. Berry, C. Cunningham, D. Hamre, M. Hofstad, L. Mallucci, K. McIntosh, D. Tyrrell, *Nature* **220**, 2 (1968)
- L.S. Sturman, K.V. Holmes, *The molecular biology of coronaviruses*, in *Advances in virus research*. (Elsevier, New York, 1983), pp. 35–112
- C.S. Goldsmith, K.M. Tatti, T.G. Ksiazek, P.E. Rollin, J.A. Comer, W.W. Lee, P.A. Rota, B. Bankamp, W.J. Bellini, S.R. Zaki, *Emerg. Infect. Dis.* **10**, 320 (2004)
- A.R. Fehr, S. Perlman, *Coronaviruses: An Overview of Their Replication and Pathogenesis* (Springer, Berlin, 2015), pp. 1–23
- B.W. Neuman, B.D. Adair, C. Yoshioka, J.D. Quispe, G. Orca, P. Kuhn, R.A. Milligan, M. Yeager, M.J. Buchmeier, *J. Virol.* **80**, 7918 (2006)
- B.W. Neuman, G. Kiss, A.H. Kunding, D. Bhella, M.F. Baksh, S. Connelly, B. Droese, J.P. Klaus, S. Makino, S.G. Sawicki, *J. Struct. Biol.* **174**, 11 (2011)
- A.C. Wong, X. Li, S.K. Lau, P.C. Woo, *Viruses* **11**, 174 (2019)
- D.S. Hui, E.E. Azhar, T.A. Madani, F. Ntoumi, R. Kock, O. Dar, G. Ippolito, T.D. Mchugh, Z.A. Memish, C. Drosten, *Int. J. Infect. Dis.* **91**, 264 (2020)
- R. Lu, X. Zhao, J. Li, P. Niu, B. Yang, H. Wu, W. Wang, H. Song, B. Huang, N. Zhu, *Lancet* **395**, 565 (2020)
- J. Cohen, *Science*, **10** (2020)
- P. Zhou, X.-L. Yang, X.-G. Wang, B. Hu, L. Zhang, W. Zhang, H.-R. Si, Y. Zhu, B. Li, C.-L. Huang, *Nature* **579**, 270 (2020)
- W. H. Organization, " WHO Statement Regarding Cluster of Pneumonia Cases in Wuhan, China. Beijing: WHO", can be found under [www.who.int](http://www.who.int). (2020)
- W. H. Organization, " Laboratory testing of human suspected cases of novel coronavirus ( nCoV) infection: interim guidance" can be found under [www.who.int](http://www.who.int)., (2020)
- W. Ji, W. Wang, X. Zhao, J. Zai, X. Li, *J. Med. Virol.* **92**, 433 (2020)
- C. COVID, " Global Cases by Johns Hopkins CSSE. Вилучено з" can be found under <https://gisanddata.maps.arcgis.com/apps/opsdashboard/index.html#/bda7594740fd40299423467b48e9ecf6>
- N. Zhu, D. Zhang, W. Wang, X. Li, B. Yang, J. Song, X. Zhao, B. Huang, W. Shi, R. Lu, N. Engl, *J. Med.* **382**, 727 (2020)
- Y. Shu, J. McCauley, *Euro Surveill.* **22**, 30494 (2017)
- B.J. Bosch, R. van der Zee, C.A. de Haan, P.J. Rottier, *J. Virol.* **77**, 8801 (2003)
- X. Liu, B. Zhang, Z. Jin, H. Yang, Z. Rao, The crystal structure of COVID-19 main protease in complex with an inhibitor N3. RCSB Protein Data Bank (2020). <https://doi.org/10.2210/PDB6L U7/PDB>
- D.P. Han, A. Penn-Nicholson, M.W. Cho, *Virology* **350**, 15 (2006)
- X.-Y. Ge, J.-L. Li, X.-L. Yang, A.A. Chmura, G. Zhu, J.H. Epstein, J.K. Mazet, B. Hu, W. Zhang, C. Peng, *Nature* **503**, 535 (2013)
- W. Li, M.J. Moore, N. Vasilieva, J. Sui, S.K. Wong, M.A. Berne, M. Somasundaran, J.L. Sullivan, K. Luzuriaga, T.C. Greenough, *Nature* **426**, 450 (2003)
- A.S. Agrawal, T. Garron, X. Tao, B.-H. Peng, M. Wakamiya, T.-S. Chan, R.B. Couch, C.-T.K. Tseng, *J. Virol.* **89**, 3659 (2015)

26. J. Zhao, K. Li, C. Wohlford-Lenane, S.S. Agnihothram, C. Fett, J. Zhao, M.J. Gale, R.S. Baric, L. Enjuanes, T. Gallagher, Proc. Natl. Acad. Sci. **111**, 4970 (2014)
27. K. Anand, J. Ziebuhr, P. Wadhvani, J.R. Mesters, R. Hilgenfeld, Science **300**, 1763 (2003)
28. I.-L. Lu, N. Mahindroo, P.-H. Liang, Y.-H. Peng, C.-J. Kuo, K.-C. Tsai, H.-P. Hsieh, Y.-S. Chao, S.-Y. Wu, J. Med. Chem. **49**, 5154 (2006)
29. B. Xia, X. Kang, Protein Cell. **2**, 282 (2011)
30. D. Needle, G.T. Lountos, D.S. Waugh, Acta Crystallogr. Sect. D: Biol. Crystallogr. **71**, 1102 (2015)
31. M.T. Ul Qamar, S.M. Alqahtani, M.A. Alamri, L.-L. Chen, J. Pharm. Anal. **10**, 313 (2020)
32. T. Pillaiyar, M. Manickam, V. Namasivayam, Y. Hayashi, S.-H. Jung, J. Med. Chem. **59**, 6595 (2016)
33. A.K. Ghosh, K. Xi, K. Ratia, B.D. Santarsiero, W. Fu, B.H. Harcourt, P.A. Rota, S.C. Baker, M.E. Johnson, A.D. Mesecar, J. Med. Chem. **48**, 6767 (2005)
34. V. Kumar, K.-P. Tan, Y.-M. Wang, S.-W. Lin, P.-H. Liang, Bioorg. Med. Chem. Lett. **24**, 3035 (2016)
35. O. Trott, A.J. Olson, J. Comput. Chem. **31**, 455 (2010)
36. R.E. Dizaji, A. Rezaie Kehkhaie, M. Taqi Khammar, R. Shirazinia, Int. J. Plant Sci. Hor. **1**, 56 (2019)
37. A. Ameri, G. Heydarirad, J. Mahdavi Jafari, A. Ghobadi, H. Rezaeizadeh, R. Chooapani, Pharm Biol. **53**, 615 (2015)
38. J. Patočka, Z. Navratilova, Pharmacology **20**, 14915 (2019)
39. M. Farahani, Q. Branch, I. Azad, Zahedan J. Res Med. Sci. **15**, 46 (2013)
40. M. Abolhassani, Arch. Med. Sci.: AMS **6**, 366 (2010)
41. G.W. Price, P.S. Gould, A. Marsh, J. Chem. Educ. **91**, 602 (2014)
42. G.M. Morris, R. Huey, W. Lindstrom, M.F. Sanner, R.K. Belew, D.S. Goodsell, A.J. Olson, J. Comput. Chem. **30**, 2785 (2009)
43. C. Notredame, D.G. Higgins, J. Heringa, J. Mol. Biol. **302**, 205 (2000)
44. P. Gouet, E. Courcelle, D.I. Stuart, F. Metoz, Bioinformatics **15**, 305 (1999)
45. S.F. Altschul, W. Gish, W. Miller, E.W. Myers, D.J. Lipman, J. Mol. Biol. **215**, 403 (1990)
46. E. Gasteiger, C. Hoogland, A. Gattiker, M.R. Wilkins, R.D. Appel, A. Bairoch, *Protein Identification and Analysis Tools on the ExPASy Server in: The Proteomics Protocols Handbook* (Springer, Berlin, 2005), pp. 571–607
47. T. Gaillard, J. Chem. Inf. Model. **58**, 1697 (2018)
48. A.C. Wallace, R.A. Laskowski, J.M. Thornton, Protein Eng. Des. Sel. **8**, 127 (1995)
49. Y. Yu, L.A. Santat, S. Choi, *Bioinformatic Package for Sequence Analysis in: Applied Mycology and Biotechnology* (Elsevier, New York, 2006), pp. 143–160
50. H. Yang, M. Yang, Y. Ding, Y. Liu, Z. Lou, Z. Zhou, L. Sun, L. Mo, S. Ye, H. Pang, Proceedings of natural academy of sciences. **100** 13190 (2003)
51. N. Chitranshi, V.K. Gupta, R. Rajput, A. Godinez, K. Pushpitha, T. Shen, M. Mirzaei, Y. You, D. Basavarajappa, V. Gupta, J. Transl. Med. **18**, 1 (2020)
52. P.H. Torres, A.C. Sodero, P. Jofily, F.P. Silva-Jr, Int. J. Mol. Sci. **20**, 4574 (2019)
53. A. Castro-Alvarez, A.M. Costa, J. Vilarrasa, Molecules **22**, 136 (2017)
54. D. Ramirez, J. Caballero, Molecules **23**, 1038 (2018)
55. N. Razzaghi-Asl, A. Ebadi, N. Edraki, A. Mehdipour, S. Shahabipour, R. Miri, J. Mol. Model. **18**, 4567 (2012)
56. J.A. Erickson, M. Jalaie, D.H. Robertson, R.A. Lewis, M. Vieth, J. Med. Chem. **47**, 45 (2004)
57. D.L. Ghorban, S.A. Madadkar, A. Shafiei, M. Mahmoudian, DARU, J. Pharm. Sci. **12**, 1 (2004)
58. C. Wu, Y. Liu, Y. Yang, P. Zhang, W. Zhong, Y. Wang, Q. Wang, Y. Xu, M. Li, X. Li, Acta Pharm. Sin. B. **91**, 602 (2020)
59. A. Ascone, R. Sakidja, Int. J. Comput Biol. Drug Des. **10**, 207 (2017)
60. R. Alikhani, N. Razzaghi-Asl, A. Ramazani, Z. Hosseinzadeh, J. Mol. Struct. **1164**, 9 (2018)
61. M. Orita, K. Ohno, M. Warizaya, Y. Amano, T. Niimi, *In Methods in Enzymology* (Elsevier, New York, 2011), pp. 383–419
62. J. Gao, Z. Tian, X. Yang, Biosci. Trends. **14**, 72 (2020)
63. Y.-C. Chang, Y.-A. Tung, K.-H. Lee, T.-F. Chen, Y.-C. Hsiao, H.-C. Chang, T.-T. Hsieh, C.-H. Su, S.-S. Wang, J.-Y. Yu, (2020), DOI <https://doi.org/10.20944/preprints202002.0242.v1>
64. J. Wang, (2020), DOI <https://doi.org/10.26434/chemrxiv.11875446.v1>
65. T. Sekhar, (2020), DOI <https://doi.org/10.20944/preprints202002.0418.v1>
66. Z. Wang, X. Chen, Y. Lu, F. Chen, W. Zhang, Biosci. Trends. **14**, 64 (2020)
67. B. R. Beck, B. Shin, Y. Choi, S. Park, K. Kang, (2020) DOI <https://doi.org/10.1101/2020.01.31.929547>
68. L. Masters, S. Eagon, M. Heying, J. Mol. Graph. Model. **96**, 107532 (2020)
69. Y. Ma, Y. Fu, S.C. Khojasteh, D. Dalvie, D. Zhang, J. Med. Chem. **60**, 8691 (2017)
70. A. a. P.-G. Grassauer, Eva, US, (2009)
71. K. Winther, A.S.V. Hansen, J. Campbell-Tofte, Botanicus **6**, 11 (2016)
72. M. Asadi-Samani, M.-T. Moradi, M. Bahmani, M. Shahrani, Int. J. PharmTech Res. **9**, 427 (2016)
73. H. Zhu, Y. Zhang, G. Ye, Z. Li, P. Zhou, C. Huang, Biol. Pharm. Bull. **32**, 68 (2009)
74. M. Naoi, M. Shamoto-Nagai, W. Maruyama, Future Neurol. **14**, FNL9 (2019)
75. S. Choubey, L.R. Varughese, V. Kumar, V. Beniwal, Pharm. Pat. Anal. **4**, 305 (2015)
76. A. Sattari, A. Ramazani, H. Aghahoseini (2020). <https://doi.org/10.21203/rs.3.rs-37994/v1>