 doi:10.20944/preprints202002.0193.v1

# Structural basis of SARS-CoV-2 3CL$^{pro}$ and anti-COVID-19 drug discovery from medicinal plants

Muhammad Tahir ul Qamar[1,2], Safar M. Alqahtani[3], Mubarak A. Alamri[3], Ling-Ling Chen[1,2,*]

[1]College of Life Science and Technology, Guangxi University, Nanning 530004, P. R. China

[2]Hubei Key Laboratory of Agricultural Bioinformatics, College of Informatics, Huazhong Agricultural University, Wuhan 430070, P. R. China

[3]Department of Pharmaceutical Chemistry, College of Pharmacy, Prince Sattam Bin Abdulaziz University, P.O. Box 11323, Alkarj, Saudi Arabia

*** To whom correspondence should be addressed.**

Prof. Ling-Ling Chen; llchen@mail.hzau.edu.cn

 doi:10.20944/preprints202002.0193.v1


## Abstract

The recent outbreak of coronavirus disease 2019 (COVID-19) caused by SARS-CoV-2 in December 2019 raised global health concerns. The viral 3-chymotrypsin-like cysteine protease ($3CL^{pro}$) enzyme, which controls coronavirus replication and is essential for its life cycle, is a proven drug discovery target in the case of severe acute respiratory syndrome coronavirus (SARS-CoV) and middle east respiratory syndrome coronavirus (MERS-CoV). Recent studies revealed that the genome sequence of SARS-CoV-2 is very similar to that of SARS-CoV. Therefore, herein, we analysed the $3CL^{pro}$ sequence, constructed a 3D homology model, and screened it against a medicinal plant library containing 32,297 potential anti-viral phytochemicals/traditional Chinese medicinal compounds. Our analyses revealed that the top nine hits may serve as potential anti-SARS-CoV-2 lead molecules for further optimisation and drug development to control COVID-19.


**Keywords:** Coronavirus, SARS-CoV-2, COVID-19, Natural products, Protein homology modelling, Molecular docking, Molecular dynamics simulation

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 14 February 2020 doi:10.20944/preprints202002.0193.v1

## 1. Introduction

The first case of a novel coronavirus was reported on December 30, 2019, in Wuhan city, Hubei province, P.R. China [1]. Swift actions were taken by the Centres for Disease Control and Prevention (CDC), Chinese health authorities, and researchers. The World Health Organization (WHO) temporarily named this pathogen 2019 novel coronavirus (2019-nCoV) [2]. On January 10, 2020, the first whole-genome sequence of 2019-nCoV was released, which helped researchers to quickly identify the virus in patients using reverse-transcription polymerase chain reaction (RT-PCR) methods [3]. On January 21, 2020, the first article was published, which revealed that 2019-nCoV belongs to the beta-coronavirus group, sharing ancestry with bat coronavirus HKU9-1, similar to SARS-coronaviruses, and despite sequence diversity its Spike protein interacts strongly with the human ACE2 receptor [1]. On January 30, 2020, the WHO announced a Public Health Emergency of International Concern (PHEIC) for the 2019-nCoV outbreak. Human-to-human transmission was confirmed. As of January 31, 51 whole-genome sequences of 2019-nCoV from different laboratories and regions have been submitted to GISAID [4]. On the February 12$^{th}$, 2020, the WHO permanently named the 2019-nCoV pathogen as SARS-CoV-2 and the causing disease as coronavirus disease 2019 (COVID-2019). By February 13, 2020, the death toll reached 1368, with 52,587 laboratory-confirmed cases and 16,067 suspected cases. The case fatality rate among infected people is presently around 2.53% (calculated as deaths / [deaths + laboratory confirmed cases]). The COVID-19 cases have also been reported in more than 20 different countries, including the UK, Singapore, Japan, Vietnam, the USA, Nepal, Australia, France, Germany, India, Malaysia, Canada and Thailand.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 14 February 2020 doi:10.20944/preprints202002.0193.v1

Coronaviruses are single-stranded positive-sense RNA viruses that possess large viral RNA genomes [5]. Recent studies showed that SARS-CoV-2 has a similar genomic organization to other beta-coronaviruses, consisting of a 5'-untranslated region (UTR), a replicase complex (orf1ab) encoding non-structural proteins (nsps), a spike protein (S) gene, and envelope protein (E) gene, a membrane protein (M) gene, a nucleocapsid protein (N) gene, 3'-UTR, and several unidentified non-structural open reading frames [3]. Although SARS-CoV-2 is classified in the beta-coronaviruses group, it is diverse from MERS-CoV and SARS-CoV. Recent studies highlighted that SARS-CoV-2 genes share <80% nucleotide identity and 89.10% nucleotide similarity with SARS-CoV genes [6, 7]. Usually, beta-coronaviruses produce a ~800 kDa polypeptide upon transcription of the genome. This polypeptide is proteolytically cleaved to generate various proteins, and proteolytic processing is mediated by papain-like protease ($PL^{pro}$) and 3-chymotrypsin-like protease ($3CL^{pro}$), of which $3CL^{pro}$ cleaves the polyprotein at 11 distinct sites to generate various non-structural proteins that are important for viral replication [8]. Thus, this protease plays a critical role in the replication of virus particles, and unlike structural/accessory protein-encoding genes located at the 3' end that exhibit excessive variability, it may serve as a potential target for anti-SARS-CoV-2 inhibitors [9]. Structure-based activity analyses and high-throughput studies have identified potential inhibitors of SARS-CoV and MERS-CoV $3CL^{pro}$ [10-12]. Medicinal plants, especially those employed in traditional Chinese medicine, have attracted significant attention because they include bioactive compounds that could be used to develop drugs against diseases with minimal side-effects [13]. Therefore, the present study was conducted to obtain structural insight into SARS-CoV-2 $3CL^{pro}$ and discover potent anti-SARS-CoV-2 compounds.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 14 February 2020 doi:10.20944/preprints202002.0193.v1

## 2. Materials and methods

### 2.1. Data collection

Whole-genome sequences of all available SARS-CoV-2 isolates were downloaded from GISAID (accession numbers and details are given in Supplementary file 1) [4]. The genome sequence of BetaCoV/Kanagawa/1/2020 (GSAID: EPI_ISL_402126) was incomplete, and the genome sequence of BetaCoV/bat/Yunnan/RaTG13/2013 (EPI_ISL_402131) was an old sequence (2013), therefore these sequences were not included in our analyses. Gene sequences of $3CL^{pro}$ were extracted from the whole-genome sequences and translated into protein sequences using the translate tool of the ExPASy server [14]. The first SARS-CoV-2 sequence (Wuhan-Hu-1; GSAID: EPI_ISL_402125) was used as a reference.

### 2.2. Sequence analyses

In order to identify similar sequences and key/conserved residues, and to infer phylogeny, multiple sequence alignment of SARS-CoV-2 $3CL^{pro}$ followed by phylogenetic tree analyses were performed using MEGA v6.0 [15]. Physicochemical parameters of SARS-CoV-2 $3CL^{pro}$ including isoelectric point, instability index, grand average of hydropathicity (GRAVY), and amino acid and atomic composition were investigated using the ProtParam tool of ExPASy [14].

### 2.3. Structural analyses

To probe the molecular architecture of SARS-CoV-2 $3CL^{pro}$, comparative homology modelling was performed using Modeller v9.11 [16]. To select closely-related templates for modelling, PSI-BLAST was performed against all known structures in the protein databank (PDB) [17]. Chimera

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 14 February 2020 doi:10.20944/preprints202002.0193.v1

v1.8.1 [18] and PyMOL [19] were used for initial quality estimation, energy minimisation, mutation analyses, and image processing.

### 2.4. Ligand database preparation and molecular docking

A comprehensive medicinal plant library containing 32,297 potential anti-viral phytochemicals and traditional Chinese medicinal compounds was generated from our previous studies [13, 20, 21] and screened against the modelled SARS-CoV-2 $3CL^{pro}$ structure. MOE v2014 [22] was used for molecular docking, ligand-protein interaction and drug likeness analyses. Protocols were performed as described in our previous studies [13, 23, 24]. The qualitative assessment of absorption, deposition, metabolism, excretion and toxicity (ADMET) profile of these hits were predicted virtually by using ADMETsar server [25].

### 2.5. Molecular dynamics simulations

Explicit solvent molecular dynamics (MD) simulations were performed to verify docking results and analyse the binding behaviour and stability of potential compounds using the SARS-CoV-2 $3CL^{pro}$ homology model. GROMOS96 and the PRODRG server were employed to run 50 ns MD simulations [26, 27] following same protocol as described in our previous studies [13, 28].

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 14 February 2020 doi:10.20944/preprints202002.0193.v1

# 3. Results and Discussion

## 3.1. Sequence and structural analyses

Multiple sequence alignment results revealed that $3CL^{pro}$ is conserved, with 100% identity among all SARS-CoV-2 genomes. Next, the SARS-CoV-2 $3CL^{pro}$ protein sequence was compared with its closest homologs (Bat-CoV, SARS-CoV, MERS-CoV, Human-CoV and Bovine-CoV). The results revealed that SARS-CoV-2 $3CL^{pro}$ clustered with bat SARS-like coronaviruses, sharing 99.02% sequence identity (Figure 1A). Furthermore, it shares 96.08%, 87.00%, 90.00% and 90.00% sequence identity with SARS-CoV, MERS-CoV, Human-CoV and Bovine-CoV homologs, respectively (Figure 1B). These finding are consistent with an initial study reporting that SARS-CoV-2 is more similar to SARS-CoV than MERS-CoV, and shares a common ancestor with bat coronaviruses [1, 3, 29]. Analysis of physicochemical parameters revealed that the SARS-CoV-2 $3CL^{pro}$ polypeptide is 306 amino acids with a molecular weight of 33,796.64 Da and a GRAVY score of -0.019 GRAVY, categorising the protein as a stable, hydrophilic molecule capable of establishing hydrogen bonds (Table 1).

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 14 February 2020 doi:10.20944/preprints202002.0193.v1

**Table 1.** Physico-Chemical Parameters of SARS-CoV-2 3CL$^{pro}$

| **Parameters** | **2019-nCoV 3CL$^{pro}$** |
|---|---|
| Mol. Weight | 33796.64 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<br>Hydrogen 2318<br>Nitrogen 402<br>Oxygen 445<br>Sulfur 22 |
| Amino Acid Composition | Ala (A) 17 5.6%<br>Arg (R) 11 3.6%<br>Asn (N) 21 6.9%<br>Asp (D) 17 5.6%<br>Cys (C) 12 3.9%<br>Gln (Q) 14 4.6%<br>Glu (E) 9 2.9%<br>Gly (G) 26 8.5%<br>His (H) 7 2.3%<br>Ile (I) 11 3.6%<br>Leu (L) 29 9.5%<br>Lys (K) 11 3.6%<br>Met (M) 10 3.3%<br>Phe (F) 17 5.6%<br>Pro (P) 13 4.2%<br>Ser (S) 16 5.2%<br>Thr (T) 24 7.8%<br>Trp (W) 3 1.0%<br>Tyr (Y) 11 3.6%<br>Val (V) 27 8.8%<br>Pyl (O) 0 0.0%<br>Sec (U) 0 0.0% |

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 14 February 2020 doi:10.20944/preprints202002.0193.v1

1

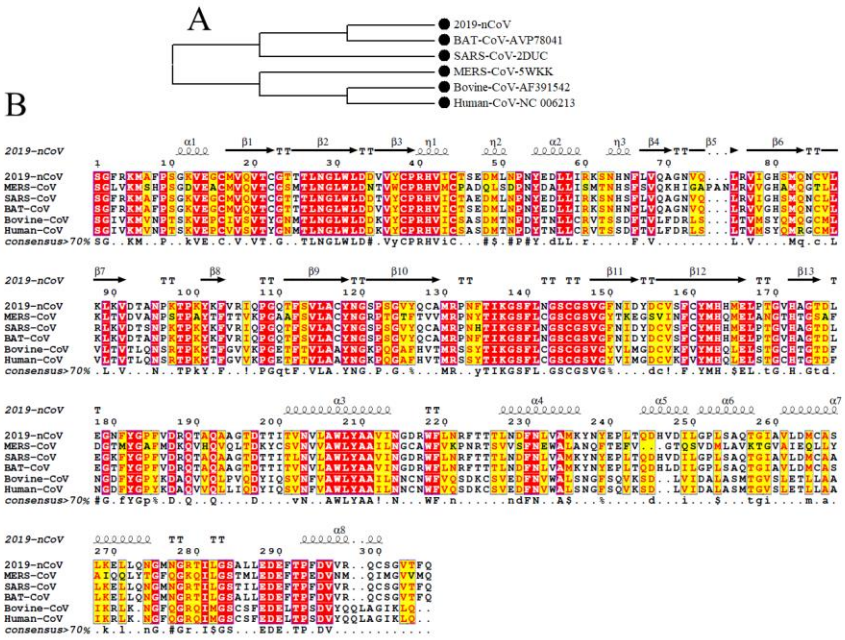
Next, for comparative modelling, BLAST [17] identified SARS-CoV $3CL^{pro}$ (PDB ID: 3M3V) as the best possible match in the PDB, with 100% query coverage, an E-value of 0.00, and 96.08% sequence identity. There are 12-point mutations (Val35Thr, Ser46Ala, Asn65Ser, Val86Leu, Lys88Arg, Ala94Ser, Phe134His, Asn180Lys, Val202Leu, Ser267Ala, Ser284Ala and Leu286Ala) between SARS-CoV and SARS-CoV-2 $3CL^{pro}$ enzymes (Supplementary file 2). Except for replacement of Leu with Ala at position 286, all other replacements conserve polarity and hydrophobicity. However, these mutations may affect protease structure and function. The 3D structure of SARS-CoV-2 $3CL^{pro}$ as therefore predicted. Firstly, a single chain monomeric model comprising all domains (Domain I = residues 8–100; Domain II = residues 101–183; Domain III = residues 200–303) was built (Supplementary file 3). N-terminal amino acids 1 to 7 form the N-finger that plays a significant role in dimerization and formation of the active site of $3CL^{pro}$. Domains I and II, collectively referred to as the N-terminal domain, includes an antiparallel β-sheet structure with 13 β-strands. The binding site for the substrate is situated in a cleft between domains I and II. A loop from residues 184 to 199 joins the N-terminal domain and Domain III, also called the C-terminal domain, which comprises an anti-parallel cluster of five α-helices. The overall molecular architecture of SARS-CoV-2 $3CL^{pro}$ is consistent with the crystal structure of SARS-CoV (PDB ID: 3M3V); the root mean square deviation (RMSD) between the homology model and the template is 0.629 Å. Structural and Ramachandran plot analyses revealed that 99% of residues are in favourable regions.

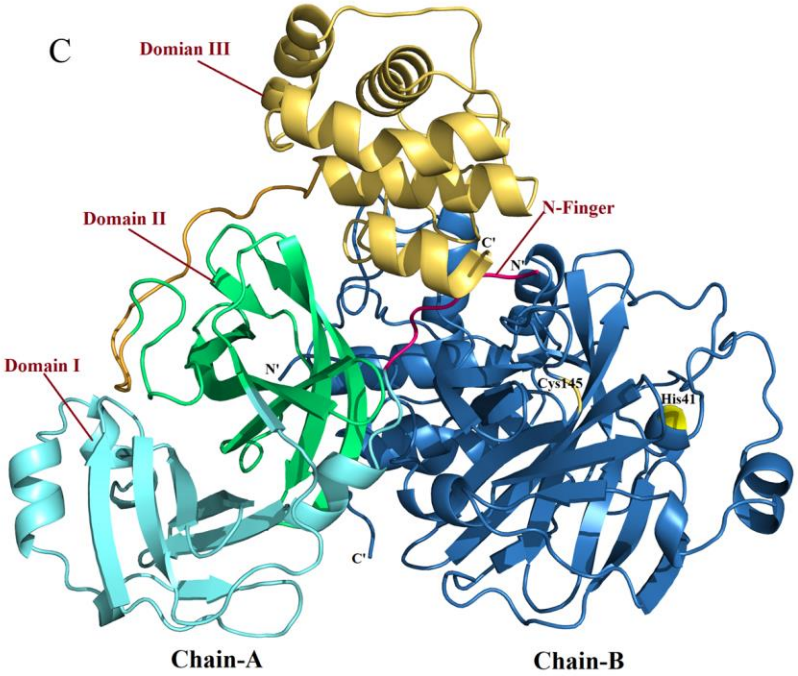
After quality assessment, individual domains were combined to form a homodimeric 3D structure, as shown in Figure 1C. Further mutational analyses showed that none of the mutations affected the overall structure of SARS-CoV-2, which fully superimposes on the SARS-CoV $3CL^{pro}$

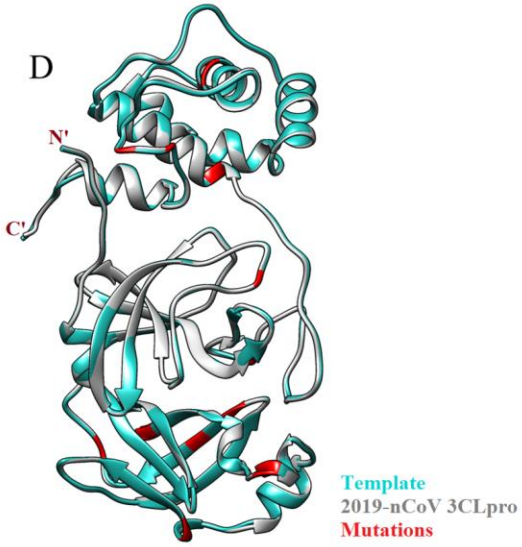
Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 14 February 2020 doi:10.20944/preprints202002.0193.v1

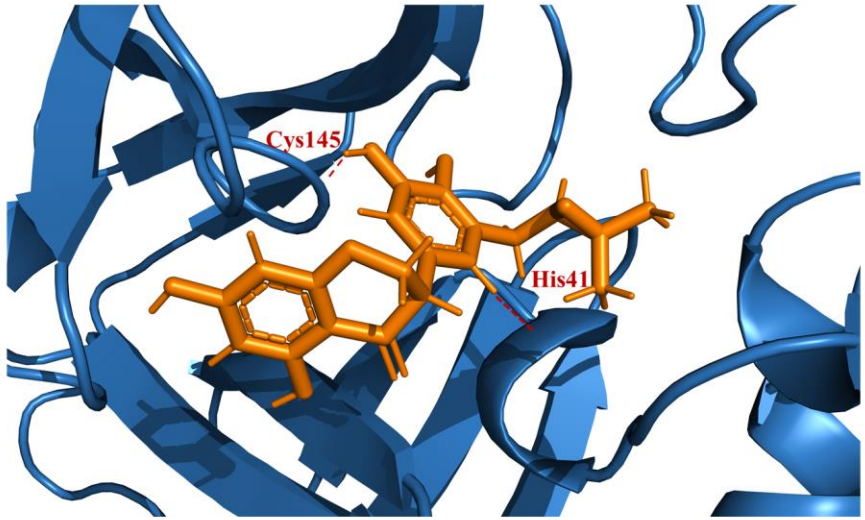
structure (Figure 1D). The results also revealed that SARS-CoV-2 has a Cys-His catalytic dyad (Cys-145 and His-41), consistent with SARS 3CL$^{pro}$ (Cys-145 and His-41), TGEV 3CL$^{pro}$ (Cys-144 and His-41) and HCoV 3CL$^{pro}$ (Cys-144 and His-41) [30]. These results revealed that the SARS-CoV-2 3CL$^{pro}$ receptor-binding pocket conformation resembles that of the SARS-CoV 3CL$^{pro}$ binding pocket and raises the possibility that inhibitors intended for SARS-CoV 3CL$^{pro}$ may also inhibit the activity of SARS-CoV-2 3CL$^{pro}$.

*Preprints* (www.preprints.org) | NOT PEER-REVIEWED | Posted: 14 February 2020 doi:10.20944/preprints202002.0193.v1



 doi:10.20944/preprints202002.0193.v1

**Figure 1.** (A) Phylogenetic tree inferred from closest homologs of SARS-CoV-2 $3CL^{pro}$. The maximum likelihood method was used to construct this tree. (B) Multiple sequence alignment of closest homologs of SARS-CoV-2 $3CL^{pro}$ sharing ≥70% sequence identity. Alignment was performed using T-Coffee [37] and the figure was generated using ESPript3 [38]. (C) Cartoon representation of the SARS-CoV-2 $3CL^{pro}$ homodimer. Chain-A (protomer-A) is in multicolour and Chain-B (protomer-B) is in dark blue. The N-finger that plays an important role in dimerization maintaining the active conformation is shown in hot pink, domain I is coloured cyan, domain II is shown in green, and domain III is coloured yellow. The N- and C-termini are labelled. Residues of the catalytic dyad (Cys-145 and His-41) are highlighted in yellow and labelled. The PyMOL educational version was used to draw this image [19]. (D) Cartoon representation of the $3CL^{pro}$ monomer model (chain/protomer-A) of SARS-CoV-2 superimposed with the SARS-CoV $3CL^{pro}$ structure. The SARS-CoV $3CL^{pro}$ template is coloured cyan, the SARS-CoV-2 $3CL^{pro}$ structure is coloured grey, and all identified mutations are highlighted in red. Chimera v1.8.1 was used to generate this image [18]. (E) Docking of 5,7,3’,4’-tetrahydroxy-2'-(3,3-dimethylallyl) isoflavone inside the receptor-binding site of SARS-CoV-2 $3CL^{pro}$, showing hydrogen bonds with the catalytic dyad (Cys-145 and His-41). The $3CL^{pro}$ structure is coloured dark blue, the 5,7,3’,4’-tetrahydroxy-2'-(3,3-dimethylallyl) isoflavone is orange, and hydrogen coloured maroon. MOE v2014 [22] was used for docking/interaction analyses, and the PyMOL educational version [19] was used to draw the docking pose image. To facilitate other researchers, the predicted 3D model has been submitted to the Protein Model Database (PMDB) [39], and anyone can download the SARS-CoV-2 $3CL^{pro}$ final structure using PMDB ID: PM0082635.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 14 February 2020 doi:10.20944/preprints202002.0193.v1

### 3.2. Molecular docking

To test this hypothesis, we docked (R)-N-(4-(tert-butyl)phenyl)-N-(2-(tert-butylamino)-2-oxo-1-(pyridin-3-yl)ethyl)furan-2-carboxamide), a potential noncovalent inhibitor of SARS-CoV $3CL^{pro}$ named ML188 [31], with the SARS-CoV-2 $3CL^{pro}$ homology model. We also docked ML188 with the SARS-CoV $3CL^{pro}$ structure (PDB ID: 3M3V) as a reference, and ML188 bound strongly to the receptor binding site of SARS-CoV $3CL^{pro}$. The inhibitor targets the Cys-His catalytic dyad (Cys-145 and His-41) along with the other residues, and the docking score (*S* = -12.27) is relatively high. However, surprisingly, ML188 did not show significant binding to the catalytic dyad (Cys-145 and His-41) of SARS-CoV-2, and the docking score (*S* = -8.31) was considerably lower (Supplementary file 4). These results suggest that the mutations may disrupt important hydrogen bonds and alter the receptor binding site, thereby affecting its ability to bind SARS-CoV inhibitors.

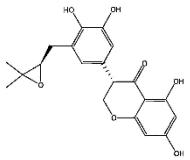
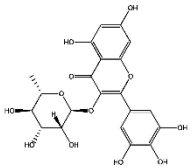
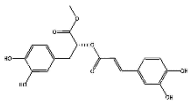
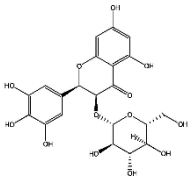
Therefore, it is essential to discover novel compounds that may inhibit SARS-CoV-2 $3CL^{pro}$ and serve as potential drug compounds. We developed a library from our previous studies [13, 20] that contains numerous natural products possessing anti-viral activity, and screened this against the SARS-CoV-2 $3CL^{pro}$ homology model. Recent drug repurposing studies proposed drugs that target SARS-CoV-2 $3CL^{pro}$ and suggested that they could be used to treat COVID-19. Herein, we selected the best of these (Nelfinavir, Prulifloxacin and Colistin) from three different drug repurposing studies [32-34] and docked them as controls in the present study (Supplementary file 5). Our analyses identified nine novel non-toxic, druggable natural products that are predicted to bind to the receptor binding site and catalytic dyad (Cys-145 and His-41) of SARS-CoV-2 $3CL^{pro}$ (Table 2; Supplementary file 6). ADMET profile of selected hits is given in Supplementary file 7. Among these, phytochemicals 5,7,3',4'-tetrahydroxy-2'-(3,3-dimethylallyl) isoflavone, an

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 14 February 2020 doi:10.20944/preprints202002.0193.v1

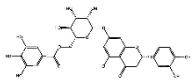
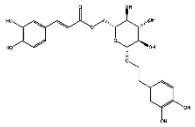
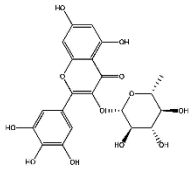
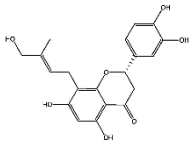
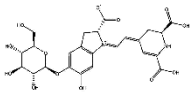
isoflavone extracted from *Psorothamnus arborescens* [35], exhibited the highest binding affinity (-29.57 Kcal/mol) and docking score (*S* = -16.35), and formed strong hydrogen bonds with the catalytic dyad residues (Cys-145 and His-41), as well as interactions with receptor-binding residues Thr24, Thr25, Thr26, Cys44, Thr45, Ser46, Met49, Asn142, Gly143, His164, Glu166 and Gln189 (Figure 1E). A literature review revealed that 5,7,3',4'-tetrahydroxy-2'-(3,3-dimethylallyl) isoflavone has been successfully used as an anti-leishmanial agent [35], and it is also found in traditional Chinese medicine records [36]. Our screened phytochemicals displayed higher docking scores, stronger binding energies, and more closer interactions with the conserved catalytic dyad residues (Cys-145 and His-41) than Nelfinavir, Prulifloxacin and Colistin. These results suggest that natural products identified in our study may prove more useful candidates for COVID-19 drug therapy.

*Preprints* (www.preprints.org) | NOT PEER-REVIEWED | Posted: 14 February 2020 doi:10.20944/preprints202002.0193.v1

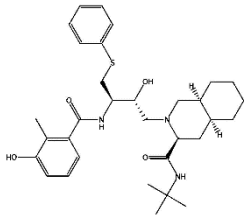
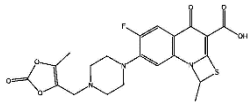
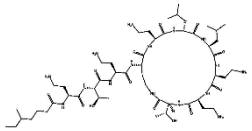
**Table 2.** Summary of top ranked phytochemicals screened against SARS-CoV-2 $3CL^{pro}$ receptor binding site with their respective structures, dock score, binding affinity, interacting and closer contact residues.

| IDs | Phytochemical Name | Plant Source | Structure | Score | Binding Affinity (Kcal/mol) | Residues interacting with Ligand through H-Bonding and other interactions |
|---|---|---|---|---|---|---|
| PubChem11610052 | 5,7,3',4'-Tetrahydroxy-2'-(3,3-dimethylallyl) isoflavone | *Psorothamnus arborescens* |  | -16.35 | -29.57 | **His41**, **Cys145**, Thr24, Thr25, Thr26, Cys44, Thr45, Ser46, Met49, Asn142, Gly143, His164, Glu166, Gln189 |
| PubChem5281673 | Myricitrin | *Myrica cerifera* |  | -15.64 | -22.13 | Gly143, **Cys145**, **His41**, Thr24, Thr25, Thr26, Leu27, Cys44, Ser46, Met49, Leu141, Asn142, Ser144, His163, Glu166, Gln189 |
| PubChem 6479915 | Methyl rosmarinate | *Hyptis atrorubens Poit* |  | -15.44 | -20.62 | **Cys145**, **His41**, Thr24, Thr25, Thr26, Cys44, Thr45, Met49, Leu141, Asn142, Gly143, Ser144, His163, His164, Glu166, Gln189 |
| NPACT00105 | 3,5,7,3',4',5'-hexahydroxy flavanone-3-O-beta-D-glucopyranoside | *Phaseolus vulgaris* |  | -14.42 | -19.10 | Met49, **Cys145**, **His41**, Thr24, Thr25, Thr26, Cys44, Ser46, Asn142, His164, Met165, Glu166, Gln189 |

*Preprints* (www.preprints.org) | NOT PEER-REVIEWED | Posted: 14 February 2020 doi:10.20944/preprints202002.0193.v1

| | | | | | | |
|---|---|---|---|---|---|---|
| PubChem 10930068 | (2S)-Eriodictyol 7-O-(6"-O-galloyl)-beta-D-glucopyranoside | *Phyllanthus emblica* |  | -14.41 | -19.47 | Thr24, Thr25, Gly143, Met49, **Cys145**, **His41**, Thr26, Cys44, Thr45, Glu166, Leu167, Gln189, Thr190, Ala191, Gln192 |
| PubChem 5273567 | Calceolarioside B | *Fraxinus sieboldiana* |  | -14.36 | -19.87 | **His41**, Gly143, **Cys145**, Glu166, Thr24, Thr25, Thr26, Leu27, Ser46, Leu50, Leu141, Asn142, Ser144, His164, Met165, Gln189 |
| PubChem 5318606 | Myricetin 3-O-beta-D-glucopyranoside | *Camellia sinensis* |  | -13.70 | -18.42 | Asn142, Glu166, **Cys145**, **His41**, Thr24, Thr25, Thr26, Thr45, Ser46, Met49, Leu141, Gly143, Ser144, His163, His164, Met165, Gln189 |
| PubChem 11111496 | Licoleafol | *Glycyrrhiza uralensis* |  | -13.63 | -19.64 | **Cys145**, **His41**, Thr24, Thr25, Thr26, Cys44, Thr45, Met49, Gly143, His163, His163, Met165, Glu166, Gln189 |
| PubChem 6123095 | Amaranthin | *Amaranthus tricolor* |  | -12.67 | -18.14 | Thr26, Glu166, **Cys145**, **His41**, Thr24, Thr25, Cys44, Cys44, Thr45, Asn142, His163, His164, Met165, Glu166, Leu167, Pro168, Gln189 |

*Preprints* (www.preprints.org) | NOT PEER-REVIEWED | Posted: 14 February 2020 doi:10.20944/preprints202002.0193.v1

| | | | | | | |
|---|---|---|---|---|---|---|
| PubChem 64143 | Nelfinavir | **Drugs used as control** |  | -12.20 | -17.31 | Met49, Met165, Glu166, Leu167, Pro168, Gly170, Gln189, Thr190, Ala191 |
| PubChem 65947 | Prulifloxacin | |  | -11.32 | -15.40 | Gln192, Leu50, Met165, Glu166, Leu167, Pro168, Arg188, Gln189, Thr190, Ala191 |
| PubChem 5311054 | Colistin | |  | -13.73 | -18.57 | Met49, Thr24, Thr25, Thr26, Thr45, Ser46, Glu47, Leu50, Asn142, Gly143, Met165, Glu166, Leu167, Pro168, Gln189, Thr190, Ala191, Gln192 |

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 14 February 2020 doi:10.20944/preprints202002.0193.v1

### 3.3. MD simulations

To further investigate the molecular docking results, the top three phytochemical complexes, namely 5,7,3',4'-tetrahydroxy-2'-(3,3-dimethylallyl) isoflavone, myricitrin, and methyl rosmarinate, were subjected to MD simulation, and RMSD, root mean square fluctuation (RMSF), radius of gyration (RoG) and hydrogen bond parameters were calculated. RMSD is an indicator of the stability of ligand-protein complexes. None of the complex showed any obvious fluctuations, and all three were stable, with average RMSD values of 1.6± 0.02 Å, 1.5± 0.02 Å and 1.7± 0.02 Å for 5,7,3',4'-tetrahydroxy-2'-(3,3-dimethylallyl) isoflavone, myricitrin, and methyl rosmarinate, respectively (Figure 2A). RMSF is an indicator of residual flexibility. Minimal fluctuations were observed for myricitrin and methyl rosmarinate, and the overall complexes remained stable throughout the simulations. The functionally important catalytic dyad residues (Cys-145 and His-41) displayed stable behaviour, and fluctuations were observed toward the C-terminal end of the SARS-CoV-2 $3CL^{pro}$ molecule (Figure 2B). RoG is an indicator of protein compactness, stability, and folding, and the results suggested normal behaviour for all three complexes; all remained compact and stable throughout the 50 ns simulations (Figure 2C). In addition, hydrogen bonds, which are the main stabilising interactions factors in proteins, suggested that the SARS-CoV-2 $3CL^{pro}$ internal hydrogen bonds remain stable throughout the simulation, with no obvious fluctuations (Figure 2D).These results confirmed our findings, and further indicate that all nine compounds may serve as potential anti-SARS-CoV-2 drug sources.

*Preprints* (www.preprints.org) | NOT PEER-REVIEWED | Posted: 14 February 2020 doi:10.20944/preprints202002.0193.v1

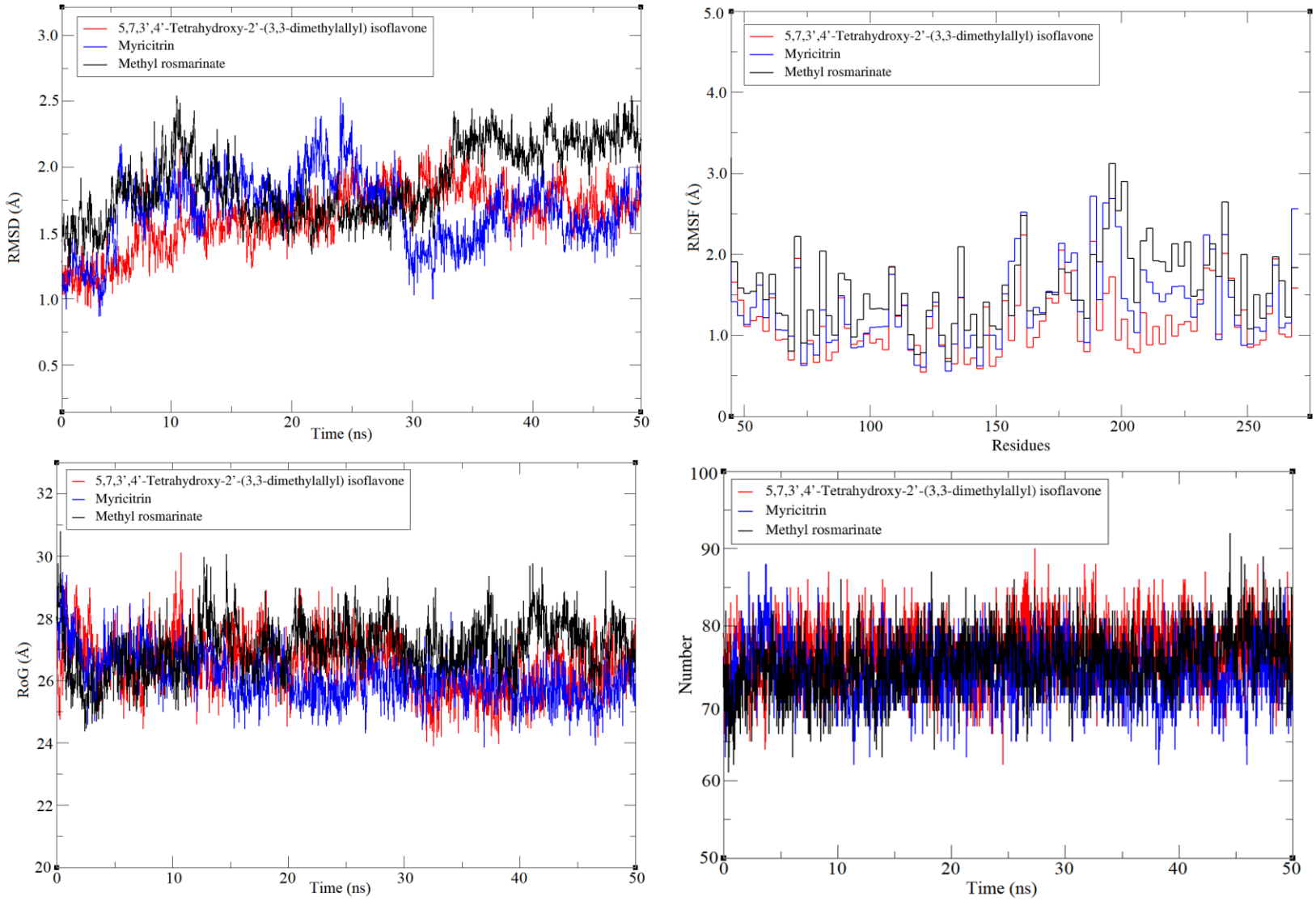



**Figure 2.** (A) Root mean square deviation (RMSD), (B) root mean square fluctuation (RMSF), (C) potential energy and (D) Hydrogen Bond interactions for all three complexes over the 50 ns simulation.

*Preprints* (www.preprints.org) | NOT PEER-REVIEWED | Posted: 14 February 2020 doi:10.20944/preprints202002.0193.v1

## 4. Conclusion

In conclusion, our study revealed that $3CL^{pro}$ is conserved in SARS-CoV-2. It is highly similar to bat SARS-like coronavirus $3CL^{pro}$, with some differences from other beta-coronaviruses. We predicted the 3D structure of the SARS-CoV-2 $3CL^{pro}$ enzyme, and the findings may help researchers working on COVID-19 drug discovery. Despite significant overall similarity with the SARS-CoV $3CL^{pro}$ structure, the SARS-CoV-2 $3CL^{pro}$ substrate binding site has some key differences, which highlights the need for rapid drug discovery to address the alarming COVID-19 outbreak. Medicinal plant compounds are already used to successfully treat numerous viral diseases. Herein, we screened a medicinal plant database containing 32,297 potential anti-viral phytochemicals and selected the top nine hits that may inhibit SARS-CoV-2 $3CL^{pro}$ activity and hence virus replication. Further *in-vitro* and *in-vivo* analyses are required to transform these potential inhibitors into clinical drugs. We anticipate that the insights obtained in the present study may prove valuable for exploring and developing novel natural anti-SARS-CoV-2 therapeutic agents in the future.

*Preprints* (www.preprints.org) | NOT PEER-REVIEWED | Posted: 14 February 2020 doi:10.20944/preprints202002.0193.v1

## Conflict of interests

The author(s) declare that they have no conflict of interest.

## Acknowledgements

This work was supported by the Hubei Provincial Natural Science Foundation of China (2019CFA014) and the Starting Research Grant for High-level Talents from Guangxi University. We also acknowledge all authors and laboratories mentioned in Supplementary file 1 for their sampling, analysis, and genome sequencing efforts. In addition, we acknowledge GISAID (https://www.gisaid.org/) for facilitating open data sharing.

## Supporting Information

**Supplementary file 1.** Acknowledgement to the authors and laboratories, sampling, analyzing and submitting the genome sequences to GISAID database.

**Supplementary file 2.** Sequence alignment between template $3CL^{pro}$ (SARS-CoV PDB ID: 3M3V) and SARS-CoV-2 $3CL^{pro}$. Brown boxes are displaying mutations (Val35Thr, Ser46Ala, Asn65Ser, Val86Leu, Lys88Arg, Ala94Ser, Phe134His, Asn180Lys, Val202Leu, Ser267Ala, Ser284Ala, Leu286Ala). Alignment was done using MEGA v6.0.

**Supplementary file 3.** (A) Cartoon representation of computationally predicted 3D structure of SARS-CoV-2 $3CL^{pro}$ (monomer), (B) Ramachandran plot displaying 99% residues are in favorable region. Figures were generated using Chimera v1.8.1.

*Preprints* (www.preprints.org) | NOT PEER-REVIEWED | Posted: 14 February 2020 doi:10.20944/preprints202002.0193.v1

**Supplementary file 4.** (A) 2D representation of ML188 binding mode with receptor binding site of SARS-CoV $3CL^{pro}$. (B) 2D representation of ML188 binding mode with receptor binding site of SARS-CoV-2 $3CL^{pro}$. MOE v2014 has been used for these analyses.

**Supplementary file 5.** (A) 2D representation of Nelfinavir binding mode with receptor binding site of SARS-CoV-2 $3CL^{pro}$. (B) 2D representation of Prulifloxacin binding mode with receptor binding site of SARS-CoV-2 $3CL^{pro}$. (C) 2D representation of Colistin binding mode with receptor binding site of SARS-CoV-2 $3CL^{pro}$. MOE v2014 has been used for these analyses.

**Supplementary file 6.** 2D representation of (A) 5,7,3',4'-Tetrahydroxy-2'-(3,3-dimethylallyl) isoflavone, (B) Myricitrin, (C) Methyl rosmarinate, (D) 3,5,7,3',4',5'-hexahydroxy flavanone-3-O-beta-D-glucopyranoside, (E) (2S)-Eriodictyol 7-O-(6''-O-galloyl)-beta-D-glucopyranoside, (F) Calceolarioside B, (G) Myricetin 3-O-beta-D-glucopyranoside, (H) Licoleafol and (I) Amaranthin binding mode with receptor binding site of SARS-CoV-2 $3CL^{pro}$. MOE v2014 has been used for these analyses.

**Supplementary file 7.** ADMET profiling enlisting absoprtion, metabloim and toxicity related drug like parameters of all 9 selected phytochemicals. ADMETsar server was used for this analysis.

 doi:10.20944/preprints202002.0193.v1

## References

[1] X. Xu, P. Chen, J. Wang, J. Feng, H. Zhou, X. Li, W. Zhong, P.J.S.C.L.S. Hao, Evolution of the novel coronavirus from the ongoing Wuhan outbreak and modeling of its spike protein for risk of human transmission, Sci China Life Sci (2020).

[2] W. Ji, W. Wang, X. Zhao, J. Zai, X.J.J.o.M.V. Li, Homologous recombination within the spike glycoprotein of the newly identified coronavirus may boost cross-species transmission from snake to human, J Med Virol (2020).

[3] N. Zhu, D. Zhang, W. Wang, X. Li, B. Yang, J. Song, X. Zhao, B. Huang, W. Shi, R.J.N.E.J.o.M. Lu, A Novel Coronavirus from Patients with Pneumonia in China, 2019, N Engl J Med (2020).

[4] Y. Shu, J.J.E. McCauley, GISAID: Global initiative on sharing all influenza data–from vision to reality, Euro Surveill 22(13) (2017).

[5] Y. Chen, Q. Liu, D.J.J.o.M.V. Guo, Coronaviruses: genome structure, replication, and pathogenesis, J Med Virol (2020).

[6] Z.-L. Shi, P. Zhou, X.-L. Yang, X.-G. Wang, B. Hu, L. Zhang, W. Zhang, H.-R. Si, Y. Zhu, B.J.b. Li, Discovery of a novel coronavirus associated with the recent pneumonia outbreak in humans and its potential bat origin, bioRxiv (2020).

[7] 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, M.-L. Yuan, Y.-L. Zhang, F.-H. Dai, Y. Liu, Q.-M. Wang, J.-J. Zheng, L. Xu, E.C. Holmes, Y.-Z. Zhang, A new coronavirus associated with human respiratory disease in China, Nature (2020).

[8] K. Anand, J. Ziebuhr, P. Wadhwani, J.R. Mesters, R.J.S. Hilgenfeld, Coronavirus main proteinase (3CLpro) structure: basis for design of anti-SARS drugs, Science 300(5626) (2003) 1763-1767.

[9] D. Needle, G.T. Lountos, D.S.J.A.C.S.D.B.C. Waugh, Structures of the Middle East respiratory syndrome coronavirus 3C-like protease reveal insights into substrate specificity, Acta Crystallogr D Biol Crystallogr 71(5) (2015) 1102-1111.

[10] 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.J.J.o.m.c. Mesecar, Design and synthesis of peptidomimetic severe acute respiratory syndrome chymotrypsin-like protease inhibitors, J Med Chem 48(22) (2005) 6767-6771.

[11] V. Kumar, K.-P. Tan, Y.-M. Wang, S.-W. Lin, P.-H.J.B. Liang, m. chemistry, Identification, synthesis and evaluation of SARS-CoV and MERS-CoV 3C-like protease inhibitors, Bioorg Med Chem 24(13) (2016) 3035-3042.

[12] T. Pillaiyar, M. Manickam, V. Namasivayam, Y. Hayashi, S.-H.J.J.o.m.c. Jung, An Overview of Severe Acute Respiratory Syndrome–Coronavirus (SARS-CoV) 3CL Protease Inhibitors: Peptidomimetics and Small Molecule Chemotherapy, J Med Chem 59(14) (2016) 6595-6628.

[13] M.T. ul Qamar, A. Maryam, I. Muneer, F. Xing, U.A. Ashfaq, F.A. Khan, F. Anwar, M.H. Geesi, R.R. Khalid, S.A.J.S.r. Rauf, Computational screening of medicinal plant phytochemicals to discover potent pan-serotype inhibitors against dengue virus, Sci Rep 9(1) (2019) 1-16.

[14] E. Gasteiger, A. Gattiker, C. Hoogland, I. Ivanyi, R.D. Appel, A.J.N.a.r. Bairoch, ExPASy: the proteomics server for in-depth protein knowledge and analysis, Nucleic Acids Res 31(13) (2003) 3784-3788.

[15] S. Kumar, M. Nei, J. Dudley, K.J.B.i.b. Tamura, MEGA: a biologist-centric software for evolutionary analysis of DNA and protein sequences, Brief Bioinform 9(4) (2008) 299-306.

[16] N. Eswar, B. Webb, M.A. Marti-Renom, M. Madhusudhan, D. Eramian, M.y. Shen, U. Pieper, A.J.C.p.i.b. Sali, Comparative protein structure modeling using Modeller, Curr Protoc Bioinformatics 15(1) (2006) 5.6. 1-5.6. 30.

[17] M. Johnson, I. Zaretskaya, Y. Raytselis, Y. Merezhuk, S. McGinnis, T.L.J.N.a.r. Madden, NCBI BLAST: a better web interface, Nucleic Acids Res 36(suppl_2) (2008) W5-W9.

[18] E.F. Pettersen, T.D. Goddard, C.C. Huang, G.S. Couch, D.M. Greenblatt, E.C. Meng, T.E.J.J.o.c.c. Ferrin, UCSF Chimera—a visualization system for exploratory research and analysis, J Comput Chem 25(13) (2004) 1605-1612.

[19] W.L.J.C.N.o.p.c. DeLano, Pymol: An open-source molecular graphics tool, CCP4 Newsletter on protein crystallography 40(1) (2002) 82-92.

[20] A. Mumtaz, U.A. Ashfaq, M.T. ul Qamar, F. Anwar, F. Gulzar, M.A. Ali, N. Saari, M.T.J.N.p.r. Pervez, MPD3: a useful medicinal plants database for drug designing, Nat Prod Res 31(11) (2017) 1228-1236.

[21] U.A. Ashfaq, A. Mumtaz, T. ul Qamar, T.J.B. Fatima, MAPS Database: Medicinal plant activities, phytochemical and structural database, Bioinformation 9(19) (2013) 993.

[22] S. Vilar, G. Cozza, S.J.C.t.i.m.c. Moro, Medicinal chemistry and the molecular operating environment (MOE): application of QSAR and molecular docking to drug discovery, Curr Top Med Chem 8(18) (2008) 1555-1572.

[23] M.T. Ul Qamar, S. Saleem, U.A. Ashfaq, A. Bari, F. Anwar, S.J.J.o.t.m. Alqahtani, Epitope-based peptide vaccine design and target site depiction against Middle East Respiratory Syndrome Coronavirus: an immune-informatics study, J Transl Med 17(1) (2019) 362.

[24] M.T. ul Qamar, A. Bari, M.M. Adeel, A. Maryam, U.A. Ashfaq, X. Du, I. Muneer, H.I. Ahmad, J.J.J.o.t.m. Wang, Peptide vaccine against chikungunya virus: immuno-informatics combined with molecular docking approach, J Transl Med 16(1) (2018) 298.

[25] F. Cheng, W. Li, Y. Zhou, J. Shen, Z. Wu, G. Liu, P.W. Lee, Y. Tang, admetSAR: a comprehensive source and free tool for assessment of chemical ADMET properties, J Chem Inf Model (2012).

[26] H.J. Berendsen, D. van der Spoel, R.J.C.p.c. van Drunen, GROMACS: a message-passing parallel molecular dynamics implementation, Comput Phys Commun 91(1-3) (1995) 43-56.

[27] D.M. Van Aalten, R. Bywater, J.B. Findlay, M. Hendlich, R.W. Hooft, G.J.J.o.c.-a.m.d. Vriend, PRODRG, a program for generating molecular topologies and unique molecular descriptors from coordinates of small molecules, J Comput Aid Mol Des 10(3) (1996) 255-262.

[28] I. Muneer, K. Tusleem, S. Abdul Rauf, H.M. Hussain, A.R.J.A.-c.d. Siddiqi, Discovery of selective inhibitors for cyclic AMP response element-binding protein: a combined ligand and structure-based resources pipeline, Anti-Cancer Drugs 30(4) (2019) 363-373.

[29] P. Zhou, X.-L. Yang, X.-G. Wang, B. Hu, L. Zhang, W. Zhang, H.-R. Si, Y. Zhu, B. Li, C.-L. Huang, H.-D. Chen, J. Chen, Y. Luo, H. Guo, R.-D. Jiang, M.-Q. Liu, Y. Chen, X.-R. Shen, X. Wang, X.-S. Zheng, K. Zhao, Q.-J. Chen, F. Deng, L.-L. Liu, B. Yan, F.-X. Zhan, Y.-Y. Wang, G.-F. Xiao, Z.-L. Shi, A pneumonia outbreak associated with a new coronavirus of probable bat origin, Nature (2020).

[30] H. Yang, M. Yang, Y. Ding, Y. Liu, Z. Lou, Z. Zhou, L. Sun, L. Mo, S. Ye, H.J.P.o.t.N.A.o.S. Pang, The crystal structures of severe acute respiratory syndrome virus main protease and its complex with an inhibitor, Proc Natl Acad Sci U S A 100(23) (2003) 13190-13195.

[31] J. Jacobs, S. Zhou, E. Dawson, J.S. Daniels, P. Hodder, V. Tokars, A. Mesecar, C.W. Lindsley, S.R. Stauffer, Discovery of non-covalent inhibitors of the SARS main proteinase 3CLpro, Probe Reports from the NIH Molecular Libraries Program [Internet], National Center for Biotechnology Information (US), Bethesda (MD), 2013.

[32] Z. Xu, C. Peng, Y. Shi, Z. Zhu, K. Mu, X. Wang, W.J.b. Zhu, Nelfinavir was predicted to be a potential inhibitor of 2019-nCov main protease by an integrative approach combining homology modelling, molecular docking and binding free energy calculation, BioRxiv (2020).

[33] Y. Li, J. Zhang, N. Wang, H. Li, Y. Shi, G. Guo, K. Liu, H. Zeng, Q.J.b. Zou, Therapeutic Drugs Targeting 2019-nCoV Main Protease by High-Throughput Screening, bioRxiv (2020).

[34] X. Liu, X.-J.J.b. Wang, Potential inhibitors for 2019-nCoV coronavirus M protease from clinically approved medicines, BioRxiv (2020).

[35] M.M. Salem, K.A.J.J.o.n.p. Werbovetz, Isoflavonoids and Other Compounds from Psorothamnus a rborescens with Antiprotozoal Activities, J Nat Prod 69(1) (2006) 43-49.

[36] J. Zhou, G. Xie, X.J.I.C.A. Yan, Encyclopedia of traditional Chinese medicines, Springer 1 (2011) 455.

[37] C. Notredame, D.G. Higgins, J.J.J.o.m.b. Heringa, T-Coffee: A novel method for fast and accurate multiple sequence alignment, J Mol Biol 302(1) (2000) 205-217.

 doi:10.20944/preprints202002.0193.v1

[38] P. Gouet, E. Courcelle, D.I. Stuart, F.J.B. M√© toz, ESPript: analysis of multiple sequence alignments in PostScript, Bioinformatics 15(4) (1999) 305-308.

[39] T. Castrignano, P.D.O. De Meo, D. Cozzetto, I.G. Talamo, A.J.N.a.r. Tramontano, The PMDB protein model database, Nucleic acids res 34(suppl_1) (2006) D306-D309.