

Journal Pre-proof

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

Muhammad Tahir ul Qamar, Safar M. Alqahtani, Mubarak A. Alamri, Ling-Ling Chen



PII: S2095-1779(20)30127-1

DOI: <https://doi.org/10.1016/j.jpha.2020.03.009>

Reference: JPHA 533

To appear in: *Journal of Pharmaceutical Analysis*

Received Date: 5 February 2020

Revised Date: 20 March 2020

Accepted Date: 21 March 2020

Please cite this article as: M.T. ul Qamar, S.M. Alqahtani, M.A. Alamri, L.-L. Chen, Structural basis of SARS-CoV-2 3CL^{pro} and anti-COVID-19 drug discovery from medicinal plants[†], *Journal of Pharmaceutical Analysis* (2020), doi: <https://doi.org/10.1016/j.jpha.2020.03.009>.

This is a PDF file of an article that has undergone enhancements after acceptance, such as the addition of a cover page and metadata, and formatting for readability, but it is not yet the definitive version of record. This version will undergo additional copyediting, typesetting and review before it is published in its final form, but we are providing this version to give early visibility of the article. Please note that, during the production process, errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

© 2020 Xi'an Jiaotong University. Production and hosting by Elsevier B.V. All rights reserved.

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. Alqahtani³, Mubarak A. Alamri³, Ling-Ling Chen^{1,2,†}

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

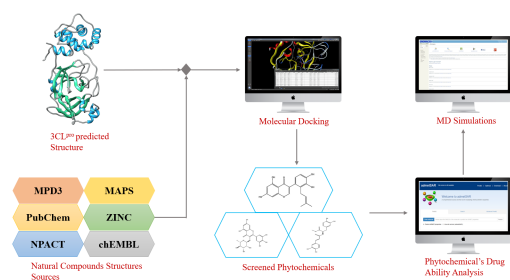
² Hubei Key Laboratory of Agricultural Bioinformatics, College of Informatics, Huazhong Agricultural University, Wuhan 430070, P. R. China

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

* This manuscript has been released as a preprint at Preprints



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 controls coronavirus replication and is essential for its life cycle. 3CL^{pro} 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 its 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 might serve as potential anti- SARS-CoV-2 lead molecules for further optimisation and drug development process to combat COVID-19.

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

1. Introduction

The first case of the novel coronavirus was reported on December 30, 2019, in Wuhan city, Hubei province, P.R. China [1]. Swift actions were taken by the Centre 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, the first article related to 2019-nCoV 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, the WHO announced a Public Health Emergency of International Concern (PHEIC) for the 2019-nCoV outbreak. Later, the 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 database [4]. On the February 12th, the WHO permanently named the 2019-nCoV pathogen as SARS-CoV-2 and the causing disease as coronavirus disease 2019 (COVID-2019). Chinese government swift actions helped them to control COVID-19 in China. However, SARS-CoV-2 quickly spread to over 150 countries. On March 11th, WHO formally recognized the COVID-19 as a pandemic. By March 19th, 2020, the global death toll reached to 9,913, with 2,42,650 laboratory-confirmed cases. The case fatality rate among infected people is varying in different countries. However, global case fatality rate is presently around 3.92% (calculated as deaths / [deaths + laboratory confirmed cases]).

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, 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 into 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. The proteolytic processing is mediated by papain-like protease (PL^{pro}) and 3-chymotrypsin-like protease (3CL^{pro}). The 3CL^{pro} cleaves the polyprotein at 11 distinct sites to generate various non-structural proteins that are important for viral replication [8]. 3CL^{pro} play a critical role in the replication of virus particles and unlike structural/accessory protein-encoding genes, it is located at the 3' end which exhibits excessive variability. Therefore, it is a potential target for anti-coronaviruses inhibitors screening [9]. Structure-based activity analyses and high-throughput studies have identified potential inhibitors for 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 formal drugs against several diseases with no or minimal side-effects [13]. Therefore, the present study was conducted to obtain structural insight into the SARS-CoV-2 3CL^{pro} and to discover potent anti-COVID-19 natural compounds.

2. Materials and methods

2.1. Data collection

Whole-genome sequences of all available SARS-CoV-2 isolates available till January 31, 2020, were downloaded from GISAID database (accession numbers and details are given in Table S1) [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 in our analysis.

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 T-Coffee [15] and the alignment figure was generated using ESPript3 [16]. 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 [17]. To select closely-related templates for modelling,

PSI-BLAST was performed against all known structures in the protein databank (PDB) [18]. Chimera v1.8.1 [19] and PyMOL educational version [20] 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 previously collected data and studies [13, 21-23] and screened against the predicted SARS-CoV-2 3CL^{pro} structure. Molecular operating environment (MOE) [24] was used for molecular docking, ligand-protein interaction and drug likeness analyses. All analyses were performed using same protocols that are already described in our previous studies [13, 25, 26]. The qualitative assessment of absorption, deposition, metabolism, excretion and toxicity (ADMET) profile of selected hits were predicted computationally by using ADMETSar server [27].

2.5. Molecular dynamics simulations

Explicit solvent molecular dynamics (MD) simulations were performed to verify docking results and to analyse the binding behaviour and stability of potential compounds using the predicted SARS-CoV-2 3CL^{pro} homology model. GROMOS96 and the PRODRG server were employed to run 50 ns MD simulations [28, 29] following same protocol as described in our previous studies [13, 30].

106

107 3. Results and Discussion

108 3.1. Sequence and structural analyses

109 Multiple sequence alignment results revealed that 3CL^{pro} was conserved, with 100% identity
 110 among all SARS-CoV-2 genomes. Next, the SARS-CoV-2 3CL^{pro} protein sequence was
 111 compared with its closest homologs (Bat-CoV, SARS-CoV, MERS-CoV, Human-CoV and
 112 Bovine-CoV). The results revealed that SARS-CoV-2 3CL^{pro} clustered with bat SARS-like
 113 coronaviruses and sharing 99.02% sequence identity (Fig. 1A). Furthermore, it shares 96.08%,
 114 87.00%, 90.00% and 90.00% sequence identity with SARS-CoV, MERS-CoV, Human-CoV and
 115 Bovine-CoV homologs, respectively (Fig. 1B). These finding were consistent with an initial
 116 study reporting that SARS-CoV-2 is more similar to SARS-CoV than MERS-CoV, and shares a
 117 common ancestor with bat coronaviruses [1, 3, 31]. Analysis of physicochemical parameters
 118 revealed that the SARS-CoV-2 3CL^{pro} polypeptide is 306 amino acids long with a molecular
 119 weight of 33,796.64 Da and a GRAVY score of -0.019, categorising the protein as a stable,
 120 hydrophilic molecule capable of establishing hydrogen bonds (Table 1).

121 Next, for comparative modelling, BLAST [32] search identified SARS-CoV 3CL^{pro} (PDB ID:
 122 3M3V) as the best possible match in the PDB, with 100% query coverage, an E-value of 0.00,
 123 and 96.08% sequence identity. There were 12-point mutations (Val35Thr, Ser46Ala, Asn65Ser,
 124 Val86Leu, Lys88Arg, Ala94Ser, Phe134His, Asn180Lys, Val202Leu, Ser267Ala, Ser284Ala
 125 and Leu286Ala) between SARS-CoV and SARS-CoV-2 3CL^{pro} enzymes (Fig. S1). Except for
 126 replacement of Leu with Ala at position 286, all other replacements conserve polarity and
 127 hydrophobicity. However, these mutations may affect 3CL^{pro} structure and function. Therefore,

the 3D structure of SARS-CoV-2 3CL^{pro} was 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 (Fig. S2). 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, which is also called the C-terminal domain and comprises an anti-parallel cluster of five α -helices. The overall molecular architecture of SARS-CoV-2 3CL^{pro} was in consistent with the crystal structure of SARS-CoV (PDB ID: 3M3V); the root mean square deviation (RMSD) between the homology model and the template was 0.629 Å. Structural and Ramachandran plot analyses revealed that 99% of residues are in favourable regions.

After quality assessment, individual chains were combined to form a homodimeric 3D structure, as shown in Fig. 1C. To facilitate other researchers, the predicted 3D model has been submitted to the Protein Model Database (PMDB) [33], and anyone can download/use the SARS-CoV-2 3CL^{pro} final structure using PMDB ID: PM0082635. Further, mutational analyses depicted none of the mutations affected the overall structure of SARS-CoV-2, which fully superimposed on the SARS-CoV 3CL^{pro} structure (Fig. 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) [34]. 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}.

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 [35], 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 ($S = -12.27$) was 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 (Fig. S3). These results indicated that the 12-point mutations identified at previous step may disrupt important hydrogen bonds and alter the receptor binding site, thereby affecting its ability to bind with the SARS-CoV inhibitor.

Therefore, it was essential to discover novel compounds that may inhibit SARS-CoV-2 3CL^{pro} and serve as potential anti-COVID-19 drug compounds. We developed a library from our previously published studies that contains numerous natural compounds possessing potential anti-viral activities and screened it against the SARS-CoV-2 3CL^{pro} homology model. Recent drug repurposing studies proposed few 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 [36, 37] and docked them as controls in the present study (Fig. S4). Our analyses identified nine novel non-toxic,

druggable natural compounds that are predicted to bind with the receptor binding site and catalytic dyad (Cys-145 and His-41) of SARS-CoV-2 3CL^{pro} (Table 2; Fig. S5). ADMET profiling of the selected hits is given in Table S2. Among these screened phytochemicals, 5,7,3',4'-tetrahydroxy-2'-(3,3-dimethylallyl) isoflavone, is an isoflavone extracted from *Psoralea argyrea* [38], that 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 significant interactions with the receptor-binding residues Thr24, Thr25, Thr26, Cys44, Thr45, Ser46, Met49, Asn142, Gly143, His164, Glu166 and Gln189 (Fig. 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 [38], and it is also found in traditional Chinese medicine records [39]. 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 suggested that natural products identified in our study may prove more useful candidates for COVID-19 drug therapy.

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 50 ns MD simulation. The root mean square deviation (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 (Fig. 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 (Fig. 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 (Fig. 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 (Fig. 2D). These results confirmed our findings, and further indicated that these compounds may serve as potential anti-COVID-19 drug sources.

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 had some key differences, which highlighted the need for rapid drug discovery to address the alarming COVID-19 pandemic. 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-COVID-19 therapeutic agents in the future.

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), the Starting Research Grant for High-level Talents from Guangxi University and Postdoctoral research platform grant of Guangxi University. We also acknowledge all authors and laboratories mentioned in Table S1 for their sampling, analysis, and genome sequencing efforts. In addition, we acknowledge GISAID (<https://www.gisaid.org/>) for facilitating open data sharing.

Supplementary data

Fig. S1. 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).

Fig. S2. (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.

Fig. S3. (A) 2D representation of ML188 binding mode with receptor binding site of SARS-CoV-2 3CL^{pro}. (B) 2D representation of ML188 binding mode with receptor binding site of SARS-CoV-2 3CL^{pro}.

Fig. S4. (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}.

Fig. S5. 2D representation of (A) 5,7,3',4'-Tetrahydroxy-2'-(3,3-dimethylallyl) isoflavone, (B) Myricitrin, (C) Methyl rosmarinic acid, (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}.

Table S1. Acknowledgement to the authors and laboratories, sampling, analysing and submitting the genome sequences to GISAID database.

Table S2. ADMET profiling enlisting absorption, metabolism and toxicity related drug like parameters of all nine selected phytochemicals.

References

- [1] X. Xu, P. Chen, J. Wang, et al., 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.* 63 (2020) 457-460.
- [2] W. Ji, W. Wang, X. Zhao, et al., Cross-species transmission of the newly identified coronavirus 2019-nCoV, *J. Med. Virol.* 92 (2020) 433-440.
- [3] N. Zhu, D. Zhang, W. Wang, et al., A novel coronavirus from patients with pneumonia in China, 2019, *N. Engl. J. Med.* 382 (2020) 727-733.
- [4] Y. Shu, J.J.E. McCauley, GISAID: Global initiative on sharing all influenza data—from vision to reality, *Euro Surveill.* 22 (2017), doi: 10.2807/1560-7917.ES.2017.22.13.30494.
- [5] Y. Chen, Q. Liu, D. Guo, Emerging coronaviruses: Genome structure, replication, and pathogenesis, *J. Med. Virol.* 92 (2020) 418-423.

- 266 [6] Z.-L. Shi, P. Zhou, X.-L. Yang, et al., Discovery of a novel coronavirus associated with the
 267 recent pneumonia outbreak in humans and its potential bat origin, *bioRxiv* (2020), doi:
 268 10.1101/2020.01.22.914952.
- 269 [7] F. Wu, S. Zhao, B. Yu, et al., A new coronavirus associated with human respiratory disease
 270 in China, *Nature* 579 (2020) 265-269.
- 271 [8] K. Anand, J. Ziebuhr, P. Wadhwani, et al., Coronavirus main proteinase (3CLpro) structure:
 272 basis for design of anti-SARS drugs, *Science* 300 (2003) 1763-1767.
- 273 [9] D. Needle, G.T. Lountos, D.S. Waugh, Structures of the middle east respiratory syndrome
 274 coronavirus 3C-like protease reveal insights into substrate specificity, *Acta Crystallogr. D Biol.*
 275 *Crystallogr.* 71 (2015) 1102-1111.
- 276 [10] A.K. Ghosh, K. Xi, K. Ratia, et al., Design and synthesis of peptidomimetic severe acute
 277 respiratory syndrome chymotrypsin-like protease inhibitors, *J. Med. Chem.* 48 (2005) 6767-
 278 6771.
- 279 [11] V. Kumar, K.P. Tan, Y.M. Wang, et al., Identification, synthesis and evaluation of SARS-
 280 CoV and MERS-CoV 3C-like protease inhibitors, *Bioorg. Med. Chem.* 24 (2016) 3035-3042.
- 281 [12] T. Pillaiyar, M. Manickam, V. Namasivayam, et al., An overview of severe acute respiratory
 282 syndrome–coronavirus (SARS-CoV) 3CL protease inhibitors: peptidomimetics and small
 283 molecule chemotherapy, *J. Med. Chem.* 59 (2016) 6595-6628.
- 284 [13] M. Tahir ul Qamar, A. Maryam, I. Muneer, et al., Computational screening of medicinal
 285 plant phytochemicals to discover potent pan-serotype inhibitors against dengue virus, *Sci. Rep.* 9
 286 (2019) 1-16.
- 287 [14] E. Gasteiger, A. Gattiker, C. Hoogland, et al., ExPASy: The proteomics server for in-depth
 288 protein knowledge and analysis, *Nucleic Acids Res.* 31 (2003) 3784-3788.
- 289 [15] C. Notredame, D.G. Higgins, J. Heringa, T-Coffee: A novel method for fast and accurate
 290 multiple sequence alignment, *J. Mol. Biol.* 302 (2000) 205-217.
- 291 [16] P. Gouet, E. Courcelle, D.I. Stuart, et al., ESPript: analysis of multiple sequence alignments
 292 in PostScript, *Bioinformatics* 15 (1999) 305-308.
- 293 [17] N. Eswar, B. Webb, M.A. MartiñRenom, et al., Comparative protein structure modeling
 294 using Modeller, *Curr. Protoc. Bioinformatics* 15 (2006), doi: 10.1002/0471250953.bi0506s15.
- 295 [18] M. Johnson, I. Zaretskaya, Y. Raytselis, et al., NCBI BLAST: a better web interface,
 296 *Nucleic Acids Res.* 36 (2008), doi: 10.1093/nar/gkn201.
- 297 [19] E.F. Pettersen, T.D. Goddard, C.C. Huang, et al., UCSF Chimera—a visualization system
 298 for exploratory research and analysis, *J. Comput. Chem.* 25 (2004) 1605-1612.

- 299 [20] W. DeLano, Pymol: An open-source molecular graphics tool, CCP4 Newsletter on protein
300 crystallography 40 (2002) 82-92.
- 301 [21] A. Mumtaz, U.A. Ashfaq, M. Tahir Ul Qamar, et al., MPD3: a useful medicinal plants
302 database for drug designing, Nat. Prod. Res. 31 (2017) 1228-1236.
- 303 [22] U.A. Ashfaq, A. Mumtaz, M. Tahir Ul Qamar, et al., MAPS Database: medicinal plant
304 activities, phytochemical and structural database, Bioinformation 9 (2013) 993-995.
- 305 [23] A. Mumtaz, U.A. Ashfaq, M. Tahir ul Qamar, et al., Addendum, Nat. Prod. Res. (2020),
306 doi: 110.1080/14786419.2020.1735129.
- 307 [24] S. Vilar, G. Cozza, S. Moro, Medicinal chemistry and the molecular operating environment
308 (MOE): application of QSAR and molecular docking to drug discovery, Curr. Top. Med. Chem.
309 8 (2008) 1555-1572.
- 310 [25] M. Tahir Ul Qamar, S. Saleem, U.A. Ashfaq, et al., Epitope-based peptide vaccine design
311 and target site depiction against middle east respiratory syndrome coronavirus: an immune-
312 informatics study, J. Transl. Med. 17 (2019) doi: 10.1186/s12967-019-2116-8.
- 313 [26] M. Tahir Ul Qamar, A. Bari, M.M. Adeel, et al., Peptide vaccine against chikungunya virus:
314 immuno-informatics combined with molecular docking approach, J. Transl. Med. 16 (2018), doi:
315 doi: 10.1186/s12967-018-1672-7.
- 316 [27] F. Cheng, W. Li, Y. Zhou, et al., admetSAR: a comprehensive source and free tool for
317 assessment of chemical ADMET properties, J. Chem. Inf. Model 52 (2012) 3099-3105.
- 318 [28] H.J.C. Berendsen, D. van der Spoel, R. van Drunen, GROMACS: a message-passing
319 parallel molecular dynamics implementation, Comput. Phys. Commun. 91 (1995) 43-56.
- 320 [29] D.M. van Aalten, R. Bywater, J.B. Findlay, et al., PRODRG, a program for generating
321 molecular topologies and unique molecular descriptors from coordinates of small molecules, J.
322 Comput. Aided Mol. Des. 10 (1996) 255-262.
- 323 [30] I. Muneer, M. Tahir Ul Qamar, K. Tusleem, et al., Discovery of selective inhibitors for
324 cyclic AMP response element-binding protein: a combined ligand and structure-based resources
325 pipeline, Anticancer Drugs 30 (2019) 363-373.
- 326 [31] P. Zhou, X.L. Yang, X.G. Wang, et al., A pneumonia outbreak associated with a new
327 coronavirus of probable bat origin, Nature 579 (2020) 270-273.
- 328 [32] J. Ye, S. McGinnis, T.L. Madden, BLAST: improvements for better sequence analysis,
329 Nucleic Acids Res. 34 (2006) doi: 10.1093/nar/gkl164.
- 330 [33] T. Castrignano, P.D. De Meo, D. Cozzetto, et al., The PMDB Protein Model Database,
331 Nucleic Acids Res. 34 (2006) doi: 10.1093/nar/gkj105.

- [34] H. Yang, M. Yang, Y. Ding, et al., The crystal structures of severe acute respiratory syndrome virus main protease and its complex with an inhibitor, *Proc. Natl. Acad. Sci USA* 100 (2003) 13190-13195.
- [35] J. Jacobs, S. Zhou, E. Dawson, et al., Discovery of non-covalent inhibitors of the SARS main proteinase 3CL^{pro}, *Probe Reports from the NIH Molecular Libraries Program*, National Center for Biotechnology Information (US), Bethesda (MD), (2010), PMID: 23658941.
- [36] Z. Xu, C. Peng, Y. Shi, et al., 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), doi: 10.1101/2020.01.27.921627.
- [37] Y. Li, J. Zhang, N. Wang, et al., Therapeutic drugs targeting 2019-nCoV main protease by high-throughput screening, *bioRxiv* (2020), doi: 10.1101/2020.01.28.922922.
- [38] M.M. Salem, K.A. Werbovetz, Isoflavonoids and other compounds from *psoralea* with antiprotozoal activities, *J. Nat. Prod.* 69 (2006) 43-49.
- [39] J. Zhou, G. Xie, X. Yan, *Encyclopedia of traditional Chinese medicines*, Springer 1 (2011), doi: 10.1007/978-3-642-16744-7_1.

Figure legends

Fig. 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 $\geq 70\%$ sequence identity. (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. (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

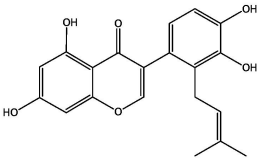
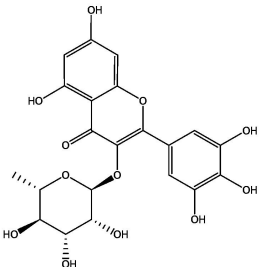
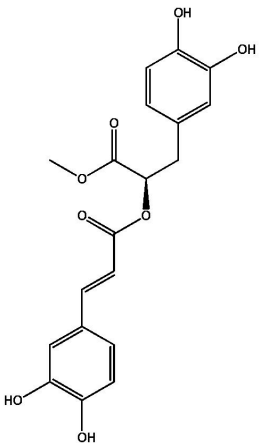
are highlighted in red. (E) Docking of 5,7,3',4'-tetrahydroxy-2'-(3,3-dimethylallyl) isoflavone inside the receptor-binding site of SARS-CoV-2 3CL^{pro}, showing hydrogen bonds with the catalytic dyad (Cys-145 and His-41). The 3CL^{pro} structure is coloured dark blue, the 5,7,3',4'-tetrahydroxy-2'-(3,3-dimethylallyl) isoflavone is orange, and hydrogen coloured maroon.

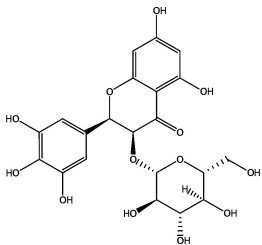
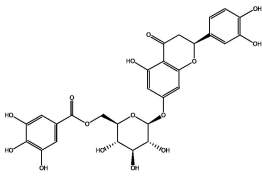
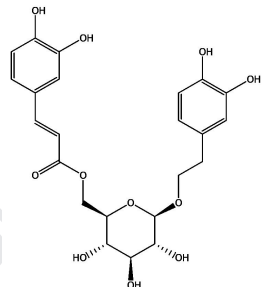
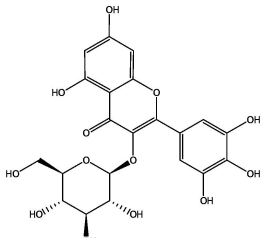
Fig. 2. (A) Root mean square deviation (RMSD), (B) root mean square fluctuation (RMSF), (C) potential energy and (D) Hydrogen Bond interactions for all three complexes over the 50 ns simulation.

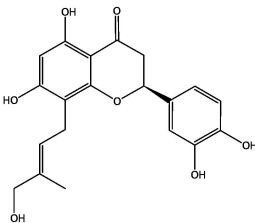
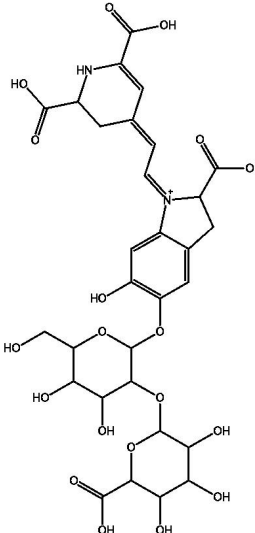
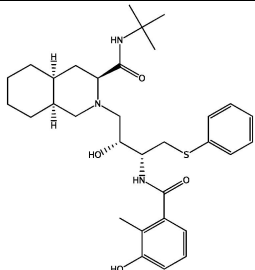
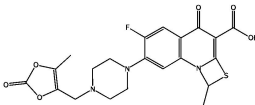
Table 1. Physicochemical parameters of SARS-CoV-2 3CL^{pro}

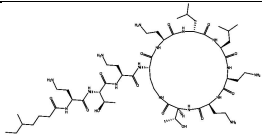
Parameters	SARS-CoV-2 3CL ^{pro}
Mol. Weight	33796.64 Dalton
No. of amino acids	306
Theoretical <i>pI</i>	5.95
Instability index (II)	27.65 (stable)
No. of Negatively Charged Residues (Asp + Glu)	26
No. of Positively Charged Residues (Arg + Lys)	22
Aliphatic Index	82.12
Grand average of Hydropathicity (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%)

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

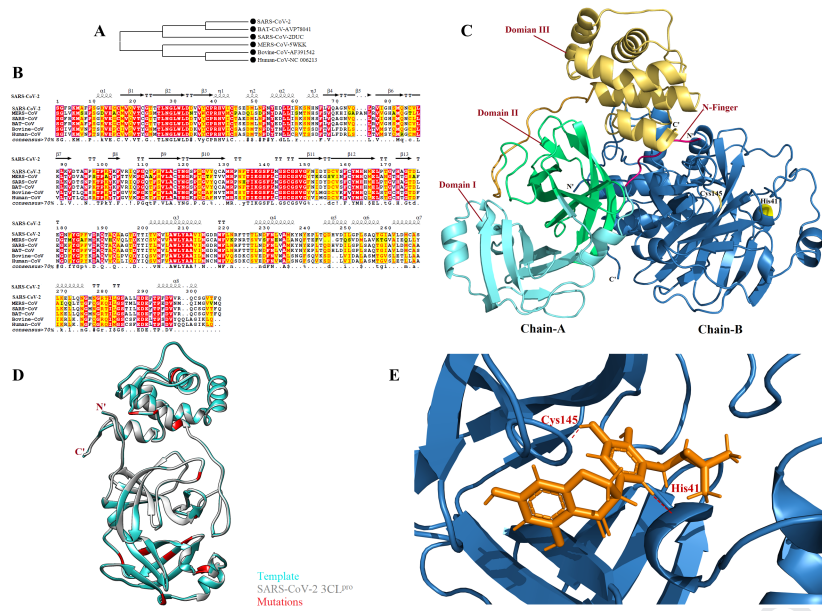
IDs	Phytochemical name	Plant source	Phytochemical structure	Docking score	Binding affinity (kcal/mol)	3CL ^{pro} residues interacting with phytochemical through H-bonding and other interactions
PubChem1610052	5,7,3',4'-Tetrahydroxy-2'-(3,3-dimethylallyl) isoflavone	<i>Psoralea arborescens</i>		-16.35	-29.57	His41 , Cys145 , Thr24, Thr25, Thr26, Cys44, Thr45, Ser46, Met49, Asn142, Gly143, His164, Glu166, Gln189
PubChem5281673	Myricitrin	<i>Myrica cerifera</i>		-15.64	-22.13	Gly143, Cys145 , His41 , Thr24, Thr25, Thr26, Leu27, Cys44, Ser46, Met49, Leu141, Asn142, Ser144, His163, Glu166, Gln189
PubChem6479915	Methyl rosmarinate	<i>Hyptis atrorubens</i> Poit		-15.44	-20.62	Cys145 , His41 , Thr24, Thr25, Thr26, Cys44, Thr45, Met49, Leu141, Asn142, Gly143, Ser144, His163, His164, Glu166, Gln189

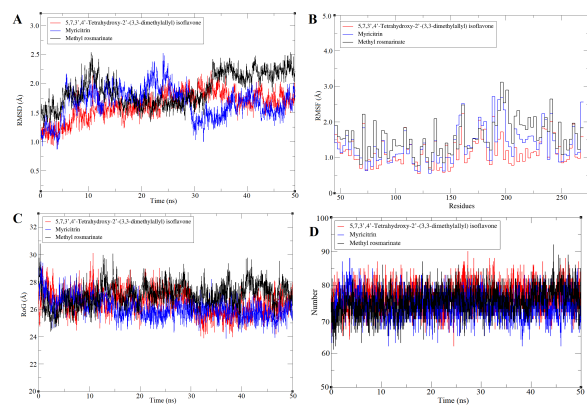
NPACT00105	3,5,7,3',4',5'-hexahydroxyflavanone-3-O-beta-D-glucopyranoside	<i>Phaseolus vulgaris</i>		-14.42	-19.10	Met49, Cys145 , His41 , Thr24, Thr25, Thr26, Cys44, Ser46, Asn142, His164, Met165, Glu166, Gln189
PubChem 10930068	(2S)-Eriodictyol 7-O-(6''-O-galloyl)-beta-D-glucopyranoside	<i>Phyllanthus emblica</i>		-14.41	-19.47	Thr24, Thr25, Gly143, Met49, Cys145 , His41 , Thr26, Cys44, Thr45, Glu166, Leu167, Gln189, Thr190, Ala191, Gln192
PubChem 5273567	Calceolarioside B	<i>Fraxinus sieboldiana</i>		-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	<i>Camellia sinensis</i>		-13.70	-18.42	Asn142, Glu166, Cys145 , His41 , Thr24, Thr25, Thr26, Thr45, Ser46, Met49, Leu141, Gly143, Ser144, His163, His164, Met165, Gln189

PubChem 11111496	Licoleafol	<i>Glycyrrhiza uralensis</i>		-13.63	-19.64	Cys145, His41, Thr24, Thr25, Thr26, Cys44, Thr45, Met49, Gly143, His163, His163, Met165, Glu166, Gln189
PubChem 6123095	Amaranthin	<i>Amaranthus tricolor</i>		-12.67	-18.14	Thr26, Glu166, Cys145, His41, Thr24, Thr25, Cys44, Cys44, Thr45, Asn142, His163, His164, Met165, Glu166, Leu167, Pro168, Gln189
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
--------------------	----------	--	--------	--------	--

*3CL^{pro} catalytic dyad (His-41 and Cys-145) residues are highlighted with bold font.





Highlights

1. SARS-CoV-2 3CL^{pro} is conserved, share 99.02% sequence identity with SARS-CoV 3CL^{pro} and together with 12 point-mutations.
2. Mutations disrupt important hydrogen bonds and alter the receptor binding site of SARS-CoV-2 3CL^{pro}.
3. Medicinal plants phytochemicals were proved potential anti-COVID-19 druggable candidates.