In Silico high-throughput virtual screening and molecular dynamics simulation study to identify 3CLpro inhibitor of COVID-19

Amreesh Parvez

School of Food and Biological Engineering, Jiangsu University, 301 Xuefu Road, Zhenjiang 212013, Jiangsu, China

Sunita Yadav

Indian Agricultural Statistics Research Institute, Library Ave, Pusa, New Delhi, India

Vivek Pandey

Faculty of Life sciences and biotechnology, South Asian University, Akbar Bhawan, Chanakyapuri, New Delhi, India

Xianghui Qi (✉ qxh@ujs.edu.cn)

School of Food and Biological Engineering, Jiangsu University, 301 Xuefu Road, Zhenjiang 212013, Jiangsu, China

Research Article

Keywords: 3CLpro, SARS-CoV, MERS-CoV, inhibitor library, molecular docking

Posted Date: May 18th, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-29546/v1>

License: This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

***In Silico* high-throughput virtual screening and molecular dynamics simulation study to identify 3CLpro inhibitor of COVID-19**

Amreesh Parvez^{1, +}, Sunita Yadav^{2, +}, Vivek Pandey³, Xianghui Qi^{1, *}

¹ School of Food and Biological Engineering, Jiangsu University, 301 Xuefu Road, Zhenjiang 212013, Jiangsu, China

² Indian Agricultural Statistics Research Institute, Library Ave, Pusa, New Delhi, India

³ Faculty of Life sciences and biotechnology, South Asian University, Akbar Bhawan, Chanakyapuri, New Delhi, India

* corresponding. qxh@ujs.edu.cn

+these authors contributed equally to this work

ABSTRACT

COVID-19 is an ongoing coronavirus pandemic with more than 260,000 deaths worldwide so far. In the event of no targeted drug and vaccine for SARS-CoV-2, causative agent of COVID-19, finding newer targets becomes indispensable. 3CLpro is a processing enzyme of viral replicase-transcriptase complex, chosen as a potential drug target. 3CLpro of COVID-19 is structurally similar to SARS-CoV and MERS-CoV 3CLpro. Here, we built up an in-house inhibitor library for 3CLpro of SARS-CoV and MERS-CoV. The library revealed 7 non-toxic compounds with 5 compounds showing significant protease and enzyme inhibition activity. Molecular docking of these compounds with 3CLpro shown lower binding energy profile of compounds 1 and 2. The compactness and flexibility analysis of compounds 1 and 2 complexes with 3CLpro showed lower stability profile as compared to unbound 3CLpro. The results cite the potential of compounds 1 and 2 as probable inhibitors of 3CLpro as a prospective therapeutic target for COVID-19.

Introduction

World is suffering from another very serious outbreak of coronavirus within a period of two decades. The 2019 coronavirus outbreak has been named as COVID-19 by the WHO to demark it from the previous Severe Acute Respiratory Syndrome (SARS) outbreak and is declared as pandemic owing to its spread across the continents. As of 8th may, around 3,800,000 cases of infection have been reported with death toll climbing to more than 260,000 worldwide (<https://coronavirus.jhu.edu/map.html>). The recent outbreak has presented a serious challenge to global health care system for handling any such outbreaks. The scientific community worldwide is working day night to develop the preventive vaccine and drugs to contain the outbreak at the earliest and save people at risk. At present, National Health Commission of the People's Republic of china has recommended ritonavir and lopinavir both are HIV-1 protease inhibitors as well as chloroquine phosphate for the COVID-19 treatment. Remdesivir, an antiviral drug against the RNA virus including Ebola¹, SARS² and Middle East Respiratory Syndrome (MERS)² has potential to inhibit COVID-19 virus³. At present, some traditional Chinese medicines (Lianhuaqingwen and Shuanghuanglian oral liquid)^{4,5} are also tested which are found to be effectively eliminating the virus. COVID-19 is a (+) sense single-stranded enveloped RNA virus codes an ~30kb long genome as reported using different sequencing methods on the patient-derived virus samples^{6,7}. COVID-19 virus belongs to the order of Nidovirales, family Coronavirinae, subfamily Coronaviridae genus *Betacoronavirus*, and subgenus *Sarbecovirus*. Previous coronavirus outbreaks have been caused by the same genus of the virus which included SARS outbreak of 2002 in China and MERS outbreak in 2012 near the Arabian Peninsula. All these diseases are marked by respiratory illness ultimately resulting in fatal pneumonia. Phylogenetic studies conducted between the causative agents of all these three outbreaks suggest of closer association between COVID-19 to SARS-CoV (79% identity) in comparison to the MERS-CoV (50% identity)⁸. However, studies also show a very close association with two other strains of the SARS-like coronaviruses, bat-SL-CoVZC45 and bat-SL-CoVZXC21⁸.

Approximately 30kb long genome of coronavirus encodes for viral replicases, structural proteins, spike, nucleocapsid, membrane and envelope protein along with it the genome codes for several non-structural proteins (nsp) which are often described as the accessory proteins or essential for various other functions⁹. Viral spike protein plays an essential role in attaching to the host cell receptor, studies suggest that the virus uses Angiotensin-converting enzyme2 as a receptor to bind to the host cell and enter into the cell^{10,11}. Spike protein and nucleocapsid are derived epitope at B cell and T cell suggests, these protein play important role in vaccines development against the COVID-19¹². Most of the viral genome is occupied by replicases (~21kb) an overlapping polyprotein with coding varying by a ribosomal frameshift of -1 resulting in ORF1a and ORF1b which encode pp1a (486kDa) and pp1ab (790kDa) respectively¹³. Proteolytic processing of this polyprotein results in release of a number of functional protein required by virus life cycle. This proteolytic cleavage is mainly accompanied by two proteases that are papain-like cysteine protease (PL_{PRO}) and 3chymotrypsin-like protease (3CL_{PRO})¹³. These two proteins upon proteolytic processing are responsible for the formation of the multi-subunit complex known as “viral replicase-transcriptase” required for viral replication, hence making these proteases indispensable for replication. Proteases thereby play an essential role in the viral life cycle at various stages, thus making these proteases an important molecule for study and developing therapeutics target. Recently, PL_{PRO} and 3CL_{PRO} of COVID-19 have been used as the therapeutics target by *in silico* as well as *in vitro* drugs screen with the FDA approved drugs¹⁴⁻¹⁷. In this study, we attempt a virtual screening of 3CLpro inhibitor as it being the main protease of COVID-19. The inhibitor library was based on SARS and MERS 3CLpro inhibitor. The toxicity and drug durability of inhibitors were assigned by Osiris and Molinspiration. Further, interaction of inhibitor with 3CLpro were analysis by molecular docking which followed by analysis of stability of 3CLpro with inhibitors by molecular simulation dynamics.

Results

Phylogenetic analysis

The new outbreak of coronavirus (COVID-19) is the highly infectious and causative agent of respiratory diseases like the SARS. COVID-19 is positive-stranded RNA viruses that contain ~30kb long genome encodes 10 open reading frames. Almost two-thirds genome of COVID-19 encodes two large ORFs (1a and 1b) viral replicase polyproteins¹³. These two polyproteins are extensively cleaved by two viral proteases into 16 mature proteins¹³. These two proteases namely, 3CL_{PRO} and PL_{PRO} are potential drug targets for the coronavirus. In respect of SARS, it is predicted that PLP and 3CLpro cleave the polyprotein at four and eleven places respectively which is predicted the same function in COVID-19. To determine the evolutionary relationship of the COVID-19 and previously discovered CoVs, we construct the phylogenetic tree based on the amino acid sequence of 3CLpro, figure 2. Our phylogenetic tree analysis revealed, 3CLpro of COVID-19 was clustered with the SARS more closely than the MERS. Although, 3CLpro from other coronavirus was distinctly related to COVID-19 and clustered in the different clades. The most striking observations were that 3CLpro of COVID-19 and SARS originated by closely related ancestors. Structural comparison of 3CLpro of COVID-19, SARS (PDB ID: 3VB3)¹⁸, and MERS (PDB ID: 5C3N)¹⁹ revealed the perfect superimposition with each other. Analysis of superimposed structure shown similar structural fold, figure 3A. The multiple sequence alignment of this protein presented 3CLpro of COVID-19 has 96% and 56% similarity with 3CLpro of SARS and MERS as shown in figure 3B. The similarity of structural and sequence of this protein gives the possibility the protease inhibitor of SARS and MERS would inhibit 3CLpro.

Molinspiration calculations:

Druglikeness and molecular properties of compounds were investigated using the Molinspiration server which is based on Lipinski rules of five. The Lipinski rules of five screens potential oral drug accordingly, the potential drug must have following properties (a) logP ≤5, (b) molecular weight ≤ 500 Dalton, (c) the number of H-bond acceptors ≤ 10, (d) number of H-bond donor ≤ 5²⁰. Molecular properties of the selected compounds were tested by Molinspiration properties as shown in table 1. For the further screening the compounds that violated more than one Lipinski rules of five were rejected. All selected compounds have a molecular weight less than 500 except compound 2. Although, high molecular weight compounds are difficult to transported and absorbed in respect to low molecular weight compounds and it also affects the action of the drug. We observed that the number of H-

bond donors as well as H-bond acceptors in these selected compounds were follow the Lipinski's rules which were less than 5 and 10 respectively.

Lipophilicity (log P) and Topological Polar Surface area (TPSA) are essential properties of the potential drug molecules, which defines the oral bioavailability of the compounds. Penetration of drug molecule through biomembranes require less than 5 logP value, the compound shown in table 1 have acceptable logP value. The higher lipophilicity value of compound helps the compound to interact with the membrane. TPSA was estimated using consideration of O- and N- polar fragments and the related hydrogen atom. It depends on the hydrogen bond-forming potential of the compounds. It is also a defined percentage of absorption using % absorption = $109 - 0.345 \times \text{TPSA}$. Compound shows in table 1 all have $\text{TPSA} \leq 140 \text{ \AA}^2$ which showed higher oral bioavailability. Another important factor is the number of rotatable bond ≤ 10 , present in all selected compound. A higher rotatable bond increases the flexibility of the compound for the interaction with protein and provides good bioavailability.

Bioactivity of all selected compounds were analyzed against, GPC (GPCR ligand), ICM (ion channel modulator), KI (Kinase inhibition), NRL (Nuclear receptor ligand), PI (Protease inhibitors), and EI (Enzyme inhibitors) showed in table 2. The compounds have more than 0.00 score are biologically high active, -0.50 to 0.00 score are moderately active and less than -0.50 score represent inactive compounds (r). Bioactivity score is summarized in table 2, in which compound 7 and 10 were biologically inactive for all proteins while compound 8 was inactive for ICM and compound 12 was active for only NRL. Compound 9 was biologically inactive for ICM and NRL proteins. Compound 1 to 5 showed positive bioactive score for all proteins except KI in which compound 1 only showed a positive score for KI other compounds shows more than -0.50 score. Compound 1 to 6 showed good bioactive scores represents that these molecules were effective against GPC, ICM, KI, NRL, PI, and EI.

Toxicology studies

The toxicity risks, drug-likeness (DL), drug score (DS), and Physico-chemical property were estimated by OSIRIS property explorer shown in table 3. Toxicity is predicted based on the screening of fragments which probably cause toxicity. Prediction of toxicity risks were presented by different colour where red colour shows a high risk, yellow colour shows moderate risk, and green colour specifies the drug conform behavior. From table 3, compound 8-10 have a high risk of mutagenicity and tumorigenicity except 10 which also has a high risk of irritation. Further, compound 11 showed the mild risk of mutagenicity and tumorigenicity whereas compound 12 was mild irritants. Compound 1-7 were established by drug-like molecules. LogS represent aqueous solubility of the molecules. In which low solubility of the compound shows poor absorption. Most of the drugs available have logs > -4.00. LogS values of 1-7 were around -4, which shows good solubility and comparable standard drug zone.

The drug score is calculated by the combination of toxicity risks, drug-likeness, cLogP, LogS and molecular weight which provide a single value for the potential drug candidate. This drug score is calculated by the following equations:

$$\text{DS} = \pi(1/2 + 1/2\text{Si})\pi\text{ti} \quad (1)$$

$$\text{Where Si} = (1 + s^{ap+b})^{-1} \quad (2)$$

Si is the contribution of cLogP, LogS, molecular weight, and drug-likeness calculated from Eq. (2), which is a spline curve. Parameters 'a' and 'b' are (1, -5), (1, 5), (0.012, -6) and (1, 0) for cLogP, LogS, molecular weight, and drug-likeness respectively. On the other hand, 'ti' represent the contribution of toxicity risk types, which have 1.0, 0.8, and 0.6 values for no risk, moderate and high risk respectively. The potential drug candidate has a positive drug score and indicates a molecule contains pharmacophoric group. All selected compounds exhibited positive drug score. The compounds which did not have toxicity risk only compound 1-3 showed moderate drug-likeness values which indicating these compounds are potential drug candidates.

Molecular Docking analysis:

The structural based virtual screenings of all the selected compounds were done using molecular docking which defines the confirmations of the ligand in 3Clpro protein. The grid was defined at the active site of the protein for the docking of the selected compounds. Most of the compounds have shown lower binding energy with the protein. Binding energy defines the affinity of the ligand and protein. Higher binding energy show weak interaction and lower binding energy show strong interactions. Almost all the compounds have shown less than -6 kcal/mol binding energy (Supplementary table 2). However, Drug likeness and Molinspiration analysis displayed only compound 1 and 2 that have drug-like properties. Docking of the compounds 1 and 2 are consistent because these compounds shows the interaction with the active site residues of the catalytic pocket of the 3Clpro protein as

shown in figure 4. However, compound 2 exhibited the lowest binding energy -7.8 kcal/mol with respect to compound 1 that have binding energy -7 kcal/mol. Docking of the compound 2 established hydrogen bond with Asn142, His164, Glu166, and Gln189 whereas compound 1 has hydrogen bond with His164 and Gln189 at the active site. Compounds 1 and 2 interact with several amino acid residues with hydrophobic as well as van der Waal bonds. The toxicity and docking study established both compound 1 and 2 have a drug like property and could be a potential inhibitor for 3CLpro protein.

3CLpro protein-ligand complex stability analysis:

The stability of protein and the protein-ligand complexes is done by using the molecular dynamics simulation to validate the structure in the context of compound 1 and 2. The RMSD values are monitored under this study by the side of several other factors like temperature, pressure, density, and potential energy. RMSD of the trajectories were calculated using only the backbone atoms. RMSD in MD simulation was calculated by Eq. (3).

$$\text{RMSD}(t_1, t_2) = [1/N \sum_{i=0}^n (r_i(t_1) - r_i(t_2))]^{1/2} \quad (3)$$

Where t_1 and t_2 represent particulate time point and time of the reference structure. r_i , i and N denoted for atom position, at the particular time, and the number of atoms. The RMSD plot shown in figure 5A, demonstrated that the stable conformations of the complexes and protein attend at 10ns and 5ns respectively. The final RMSD of the complexes were high with respect to unbound protein. Bound form of the Protein with the compounds have high movement during the simulation. The final RMSD of unbound protein was $1.5\text{\AA}$ whereas complex with compound 1 and 2 were $2.7\text{\AA}$ and $2.2\text{\AA}$ respectively.

To understand the flexibility of amino acid residues, RMSF of protein was calculated by eq, (4).

$$\text{RMSF} = \sqrt{3B/8\pi^2} \quad (4)$$

RMSF plot was shown in figure 5B which indicates complexes RMSF were higher than unbound protein. Higher RMSF values indicate higher flexible regions of the protein. The present study indicates that the protein in complexes having more flexible regions than the unbound proteins. Higher flexibility of complexes suggests that the ligands made backbone more flexible as compare to unbound protein.

Discussion

The recent COVID-19 is a serious threat for the whole world. Outbreaks in 2002 and 2012 SARS outbreak lead to several deaths and in 2020 COVID-19 has already resulted in more than 260,000 deaths world-wide till 8th April. Non-availability of any specific drug for treatment has put a great emphasis on the new drug discovery to target COVID-19. In our study, we have done the extensive virtual screening and molecular dynamic of 3CLpro protein. Only 7 compounds were satisfied with Lipinski's rule and toxicity study. However, out of 7, only 5 compounds have bioactivity with GPCR, NRL, KI, PI, and ICM. These compounds were docked with 3CLpro, where compound 1 and 2 had low binding energy with respect to the other compounds exhibited good interaction with protein. We carried out MD simulation of unbound 3CLpro and complex with compound 1 and 2. The RMSD and RMSF complexes of compound 1 and 2 with 3CLpro loss stability respect to the unbound 3CLpro. The flexibility of the 3CLpro complex with compound suggested compound 1 and 2 would be a useful inhibitor of 3CLpro which would be helpful drugs for the prevention of COVID-19.

Methods

Sequence retrieval and structure modeling

The sequence of 3CLpro of COVID-19 was retrieved from NCBI (<https://www.ncbi.nlm.nih.gov/gene/?term=wuhan+coronavirus>). The multiple sequence alignment of 3CLpro was done by ClustelOmega²¹ and analyze by BoxShade (https://embnet.vital-it.ch/software/BOX_form.html). 3CLpro sequence from several viruses were retrieved from NCBI and UniProt (<https://www.uniprot.org/>) (Supplementary Table1), which were used in phylogenetic tree construction using maximum like hood method with 1000 bootstraps using Mega²². Homology modeling is important techniques in case of structure-based drug designing applications, especially when a crystallographic structure is unavailable²³. Tertiary structure of 3CLpro was predicted by using the SWISS-MODEL²⁴ that is Comparative structure modeling technique by using (PDB id : 3VB3)

as a template. The model structure was validated by PROCHECK²⁵. The discovery studio visualizer was used for the visualization and superimposition of the 3CLpro structures.

Docking protocol and Drug scanning

The ligand library of protease inhibitors was created from PDB database from the structure of protease inhibitors of SARS and MERS (Supplementary Table2). Active site of the protein is predicted from the ProDRG server²⁶. Docking was performed by AutoDock 4.0 program²⁷, using the implemented empirical free energy function and the Lamarckian Genetic Algorithm (LGA)²⁸. The grid maps were calculated using Auto Grid. In all dockings, a grid map with 44× 44 × 44 points and grid-point spacing from the centre 12.053 x -0.797 x 19.586 was applied. The drug scans of the selected compounds were screened by Molinspiration (<https://www.molinspiration.com/cgi-bin/properties>) and Osiris Property Explorer. Molinspiration analysed Lipinski's Rule of Five²⁰ and drug-likeness such as LogP, molecular weight, protease, enzyme inhibition, and other property. Osiris Property Explorer analyses drug-likeness and toxicity of the compounds.

Molecular dynamics simulations

The structure of 3CLpro was solvated in a cubic box with TIP3P water molecules using AMBER99sb-ILDN. The systems were neutralized by the addition of counter ions (Na⁺ and Cl⁻). The Steepest Descents energy minimization was performed for 1000 steps. Then, the system was equilibrated by two steps: pressure and temperature equilibration for 1000ps. After the equilibration of systems, the 3CLpro was simulated for 20ns. Then, molecular dynamics simulation data were calculated and analyzed by Gromacs²⁹ and qtGrace³⁰ (<https://sourceforge.net/projects/qtgrace/>) in respect of RMSD (root-mean-square deviation) and RMSF (root-mean-square fluctuation). The methodological workflow is depicted in figure 1.

References

1. Mulangu, S., Dodd, L.E., Davey Jr, R.T., Tshiani Mbaya, O., Proschan, M., Mukadi, D., Lusakibanza Manzo, M., Nzolo, D., Tshomba Oloma, A., Ibanda, A. and Ali, R. A randomized, controlled trial of Ebola virus disease therapeutics. *N Engl J Med*, **381**, 2293-2303, DOI: [10.1056/NEJMoa1910993](https://doi.org/10.1056/NEJMoa1910993) (2019).
2. Sheahan, T.P., Sims, A.C., Graham, R.L., Menachery, V.D., Gralinski, L.E., Case, J.B., Leist, S.R., Pyrc, K., Feng, J.Y., Trantcheva, I. and Bannister, R., Broad-spectrum antiviral GS-5734 inhibits both epidemic and zoonotic coronaviruses. *Sci Trans Med*, **9**, DOI: [10.1126/scitranslmed.aal3653](https://doi.org/10.1126/scitranslmed.aal3653) (2017).
3. Wang, M., Cao, R., Zhang, L., Yang, X., Liu, J., Xu, M., Shi, Z., Hu, Z., Zhong, W. and Xiao, G. Remdesivir and chloroquine effectively inhibit the recently emerged novel coronavirus (2019-nCoV) in vitro. *Cell Res*. **30**, 269-271, DOI: <https://doi.org/10.1038/s41422-020-0282-0> (2020).
4. Ni, L., Zhou, L., Zhou, M., Zhao, J. & Wang, D. W. J. F. o. M. Combination of western medicine and Chinese traditional patent medicine in treating a family case of COVID-19 in Wuhan. *Front. Med* 1-5, DOI: <https://doi.org/10.1007/s11684-020-0757-x> (2020).
5. Runfeng, L., Yunlong, H., Jicheng, H., Weiqi, P., Qin Hai, M., Yongxia, S., Chufang, L., Jin, Z., Zhenhua, J., Haiming, J. and Kui, Z. Lianhuaqingwen exerts anti-viral and anti-inflammatory activity against novel coronavirus (SARS-CoV-2). *Pharmacol Res*104761, DOI: <https://doi.org/10.1016/j.phrs.2020.104761> (2020).
6. Wu, F., Zhao, S., Yu, B., Chen, Y.M., Wang, W., Song, Z.G., Hu, Y., Tao, Z.W., Tian, J.H., Pei, Y.Y. and Yuan, M.L. A new coronavirus associated with human respiratory disease in China. *Nature* **579**, 265-269, DOI: <https://doi.org/10.1038/s41586-020-2008-3> (2020).
7. Zhou, P., Yang, X.L., Wang, X.G., Hu, B., Zhang, L., Zhang, W., Si, H.R., Zhu, Y., Li, B., Huang, C.L. and Chen, H.D. A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature* **579**, 270-273 DOI : <https://doi.org/10.1038/s41586-020-2012-7> (2020).

8. Lu, R., Zhao, X., Li, J., Niu, P., Yang, B., Wu, H., Wang, W., Song, H., Huang, B., Zhu, N. and Bi, Y. Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding. *Lancet* 395, 565-574, DOI: [https://doi.org/10.1016/S0140-6736\(20\)30251-8](https://doi.org/10.1016/S0140-6736(20)30251-8) (2020).
9. Tan, Y.-J., Lim, S. G. & Hong, W. J. A. r. Characterization of viral proteins encoded by the SARS-coronavirus genome. *Antiviral Res* 65, 69-78, DOI: <https://doi.org/10.1016/j.antiviral.2004.10.001> (2005).
10. Li, W., Moore, M.J., Vasilieva, N., Sui, J., Wong, S.K., Berne, M.A., Somasundaran, M., Sullivan, J.L., Luzuriaga, K., Greenough, T.C. and Choe, H. Angiotensin-converting enzyme 2 is a functional receptor for the SARS coronavirus. *Nature* 426, 450-454, DOI: [10.1038/nature02145](https://doi.org/10.1038/nature02145) (2003).
11. Liu, Z., Xiao, X., Wei, X., Li, J., Yang, J., Tan, H., Zhu, J., Zhang, Q., Wu, J. and Liu, L. Composition and divergence of coronavirus spike proteins and host ACE2 receptors predict potential intermediate hosts of SARS-CoV-2. *J Med Virol* 92, no. 6: 595-601, DOI: [10.1002/jmv.25726](https://doi.org/10.1002/jmv.25726) (2020).
12. Ahmed, S. F., Quadeer, A. A. & McKay, M. R. J. V. Preliminary identification of potential vaccine targets for the COVID-19 coronavirus (SARS-CoV-2) based on SARS-CoV immunological studies. *Viruses*. 12, 254, DOI: <https://doi.org/10.3390/v12030254> (2020).
13. Snijder, E.J., Bredenbeek, P.J., Dobbe, J.C., Thiel, V., Ziebuhr, J., Poon, L.L., Guan, Y., Rozanov, M., Spaan, W.J. and Gorbalenya, A.E. Unique and conserved features of genome and proteome of SARS-coronavirus, an early split-off from the coronavirus group 2 lineage. *J mol biol* 331(5), 991-1004, DOI: [https://doi.org/10.1016/S0022-2836\(03\)00865-9](https://doi.org/10.1016/S0022-2836(03)00865-9) (2003).
14. Arya, R., Das, A., Prashar, V. & Kumar, M. Potential inhibitors against papain-like protease of novel coronavirus (SARS-CoV-2) from FDA approved drugs. *ChemRxiv*, DOI: doi.org/10.26434/chemrxiv.11860011.v1 (2020).
15. Liu, X. & Wang, X. J. Potential inhibitors against 2019-nCoV coronavirus M protease from clinically approved medicines. *J Genet Genomics* 47, 119-121, DOI: [10.1016/j.jgg.2020.02.001](https://doi.org/10.1016/j.jgg.2020.02.001) (2020).
16. Xu, Z., Peng, C., Shi, Y., Zhu, Z., Mu, K., Wang, X. and Zhu, W. 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*, DOI: <https://doi.org/10.1101/2020.01.27.921627> (2020).
17. ZZhavoronkov, A., Aladinskiy, V., Zhebrak, A., Zagribelnyy, B., Terentiev, V., Bezrukov, D.S., Polykovskiy, D., Shayakhmetov, R., Filimonov, A., Orekhov, P. and Yan, Y. Potential COVID-2019 3C-like Protease Inhibitors Designed Using Generative Deep Learning Approaches. *ChemRxiv*. v2, DOI: <https://doi.org/10.26434/chemrxiv.11829102> (2020).
18. Sievers, F., Wilm, A., Dineen, D., Gibson, T.J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Söding, J. and Thompson, J.D. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol syst biol* 7(1), DOI: <https://doi.org/10.1038/msb.2011.75> (2011).
19. Kumar, S., Stecher, G. and Tamura, K. MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol biol evol.* 33(7), 1870-1874, DOI: <https://doi.org/10.1093/molbev/msw054> (2016).
20. Sivasubramanian, A., Sircar, A., Chaudhury, S. and Gray, J.J. Toward high-resolution homology modeling of antibody Fv regions and application to antibody-antigen docking. *Proteins*. 74(2), 497-514, DOI: [10.1002/prot.22309](https://doi.org/10.1002/prot.22309) (2009).
21. Schwede, T., Kopp, J., Guex, N. & Peitsch, M. C. J. N. a. r. SWISS-MODEL: an automated protein homology-modeling server. *Nucleic Acids Res* 31, 3381-3385, DIO: <https://doi.org/10.1093/nar/gkg520> (2003).
22. Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. J. J. o. a. c. PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.* 26, 283-291, DOI: <https://doi.org/10.1107/S0021889892009944> (1993).

23. Medeiros, D. d. J., Cortopassi, W. A., Costa França, T. C. & Pimentel, A. S. J. J. o. C. ITP adjuster 1.0: A new utility program to adjust charges in the topology files generated by the PRODRG server. *J Chem.* **2013**, DOI: [10.1155/2013/803151](https://doi.org/10.1155/2013/803151) (2012).
24. Sharma, V., Pattanaik, K. K., Jayprakash, V., Basu, A. & Mishra, N. J. B. A utility script for automating and integrating AutoDock and other associated programs for virtual screening. *Bioinformation* **4**, 84, DOI: <https://doi.org/10.6026/97320630004084> (2009).
25. Allouche, A. R. J. J. o. c. c. Gabedit—A graphical user interface for computational chemistry softwares. *J Comput Chem.* **32**, 174-182, DOI: <https://doi.org/10.1002/jcc.21600> (2011).
26. Lipinski, C. A. "Lead-and drug-like compounds: the rule-of-five revolution." *Drug Discov. Today Technol.* **1**, 337-341, DOI: <https://doi.org/10.1016/j.ddtec.2004.11.007> (2004).
27. Van Der Spoel, D., Lindahl, E., Hess, B., Groenhof, G., Mark, A.E. and Berendsen, H.J. GROMACS: fast, flexible, and free. *J Comput Chem.* **26**, 1701-1718, DOI: [10.1002/jcc.20291](https://doi.org/10.1002/jcc.20291) (2005).
28. Turner, P.J. Grace-5.1. 22/qtGrace v 0.2. 4, 2018.
29. Chuck, C.P., Chen, C., Ke, Z., Wan, D.C.C., Chow, H.F. and Wong, K.B. "Design, synthesis and crystallographic analysis of nitrile-based broad-spectrum peptidomimetic inhibitors for coronavirus 3C-like proteases." *Eur J Med Chem* **59**, 1-6, DOI: <http://dx.doi.org/10.1016/j.ejmech> (2012).
30. Ho, B.L., Cheng, S.C., Shi, L., Wang, T.Y., Ho, K.I. and Chou, C.Y. Critical assessment of the important residues involved in the dimerization and catalysis of MERS coronavirus main protease. *PLoS One* **10**(12), DOI:[10.1371/journal.pone.0144865](https://doi.org/10.1371/journal.pone.0144865) (2015) ., 2015.

Acknowledgements

This work was supported by the National Natural Science Foundation of China (Grant Nos. 31972042 and 31950410550), China Postdoctoral Science Foundation (Grant Nos. 2019M661759, 2019T120402 and 2017M621657) and National Key R & D Program of China (Grant No. 2017YFC1600806). The authors are grateful to the Jiangsu University for providing financial assistance under postdoctoral fellowship program

Author contributions statement

AP and SY designed the study. AP performed all experiments. SY and AP drafted manuscript and analyzed the data. VP and XQ helped to revise the manuscript. XQ conceived and supervised the overall study. All authors have read, contributed to manuscript revision, and approved the submitted version. All authors reviewed the manuscript.

Additional information

Competing Interests: The authors declare no competing interests.

Figure 1: Schematic diagram of workflow

Figure 2: Phylogenetic analysis of 3C-like protease (3CL_{pro}). 3CL_{pro} from Covid-19 shown in red colour and the SARS and MERS are shown as green and violet colour respectively. The trees were generated with the maximum likelihood method with 1000 bootstraps and are based on almost complete sequences.

Figure 3: Structural and sequence compression of 3CLpro. (A) Structure of 3CLpro from COVID-19, SARS-CoV and MERS-CoV, (B) Multiple sequence alignment of different three protein sequences related to 3CLpro.

A

B

Figure 4: Molecular docking analysis of the 3CL_{pro} with compound 1 (A) and compound 2 (B), showing the confirmations of ligands and hydrogen bonds interaction with 3CL_{pro}.

A **B**

Figure 5: Molecular dynamics simulation analysis. (A) represent the RMSD plot and the (B) represent the RMSF of the protein with and without compound bound and unbound form shown in different colour as red, green and blue colour respectively.

Table 1: Lipinski parameters and physico-Chemical property calculation ^a.

Compound	miLogP	TPSA	Natoms	MW	nON	nOHNH	VIOL	ROTB	VOL
Compound 1	3.61	78.09	30	473.42	6	2	0	7	397.36
Compound 2	2.21	136.62	39	536.67	9	5	1	14	513.04

Compound 3	3.23	116.76	29	403.48	8	4	0	11	379.42
Compound 4	3.23	116.76	29	403.48	8	4	0	11	379.42
Compound 5	3.38	104.73	29	398.46	7	3	0	11	371.63
Compound 6	2.66	116.76	28	391.47	8	4	0	9	368.24
Compound 7	0.12	40.46	7	102.13	2	2	0	2	106.25
Compound 8	3.45	80.12	30	419.51	7	1	0	7	368.74
Compound 9	4.18	80.12	33	473.58	7	1	0	7	397.5
Compound 10	1.73	17.07	8	106.12	1	0	0	1	103.03
Compound 11	1.09	55.4	15	207.23	4	1	0	5	193.56
Compound 12	0.49	40.46	8	118.18	2	2	0	2	128.49

^a miLogP: N-octanol/water partition coefficient logarithm, TPSA: Topological polar surface area, , Natom: Number of atoms, MW: Molecular weight, nON: Number of hydrogen bond acceptors, nOHNH: Number of hydrogen bond donors, VIOL: Number of violations, ROTB: Number of rotatable bonds, VOL: Volume

Table 2: Drug likeness^a calculations of compounds

Compound	GPCRL	ICM	KI	NRL	PI	EI
Compound 1	0.56	0.3	0.16	0.5	0.87	0.39
Compound 2	0.35	-0.04	-0.1	-0.19	0.72	0.21
Compound 3	0.45	0.12	-0.04	0.14	0.74	0.35
Compound 4	0.45	0.12	-0.04	0.14	0.74	0.35
Compound 5	0.37	0.17	-0.13	0.06	0.89	0.28
Compound 6	0.5	0.32	-0.05	-0.1	1.11	0.42
Compound 7	-3.22	-2.9	-3.61	-2.75	-3.51	-2.33
Compound 8	-0.13	-0.57	-0.23	-0.43	-0.36	-0.22
Compound 9	-0.18	-0.59	-0.17	-0.51	-0.34	-0.24
Compound 10	-3.43	-2.75	-3.38	-3.13	-3.58	-2.99

Compound 11	0.01	0.16	-0.53	-0.41	0.35	0.16
Compound 12	-2.84	-2.25	-3.5	2.31	-2.72	-2.18

^a GPCRL: GPCR ligand, ICM: Ion channel modulator, KI: Kinase inhibitor, NRL: Nuclear receptor ligand, PI: Protease inhibitor, EI: Enzyme inhibitor

Table 3: Phsico-chemical Properties calculations of the selected compounds.

Compound	Toxicity Risks				Physicochemical Properties			
	MUT	TUMO	IRRI	RE	CLP	SOL	DL	DS
Compound 1	●	●	●	●	2.79	-4.22	-6.78	0.32
Compound 2	●	●	●	●	2	-4.32	1.2	0.5

Compound 3	●	●	●	●	1	-3.61	-29.92	0.39
Compound 4	●	●	●	●	1	-3.61	-29.92	0.39
Compound 5	●	●	●	●	1.86	-3.73	-29.5	0.39
Compound 6	●	●	●	●	0.39	-0.41	-2.92	0.52
Compound 7	●	●	●	●	0.88	-3.17	-55.19	0.41
Compound 8	●	●	●	●	3.06	-4.56	-1.72	0.13
Compound 9	●	●	●	●	3.91	-5.47	-1.02	0.11
Compound 10	●	●	●	●	1.59	-1.94	-4.04	0.11
Compound 11	●	●	●	●	1.08	-2.31	-29.38	0.3
Compound 12	●	●	●	●	0.68	-1.11	-5.87	0.39

● : Not Toxic, ● Moderate Toxic, ● Highly Toxic, MUT: Mutagenic, TUM: Tumourigenic, IRRI: Irritant, RE: Reproductive effect, CLP: cLogP, SOL: Solubility, DL: Drug likeness, DS: Drug score

2/3

3/3

Supplementary Table 1: Accession Number of 3CLpro protein of different Coronavirus used in phylogenetic tree construction.

S. No.	Accession No.	Coronavirus Strains
1	P0C6Y4	<i>Porcine epidemic diarrhea virus (strain CV777) (PEDV)</i>
2	P0C6W0	<i>Bat coronavirus 512/2005 (BtCoV) (BtCoV/512/2005)</i>
3	A0A220YLK8	<i>Myotis lucifugus coronavirus</i>
4	B1PHJ3	<i>Miniopterus bat coronavirus 1</i>
5	B1PHK1	<i>Miniopterus bat coronavirus HKU8</i>
6	P0C6X5	<i>Human coronavirus NL63 (HCoV-NL63)</i>
7	A0A0U2GML5	<i>Camel alphacoronavirus</i>
8	A8JNZ8	<i>Rhinolophus bat coronavirus HKU2</i>
9	A0A096XNJ3	<i>Lucheng Rn rat coronavirus</i>
10	D9J1Z2	<i>Mink coronavirus strain WD1127</i>
11	009PH5	<i>Feline coronavirus</i>
12	K7XSF6	<i>Canine coronavirus</i>
13	P0C6Y5	<i>Porcine transmissible gastroenteritis coronavirus (strain Purdue) (TGEV)</i>
14	V5TGG2	<i>Bottlenose dolphin coronavirus HKU22</i>
15	G3FER5	<i>Duck coronavirus</i>
16	A0A2Z4EVW7	<i>Sparrow deltacoronavirus</i>
17	A0A2Z5XXU9	<i>Quail coronavirus UAE-HKU30</i>
18	YP_009725301.1	Covid-19
19	P0C6X7	<i>Human SARS coronavirus (SARS-CoV) (Severe acute respiratory syndrome coronavirus)</i>
20	P0C6W2	<i>Bat coronavirus HKU3 (BtCoV) (SARS-like coronavirus HKU3)</i>
21	M1S8Z6	<i>Human betacoronavirus 2c England-Qatar/2012</i>
22	T2BAX6	<i>Middle East respiratory syndrome-related coronavirus</i>
23	P0C6W3	<i>Bat coronavirus HKU4 (BtCoV) (BtCoV/HKU4/2004)</i>
24	U5KNN1	<i>Betacoronavirus Erinaceus/VMC/DEU/2012</i>
25	A0A096XNG9	<i>Longquan RI rat coronavirus</i>
26	P0C6X2	<i>Human coronavirus HKU1 (isolate N1) (HCoV-HKU1)</i>
27	C6GHR8	<i>Rat coronavirus Parker</i>
28	H9AA28	<i>Rabbit coronavirus HKU14</i>
29	A0A2H4MY9	<i>Rodent coronavirus</i>
30	A8R4C0	<i>Equine coronavirus</i>
31	P0C6W9	<i>Bovine coronavirus (strain Mebus) (BCoV) (BCV)</i>
32	A0A309R4C4	<i>Yak coronavirus</i>

Supplementary Table 2: SMILES, IUPAC, and docking binding energy of the compounds used in virtual screening

Compound	SMILES of compounds	IUPAC name of compound	Docking binding energy
Compound1	<chem>BrC1ccc(cc1)C(=O)N2C[C@H]3CCCC[C@H]3C[C@H]2CN[C@@H](Cc4c[nH]cn4)C=O</chem>	(2S)-2-[[[(3S,4aR,8aS)-2-(4-bromobenzoyl)-3,4,4a,5,6,7,8,8a-octahydro-1H-isoquinolin-3-yl]methylamino]-3-(1H-imidazol-4-yl)propanal	-7.5
Compound2	<chem>CC(C)C[C@H](NC(=O)CCc1ccccc1)C(=O)N[C@@H](C[C@@H]2CCNC2=O)[C@@H](O)C(=O)NCc3cccc3</chem>	(2S)-N-[(1S,2R)-3-(benzylamino)-2-hydroxy-3-oxo-1-[[[(3S)-2-oxopyrrolidin-3-yl]methyl]propyl]-4-methyl-2-(3-phenylpropanoylamino)pentanamide	-7.8
Compound3	<chem>CC(C)C[C@H](NC(=O)OCc1ccccc1)C(=O)N[C@@H](CO)C[C@@H]2C=CNC2=O</chem>	benzyl N-[(1S)-1-[(1S)-1-(hydroxymethyl)-2-[(3R)-2-oxo-1,3-dihydropyrrol-3-yl]ethyl]carbamoyl]-3-methylbutyl]carbamate	-7.3
Compound4	<chem>CC(C)C[C@H](NC(=O)OCc1ccccc1)C(=O)N[C@@H](CO)C[C@@H]2C=CNC2=O</chem>	benzyl N-[(1S)-1-[(1S)-1-(hydroxymethyl)-2-[(3R)-2-oxo-1,3-dihydropyrrol-3-yl]ethyl]carbamoyl]-3-methylbutyl]carbamate	-7.4
Compound5	<chem>C[C@@H](NC(=O)[C@H](Cc1ccccc1)NC(=O)OCc2ccccc2)C(=O)CCO</chem>	benzyl N-[(1S)-1-benzyl-2-[(1R)-4-hydroxy-1-methyl-2-oxo-butyl]amino]-2-oxo-ethyl]carbamate	-7.1
Compound6	<chem>CC(C)(C)OC(=O)N[C@@H](C[C@@H]1CCNC1=O)[C@@H](O)C(=O)NCc2cccc2</chem>	tert-butyl N-[(1S,2R)-3-(benzylamino)-2-hydroxy-3-oxo-1-[[[(3S)-2-oxopyrrolidin-3-yl]methyl]propyl]carbamate	-7.3
Compound7	<chem>C\C=C\CO\CO</chem>	(Z)-2-methylbut-2-ene-1,4-diol	-3.9
Compound8	<chem>CCC(=O)Nc1ccc(cc1)N(Cc2ccsc2)C(=O)Cn3nnc4cccc34</chem>	N-[4-[[2-(benzotriazol-1-yl)acetyl]-(3-thienylmethyl)amino]phenyl]propanamide	-7.6

Compound9	<chem>O=C(Cn1nnc2ccccc12)N(Cc3ccsc3)c4ccc(NC(=O)c5cccs5)cc4</chem>	N-[4-[[2-(benzotriazol-1-yl)acetyl]-(3-thienylmethyl)amino]phenyl]thiophene-2-carboxamide	-8.1
Compound10	<chem>O=Cc1ccccc1</chem>	benzaldehyde	-4.9
Compound11	<chem>CC(=O)CNC(=O)OCc1cccc1</chem>	benzyl N-acetylcarbamate	-5.8
Compound12	<chem>C[C@H](O)CC(C)(C)O</chem>	(4S)-2-methylpentane-2,4-diol	-3.8
Compound13	<chem>[S][Zn+2]=S</chem>	lambda1-sulfanyl(thioxo)zinc(2+)	-2
Compound14	<chem>CCOC(=O)CC[C@@H](C[C@@H]1CCNC1=O)N[C@@H](O)[C@H](CC(C)C)NC(=O)[C@@H](C)NC(=O)c2cc(C)on2)C(C)C</chem>	(2S)-N-[(1S)-1-[(R)-[(1S)-4-ethoxy-4-oxo-1-[(3S)-2-oxopyrrolidin-3-yl]methyl]butyl]amino]-hydroxy-methyl]-3-methyl-butyl]-3-methyl-2-[(2S)-2-[(5-methylisoxazole-3-carbonyl)amino]propanoyl]amino]butanimidic acid	-6.9
Compound15	<chem>CC(=O)Nc1ccc(cc1)N([C@@H](C(=O)NC(C)(C)C)c2ccsc2)C(=O)Cn3nnc4cccc34</chem>	(2R)-2-(4-acetamido-N-[2-(benzotriazol-1-yl)acetyl]anilino)-N-tert-butyl-2-(3-thienyl)acetamide	-8.2
Compound16	<chem>CCOC(=O)CC[C@H](C[C@@H]1CCNC1=O)NC(=O)[C@H](CCC(C)C)CC(=O)[C@@H](NC(=O)c2cc(C)on2)C(C)C</chem>	ethyl (4R)-4-[(2R,5S)-2-isopentyl-6-methyl-5-[(5-methylisoxazole-3-carbonyl)amino]-4-oxo-heptanoyl]amino]-5-[(3S)-2-oxopyrrolidin-3-yl]pentanoate	-7
Compound17	<chem>CCOC(=O)CC[C@H](C[C@@H]1CCN=C1O)NC(=O)[C@H](CC=C(C)C)CC(=O)[C@@H](NC(=O)[C@H](CO)NC(=O)OC(C)(C)C)C(C)C</chem>	ethyl (4R)-4-[(2R,5S)-5-[(2S)-2-(tert-butoxycarbonylamino)-3-hydroxypropanoyl]amino]-6-methyl-2-(3-methylbut-2-enyl)-4-oxo-heptanoyl]amino]-5-[(4S)-5-hydroxy-3,4-dihydro-2H-pyrrol-4-yl]pentanoate	-6.8
Compound18	<chem>CCOC(=O)[C@@H](O)CC(=O)N(CCC(=O)N)NC(=O)[C@H](Cc1ccccc1)N=C(O)[C@H](CC(C)C)NC(=O)OC</chem>	(2S)-N-[(1S)-2-[2-(3-amino-3-oxo-propyl)-2-[(3S)-4-ethoxy-3-hydroxy-4-oxo-butanoyl]hydrazino]-1-benzyl-2-oxo-ethyl]-2-(benzyloxycarbonylamino)-4-methyl-pentanimidic acid	-7

Compound19	<chem>CCC(C)(C)NC(=O)[C@H](N(C(=O)Cn1nnc2ccccc12)c3ccc(NC(=O)C)cc3)c4cccn4C</chem>	(2R)-2-(4-acetamido-N-[2-(benzotriazol-1-yl)acetyl]anilino)-N-(1,1-dimethylpropyl)-2-(1-methylpyrrol-2-yl)acetamide	-7.7
Compound20	<chem>CC(C)C[C@H](NC(=O)[C@@H](NC(=O)c1cc(C)on1)C(C)C)C(=O)N[C@H](CC[C@@H](O)OCc2ccccc2)[C@H]3CCNC3=O</chem>	N-[[[(1S)-2-[[[(1S)-1-[[[(1R,4S)-4-benzyloxy-4-hydroxy-1-[[[(3R)-2-oxopyrrolidin-3-yl]methyl]butyl]carbamoyl]-3-methyl-butyl]carbamoyl]-2-methyl-propyl]amino]-1-methyl-2-oxo-ethyl]-5-methyl-isoxazole-3-carboxamide	-7.8
Compound21	<chem>CCOC(=O)CC[C@@H](C[C@@H]1CCNC1=O)N[C@@H](O)[C@H](CC(C)C)N=C(O)[C@@H](NC(=O)[C@@H](C)NC(=O)c2cc(C)on2)C(C)C</chem>	(2S)-N-[(1S)-1-[(R)-[[[(1S)-4-ethoxy-4-oxo-1-[[[(3S)-2-oxopyrrolidin-3-yl]methyl]butyl]amino]-hydroxy-methyl]-3-methyl-butyl]-3-methyl-2-[[[(2S)-2-[(5-methylisoxazole-3-carbonyl)amino]propanoyl]amino]butanimidic acid	-7.3