Deep learning based drug screening for novel coronavirus 2019-nCov

Haiping Zhang[1], Konda Mani Saravanan[1], Yang Yang[2],Md. Tofazzal Hossain[1,5], Junxin Li[3], Xiaohu Ren[4],Yanjie Wei[1]*

[1]Center for High Performance Computing, Joint Engineering Research Center for Health Big Data Intelligent Analysis Technology

Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen, Guangdong, PR China 518055

[2]Shenzhen Key Laboratory of Pathogen and Immunity, Guangdong Key Laboratory for Diagnosis and Treatment of Emerging Infectious Diseases, State Key Discipline of Infectious Disease, Second Hospital Affiliated to Southern University of Science and Technology, Shenzhen Third People's Hospital, Shenzhen, 518112, China

[3]Shenzhen Laboratory of Human Antibody Engineering, Institute of Biomedicine and Biotechnology, Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, 1068 Xueyuan Boulevard, University City of Shenzhen, XiliNanshan, Shenzhen, 518055, China

[4]Institute of Toxicology, Shenzhen Center for Disease Control and Prevention, No 8 Longyuan Road, Nanshan District, Shenzhen, 518055, China

[5]University of Chinese Academy of Sciences, No.19(A) Yuquan Road, Shijingsh an District,Beijing, P.R.China 100049

Corresponding Author: yj.wei@siat.ac.cn

Ethical issues

Compliance with ethical standards

Conflict of Interest

The authors have declared that no competing interests exist.

Funding

This work was partly supported by the National Key Research and Development Program of China under grant No. 2018YFB0204403 and 2016YFB0201305, National Science Foundation of China under grant no. U1813203 and 61433012; the National Natural Youth Science Foundation of China (grant no. 31601028), the Shenzhen Basic Research Fund under grant no. JCYJ20180507182818013, GGFW2017073114031767 and JCYJ20170413093358429, the China Postdoctoral Science Foundation under grant no. 2019M653132, CAS Key Lab under grant no. 2011DP173015. We would also like to thank the funding support by the Shenzhen Discipline Construction Project for Urban Computing and Data Intelligence, Youth Innovation Promotion Association, CAS to Yanjie Wei.

Abstract

A novel coronavirus called 2019-nCoV was recently found in Wuhan, Hubei Province of China, and now is spreading across China and other parts of the world. 2019-nCoV spreads more rapidly than SARS-CoV. Unfortunately, there is no drug to combat the virus. It is of high significance to develop a drug that can combat the virus effectively before the situation gets worse. It usually takes a much longer time to develop a drug using traditional methods. For 2019-nCoV, it is now better to rely on some alternative methods to develop drugs that can combat such a disease effectively since 2019-nCoV is highly homologous to SARS-CoV. In this paper, we first collected virus RNA sequences from the GISAID database, translated the RNA sequences into protein sequences, and built a protein 3D model using homology modeling. Coronavirus main protease is considered to be a major therapeutic target, thus this paper focused on drug screening based on the modeled 2019-nCov_main_protease structure. The deep learning based method DFCNN, developed by our group, can identify/rank the protein-ligand interactions with relatively high accuracy. DFCNN is capable of performing virtual screening quickly since no docking or molecular dynamic simulation is needed. DFCNN identifies potential drugs for 2019-nCoV protease by performing drug screening against 4 chemical compound databases. Also, we performed drug screening for all tripeptides against the binding site of 2019-nCov_main_protease since peptides often show better stability, more bio-availability and negligible immune responses. In the end, we provided the list of possible chemical ligands and peptide drugs for experimental validation.

Keywords

Coronavirus; Deep learning; Drug screening; homology modeling; main protease

Introduction

In the last two decades, China faced several outbreaks such as an avian influenza in 1997 [1], the severe acute respiratory syndrome (SARS) in 2003 [2], and a severe fever with thrombocytopenia syndrome (SFTS) in 2010 [3]. The most recent outbreak of the viral pneumonia was first disclosed by the Wuhan Municipal Health Commission [4] on December 30, 2019. On December 31, 2019, the World Health Organization (WHO) was alarmed about the outbreak of pneumonia by the Chinese Officials [5]. The novel coronavirus (2019-nCoV) was isolated from 27 patients who were initially reported and the number of patients was subsequently revised to 5997as of Jan 28, 2020, with 132 death [6]. The current 2019-nCoV outbreak has some common features like the SARS outbreak: both happened in winter, linked to live animal markets, and caused by unknown coronaviruses [7, 8]. Fever, cough, and shortness of breath are the symptoms in common cases whereas pneumonia, severe acute respiratory syndrome and kidney failure are being reported as the symptoms in severe cases [5]. Most of the 2019-nCoV patients are linked to the Huanan Seafood Wholesale Market where several wildlife animals including bats, snakes as well as poultry are sold. So far, no specific wildlife animal is identified as the host of the novel coronavirus. Bat is considered as the native host of the novel coronavirus (2019-nCoV) although there may have other hosts in transmission from bats to humans [8].

The Spring Festival travel rush has accelerated the spread, so it is of top priority to prevent the spread, develop a new drug to combat it, and cure the patients in time. Knowledge of current 2019-nCoV can be learned from previous SARS-CoV. The genetic sequences of 2019-nCoV have shown similarities to SARS-CoV (79.5%) [9]. The S protein and 2019-nCov_main_protease are potential drug targets. The S protein is the main target of neutralizing antibodies, and antibodies binding with this protein have the potential to stop the virus entry into host cells [10]. The main protease (also known as 3C-like proteinase) catalyzes a chemical reaction which is important in SARS coronavirus replicase polyprotein processing [11, 12]. Based on these target information, computer aided drug design can facilitate the development of a novel drug. The neutralizing antibodies against S protein of SARS have been obtained from human patients and the anti-SARS-CoV S antibody triggered fusogenic conformational changes [10]. This provides an important clue to prevent virus entry into host cells by antibodies or peptides. The 2019-nCov_main_protease inhibitors also have potential to prevent coronavirus maturation, and series of unsaturated esters inhibitors against 2019-nCov_main_protease of SARS-CoV was deposited in PDB database (Crystal structures of SARS-Cov main protease complexed with a series of unsaturated esters, PDB ID: 3TIT). One can also use these previous SARS inhibitors to design the inhibitor against 2019-nCoV. Based on the increasing protein-ligand complex structures, the deep learning algorithms for identifying/predicting potential binding compounds for a given target became possible [13, 14]. In addition to small molecular chemical compounds, scientists also rely on peptide/antibody to combat the virus due to stronger binding affinity.

In this work, we use a deep learning based method, DFCNN [14], to identify potential drugs (peptides or small ligands) against the protein target of the 2019-nCoV virus. It should be noted that the DFCNN is extremely fast in virtual drug screening. We believe the proposed potential drug list can help the experimental scientist to locate drug that can symptom relives or even cures the disease rapidly.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 5 February 2020 doi:10.20944/preprints202002.0061.v1

Peer-reviewed version available at Interdisciplinary Sciences: Computational Life Sciences 2020; doi:10.1007/s12539-020-00376-6

Methods

The overall workflow of the proposed method is shown in Figure 1. The DFCNN is a densely fully connected neural network, and the densely network (similar to DenseNet, but replace the convolution layer to fully connected layer) allows deep layer without the gradient vanishing problem. The deeper layers make it can learn more abstract features from the data. The training data of DFCNN is from PDBBIND database [15], for which we define the crystal protein-ligand PDB complexes as positive and cross-docking complexes as negative. The detail process to build the deep learning model is described in our previous paper [14]. DFCNN model has two advantages over many other methods such as independent of docking simulation and the training dataset includesnonbinding decoys. The independent of the docking simulation makes it extremely fast, while the inclusionof nonbinding decoys during training makes the model robust in the real application scenarios.

Sequence analysis of virus proteins from different patients

We retrieved the virus RNA sequences from GISAID database [16]. The amino acid sequence is translated from the RNA sequence by Translate web tool (https://web.expasy.org/translate/). The alignment focuses on the interested S protein and ligand binding region of 2019-nCov_main_protease (2019-nCov_main_protease). We used 18 patients' sequences in this work (EPI_ISL_402119.fasta; EPI_ISL_402120.fasta; EPI_ISL_402121.fasta; EPI_ISL_402123.fasta; EPI_ISL_402124.fasta; EPI_ISL_402125.fasta; EPI_ISL_402127.fasta; EPI_ISL_402128.fasta; EPI_ISL_402129.fasta; EPI_ISL_402130.fasta; EPI_ISL_402132.fasta; EPI_ISL_403928.fasta; EPI_ISL_403929.fasta; EPI_ISL_403930.fasta; EPI_ISL_403931.fasta; EPI_ISL_403962.fasta; EPI_ISL_403963.fasta; EPI_ISL_404227.fasta; EPI_ISL_404228.fasta). Details of the sequences and acknowledgement to the authors who submitted the data to the server is presented in the Table S1.

Homology modeling of 2019-nCov_main_protease

Homology modeling is performed for the 2019-nCov_main_protease. The SARS coronavirus main protease was used as a template (PDB ID: 3TNT). The sequence alignment score is 96.0784 by CLUSTAL 2.1 (https://www.genome.jp/tools-bin/clustalw) [17]. Detailed analysis of the modeled structure is presented in the results section.

Virtual screening against Chimdiv database

The homology model of the ligand binding region of 2019-nCov_main_protease is used as the target structure. We define the residues with a cutoff distance of 1 nm from the known ligand as a pocket (ligand from the template PDB 3TNT is used). The ligand database is taken from the chimdiv company (https://www.chemdiv.com/) which contains around 1000,000 compounds. We first used the DFCNN model to perform large scale virtual screening. The mean and deviation of the training dataset were used during data normalization for a more stable performance. In the second stage, the top prediction by DFCNN model was chosen for an autodock vina based docking simulation. The docking

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 5 February 2020 doi:10.20944/preprints202002.0061.v1

Peer-reviewed version available at Interdisciplinary Sciences: Computational Life Sciences 2020; doi:10.1007/s12539-020-00376-6

result was visualized and examined by the discovery studio visualizer [18]. Finally, we provide a proposed compound list that has the potential to bind to this protein pocket.

Virtual screening against Targetmol-Approved_Drug_Library, Targetmol-Natural_Compound_Library, and Targetmol-Bioactive_Compound_Library

The Targetmol-Approved_Drug_Library, Targetmol-Natural_Compound_Library, and Targetmol-Bioactive_Compound_Library contain about 2040, 1680, and 5370 compounds respectively. We have applied DFCNN model to perform virtual screening against these 3 libraries for 2019-nCov_main_protease. The compounds with high DFCNN scores are recommended as the potential inhibitors for further experimental validation.

Virtual screening against tripeptide database

Tri-amino acid peptide database is firstly built, with a total size of 8000. Each amino acid in the tripeptide database was converted into a molecule vector by Mol2vec [19]. For each peptide, the sum of its amino acid vector was used to represent this peptide's vector. Protein pocket is defined as residues with a cutoff distance of 1 nm from the known ligand. The pocket is then converted into Vector. The pocket and peptide vector are then concatenated into one line as input with a maximum dimension of 600. The process is similar to our previous work. We will use the same model as DFCNN, a densely fully connected model that is trained by a protein-ligand dataset from the PDB bind database. Since the ligand and peptides are composed of chemical groups, the model trained on the protein-ligand compound should also be suitable for protein-small peptide interaction.

Results

Sequence alignment and homology modeling

18 patient's RNA sequences are translated into protein sequences. We have checked the mutations in the pocket region of 2019-nCov_main_protease, and the sequences have 100% similarity for the virus from 18 different patients. This indicates the virus is currently stable in this region, and it is suitable for designing drugs. The alignment of S-protein epitope regions also shows high conservation among the patients (Figure S1). 2019-nCoV_main_protease is also aligned to SARS-CoV protease by CLUSTAL. The aligned sequence is shown in Figure 2. The figure indicates high similarity between 2019-nCov and SARS-CoV, which is consistent with the findings by Xu et al (2020) [8]. Using SARS coronavirus main protease, a theoretical protein model is built for 2019-nCoV_main_protease. There are only four mutations (T35V, A46S, S94A and K180N) between SARS_coronavirus_main_protease and 2019-nCoV main_protease shown in Figure 3. In Figure 3, the mutated residues are marked with blue color. As shown in Figure 3(D), the binding pocket and interaction pattern is extracted from the modeled structure and will be used for the deep learning based drug screening.

Virtual screening against 4 small compound databases

Chemdiv dataset, widely used for large scale virtual screening, contains a large amount (~1000,000) of drug-like compounds or drug leads. The potential drug candidates selected from the Chemdiv dataset are shown in Table 1. The top 100 predictions by DFCNN in the

database are shown in Table S2. Three other well known compound libraries were screened in this paper, including Targetmol-Approved_Drug_Library, Targetmol-Natural_Compound_Library, and Targetmol-Bioactive_Compound_Library. The natural compound library is also very popular for ligand virtual screening. It is worth to test whether there is any natural compound that can combat the virus by inhibiting 2019-nCov_main_protease. Table 2 shows the screening result for Targetmol-Natural compound library. The compounds with a DFCNN score higher than 0.997 are listed in Table 2, and it is found that Adenosine, Vidarabine, Mannitol, Dulcitol, D-Sorbitol, D-Mannitol, Allitol, Sodium_gluconate are the top predictions (Table 2). Natural products are often active ingredients of known herb medicine, and relatively safe because of long history usage. If it is proved by an experiment that is effective to the target, patients can easily access it by taking corresponding herb medicine. The list of compounds with a score higher than 0.990 is provided in Table S3.

The screening result for Targetmol-Approved Drug library is shown in Table 3. The compounds with a DFCNN score higher than 0.997 are listed in Table 3. We randomly considered drugs from potential drugs list and performed a systematic literature search. Interestingly, we found most of the drugs in the list such as meglumine,Ganciclovir and Vidarabine respectively show antiviral activity. It is found that Meglumine, Vidarabine, Adenosine, D-Sorbitol, D-Mannitol, Sodium_gluconate, Ganciclovir and Chlorobutanol respectively are top predictions according to the DFCNN score (Table 3). The list of compounds above score 0.990 is provided in Table S4. The screening result for Targetmol-Bioactive_Compound_Library is shown in Table 4. The compounds with a DFCNN score higher than 0.997 is listed in Table 4. Bioactive compounds are a type of chemicals that can found in plants and some foods and have been studied in the prevention of various diseases. It is worth to check whether any of them can act on the target protein. We found compounds such as Vidarabine, Adenosine, Dulcitol, D-Sorbitol, D-Mannitol, Ganciclovir and 5'-Deoxyadenosine are the top predictions in the Targetmol-Bioactive compounds (Table 4). The list in Table 4 has narrowed down the hit compounds for later drug development stages, such as docking, molecular dynamics simulation, or even directly experimental validation for finding bioactive compounds against 2019-nCov_main_protease. The list of compounds above score 0.99 is provided in Table S5.

Virtual screening against tripeptides

Peptides have the potential to exert higher binding affinity and specificity than small compounds meanwhile small peptides are easier to be synthesized compared with small compounds and antibodies. Since the known ligands of SARS_main_protease are compounds similar to tripeptides and the combination of 20 amino acids for tripeptide is also affordable for our method, we decide to perform virtual screening on the whole tripeptides. The screened tripeptides with a DFCNN score higher than 0.995 for the 2019-nCov_main_protease is shown in Table 5. A higher value indicates the peptide can most likely bind with the pocket of the 2019-nCov_main_protease. Our method found that the peptides formed by I, K, P amino acids have the highest possibility to bind in the pocket. The combinations by G, K, L or G, K, K or K, P, V are also found to be favorable binding partners predicted by DFCNN (Table 5). Since the tripeptides are relatively easy to produce, many of

the top predictions can be validated by the experimental techniques in a very fast and less expensive manner. The list of tripeptides above score 0.99 is provided in Table S6.

Conclusion

Designing small compound or peptide drugs to cure the 2019-nCoV is extremely urgent. With the extremely high speed and relatively high accuracy, our DFCNN model for protein-ligand interaction analysis is suitable to overcome the challenge of screening tens of thousands of drugs in a short time in a certain emergency situations, such as 2019-nCov outbreak. In this paper, we have performed a deep learning based drug screening for 2019-nCov_main_protease, and provided potential compound and tripeptide lists for 2019-nCov_main_protease. The inhibitor candidate lists provided in the paper can help to facilitate the 2019-nCov_main_protease drug development. In such an emergency situation, this paper focused more on drug efficacy and availability. For instance, we performed virtual screening against on-market drugs, since it could be used immediately. The design of peptide is for a similar reason since it is easier to be synthesized compared with small compounds and antibodies. Future works will focus on virtual screening longer length peptides against 2019-nCov_main_protease, and also try to develop a scoring function and reinforcement algorithm to redesign the antibodies of SARS S protein into antibodies or peptides of 2019-nCov S protein. We believe that the deep learning based drug screening technique can play an important role in designing drugs for combating such coronaviruses.

References

1. Chan PKS (2002) Outbreak of Avian Influenza A(H5N1) Virus Infection in Hong Kong in 1997. Clin Infect Dis. https://doi.org/10.1086/338820
2. Guan Y, Zheng BJ, He YQ, et al (2003) Isolation and characterization of viruses related to the SARS coronavirus from animals in Southern China. Science (80-). https://doi.org/10.1126/science.1087139
3. Yu XJ, Liang MF, Zhang SY, et al (2011) Fever with thrombocytopenia associated with a novel bunyavirus in China. N Engl J Med. https://doi.org/10.1056/NEJMoa1010095
4. (2020) Pneumonia of unknown cause in China. WHO. https://www.who.int/csr/don/05-january-2020-pneumonia-of-unkown-cause-china/en/
5. Hui DS, I Azhar E, Madani TA, et al (2020) The continuing 2019-nCoV epidemic threat of novel coronaviruses to global health — The latest 2019 novel coronavirus outbreak in Wuhan, China. Int. J. Infect. Dis.
6. (2019) No Title. http://114.116.95.250/2019-nCoV/index.html
7. Lu H, Stratton CW, Tang Y (2020) Outbreak of Pneumonia of Unknown Etiology in Wuhan China: the Mystery and the Miracle. J Med Virol. https://doi.org/10.1002/jmv.25678
8. Xintian Xu, Ping Chen, Jingfang Wang, Jiannan Feng, Hui Zhou, Xuan Li, Wu Zhong PH (2020) 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
9. Zhou P, Yang X-L, Wang X-G, et al (2020) Discovery of a novel coronavirus associated with the recent pneumonia outbreak in humans and its potential bat origin. bioRxiv. https://doi.org/10.1101/2020.01.22.914952

10. Walls AC, Xiong X, Park YJ, et al (2019) Unexpected Receptor Functional Mimicry Elucidates Activation of Coronavirus Fusion. Cell. https://doi.org/10.1016/j.cell.2018.12.028
11. Goetz DH, Choe Y, Hansell E, et al (2007) Substrate specificity profiling and identification of a new class of inhibitor for the major protease of the SARS Coronavirus. Biochemistry. https://doi.org/10.1021/bi0621415
12. Kim Y, Lovell S, Tiew K-C, et al (2012) Broad-Spectrum Antivirals against 3C or 3C-Like Proteases of Picornaviruses, Noroviruses, and Coronaviruses. J Virol. https://doi.org/10.1128/jvi.01348-12
13. Zhang H, Liao L, Saravanan KM, et al (2019) DeepBindRG: a deep learning based method for estimating effective protein–ligand affinity. PeerJ. https://doi.org/10.7717/peerj.7362
14. Zhang H, Liao L, Cai Y, et al (2019) IVS2vec: A tool of Inverse Virtual Screening based on word2vec and deep learning techniques. Methods. https://doi.org/10.1016/j.ymeth.2019.03.012
15. Liu Z, Li Y, Han L, et al (2015) PDB-wide collection of binding data: Current status of the PDBbind database. Bioinformatics. https://doi.org/10.1093/bioinformatics/btu626
16. Shu Y, McCauley J (2017) GISAID: Global initiative on sharing all influenza data – from vision to reality. Eurosurveillance
17. Larkin MA, Blackshields G, Brown NP, et al (2007) Clustal W and Clustal X version 2.0. Bioinformatics. https://doi.org/10.1093/bioinformatics/btm404
18. Accelrys: Materials Studio is a Software Environment for Molecular Modeling (2009) Dassault Systèmes BIOVIA,. Discovery
19. Jaeger S, Fulle S, Turk S (2018) Mol2vec: Unsupervised Machine Learning Approach with Chemical Intuition. J Chem Inf Model. https://doi.org/10.1021/acs.jcim.7b00616

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 5 February 2020 doi:10.20944/preprints202002.0061.v1

Peer-reviewed version available at Interdisciplinary Sciences: Computational Life Sciences 2020; doi:10.1007/s12539-020-00376-6

Figure 1. The overall workflow of virtual screening of small chemical compounds and tripeptides against the 2019-nCov_main_protease.

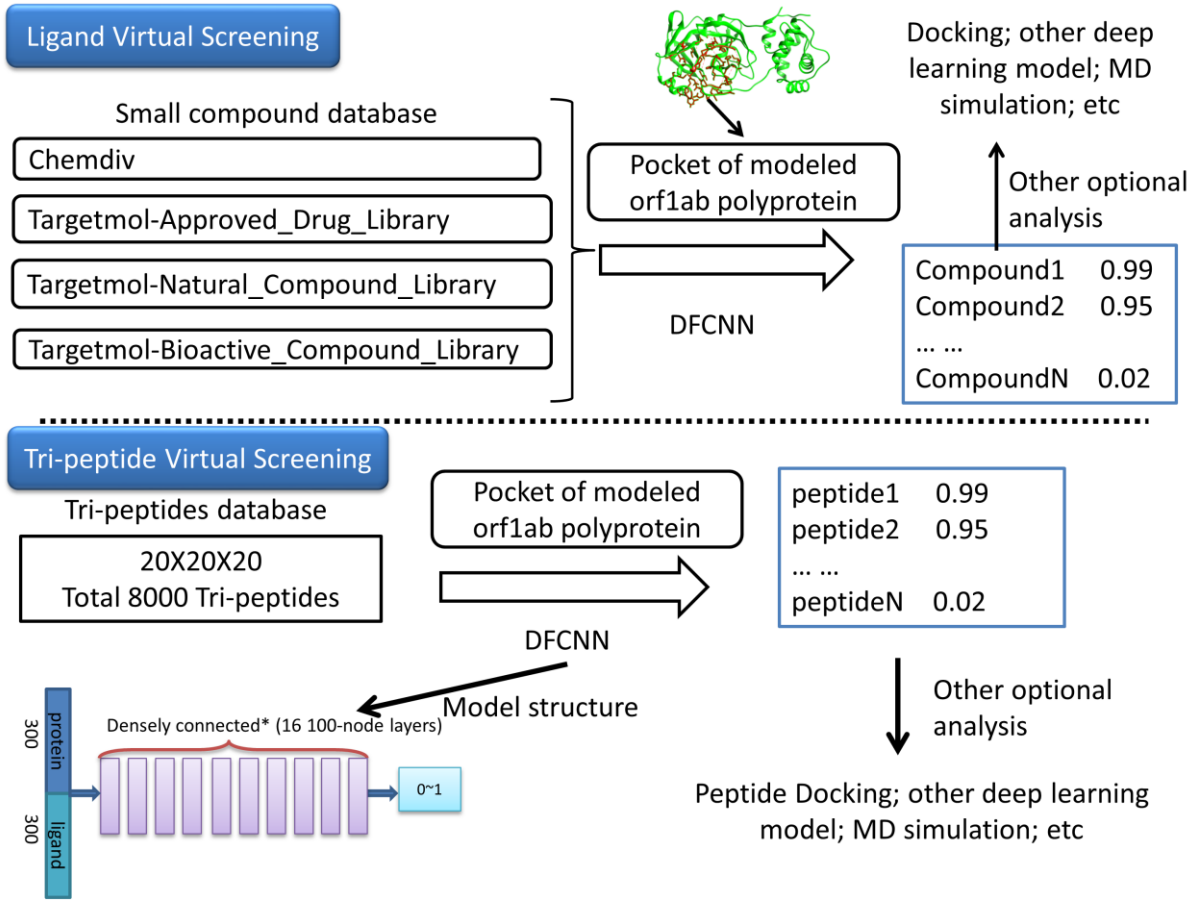
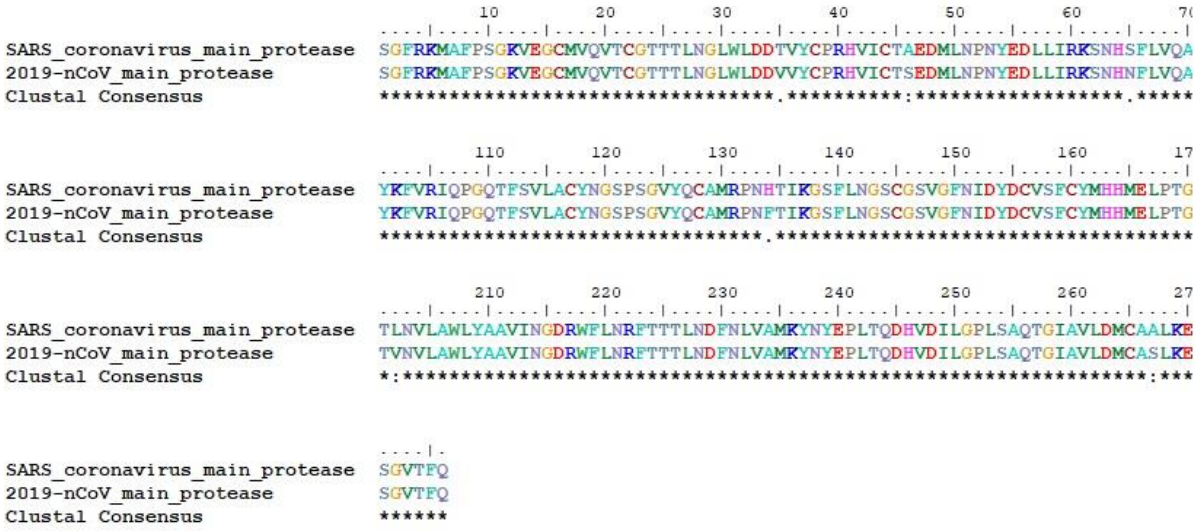


Figure 2. The sequence alignment of SARS_coronaivrus_main_protease and 2019-nCov_main_protease.



Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 5 February 2020 doi:10.20944/preprints202002.0061.v1

Peer-reviewed version available at Interdisciplinary Sciences: Computational Life Sciences 2020; doi:10.1007/s12539-020-00376-6

Figure 3. The structural model of 2019-nCov_main_protease and its template. In panels A and B, the modeled 2019-nCov_main_protease and SARS_main_protease are shown with the mutated four residues marked with blue color. The ligand from the PDB 3TNT is transferred to the modeled structure (Panel C) and based on residue distance from the transferred ligand, we define the pocket (Panel D). The interaction between the ligand and the modeled 2019-nCov_main_protease is also shown (Panel D).

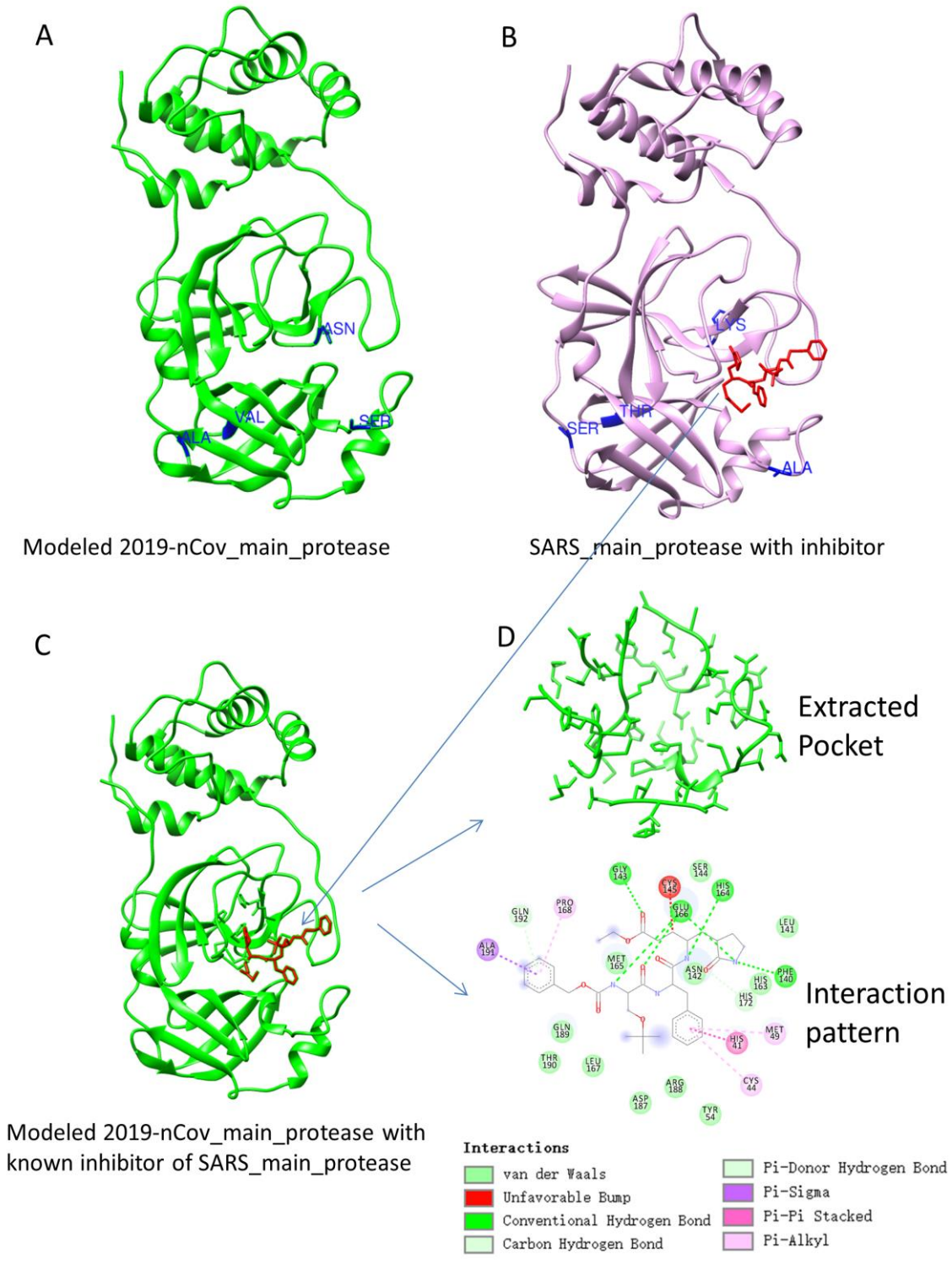


Table 1. The selected compounds that may inhibit 2019-nCov_main_protease, based on the DFCNN score andautodockvina score.

Chemdiv ID	**Vina score（kcal/mol）**	**DeepBindVec**	**Recommendation**
C998-0189	-8.5	>0.995	Recommended
C998-0197	-7.9	>0.995	Can Try
C998-0090	-7.8	>0.995	Can Try
C998-0948	-7.7	>0.995	Recommended
C998-1046	-7.6	>0.995	Recommended
D076-0195	-7.3	>0.995	Recommended

Table 2. The potential drug candidates selected from the Targetmol-Natural compound library.

Natural Compound	**DFCNN score**
Adenosine;Vidarabine;Mannitol;Dulcitol;D-Sorbitol;D-Mannitol;Allitol;Sodium_gluconate	score>=0.999
L(-)-sorbose;D-(-)-Fructose;Guanosine;Inosine;Trichostatin_A;D-(-)-Ribose;DL-Xylose;Cordycepin;β-Glycerophosphate_disodium_salt_hydrate;Xanthosine;Zeatin;N6-methyladenosine;Atractylodin;Tubercidin;Glucosamine_sulfate;Panthenol;Dexpanthenol;Ubenimex;Phospho(enol)pyruvic_acid_monopotassium	0.999>Score>=0.998
Aztreonam;Cytidine;Cytarabine;D-Saccharic_acid_potassium_salt;D-Glucose_6-phosphate_sodium_salt;Quinic_acid;2'-Deoxyadenosine_monohydrate;N-Sulfo-glucosamine_sodium_salt;2'-Deoxyguanosine_monohydrate	0.998>Score>=0.997

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 5 February 2020 doi:10.20944/preprints202002.0061.v1

Peer-reviewed version available at *Interdisciplinary Sciences: Computational Life Sciences* **2020**; doi:10.1007/s12539-020-00376-6

Table 3. The potential drug candidates selected from the Targetmol-Approved Drug library

Approved Drug name	**DFCNN score**
Meglumine;Vidarabine;Adenosine;D-Sorbitol;D-Mannitol;Sodium_gluconate;Ganciclovir;Chlorobutanol	score>=0.999
AICAR_(Acadesine);Mylosar;Inosine;D-Pantothenic_acid_sodium_salt;DL-Xylose;Ethambutol_dihydrochloride;Glucosamine;Myclobutanil;Sodium_etidronate;Fludarabine;Gemcitabine;Emtricitabine;Tubercidin;Bestatin_hydrochloride;Panthenol;Dexpanthenol;Cladribine;Entecavir;Ubenimex	0.999>Score>=0.998
Entecavir_hydrate;Procarbazine_hydrochloride;Aztreonam;Disopyramide;Benznidazole;Clofarabine;Bucetin;Nifuroxazide;Triflupromazine_hydrochloride;Doxifluridine;Cytarabine;Cefdinir;Bupropion_hydrochloride;Fluoxetine;Tenofovir;Pentostatin;Fluoxetine_hydrochloride;Imazalil;Atenolol	0.998>Score>=0.997

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 5 February 2020 doi:10.20944/preprints202002.0061.v1

Peer-reviewed version available at Interdisciplinary Sciences: Computational Life Sciences 2020; doi:10.1007/s12539-020-00376-6

Table 4. The potential drug candidates selected from the Targetmol-Bioactive compounds.

Bioactive Compound	**DFCNN score**
Vidarabine;Adenosine;Dulcitol;D-Sorbitol;D-Mannitol;Ganciclovir;5'-DEOXYADENOSINE	score>=0.999
Nelarabine;Tosedostat;Fosfomycin_Tromethamine;AICAR_(Acadesine);Mylosar;Guanosine;Inosine;Crotonoside;D-(-)-Ribose;Cordycepin;β-Glycerophosphate_disodium_salt_hydrate;Zeatin;Ethambutol_dihydrochloride;5-Iodotubercidin;Myclobutanil;Sodium_etidronate;Atractylodin;Fludarabine;Heterophyllin_B;Gemcitabine;Emtricitabine;Disodium_clodronate_tetrahydrate;Ostarine;Tubercidin;Bestatin_hydrochloride;Panthenol;Dexpanthenol;FCCP;Cladribine;Z-VAD(OMe)-FMK;WP1066;Entecavir;Ubenimex;Batimastat;ML264;GSK4112;Degrasyn;Cefcapene_Pivoxil_Hydrochloride;Phospho(enol)pyruvic_acid_monopotassium;A-804598;SR3335;IPTG	0.999>Score>=0.998
KYA1797K;Mizoribine;5-Hydroxy-1,7-diphenyl-6-hepten-3-one;ATPO;Entecavir_hydrate;Aztreonam;NXY-059;D-Pantothenic_acid;Bay_11-7085;Disopyramide;Benznidazole;SB_297006;Imidafenacin;Clofarabine;Bucetin;Nifuroxazide;Triflupromazine_hydrochloride;Doxifluridine;Selegiline_hydrochloride;Cytarabine;Cytidine;BGP-15;Cefdinir;Bupropion_hydrochloride;UK-371804;Fluoxetine;D-Saccharic_acid_potassium_salt;D-Glucose_6-phosphate_sodium_salt;J147;Tenofovir;N-Sulfo-glucosamine_sodium_salt;Pentostatin;Fluoxetine_hydrochloride;Nifurtimox;Imazalil;5-Fluorouridine;Atenolol;Repertaxin;ACY-738	0.998>Score>=0.997

Table 5. The predicted tripeptide that have high possibility (DFCNN score >=0.99) to bind with the pocket of 2019-nCov_main_protease by DFCNN score.

Peptide sequence	**DFCNN score**
IKP;IPK;KIP;KPI;PIK;PKI	Score>=0.997
GKL;LGK;LKG;KGL;KLG;GKK;KGK;KKG;AKK;KAK; KKA;KPV;KVP;PKV;PVK;VKP;VPK	0.997>Score>=0.996
GKI;IGK;IKG;KGI;KIG;LKP;LPK;KLP;KPL;PLK;PKL;L LK;LKL;KLL	0.996>Score>=0.995