J. Mol. Chem. 2021, 1(1), 105 Article

Chemical Biology
LETTERS

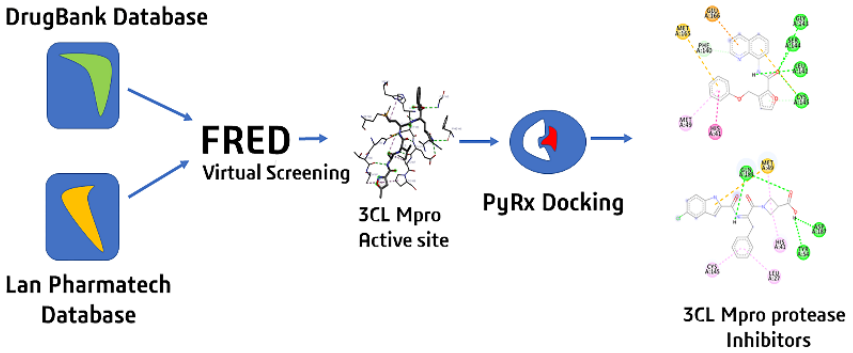
Virtual screening of potential inhibitors against SARS-CoV-2 main proteases (M^{pro}) by dual docking with FRED and AutoDock Vina programs

Prasan Tangyuenyongwatana,[1*] Wandee Gritsanapan[2]

[1]College of Oriental Medicine, Rangsit University, Pathumthani, 12000, Thailand. [2]Phyto Product Research, 165 Soi Suwandee 3, Rimklongprapra Road, Bangsue, Bangkok 10800, Thailand.

Submitted on: 7-Feb-2021, Accepted and Published on: 16-Feb-2021

ABSTRACT



At this time, the COVID-19 (SAR-CoV-2) pandemic is still ongoing and considered the most serious global outbreak and we need drugs for treatment this virus. This study focused on searching existing drugs or compounds from PubChem database to provide a faster route for compounds to combat with COVID-19. Two databases from DrugBank (13,575 compounds) and Lan Pharmatech (72,350 compounds) were subjected to FRED docking program. Top five from the docking results of both databases had the binding energy ranging from -11.63 to -12.60 and -11.62 to -13.30 kcal/mol, respectively. These two sets of compounds were subjected to AutoDock Vina for fine tuning docking. The two best compounds from DrugBank interacted with catalytic Cys145-His41 dyad in 3CL M^{pro} protease, suggesting good binding pattern and energies. For Lan Pharmatech, two compounds demonstrated better protein-ligand interactions than the others. This dual docking protocol demonstrated fast and high efficiency in searching 3CL M^{pro} protease inhibitors.

Keywords: SAR-CoV-2, 3CL M^{pro} protease, virtual screening, FRED, AutoDock Vina

INTRODUCTION

In Wuhan city on 31st December 2019, a new strain of coronavirus (2019-nCOV) or COVID-19 was revealed to the world.[1,2] More than a year later, the COVID-19 pandemic caused by SAR-CoV-2 is still ongoing and considered the most serious global outbreak, infecting more than 90 million and killing more than 2 million globally.[3] The higher death rate than normal flue is attributed to SAR-CoV-2's ability to infect both upper and lower respiratory systems, which enabled a faster spread of the virus and higher severity of symptoms. The mechanism of infection involves the viral binding onto angiotensin-converting enzyme 2 (ACE-2) in the alveoli of the lung, hijacking the enzyme as the vehicle to enter the host cell.[4]

The components for virus assembly normally compose of four structural proteins which are membrane (M) protein, spike (S) protein, nucleocapsid (N) protein and envelope (E) protein.[5] SARS-CoV-2 replication comprises the synthesis of two huge polyproteins, pp1a and pp 1ab, which are still inactive until enzyme protease (3CL M^{pro}) or main protease cleaves them into smaller functional proteins.[6] The main protein 3CL M^{pro} is a promising target for finding the specific inhibitors which can prevent the production of infectious viral. 3CL M^{pro} active site contains two catalytic residues, Cys145 and His41 which are buried in the cavity located on the surface of the enzyme.[7] Until now, we need a drug that can be used to cure the disease in this critical crisis.

Computational virtual screening of compounds in present database or repurposing virtual screening[8] of exist drugs is a quick strategy for finding lead compounds for *in vitro* and *in vivo* activity testing. Structural-based virtual screening uses structural

*Corresponding Author Dr. Prasan Tangyuenyongwatana
Tel: +668 58124285; Email: Prasan.t@rsu.ac.th

Cite as: *J. Mol. Chem. 2021, 1(1), 105.*

complementarity docking results from interaction between ligand and receptor protein.[9] FRED (Fast Rigid Exhaustive Docking, OpenEye Scientific Software, Santa Fe, NM, USA) is a structural-based virtual screening docking program between protein-ligand using a multi-conformer database and receptor file as input and deliver the best output molecules that bind to the receptor. FRED performs multiconformer molecules docking into a receptor using an exhaustive search that thoroughly explores rotations and translations of each conformer of the ligand within the active site. After the exhaustive search finish, the top scoring poses are further adjusted and given a final score. For this algorithm, the average docking time for FRED is as short as a 1-2 seconds per ligand that makes FRED the best choice for virtual screening of big databases. FRED jobs can also be easily distributed over multiple computers/processors to further reduce docking time.[10-11] In this research, the dual docking strategy was applied by using FRED as fast structural-based virtual screening against two databases, DrugBank (13,575 compounds) and Lan Pharmatech (72,350 compounds). These two databases were selected from PubChem databases based on size and the available of compounds. After obtaining top scores docking from FRED, the top five compounds of each database were subjected to AutoDock Vina in PyRx 0.8 virtual screening tool for fine tuning docking.[12] The protein-ligand interactions of docking items were performed in Discovery Studio Visualizer and compared the results with x-ray protein structures.

Materials and Methods

Preparation of ligand databases and protein

The crystalline structure of COVID-19 main protease (M^{pro}, PDB ID: 6LU7) was obtained from RCSB protein databank. The water molecules in 6LU7 were removed and the protein structure was prepared by using Make Receptor version 3.5.0.4 program or PDB2RECEPTOR command in command line window (OpenEye Scientific Software, Santa Fe, NM, USA). The output receptor file was obtained in "oeb.gz" format. The ligand databases were downloaded as "sdf" files from PubChem which were DrugBank (13,575 compounds) and Lan Pharmatech (72,350 compounds). The compounds were subjected to OMEGA program (OpenEye Scientific Software, Santa Fe, NM, USA) to generate multi-conformer OEBinary file in 3D format that was specific for FRED docking.[13]

Procedure for Molecular Docking of Ligands and Protein.

The FRED 3.2.0.2 docking suit used a command line procedure to set the docking work. Chemgauss 4 was used as a scoring function in FRED and the first 500 top ranking scores were recorded. All required files which were receptor and database files must stay in the same folder. In this study, the computer system has multiprocessors, it can be speed up the docking process faster than single processor by adding a multiple processors command and we used 4 processors. After the docking process end, the docking results were obtained in "oeb.gz" file format which can be opened with VIDA 4.4.0 application (OpenEye Scientific Software, Santa Fe, NM, USA). Top five compounds from each docking were chosen with corresponding binding energies. Then, they were subjected to AutoDock Vina in PyRx 0.8 virtual screening tool for fine tune docking. The center of co-ordinate of the binding site equal to X = -12.3306, Y = 12.3425, Z = 69.1374 and the box dimensions were 25 × 25 × 25 Å. The exhaustiveness values of all docking were set to 8. The poses of protein-ligand interactions of docking results were performed in Discovery Studio Visualizer 17.2.0 (Dassault Systèmes, French). The colors in the 2D protein-ligand interaction in the pictures were assigned as follows: Light green = weak hydrogen bond, Pink = hydrophobic interaction alkyl-alkyl or pi-alkyl, Purple = pi-pi interaction, Red = unfavorable hydrogen bond, Green = conventional hydrogen bonding and Yellow = pi-sulfur

Results and Discussion

Virtual screening of databases

Virtual screening with FRED docking against two databases obtained two sets of output data. For DrugBank and Lan Pharmatech databases, the first 500 top ranking had binding energy ranging from -8.95 to -12.60 kcal/mol and -9.55 to -13.30 kcal/mol, respectively. Top five ranking of both databases were selected and listed in the Table 1 and Table 2. Top two ranking of each database were depicted in Figure 1.

FRED uses exhaustive docking to generate all compounds within the active site. The steps of the procedure were collected of list all possible poses of the ligand around the active site by rigidly rotating and translating each conformer within the active site and screen the resulting pose ensemble by discarding poses that unqualified within the larger of the two volumes specified by the receptor file's shape potential grid and a contour level. Next, all remaining poses were ranked by using Chemgauss 4 scoring functions and force field refinement was performed by a full coordinate optimization of the pose using the MMFF94 force field.[14] This protocol made the average docking time per ligand of FRED as short as 1.5 second per ligand in our computing system.

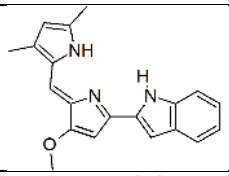
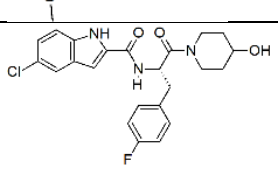
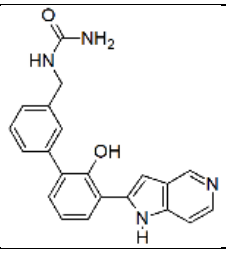
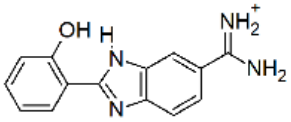
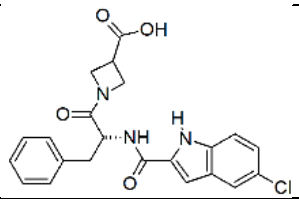
Refine the FRED results by AutoDock Vina docking

AutoDock Vina is another structure-based program that used Broyden-Fletcher-Goldfarb-Shanno (BFGS) method for the local optimization, which is an efficient quasi-Newton method as lead searching strategy. This method needs more time for virtual screening of big databases and its docking speed is slower than that of FRED. For this reason, AutoDock Vina was used to find the protein-ligand binding poses in the later step to fine tune the results. All compounds from Table 1 and 2 were subjected to AutoDock Vina docking engine in PyRx 0.8 virtual screening tool and the docking results are shown in Table 3.

The 3CL M^{pro} protease enzyme has a catalytic Cys-His dyad instead of a canonical Ser(Cys)-His-Asp(Glu) triad.[15] The active site of 3CL M^{pro} contained two catalytic residues, Cys145 and His41, which were buried in the cavity located on the surface of the enzyme. However, when the AutoDock Vina docking explored the interaction between x-ray ligand and amino acid residues around active site of 6LU7 protease enzyme, the interaction between protein-ligand showed only His41 of catalytic Cys-His dyad (Figure 2). The other amino acids surrounded the ligand were MET49, PHE140, GLY143, HIS164,

MET165, GLU166, LEU167, PRO168, HIS172, GLN189, THR190, and ALA191.

Table 1. FRED top five ranking compounds from DrugBank database.

Rank	PubChem ID	Structure	Binding energy (kcal/mol)
1	347828476		-12.60
2	46508116		-12.25
3	99443718		-12.24
4	46505811		-12.01
5	46506135		-11.63

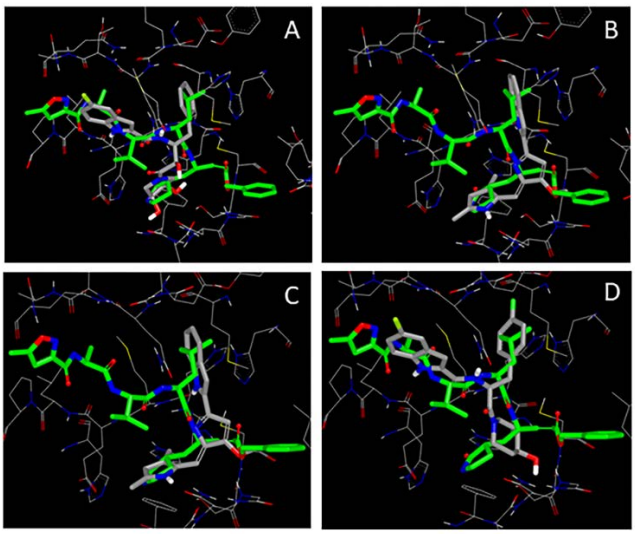
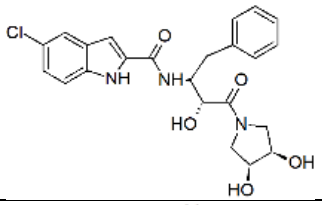
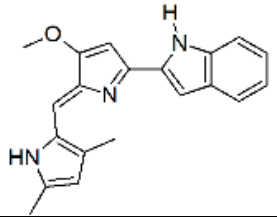
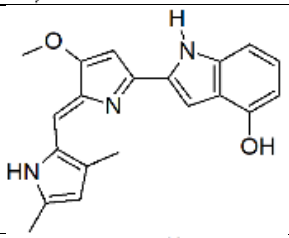
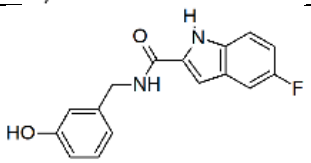
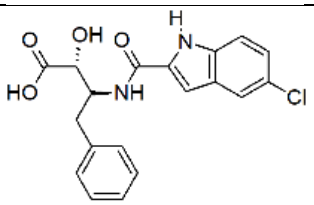


Figure 1. FRED docking pose of Lan Pharmatech rank 1 (A), rank 2 (B) and DrugBank rank 1 (C), rank 2 (D). 6LU7 ligand (green)

When the AutoDock Vina docking explored the FRED results from the two databases, the docking results showed that the binding energy of each compound was ranked differently comparing to the FRED results (Table 3). This was likely due to the difference in docking algorithm and scoring function of each program. In docking analysis, the catalytic Cys-His dyad residues were focused first and followed by the surrounded amino acid residues comparing to 6LU7 protease enzyme.

Protein-ligand interactions of the 2D and 3D structures of each compound were performed by Discovery studio visualizer.

Table 2. FRED top five ranking compounds from Lan Pharmatech database.

Rank	PubChem ID	Structure	Binding energy (kcal/mol)
1	433862728		-13.30
2	433809369		-12.61
3	433804071		-12.61
4	433795320		-11.76
5	433845243		-11.62

(Figure 3-4). The DrugBank database compounds (46508116, and 99443718) had binding energies -8.6, and -8.4 kcal/mol, respectively, and showed interaction with catalytic Cys-His dyad in binding cavity (Figure 3). On the other hand, 347828476 and 46505811 with moderate binding energies (-7.6 and -7.7 kcal/mol) demonstrated only one crucial catalytic CYS145 residue.

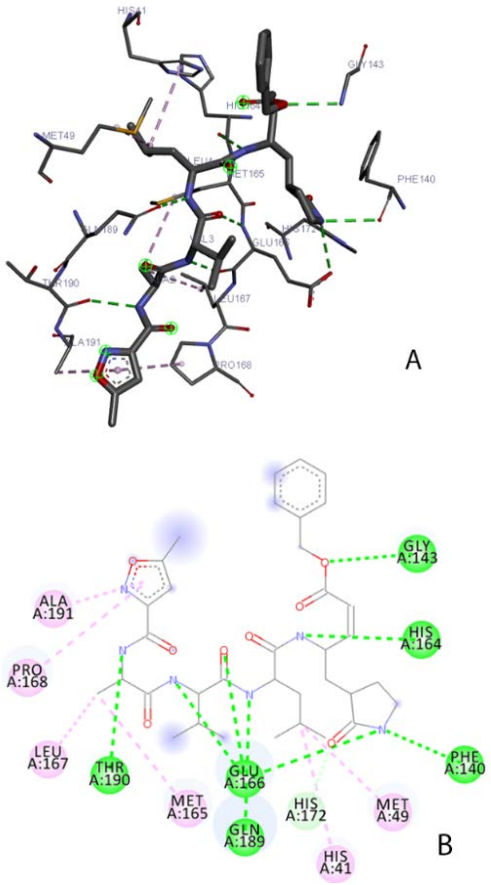


Figure 2. 3D structure of the protein-ligand (6LU7) interaction (A) and 2D animated pose showing interaction (B).

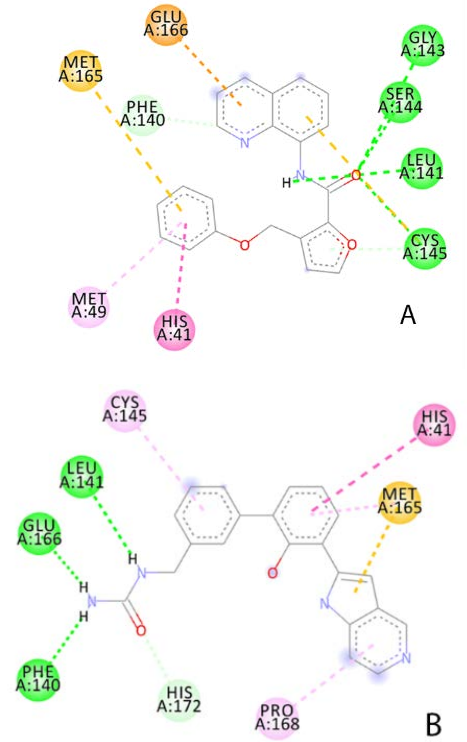


Figure 3. 2D animated pose showing interaction of 46508116 (A), and 99443718 (B) [DrugBank].

Table 3. AutoDock Vina docking results

Data bases	Top rank	PubChem ID	Binding energy (kcal/mol)
Drug Bank	1	347828476	-7.6
	2	46508116	-8.6
	3	99443718	-8.4
	4	46505811	-7.7
	5	46506135	-6.9
Lan Pharma tech	1	433862728	-7.0
	2	433809369	-7.6
	3	433804071	-7.8
	4	433795320	-7.3
	5	433845243	-7.8

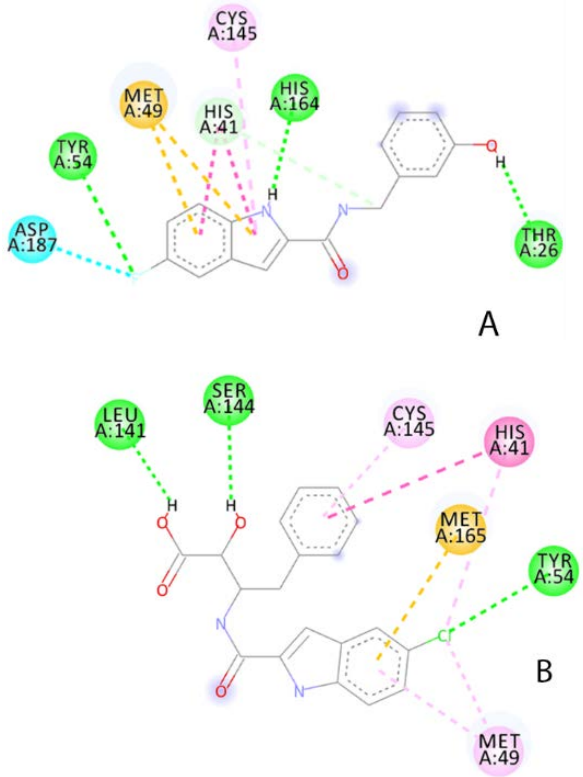


Figure 4. 2D animated pose showing interaction of 433795320 (A), and 433845243 (B) [Lan Pharmatech].

For Lan Pharmatech database, all compounds (433862728, 433809369, 433804071, 433795320, and 433845243) had binding energies ranging from -7.0 to -7.8 kcal/mol and exhibited interaction with catalytic Cys-His dyad in binding cavity (Figure 4). The analysis of docking results based on the protein-ligand interactions (van der Waals, hydrogen bonding, and electrostatic forces) and binding energies.[16] The two DrugBank database compounds that were good candidates for further development as 3CL M^{pro} inhibitors were compound 46508116 and compound 99443718 with binding energy of -8.6 and -8.4 kcal/mol, respectively. Compound 46508116 showed pi-pi stacked interaction between HIS41 and phenyl ring of ligand. In addition, CYS145 had hydrogen bonding with oxygen of carbonyl of amide group in the molecule plus pi-sulfur with aromatic ring. Moreover, GLY143, SER144, LEU141 also exhibited hydrogen

bonding with oxygen of carbonyl of amide group of the substance. Compound 99443718 showed a good pattern of interaction of pi-pi stacked interaction between HIS41 and phenol ring of the structure while CYS145 demonstrated pi-alkyl interaction with phenyl ring. Three conventional hydrogen bonding occurred between PHE140, LEU141, GLU166 and hydrogen atom of urea terminal group of the compound.[17] Whereas compound 46506135 demonstrated moderate affinity and the remained compounds, 347828476 and 46505811, showed only one crucial catalytic residue, resulting in a lower affinity than the former ones.

Five candidates of Lan Pharmatech database showed interaction with catalytic Cys-His dyad in receptor site. The good candidates included compounds 433795320 and 433845243 with binding energy of -7.3, and -7.8 kcal/mol, respectively. Beside the interaction with Cys145-His41, compound 433845243 exhibited additional conventional hydrogen bonds with TYR54, LEU141, SER144 and pi-sulfur interaction with MET165. These overall interactions led to the stability of the protein-ligand binding. The medium interaction candidates were compounds 433809369, 433804071, and 433862728. The forces included four to five number of amino acid residues in receptor site interacted as pi-sigma bond or pi-alkyl bond for the first two compounds. For compounds 433862728 and 433862728, there were unfavorable hydrogen bonding interactions in the receptor cavity, classifying them in the medium class group.

Conclusions

This study aimed to search promising database to repurpose existing drugs or compounds that may provide a quicker route to available compounds to fight COVID-19. The selected compounds may not be specifically made for treatment of SAR-CoV-2 virus; however, these compounds may be of benefit in relieving the virus symptom or decreasing virus spreading. Two medium size databases, DrugBank and Lan Pharmatech databases, from PubChem were chosen and subjected to FRED virtual screening. This program can be used to screen a large database with high efficiency and less time consuming. From the docking results, the top five compounds from DrugBank and Lan Pharmatech databases with binding energy range from -11.63 to -12.60 and -11.62 to -13.30 kcal/mol, respectively, were transferred to AutoDock Vina program for fine tuning docking. The good candidates from both databases showed interaction with catalytic Cys145-His41 dyad in receptor site of 3CL M^{pro} protease. These compounds were 46508116 and 99443718 from DrugBank database and 433795320 and 433845243 from Lan Pharmatech database. This study used dual docking protocol that demonstrated the fast and high efficiency databases screening and this method can be used as an alternative method to find inhibitors to potentially treat the SAR-CoV-2 virus and other targets.

Acknowledgments

This research project was supported by College of Oriental of Medicine, Rangsit University, Thailand for computer resources in performing the research. We thank Dr. Panupon Khumsupan for his kind proofreading of the manuscript.

Conflict of Interest

Authors declared that there is no conflict of interest for publication of this work.

References

1. D. Bhowmik, R. Nandi, R. Jagadeesan, et al. Identification of potential inhibitors against SARS-CoV-2 by targeting proteins responsible for envelope formation and virion assembly using docking based virtual screening, and pharmacokinetics approaches. *Infect. Genet. Evol.* **2020**, 84, 104451.
2. P. Zhou, X. Lou Yang, X.G. Wang, et al. A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature* **2020**, 579 (7798), 270–273.
3. B.S. Chhikara, B. Rathi, J. Singh, P. FNU. Corona virus SARS-CoV-2 disease COVID-19: Infection, prevention and clinical advances of the prospective chemical drug therapeutics. *Chem. Biol. Lett.* **2020**, 7 (1), 63–72.
4. B. Havranek, S.M. Islam. An in silico approach for identification of novel inhibitors as potential therapeutics targeting COVID-19 main protease. *J. Biomol. Struct. Dyn.* **2020**, 1–12.
5. S.D. Dent, D. Xia, J.M. Wastling, et al. The proteome of the infectious bronchitis virus Beau-R virion. *J. Gen. Virol.* **2015**, 96 (12), 3499–3506.
6. T. Muramatsu, C. Takemoto, Y.T. Kim, et al. SARS-CoV 3CL protease cleaves its C-terminal autoprocessing site by novel subsite cooperativity. *Proc. Natl. Acad. Sci. U. S. A.* **2016**, 113 (46), 12997–13002.
7. D.W. Kneller, G. Phillips, H.M. O'Neill, et al. Structural plasticity of SARS-CoV-2 3CL Mpro active site cavity revealed by room temperature X-ray crystallography. *Nat. Commun.* **2020**, 11 (1).
8. N. Issa, O. Peters, S. Byers, S. Dakshanamurthy. RepurposeVS: A Drug Repurposing-Focused Computational Method for Accurate Drug-Target Signature Predictions. *Comb. Chem. High Throughput Screen.* **2015**, 18 (8), 784–794.
9. J. Vázquez, M. López, E. Gibert, E. Herrero, F.J. Luque. Merging Ligand-Based and Structure-Based Methods in Drug Discovery: An Overview of Combined Virtual Screening Approaches. *Molecules* **2020**, 25 (20), 4723.
10. M. McGann. FRED pose prediction and virtual screening accuracy. *J. Chem. Inf. Model.* **2011**, 51 (3), 578–596.
11. J.O. Ebalunode, Z. Ouyang, J. Liang, W. Zheng. Novel approach to structure-based pharmacophore search using computational geometry and shape matching techniques. *J. Chem. Inf. Model.* **2008**, 48 (4), 889–901.
12. S. Dallakyan, A.J. Olson. Small-molecule library screening by docking with PyRx. *Methods Mol. Biol.* **2015**, 1263, 243–250.
13. P.C.D. Hawkins, A.G. Skillman, G.L. Warren, B.A. Ellingson, M.T. Stahl. Conformer generation with OMEGA: Algorithm and validation using high quality structures from the protein databank and cambridge structural database. *J. Chem. Inf. Model.* **2010**, 50 (4), 572–584.
14. T.A. Halgren. MMFF VII. Characterization of MMFF94, MMFF94s, and other widely available force fields for conformational energies and for intermolecular-interaction energies and geometries. *J. Comput. Chem.* **1999**, 20 (7), 730–748.
15. Q. Li, C. Kang. Progress in Developing Inhibitors of SARS-CoV-2 3C-Like Protease. *Microorganisms* **2020**, 8 (8), 1250.
16. R. Singh, A. Gautam, S. Chandel, et al. Protease inhibitory effect of natural polyphenolic compounds on SARS-CoV-2: An in silico study. *Molecules* **2020**, 25 (20).
17. D.W. Kneller, S. Galanie, G. Phillips, et al. Malleability of the SARS-CoV-2 3CL Mpro Active-Site Cavity Facilitates Binding of Clinical Antivirals. *Structure* **2020**, 28 (12), 1313-1320.e3.