Sains Malaysiana 50(5)(2021): 1473-1484
http://doi.org/10.17576/jsm-2021-5005-26

Molecular Docking Unveils Prospective Inhibitors for the SARS-COV-2 Main Protease

(Mengedok Molekul Mendedahkan Prospektif Perencat untuk Protease Utama SARS-COV-2)

FAWAD AHMAD, SAIMA IKRAM, JAMSHAID AHMAD*, IRSHAD UR REHMAN, SAEED ULLAH KHATTAK, SADIA BUTT & MARYAM MUSHTAQ

ABSTRACT

The recent emergence of a novel coronavirus strain (SARS-CoV-2) has stimulated global efforts to identify potential drugs that target proteins expressed by this novel coronavirus. Among these, the main protease of SARS-CoV-2 (3CL-protease (3CL^{Pro}), also known as (M^{Pro}) is one of the best choices for the scientists to target. 3CL^{Pro} is involved in the processing of polyproteins into mature non-structural viral proteins. An X-ray crystallographic structure (PDB ID 6LU7) of this protein was obtained from the PDB database. ChemDiv libraries of ~80,000 antiviral and ~13,000 coronavirus-targeting molecules were screened against the 3D structure of 3CL^{Pro} of SARS-CoV-2. We have identified a panel of molecules that showed an activity and potentially block the active site of the SARS-CoV-2 main protease. These molecules can be investigated further to develop effective virus-inhibiting molecules to treat this highly distressing disease, causing extreme unrest across the globe.

Keywords: Antivirals; coronavirus library; COVID-19; SARS-CoV-2; virtual screening

ABSTRAK

Kemunculan strain koronavirus baru-baru ini (SARS-CoV-2) telah mendorong usaha global untuk mengenal pasti ubat berpotensi yang mensasarkan protein yang diekspres oleh novel koronavirus ini. Antaranya, protease utama SARS-CoV-2 (3CL-protease (3CL^{Pro}), juga dikenali sebagai (M^{Pro}) adalah salah satu pilihan terbaik untuk disasarkan oleh para saintis. 3CL^{Pro} terlibat dalam pemprosesan poliprotein kepada virus protein matang tak-berstruktur. Struktur kristalografi sinar-X (PDB ID 6LU7) protein ini diperoleh dari pangkalan data PDB. Perpustakaan ChemDiv dengan ~80,000 antivirus dan ~13,000 molekul penyasaran koronavirus disaring terhadap struktur 3D 3CL^{Pro} SARS-CoV-2. Kami telah mengenal pasti panel molekul yang menunjukkan aktiviti dan berpotensi menyekat tapak aktif protease utama SARS-CoV-2. Molekul ini boleh dikaji lebih lanjut untuk mengembangkan molekul perencatan-virus yang berkesan untuk merawat penyakit yang menyebabkan keresahan melampau di seluruh dunia ini.

Kata kunci: Antivirus; COVID-19; perpustakaan koronavirus; saringan maya; SARS-CoV-2

INTRODUCTION

Severe acute respiratory syndrome (SARS) caused by the novel coronavirus-2 (CoV-2) is a major public health concern. Initially reported in China in December 2019, the virus now has a global presence (Huang et al. 2020). On the 11th of March, 2020, the World Health Organization (WHO) declared that the SARS coronavirus disease-19 (SARS-COVID-19) is a pandemic (Wang et al. 2020; Wu et al. 2020a). Initial symptoms include mild fever, illness, dyspnea, and dry cough, which subsequently progresses to severe forms of pneumonia (Ren et al. 2020).

Coronaviruses are enveloped viruses belonging to the family *Coronoviridae* (Wu et al. 2020b). It is further subdivided into four genera: *Alpha-, Beta-, Gamma-,* and *Delta-coronavirus* (Li 2016; Wu et al. 2020b). Members of the *Alpha-* and *Beta-coronaviruses* cause infections

in mammals. *Gamma-* and *Delta-coronaviruses* are predominantly found in birds (Cui et al. 2019; Tang et al. 2015; Wu et al. 2020c; Zhou et al. 2020). SARS-CoV-2 belongs to the *Beta-coronavirus* genus. Other members of the same genus include HCoV-OC43, HCoV-HKU1, SARS-CoV (Severe Acute Respiratory Syndrome), and MERS-CoV (Middle East Respiratory Syndrome) (Tang et al. 2015; Wu et al. 2020b, 2020c).

The SARS-CoV-2 genome is a 29.8 kb positive-sense single-stranded RNA and has 14 open reading frames (ORFs). They encode information for 27 proteins (Wu et al. 2020c). At the 5'-end of the genome, there are two ORFs (ORF-1a and ORF-1ab) encoding two long stretches of polyproteins (pp-1a and pp-1ab) and contains information for 15 non-structural proteins (nsp 1-10 and nsp12-16) (Wu et al. 2020c). The vital non-structural proteins are nsp3 (multi-domain protein with a PL-pro domain), nsp5 (3CL-chymotrypsin-like), nsp9 (helicase potentially involved in viral replication), nsp12 (RNA-dependent RNA polymerase), and nsp13 (helicase). The 3'-end of the genome encodes information for four structural and eight accessory proteins. The structural proteins are spike surface glycoproteins (S), envelope (E), matrix (M), and nucleocapsid (N) proteins. The accessory proteins are 3a, 3b, p6, 7a, 7b, 8b, 9b, and orf14 (Wu et al. 2020c). Because they have crucial functions in viral replication, these proteins are ideal targets to screen different inhibitors for potential treatment.

Due to an inadequate understanding of SARS-CoV-2 transmission and pathogenesis, social distancing is currently the main method opted by authorities to prevent the virus from spreading. Vaccine and drug development seem to be distant goals. In the current global health emergency, screening anti-viral drug databases to identify potential inhibitors appears to be a convenient approach. The current study reports potential molecules that can inhibit the function of vital viral proteins. Two ChemDiv libraries: antiviral (~80,000 molecules) and coronavirus (~13,873 molecules), were investigated to identify potential drugs that target the SARS-CoV-2 3CL protease ($3CL^{Pro}$). These potential molecules can then be investigated further to develop effective virus-inhibiting molecules to treat this highly distressing illness. These molecules interacted strongly with key amino acid residues of the target protein.

MATERIALS AND METHODS

A crystal structure of the main protease of SARS-CoV-2 was retrieved from the Protein Data Bank (PDB ID 6LU7). The three-dimensional structure of this protein was imported and visualized in Maestro. The protein was prepared using the protein preparation wizard. During this process, hydrogen atoms were added to the protein, side-chains and loops were fixed, and disulfide bonds were created. To mimic the physiological environment, the protonation state of amino acids was assigned at pH7.4 using PROPKA (Sastry et al. 2013). The OPLS2005 force field was used to optimize the protein structure (Bas et al. 2008; Shivakumar et al. 2010). We used two large ChemDiv compound libraries: An antiviral library containing ~80,000 antiviral compounds and a coronavirus Library containing ~13,873 compounds.

The database was downloaded from the Chemdiv website (https://www.chemdiv.com), for which access was granted on prior request. Both of these libraries are already passed through PAINS, REOS (rapid elimination of swill) & MedChem filters. After downloading, they were prepared individually in Maestro using the LigPrep module. After preparation, and allowing for the generation of potential conformations, the total number of antiviral and coronavirus molecules submitted to the module was ~103,225 and ~29,084, respectively. Epik was used to generate the protonation state of the ligands at pH7.4 (Shelly et al. 2007). A grid box was generated by selecting the native ligand.

For molecular docking simulation, the Glide/SP module of Maestro was used (Friesner et al. 2004). During the grid box generation for the receptor, side-chain rotations of certain amino acids (Thr 25, Cys44, Thr 45, Ser 46, Cys 85, Ser 144, Cys 145, Ser 147, Tyr 161, Thr 169, and Thr 190) were allowed for flexible docking. Five poses per ligand were selected during Glide/SP docking. The top twelve molecules from both libraries were re-docked using Induced Fit Docking (IFD) algorithm to analyze the energy differences of selected compounds in the binding pocket of SARS-CoV-2 $3CL^{Pro}$.

RESULTS AND DISCUSSION

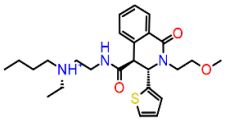
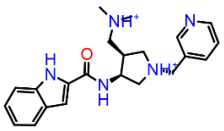
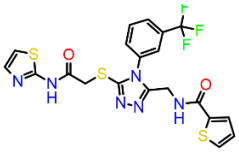
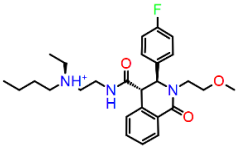
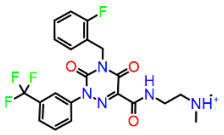
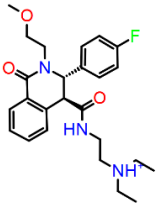
SARS is a respiratory system disease caused by the newly identified coronavirus strain. According to the WHO weekly epidemiological update published on December 22nd, 2020, a total of 75 million reported cases and 1.6 million deaths have been registered globally that were associated with COVID-19 (https://www.who.int/publications/m/item/weekly-epidemiological-update---22-december-2020). COVID-19 is spreading rapidly and does not respect geographical boundaries. There is an urgent need for an effective antiviral drug therapy against COVID-19 amid this global pandemic emergency.

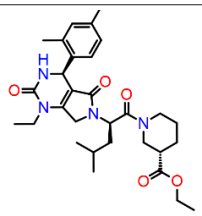
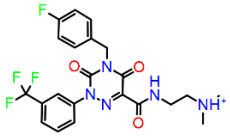
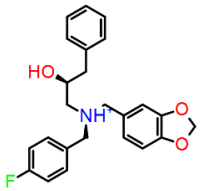
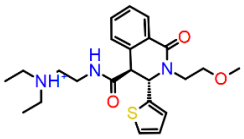
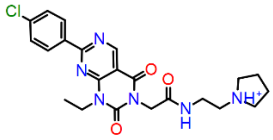
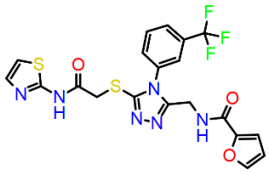
As was discussed earlier, the SARS-CoV2 genome encodes for various structural and non-structural proteins. One of these is the main protease $3CL^{Pro}$, which

is essential for processing polyproteins into other non-structural coronavirus proteins inside the host. This protein is being targeted by different research groups across the world as a potential drug target, and most studies have focused on finding its inhibitor among the Food and Drug Administration (FDA)-approved drugs. A crystal structure of this main protease bound to its peptide-like inhibitor was determined using X-ray crystallography with a resolution of 2.16 Å (Jin et al. 2020). In this study, the SARS-CoV2 essential protein (6LU7) was targeted, and compounds from two libraries were screened against it to find potential novel inhibitors.

After screening the antiviral and coronavirus libraries containing ~80,000 and ~13,873 compounds, respectively, we identified molecules that showed a significant binding affinity to the viral protein. Molecules were analyzed based on docking energy, ligand efficiency, and other ligand-protein interactions. The top 12 molecules were selected based on the Glide/SP scores, and these are listed in Table 1 along with their two-dimensional (2D) structures.

TABLE 1. The top 12 molecules and their 2D structures identified from screening the antiviral library for $3CL^{Pro}$ inhibitors. These are arranged based on their docking energy for both SP and IFD docking in kcal/mol. The last column presents the ligand efficiency of the corresponding ligand in kcal/mol

Molecules	2D structures	Docking score SP (kcal/mol)	Docking score IFD (kcal/mol)	Ligand efficiency SP (kcal/mol)
52647		-9.50	-7.60	-0.29
36600		-9.39	-11.67	-0.33
9194		-9.30	-8.45	-0.27
10805		-9.10	-7.45	-0.26
32077		-9.09	-7.04	-0.26
41756		-9.06	-8.06	-0.28

50951		-9.03	-10.49	-0.23
32299		-9.01	-6.54	-0.26
34813		-8.96	-8.24	-0.30
52455		-8.96	-7.37	-0.29
26435		-8.95	-5.35	-0.27
44953		-8.93	-8.39	-0.26

Molecules with docking scores greater than -8.00 kcal/mol were identified. The hydrogen bonding (H-bonding) and other potential non-covalent interactions within the binding pockets of the top 12 molecules were analyzed (available in the supplementary data). The antiviral molecule 52647 docked within the binding pocket of 6LU7, has a docking score of -9.50 kcal/mol and a ligand efficiency of -0.29 kcal/mol. Four hydrogen bonds (H-bonds) were observed with GLY143, GLU166, and GLN189. Similar interactions were observed for the antiviral molecule 36600, which has one additional Pi-cation interaction with HIS41. This molecule has a docking score of -9.33 kcal/mol and a ligand efficiency of -0.33 kcal/mol. H-bonds were observed in GLY143, SER144, and CYS145 for the antiviral molecule 9194, and its docking score and ligand efficiency was

-9.30 kcal/mol and -0.27 kcal/mol, respectively (Figure 1). Similar types of interactions and docking scores were observed for FDA-approved drugs (Durdagi et al. 2020). Supplementary Figure S1 shows an interaction diagram of the nine other antiviral molecules. Detailed information of the top 350 antiviral molecules and their structures are available in CSV and MOL2 formats, respectively, and can be downloaded from a folder linked to in the supplementary data.

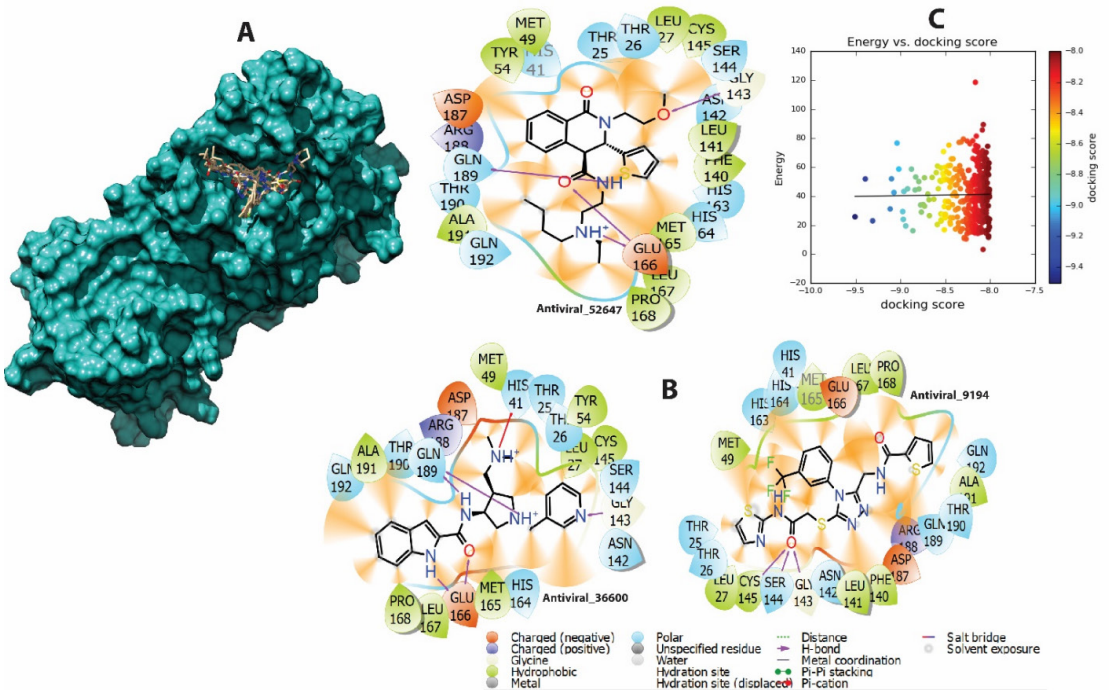
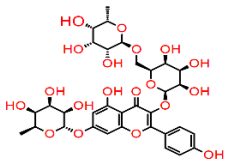
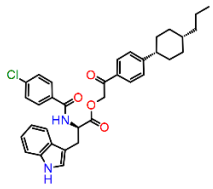
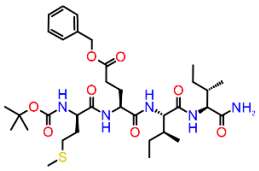
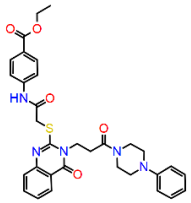
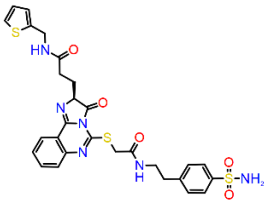
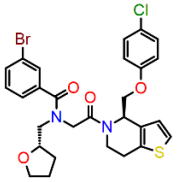
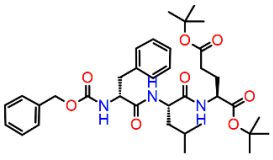
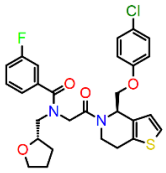
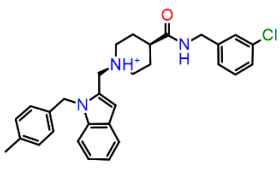


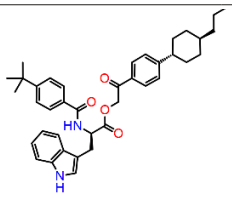
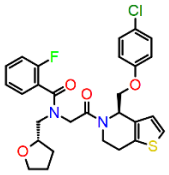
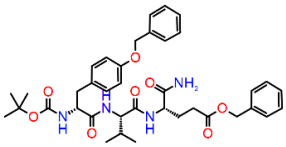
FIGURE 1. A) Surface representation of the $3CL^{Pro}$ protein and how the top 12 molecules docked into the binding pocket. B) 2D interactions of the top 3 antiviral molecules within the protease pocket. C) A scatter plot of the top 350 molecules with binding energies greater than -8.00 kcal/mol. A line of best fit is shown

Screening the coronavirus library also yielded compounds that exhibited significant binding energy with the main protease. The top 12 compounds were analyzed and selected based on the same criteria as that used on the antiviral library. The docking energies, ligand efficiencies, and 2D structures of these molecules are shown in Table 2.

TABLE 2. The top 12 molecules and their 2D structures identified by screening the coronavirus library, arranged based on docking energy for both SP and IFD docking in kcal/mol. The last column presents the ligand efficiency of the corresponding ligands in kcal/mol

Molecules	2D structures	Docking score SP (kcal/mol)	Docking score IFD (kcal/mole)	Ligand efficiency SP (kcal/mol)
N039_0006		-9.30	-17.29	-0.17

K935_0038		-9.15	-9.24	-0.21
4464_0994		-9.07	-8.07	-0.18
K284_3093		-9.07	-9.36	-0.21
K279_0897		-8.98	-11.45	-0.21
V019-1118		-8.94	-10.27	-0.24
8003-5413		-8.92	-11.20	-0.19
V023-5932		-8.87	-8.74	-0.24
G907-0944		-8.76	-6.71	-0.25

K935-0052		-8.75	-8.27	-0.19
V019-1069		-8.75	-10.27	-0.23
4465-2465		-8.69	-7.65	-0.174

The compound N039-0006 showed significant H-bonding with GLN189, THR26, MET49, and GLU166 amino acids, and has a docking score of -9.30 kcal/mol and a ligand efficiency of -0.17 kcal/mol. The compound K935-0038 formed 2 H-bonds with SER46 and GLY143, has a docking score of -9.15 kcal/mol, and a ligand efficiency of -0.21 kcal/mol. The compound 4464-0994 also formed H-bonds with GLU166 and GLN189, has a docking score of -9.07 kcal/mol, and a ligand efficiency of -0.18 kcal/mol. Figure 2 shows receptor-ligand interactions for the top three compounds. Diagrams showing the ligand interactions for the remaining nine molecules are available in the Supplementary Figure S2. Moreover, detailed information about the top 114 interacting molecules and their corresponding structures in mol2 format are available from a folder linked to in the supplementary data.

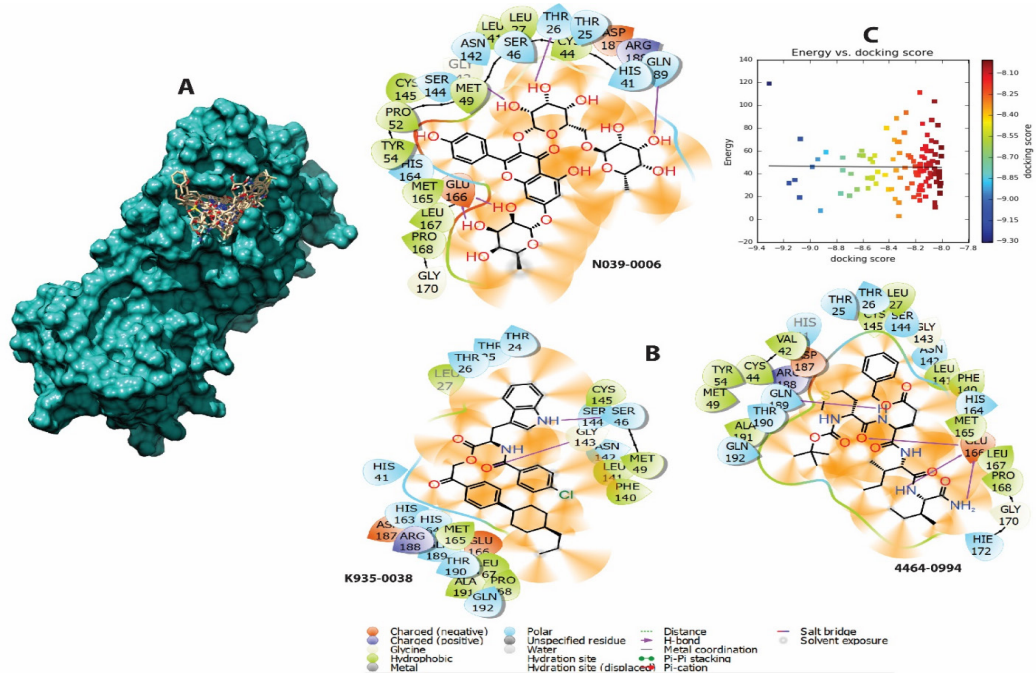


FIGURE 2. A) Surface representation of the 3CL^{Pro} protein and how the top 12 molecules docked into the binding pocket. B) 2D interaction of the top three antiviral molecules within the binding pocket. C). A scatter plot of the top 114 molecules with binding energies greater than -8.00 kcal/mol. A line of best fit is shown

While re-docking the top twelve selected compounds from both the libraries, we used IFD algorithm and found no significant difference in their energies compared to the SP docking algorithm. However, the compound N039_0006 from the coronavirus library has shown marked difference in the binding energy in IFD algorithm. Several compounds, such as flavonoids from medicinal plants, have been reported to have antiviral bioactivities (Thayil et al 2016; Zakaryan et al. 2017). Our screened compounds showed similar interactions as these previously reported compounds. Consequently, our docking models have identified new potential inhibitors targeting the SARS-CoV2 main protein.

Conclusion

In this study, we virtually screened two libraries downloaded from ChemDiv (Antiviral and Coronavirus) against the COVID-19 main protease. These compounds were docked against the PDB structure (PDB Code 6LU7). The compounds were then arranged according to their docking scores, which are represented as binding energies. The top 12 molecules with docking scores greater than -8.00 kcal/mol were considered for analysis. Based on these docking energies, the molecules were re-docked using Glide/IFD analysis, and the following COVID-19 main protease inhibitors were identified. From the antiviral library 52647, 36600, 9194, 10805, 32077, 41756, 50951, and 32299; and from the coronavirus library N039_0006, K935_0038, 4464_0994, and K284_3093. These compounds can be further tested as potential drugs against the SARS-CoV-2 3CL-protease.

Acknowledgements

We are thankful to the Schrodinger Team for providing us with the license to use the Schrodinger Suite, which enabled us to successfully complete this study. We are extremely thankful to the Wiley Editing Services for editing, revising, proofreading and formatting this manuscript for us.

References

Bas, D.C., Rogers, D.M. & Jensen, J.H. 2008. Very fast prediction and rationalization of pKa values for protein-ligand complexes. *Proteins: Structure, Function, and Bioinformatics* 73(3): 765-783.

Cui, J., Li, F. & Shi, Z.L. 2019. Origin and evolution of pathogenic coronaviruses. *Nat. Rev. Microbiol.* 17(3): 181-192.

Durdagi, S., Aksoydan, B., Dogan, B., Sahin, K. & Shahraki, A. 2020. Screening of clinically approved and investigation drugs as potential inhibitors of COVID-19 main protease: A virtual drug repurposing study. *ChemRxiv. Preprint.* https://doi.org/10.26434/chemrxiv.12032712.v1.

Friesner, R.A., Banks, J.L., Murphy, R.B., Halgren, T.A., Klicic, J.J., Mainz, D.T., Repasky, M.P., Knoll, E.H., Shelley, M., Perry, J.K., Shaw, D.E., Francis, P. & Shenkin, P.S. 2004. Glide: A new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. *Journal of Medicinal Chemistry* 47(7): 1739-1749.

Huang, C., Wang, Y., Li, X., Ren, L., Zhao, J., Hu, Y., Zhang, L., Fan, G., Xu, J., Gu, X., Cheng, Z., Yu, T., Xia, J., Wei, Y., Wu, W., Xie, X., Yin, W., Li, H., Liu, M., Xiao, Y., Gao, H., Guo, L., Xie, J., Wang, G., Jiang, R., Gao, Z., Jin, Q., Wang, J. & Cao, B. 2020. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *Lancet.* 395(10223): 497-506.

Jin, Z., Du, X., Xu, Y., Deng, Y., Liu, M., Zhao, Y., Zhang, B., Li, X., Zhang, L., Peng, C., Duan, Y., Yu, J., Wang, L., Yang, K., Liu, F., Jiang, R., Yang, X., You, T., Liu, X., Yang, X., Bai, F., Liu, H., Liu, X., Guddat, L.W., Xu, W., Xiao, G., Qin, C., Shi, Z., Jiang, H., Rao, Z. & Yang, H. 2020. Structure of Mpro from COVID-19 virus and discovery of its inhibitors. *Nature* 582(7811): 289-293.

Li, F. 2016. Structure, function, and evolution of coronavirus spike proteins. *Annu. Rev. Virol.* 3(1): 237-261.

Ren, L.L., Wang, Y.M., Wu, Z.Q. & Wang, J.W. 2020. Identification of a novel coronavirus causing severe pneumonia in human: A descriptive study. *Chin. Med. J. (Engl)* 133(9): 1015-1024.

Sastry, G.M., Adzhigirey, M., Day, T., Annabhimoju, R. & Sherman, W. 2013. Protein and ligand preparation: Parameters, protocols, and influence on virtual screening enrichments. *Journal of Computer-Aided Molecular Design* 27(3): 221-234.

Shelley, J.C., Cholleti, A., Frye, L.L., Greenwood, J.R., Timlin, M.R. & Uchimaya, M. 2007. Epik: A software program for pK a prediction and protonation state generation for drug-like molecules. *Journal of Computer-Aided Molecular Design* 21(12): 681-691.

Shivakumar, D., Williams, J., Wu, Y., Damm, W., Shelley, J. & Sherman, W. 2010. Prediction of absolute solvation free energies using molecular dynamics free energy perturbation and the OPLS force field. *Journal of Chemical Theory and Computation* 6(5): 1509-1519.

Tang, Q., Song, Y., Shi, M., Cheng, Y., Zhang, W. & Xia, X.Q. 2015. Inferring the hosts of coronavirus using dual statistical models based on nucleotide composition. *Sci. Rep.* 5: 17155.

Thayil, S.M. & Thyagarajan, S. 2016. PA-9: A flavonoid extracted from *Plectranthus amboinicus* inhibits HIV-1 protease. *International Journal of Pharmacognosy and Phytochemical Research* 8(6): 1020-1024.

Wang, Z., Yang, B., Li, Q., Wen, L. & Zhang, R. 2020. Clinical features of 69 cases with coronavirus disease 2019 in Wuhan, China. *Clin. Infect. Dis.* 71(15): 769-777.

Wu, D., Wu, T., Liu, Q. & Yang, Z. 2020a. The SARS-CoV-2 outbreak: What we know. *Int. J. Infect Dis.* 94: 44-48.

Wu, F., Zhao, Su., Yu, B., Chen, Y-M., Wang, W., Song, Z-G., Hu, Y., Tao, Z-W., Tian, J-H., Pei, Y-Y., Yuan, M-L., Zhang, Y-L, Dai, F-H., Liu, Y., Wang, Q-M., Zheng, J-J., Xu, L., Holmes, E.C. & Zhang, Y-Z. 2020b. A new coronavirus associated with human respiratory disease in China. *Nature* 579(7798): 265-269.

Wu, A., Peng, Y., Huang, B., Ding, X., Wang, X., Niu, P., Meng, J., Zhu, Z., Zhang, Z., Wang, J., Sheng, J., Quan, L., Xia, Z., Tan, W., Cheng, G. & Jiang, T. 2020c. Genome composition and divergence of the novel coronavirus (2019-nCoV) Originating in China. *Cell Host Microbe* 27(3): 325-328.

Zakaryan, H., Arabyan, E., Oo, A. & Zandi, K. 2017. Flavonoids: Promising natural compounds against viral infections. *Archives of Virology* 162(9): 2539-2551.

Zhou, P., Yang, X-L., Wang, X-G., Hu, B., Zhang, L., Zhang, W., Si, H-R., Zhu, Y., Li, B., Huang, C-L., Chen, H-D., Chen, J., Luo, Y., Guo, H., Jiang, R-D., Liu, M-Q., Chen, Y., Shen, X-R., Wang, X., Zheng, X-S., Zhao, K., Chen, Q-J., Deng, F., Liu, L-L., Yan, B., Zhan, F-X., Wang, Y-Y., Xiao, G-F. & Shi, Z-L. 2020. A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature* 579(7798): 270-273.

Fawad Ahmad, Saima Ikram, Jamshaid Ahmad*, Irshad ur Rehman & Saeed Ullah Khattak
Centre of Biotechnology & Microbiology
University of Peshawar
Peshawar
Pakistan

Sadia Butt
Shaheed Benazir Bhutto Women University
Peshawar
Pakistan

Maryam Mushtaq
Northwest General Hospital & Research Center
Peshawar
Pakistan

*Corresponding author; email: jamshaidbiotech@yahoo.com

Received: 16 January 2021
Accepted: 17 March 2021

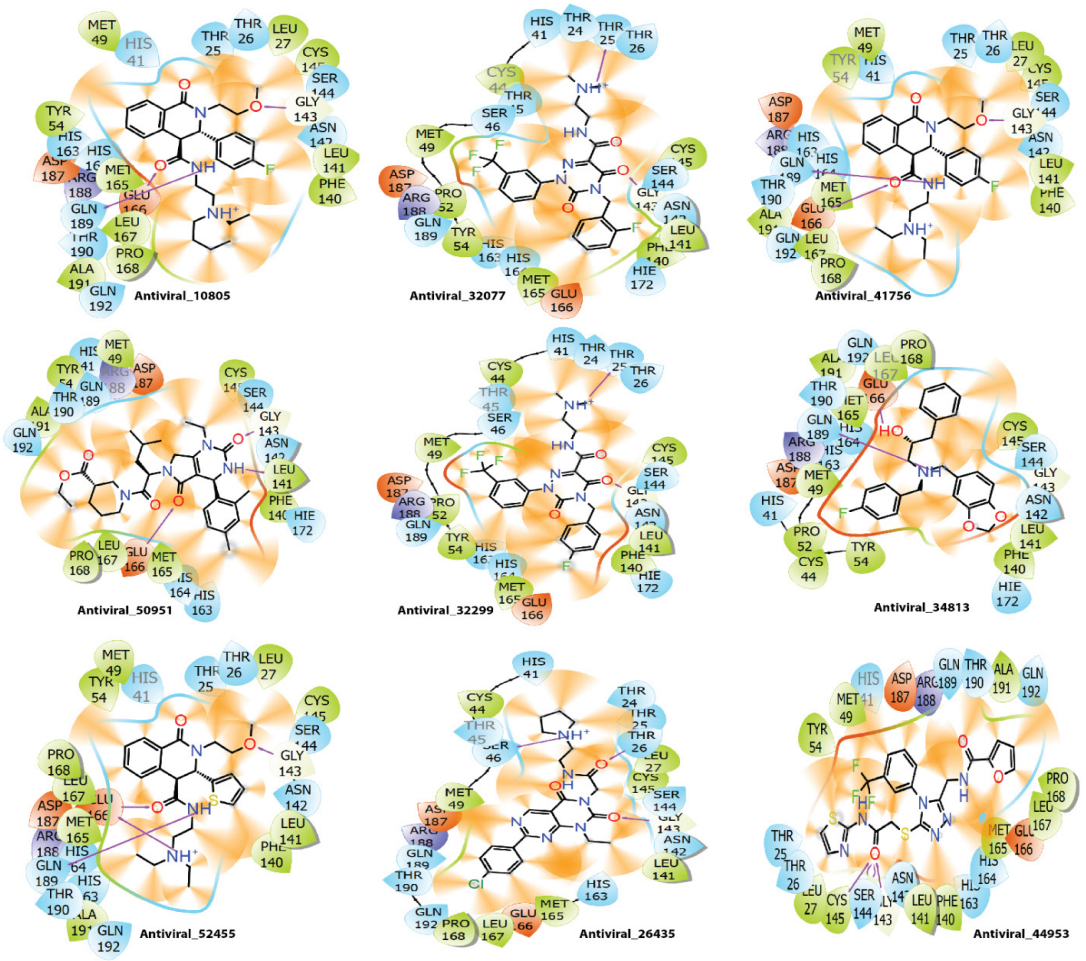


FIGURE S1. Interaction diagrams of the 9 remaining molecules among the 12 selected from the antiviral library, which were selected based on their docking scores

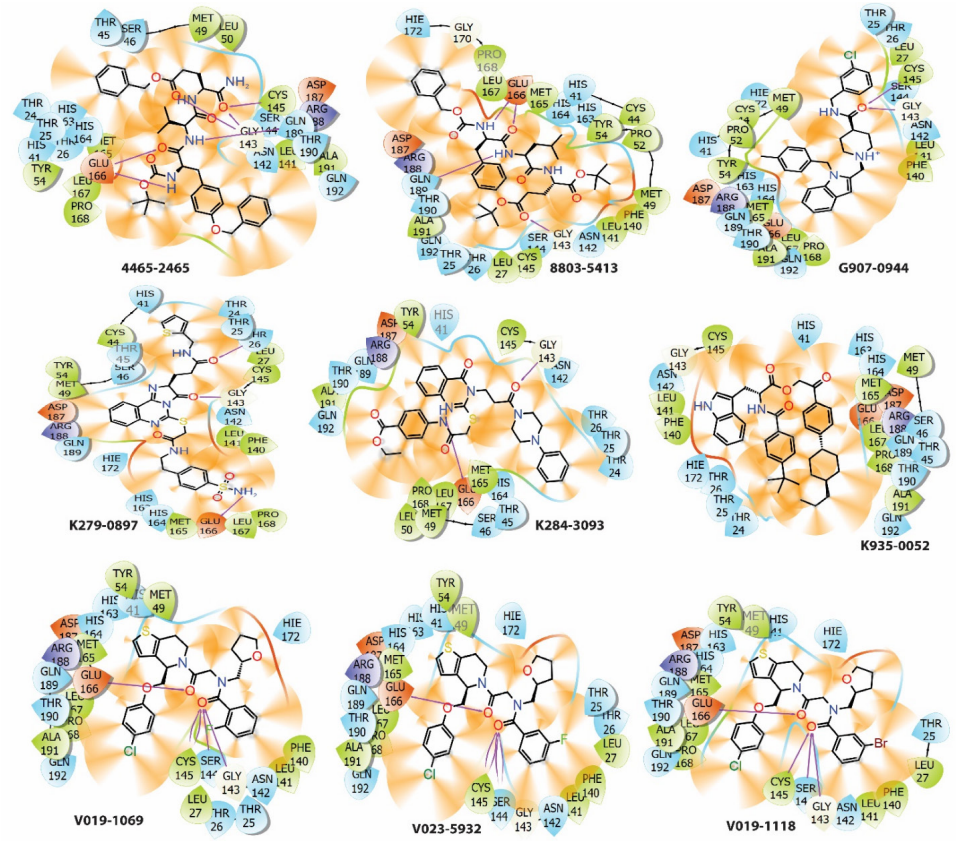


FIGURE S2. Interaction diagrams of the 9 remaining molecules among the 12 selected from the coronavirus library

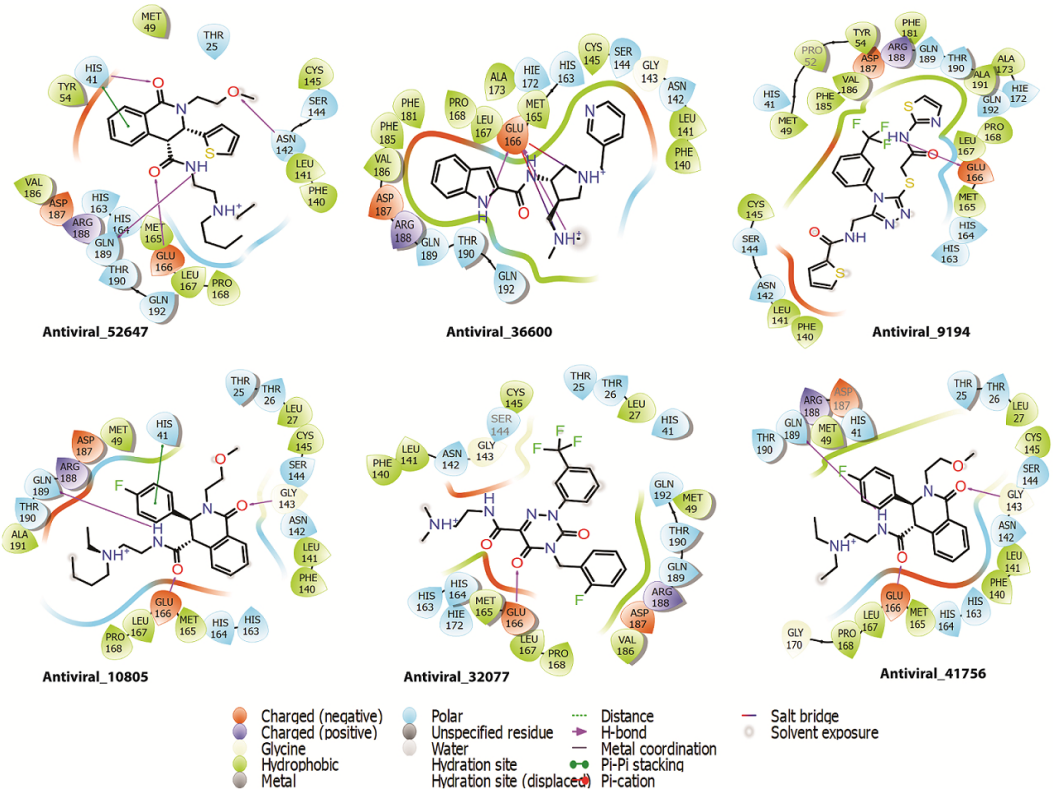


FIGURE S3. Interaction diagrams of the first 6 molecules among the 12 selected from the antiviral library using IFD algorithm for re-docking

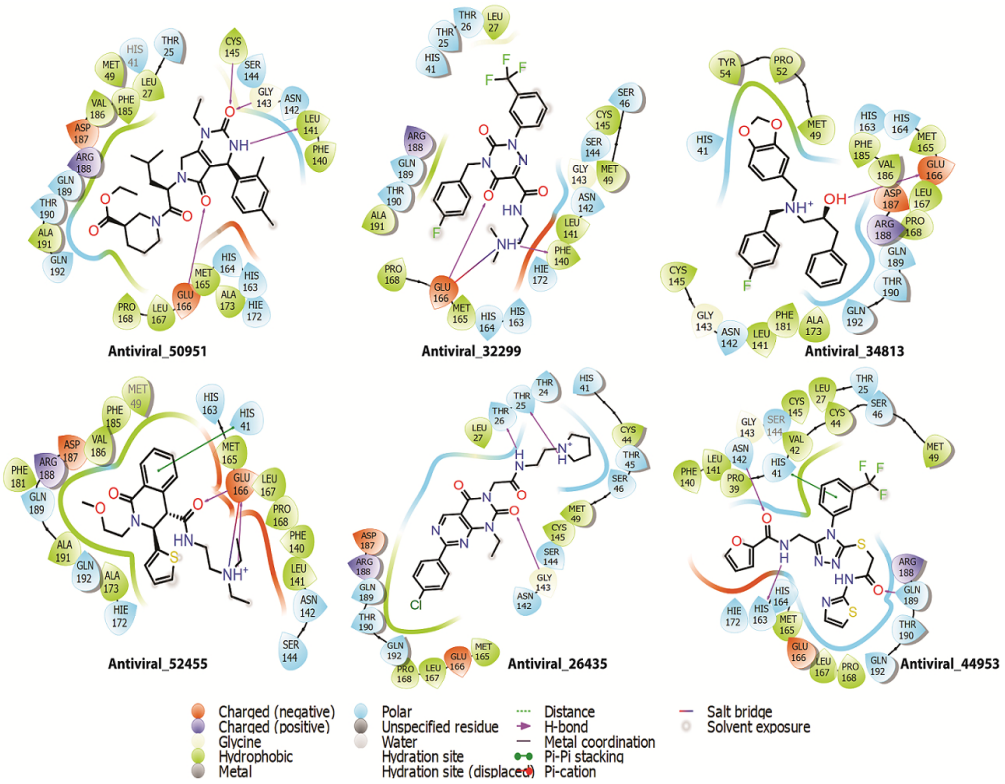


FIGURE S4. Interaction diagrams of the remaining 6 molecules among the 12 selected from the antiviral library using IFD algorithm for re-docking

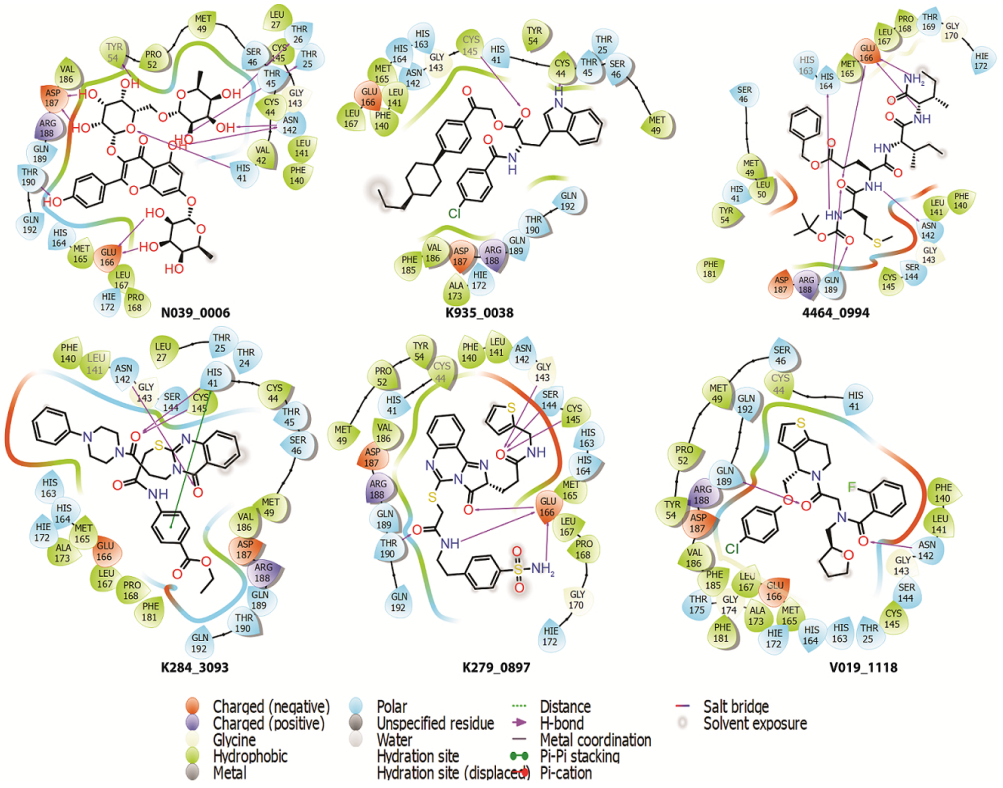


FIGURE S5. Interaction diagrams of the first 6 molecules among the 12 selected from the coronavirus library using IFD algorithm for re-docking

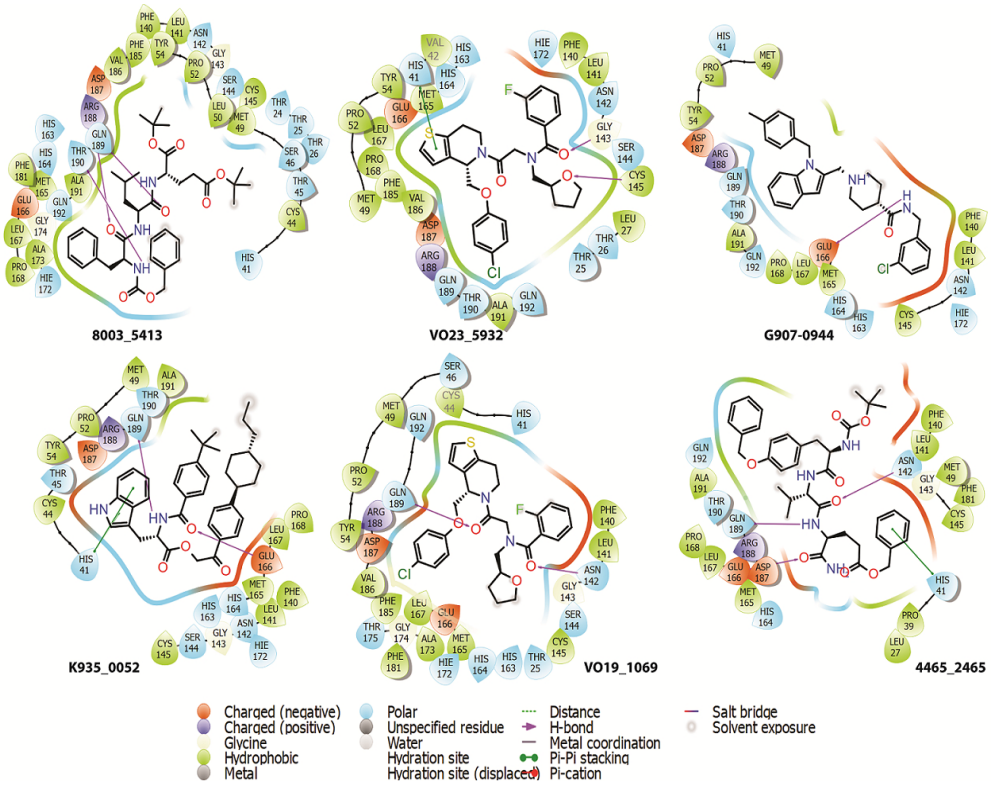


FIGURE S6. Interaction diagrams of the remaining 6 molecules among the 12 selected from the coronavirus library using IFD algorithm for re-docking

The csv files containing the details of the ligand binding parameters and the structures of the molecules can be downloaded from the folder shared at the following URL:

https://drive.google.com/drive/folders/1EwKp6RpAY27KTG8cG6GeehTvRObl6QrB