In silico Identification of Potent COVID-19 Main Protease Inhibitors from FDA Approved Antiviral Compounds and Active Phytochemicals through Molecular Docking: A Drug Repurposing Approach

Vaishali Chandel[1], Sibi Raj[1], Brijesh Rathi[2], Dhruv Kumar[1,*]

[1]Amity Institute of Molecular Medicine & Stem Cell Research (AIMMSCR), Amity University Uttar Pradesh, Sec-125, Noida-201313, India
[2]Department of Chemistry, Hansraj College University of Delhi, Delhi-110007 India

*Corresponding Author: Dr. Dhruv Kumar, J3-112, Amity Institute of Molecular Medicine & Stem Cell Research (AIMMSCR), Amity University Uttar Pradesh, Sec-125, Noida-201313, India Tel: 7082436598

Email: dhruvbhu@gmail.com, dkumar13@amity.edu

Abstract

The Coronavirus COVID-19 Main Proteases play critical role in the propagation of the Novel Coronavirus (COVID-19). Therefore, it is very important to identify a potential inhibitor against COVID-19 Main Proteases to inhibit the propagation of the Novel Coronavirus (COVID-19). We have applied a bioinformatics approach of drug repurposing to identify possible potent inhibitors against Novel Coronavirus (COVID-19) through targeting COVID-19 Main Protease from FDA approved antiviral compounds and from the library of active phytochemicals. The compounds were screened using PyRx virtual screening tool. Total 19 best compounds were identified after screening, based on their highest binding affinity with respect to the other screened compounds. Out of 19, 6 best compounds were further screened based on their binding affinity and best ADME properties. Nelfinavir exhibited highest binding energy -8.4 Kcal/mol and strong stability to interacted with the amino acid residues present on active site of COVID-19 Main Protease. In addition to Nelfinavir (-8.4), Rhein (-8.1), Withanolide D (-7.8), Withaferin A (-7.7), Enoxacin (-7.4), and Aloe-emodin (-7.4) also showed good binding affinity and best ADME properties. Our findings suggest that these compounds can be used as potential inhibitors against COVID-19 Main Protease, which could

be helpful in inhibiting the propagation of the Novel Coronavirus (COVID-19). Moreover, further investigation and validation of these inhibitors against Coronavirus would be very helpful to bring these molecules at the clinical settings.
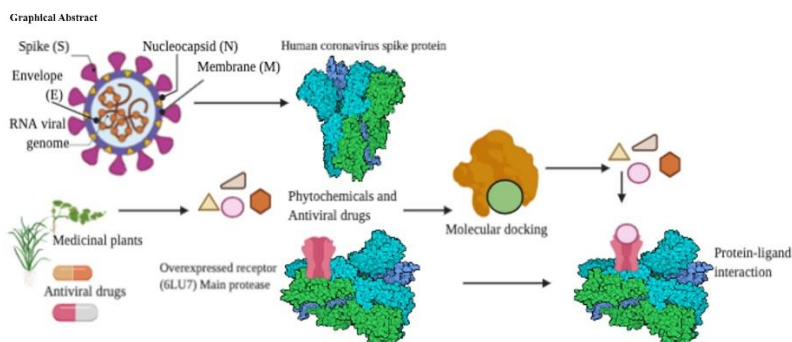
Key words: Novel Coronavirus, COVID-19, Protease, Molecular Docking, Drug Designing, ADME, Drug Repurposing

1. Introduction

The current outbreak of the novel coronavirus (CoVs) starting from December 2019 has been widely spread from the Hubei province of china to many other countries. The WHO emergency committee on January 2020 has declared a global health emergency based on the rate of increasing spread of the viral infection also estimated the fatality rate of about 4% [1]. Collaborative efforts from scientists worldwide are underway to understand the rapid spreading of the novel corona virus and to develop effective interventions to control and prevent the viral spread. Coronaviruses are positive-single stranded, enveloped large RNA viruses that majorly infect humans as well as a wide range of animals [2]. Tyrell and Bonne had first reported coronavirus in 1966, who cultivated the viruses from patients with common cold [3]. Depending on their morphological features representing a spherical virion with a core shell and projections in the surface resembling a solar corona, the viruses were termed as corona. Corona in Latin means crown based on their shape. Coronaviruses have four subfamilies which includes alpha-, beta-, gamma- and delta subtypes [4]. Alpha and beta coronaviruses are reported to be originated from mammals such as bats while gamma and delta viruses originate from mostly pigs and birds [5]. CoVs are large family of viruses that are common in various other species including camels, cattle, cats and bats. They range in the size of 60-80 nm with genome size varying between 26 and 32 kb. Recombination rates of CoVs are very high due to the ability to develop constant transcription errors and RNA Dependent RNA Polymerase (RdRP) jumps [6]. Most of the RNA content they have encode viral polymerase, RNA synthesis materials, and two large nonstructural polyproteins that are not involved in host response modulation (ORF1a-ORF1b). The other one third of the genome portion codes for four structural proteins (spike (S), envelope (E), membrane (M) nucleocapsid (N), and the other helper proteins [7]. CoVs have high mutation rates with the capability of causing infections in respiratory, gastrointestinal, hepatic and neurologic systems. The viruses are highly pathogenic in nature as they are also associated with severe acute respiratory syndrome (SARS). The novel corona virus (nCoV) emerged in 2019 has been the focus of global attention due to the pneumonia epidemic of unknown cause. The first case of pneumonia was reported on December 12, 2019 where possible pneumonia and corona virus infection were ruled out by clinicians.

The first step in viral infection is the interaction of spike protein with sensitive human cells [8]. After entering to the cells CoVs adapt to their human hosts by genome encoding and facilitating the expression of the genes that encode necessary accessory proteins. Genome alteration by CoVs are done through the mechanism of recombination, gene exchange, gene insertion or deletion [9]. The CoV subfamily is rapidly expanding with new generation sequencing applications that improve the detection and definition of novel CoV species. CoV classification is continually changing. According to the most recent classification of The International Committee on Taxonomy of Viruses (ICTV), there are four genera of thirty-eight unique species. SARS-CoV and middle east respiratory syndrome (MERS-CoV) attach to the host cell respectively and binds to the cellular receptor angiotensin converting enzyme 2 2 (SARS-CoV associated) and cellular receptor of dipeptidyl peptidase 4 (MERS-CoV associated) [10]. After infecting the host, the viral RNA manifests itself in the cytoplasm of the cell. Genomic RNA is modified through the process of encapsulation and polyadenylation and encodes several structural and non-structural genes. Proteases exhibiting chymotrypsin-like activity splits these polyproteins which drives the production of (-) RNA through replication as well as transcription [11]. During the replicative machinery the full length (-) RNA copies of the genome are used as a template for full length (+) RNA genomes. During the process of transcription, rna encoding all structural proteins and a subset of 7-9 sub-genomic RNAs are produced by discontinuous transcription. In the cytoplasm, viral nucleocapsids are combined from genomic RNA and R protein and then are budded into the lumen of the endoplasmic reticulum. Through the process of exocytosis virions are then released from the infected cell.

The released viruses then are capable to infect kidney cells, liver cells, intestines, and T lymphocytes, as well as the lower respiratory tract, where they form the main symptoms and signs. In notion of this, three patients having SARS-CoV infection were found to have CDT lymphocytes lower than 200 cells/mm3 [12]. MERS-CoV is able to affect human dendritic cells and macrophages in-vitro. T lymphocytes are also a target for the pathogen due to the characteristic CD26 rosettes. This virus can make the antiviral T-cell response irregular due to the stimulation of T-cell apoptosis, thus causing a collapse of the immune system.

Main protease domain (Mpro) has been reported to be a conserved target, in favour to design new inhibitors throughout the entire coronavirinae subfamily. The two-third region of 5' in the coronavirus genome consists of the of open reading frame I which encodes two large polypeptides of the replicase machinery: pp1a, and through ribosomal frameshift, pp1ab1. Two proteases encoded in the 5' region of ORF 1: papain-like protease (PLP) and 3C-like protease (3CL or Nsp5) cotranslationally cleaves the two polypetides into mature nonstructural proteins (NSPs) [13]. 3 CL protease is more commonly known as Mpro as it has a dominant role in the posttranslational processing of the replicase protein. Significant homology of Mpros in primary amino acid sequence as well as in 3D architechture has been reported in different human and animal CoVs. Also, they have a similar substrate binding pocket with a requirement for glutamine at P1 position and a preference for leucine/methionine at P2 position. This strong structural basis provides a loop hole to design a wide-spectrum anti CoV inhibitors. In general, there are few or no treatment options for viral diseases that occur suddenly and spread at a higher frequency. In notion of this note, there are no vaccine or effective treatment available to prevent the novel COVID-19 infection. Compounds are being tested in vitro against the COVID-19 infections and in humans for SARS-CoV and MERS-CoV trials. Studies analyzing the antiviral activity of type I interferon and type II interferon have provided significant results on interferon-beta (IFNb) as the most potent interferon that showed reduced in vitro replication of MERS-CoV. A report from South Korea has reported that the use of Lopinavir/Ritonavir (LPV/RTV- Anti-HIV drugs), EJMO 5 pegylated interferon and ribavirirn in combination showed successful viral clearance in human patients [14]. Although, remdesivir another antiviral drug was reported to be successful in United States of America. In vitro studies with this drug showed the termination of RNA transcription at early stage. There are publications demonstrating that remdesivir has a strong antiviral activity in epithelial cell cultures against SARS-CoV, MERS-CoV and related zoonotic bat CoVs. Several measures have to be taken in order to prevent the epidemic at a larger rate, such as early diagnosis, reporting isolation., supportive treatments, avoiding unnecessary panics. Also, basic preventive measures such as regular handwashing, using disinfectant solutions, avoiding contact with patients in order to prevent the spread of viruses by droplets. Healthcare staff should be informed about taking personal protective measures such as the use of gloves, eye masks and N95 masks during the examination of patients with a suspected history of COVID-19. The public services and facilities should provide decontaminating reagents for cleaning hands on a routine basis. Physical contact with wet and contaminated objects should be considered in dealing with the virus, especially agents such as fecal and urine samples that can potentially serve as an alternative route of transmission. A huge variety of biological active agents have been identified for their diverse therapeutical functions. Deep understanding of phytochemicals for antiviral activities have assumed greater importance in the last few decades [15]. A wide variety of active phytochemicals, including the flavonoids, terpenoids, organosulfur compounds, limonoids, lignans, sulphides, polyphenolics, coumarins, saponins, chlorophyllins, furyl compounds, alkaloids, polyines, thiophenes, proteins and peptides have been found to have therapeutic applications against different genetically and functionally diverse viruses. The antiviral mechanism of these agents may be explained on basis of their antioxidant activities, scavenging capacities, inhibiting DNA, RNA synthesis, inhibition of the viral entry, or

inhibiting the viral reproduction etc. Large number candidate substances such as phytochemicals and their synthetic derivatives have been identified by a combination of in vitro and in vivo studies in different biological assays.

The preliminary studies done till date are not approved for the therapeutic use against COVID-19 infected patients. Liu et al. (2020) have successfully crystallised the main protease (MPro) of COVID-19, PDB-ID 6LU7, which is now accessible to the globe. 6LU7 represents a potential target for the inhibition of CoV replication. Therefore, in our study we have identified 19 potential inhibitors and 6 best compounds (Nelfinavir, Rhein, Withaferin A, Withanolide D, Enoxacin and aloe-emodin) as potential inhibitor COVID-19 major protease. These inhibitors can be repurposed against COVID-19 major protease to control the spread of Coronavirus.

2. Material and Methods

Data sources

In this study, a dataset of 100 FDA approved antiviral compounds and 1000 active phytochemicals were obtained from FDA and Indian Medicinal Plants, Phytochemistry, and Therapeutics database [16, 17].

Preparation of receptor/membrane transporter

The atomic coordinates of the protein, COVID-19 (PDB ID- 6LU7) was downloaded from the RCSB PDB (protein data bank) database. Before analysis or docking, the charge assignment, solvation parameters and fragmental volumes to the protein was done using the Autodock Tool 4 (ADT) [18]. The protein molecule was further optimized using Autodock Tool for the molecular docking.

Preparation of ligands

The 3D SDF structure of all the compounds were downloaded from Pubchem database [19] and 2D ligand structures were designed using Chemsketch program (Figure-3). The ligands were optimized using Avogadro and converted into the PDB format with the help of Open Babel. In order to further simplify the analysis, ligands were first optimized and converted to PDBQT format using the graphical user interface version of PyRx virtual screening tool-python prescription 0.8.

Compound screening using PyRx program

Molecular screening of all the compound libraries was performed using PyRx software by autodock wizard as the engine for docking [20, 21]. During the docking period, the ligands were considered to be flexible and the protein was considered to be rigid. The configuration file for the grid parameters was generated using Auto Grid engine in Pyrex. The application was also used to know/predict the amino acids in the active site of the protein that interact with the ligands. The results less than 1.0Å in positional root-mean-square deviation (RMSD) were considered ideal and clustered together for finding the favourable binding. The highest binding energy (most negative) was considered as the ligand with maximum binding affinity.

Analysis and visualization

Visual analysis of the docking site was performed using Pymol version 2.3.4 and the results were validated using Autodock-Vina [22].

Active site identification

Active site, amino acid residues in 6LU7 was identified using an online software using protein-ligand indentifier profiler (PLIP) BIOTEC Du Dresden [23].

ADME analysis

On the basis of canonical SMILES of the selected ligands obtained from pubchem, ADME properties of the studied compound were calculated using online SwissADME program [24]. The major parameters for ADME associated properties such as Lipinski's rule of five, the solubility of the drug, pharmacokinetic properties and drug likeliness were considered. The values of the observe properties are presented in Table-3.

3.   Results and Discussion

Coronavirus belongs to a group of viruses which can infect humans and vertebrate animals. It has killed thousands of people around the globe with an increase in death rate every single day. The infection hampers liver, respiratory, central nervous system, and digestive of humans and animals. Our study was focused on the drug repurosing against the main protease in coronavirus (3CLpro/Mpro), (PDB-ID:6LU7), as a potential therapeutic target for the treatment of coronavirus. 6LU7 is the major protease (Mpro) in COVID-19 that has been repositioned and structured in PDB recently and is accessible to everyone in the world (Figure-1). The Mpro in coronavirus is very important for the proteolytic maturation of the virus. Mpro has been examined as a potential target to prevent the spread of infection by inhibiting viral polyprotein cleavage through blocking active sites of the protein (Figure-2). With this new discovery of Mpro structure in COVID-19, has provided an immense opportunity to identify potential drug candidates for the treatment of coronavirus [25]. In our study, we have applied a computational approach of drug repurosing in order to identify a specific therapeutic agent against COVID-19. We have created a database of 100 FDA approved antiviral compounds and 1000 active phytochemicals from plants. The compounds were screened using PyRx virtual tool and out of the 1000, 19 compounds were selected based on their best binding affinity with COVID-19 major protease (6LU7). Molecular docking was performed with the 19 compounds against COVID-19 (Mpro) structure.

Molecular docking is a computational method which aims to identify non-Covalent binding between protein (receptor) and a small molecule (ligand/Inhibitor). Docking predicts the mode of interaction between a target protein and a small ligand for an established binding site. Binding energy suggests the affinity of a specific ligand and strength by which a compound interacts with and binds to the pocket of a target protein. A compound with a lower binding energy is preferred as a possible drug candidate. In order to understand the effect of active antiviral and phytochemicals compounds on COVID-19, molecular docking of 19 active phytochemicals and FDA approved antiviral compounds were selected after screening from PyRx, was performed against COVID-19 (Table-2). Docking results of COVID-19 major protease (6LU7) with selected 6 compounds (Nelfinavir, Rhein, Withanolide D, Withaferin A, Enoxacin, Aloe-emodin) out of the selected 19 showed best docking score and were found to be best molecules at the target site of the protein. Out of the 6 compounds, Nelfinavir exhibited

the best docked score (-8.4 Kcal/mol) with COVID-19 protein. TRP207, ILE281, LEU282, PHE3, PHE291, GLN127, ARG4, GLY283, GLU288, LYS5, LYS137, TYR126, GLY138, TYR126, SER139 and VAL135 are the amino acid residues participating in the interaction at the binding pocket of COVID-19 (Figure-4). Nelfinavir is one of the recently identified anti-retroviral drug against HIV. It is a protease inhibitor used to limit the replication of the virus and boost immune function in individuals affected with HIV. Nelfinavir is a nucleoside reverse transcriptase inhibitors (NRTIs) which is evaluated as first-line therapy in HIV patients [26] (Figure-5). Withanolides are a group of naturally occurring steroid which are oxygenated and are present in medicinal plants of solanaceae family. Formulations of withanolides have been exploited in various pharmacological activities including immunomodulatory, antioxidant, antibacterial, antiviral, antitumor, angiogenesis inhibitor, hypnosedative and antiarthritic [27]. Withanolide D exhibited (-7.8Kcal/mol) binding affinity with 6LU7. LYS102, PHE103, VAL104, ARG105, ILE106, GLN107, GLN110, PHE294, PHE8, ASN151, TYR154, ASP153 are the amino acid residues participating in the interaction at the binding pocket of 6LU7 (Figure-6). Rhein showed (-8.1Kcal/mol) binding affinity with 6LU7. LYS102, VAL104, ILE106 GLN110, THR29, THR111, PHE294, ASP295, GLN127, PHE8, ASN151, ILE152, ASP153, SER158 are the amino acid residues participating in the interaction at the binding pocket of 6LU7. Rhein (4, 5-dihydroxyanthraquinone-2-carboxylic acid) is extensively found in several medical herbs, such as *Cassia tora* L., *Rheum palmatum* L., *Aloe barbadensis* Miller., and *Polygonum multiflorum* Thunb, which have been used medically in China for over a decade. It is a lipophilic anthraquinone and has many anti-inflammatory effects, including anticancer, hepatoprotective, antioxidant, nephroprotective, pharmacological effects, and antimicrobial activities [28](Figure-7). Withaferin A exhibited (-7.7Kcal/mol) binding affinity with 6LU7. PHE294, THR292, ASP295, ASP153, SER158, LYS102, PHE103, GLU178, ARG105, ILE106, GLN110, THR111, GLN178, VAL108 are the amino acid residues participating in the interaction at the binding pocket of 6LU7. Withaferin A is an active component and phytoconstituent of *Withania somnifera.* It has been exploited as a therapeutic potential and significantly validated for various pharmacological activities including neurological, cardioprotective, immunomodulatory, anti-cancer, anti-stress, neuroprotective activities (Figure-8) (pharmacological and analytical aspects of withaferin A). Enoxacin exhibited (-7.4Kcal/mol) binding affinity with 6LU7. ASP295, PHE294, THR292, GLY109, THR111, ILE106, VAL104, ASN151, ASP153, GLN110, PHE112, ILE152, PHE8, PHE112 are the amino acid residues participating in the interaction at the binding pocket of 6LU7. Enoxacin belongs to the class of 4-quinolone. It has been recently identified as anti-cancer, antiinflmmatory and anti-bacterial agent [29](Figure-9). Aloe-emodin exhibited (-7.4Kcal/mol) binding affinity with 6LU7. Aloe-emodin belongs to class of anthraquinones and possesses multiple anti-carcinogenic, anti-proliferative and anti-viral actions on humans [30](Figure-10). The molecular docking analysis in the present study showed the inhibition potential of 6 compounds, ranked by affinity (ΔG); Nelfinavir > Rhein > Withanolide D> Withaferin A> Enoxacin>Aloe-Emodin.

Lipinski's rule of five is a major criterion to evaluate drug likeliness and if a particular chemical compound with a certain biological and pharmacological activity has physical and chemical properties that would make it a likely orally active drug in humans. Lipinski's rule determines the molecular properties which are important for a drug's. pharmacokinetics in the human body such as absorption, distribution, metabolism, and excretion (ADME) [32]. Lipinski's rule of five states that (i) a molecular mass less than 500 daltons, (ii) no more than 5 hydrogen bond donors, (iii) no more than 10 hydrogen bond acceptors, (iv) an octanol–water partition coeffcient log P not greater than 5. Three or more than 3 violations do not fit into the criteria of drug likeliness and it is not considered in order to proceed with drug discovery.

ADME studies of selected 19 compounds showed that out of 19, 15 virtual hits were successful at passing through these ADME test filters (Table-3).

Conclusion

The drug repurposing approach would be the fast and most approapriate option to find therapeutic solutions for the Novel Coronavirus (COVID-19). The bioinformatics approach can be a very useful tool to identify potent inhibitors against the the Novel Coronavirus. In this study, we have used PyRx and Autodock-Vina to identify potent FDA approved inhibitors against COVID-19 Main Proteases which play crucial role in Coronavirus propagation. We have identified 19 potent inhibitors out of thousands compounds and found Nelfinavir, Rhein, Withanolide D, Withaferin A, Enoxacin and Aloe-emodin as most appropriate inhibitors against against COVID-19 Main Proteases. Our findings suggest the protentioal inhibitors agaianst COVID-19 Main Proteases, which can be further explored to test against Coronavirus (COVID-19) in pre-clinical and clinical settings.

Acknowledgements

Conflict of Interest

Authors declare no conflict of Interest

References:

[1]   M. Cascella, M. Rajnik, A. Cuomo, S. C. Dulebohn, and R. Di Napoli, Features, Evaluation and Treatment Coronavirus (COVID-*19)*. 2020.

[2]   A. R. Fehr and S. Perlman, "Coronaviruses: An overview of their replication and pathogenesis," in *Coronaviruses: Methods and Protocols*, 2015.

[3]   A. F. Bradburne, M. L. Bynoe, and D. A. J. Tyrrell, "Effects of a 'New' Human Respiratory Virus in Volunteers," *Br. Med. J.*, 1967.

[4]   C. I. Paules, H. D. Marston, and A. S. Fauci, "Coronavirus Infections-More Than Just the Common Cold," *JAMA - Journal of the American Medical Association*. 2020.

[5]   A. Banerjee, K. Kulcsar, V. Misra, M. Frieman, and K. Mossman, "Bats and coronaviruses," *Viruses*. 2019.

[6]   S. K. P. Lau *et al.*, "DisCoVery and Sequence Analysis of Four Deltacoronaviruses from Birds in the Middle East Reveal Interspecies Jumping with Recombination as a Potential Mechanism for Avian-to-Avian and Avian-to-Mammalian Transmission," *J. Virol.*, 2018.

[7]   K. R. Hurst, L. Kuo, C. A. Koetzner, R. Ye, B. Hsue, and P. S. Masters, "A Major Determinant for Membrane Protein Interaction Localizes to the Carboxy-Terminal Domain of the Mouse Coronavirus Nucleocapsid Protein," *J. Virol.*, 2005.

[8]   S. Belouzard, J. K. Millet, B. N. Licitra, and G. R. Whittaker, "Mechanisms of coronavirus cell entry mediated by the viral spike protein.," *Viruses*. 2012.

[9]   R. L. Graham and R. S. Baric, "Recombination, Reservoirs, and the Modular Spike: Mechanisms of Coronavirus Cross-Species Transmission," *J. Virol.*, 2010, doi: 10.1128/jvi.01394-09.

[10]   Z. Song *et al.*, "From SARS to MERS, thrusting coronaviruses into the spotlight," *Viruses*. 2019.

[11]   S. G. Sawicki, D. L. Sawicki, and S. G. Siddell, "A Contemporary View of Coronavirus Transcription," *J. Virol.*, 2007,

[12]   J. Tynell *et al.*, "Middle east respiratory syndrome coronavirus shows poor replication but significant

induction of antiviral responses in human monocyte-derived macrophages and dendritic cells," *J. Gen. Virol.*, 2016.

[13] K. P. Lim, L. F. P. Ng, and D. X. Liu, "Identification of a Novel Cleavage Activity of the First Papain-Like Proteinase Domain Encoded by Open Reading Frame 1a of the Coronavirus Avian Infectious Bronchitis Virus and Characterization of the Cleavage Products," *J. Virol.*, 2000.

[14] J. Cohen, "Can an anti-HIV combination or other existing drugs outwit the new coronavirus?," *Science.,* 2020.

[15] R. Naithani *et al.*, "Antiviral Activity of Phytochemicals: A Comprehensive Review," *Mini-Reviews Med. Chem.*, 2008.

[16] S. Choudhuri, J.A. Symons, J. Deval, "Innovation and trends in the development and approval of antiviral medicines:1987-2017 and beyond." *Antiviral res.,*2018.

[17] K. Mohanraj, B. Shanmugun, R.P. Vivel-ananth, R.P. Bharath Chand, S.R. Aparna, P. Mangalapandi, A. Samal, "IMPAAT: A curated database of Indian Medicinal Plants, Phytochemistry, And Therapeutics." *Sci rep.* 2018.

[18] G. Morris, R. Huey, W. Lindstrom, et al. AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. *J. Comput. Chem.* 2009.

[19] N.M. O'Boyle, M. Banck, C.A. James, et al. Open Babel: An Open chemical toolbox. *J. Cheminform.* 2011.

[20] S.Dallakyan, A.J. Olson. " Small-molecule library screening by docking with PyRx." *Methods mol biol.* 2015.

[21] N.S. Pagadala, K.Syed, T.Jack." Software for molecular docking : a review." *Biophys Rev.* 2017.

[22] D. Seeliger, B.L.De. Groot. " Ligand docking and binding site analysis with PyMOL and Autodock/vina." *J Comput Aided Mol Des.*" 2010

[23] S. Salentin, V. Joanchim Haupt, M.F. Adasme, M. Schroeder. "PLIP: fully automated protein-ligand interaction profiler." *Nucleic acid research.*2015.

[24] A. Diana, O. Michielin, V.Zoete." SwissADME: a free web tool to evaluate pharmacokinetics, drug-likeness and medicinal chemistry friendliness of small molecules." *Sci rep.* 2017.

[25] S. Khaerunnisa , H. Kurniawan, R. Awaluddin, S. Suhartati, S. Soetjipto. "Potential Inhibitor of COVID-19 Main Protease (Mpro) from Several Medicinal Plant Compounds by Molecular Docking Study." *Preprints.* 2020.

[26] A.E.Bardsley, G.L. Plosker." Nelfinavir: an update on its use in HIV infection." *Drugs.* 2000.

[27] S. Amritpal, D. Sanjeev, S. Hermanpreet, S. Jaswinder, K. Shankar." Withanolides: Phytoconstituents with significant pharmacological activities." *International journal of green pharmacy.* 2010.

[28] Y.X. Zhou, W. Xia, W. Yue, C. Peng, K.Rehman, H.Zhang."Rhein: A review of pharmacological activities." Evid Based Complement Alternat Med.

[29] J.M. Henwood, J.P.Monk, " Enoxacin. A review of its antibacterial activity, pharmacokinetic properties and therapeutic use."*Drugs,*1988.

[30] B. Sanders, A.M. Ray, S. Goldberg, T. Clark, H. R. McDaniel, S. E. Atlas, A. Farooqi, J. Konefal, L. C. Lages, J. Lopez, A. Rasul, E. Tiozzo, J.M. Woolger, and J. E. Lewis. " Anti-cancer effects of aloe-emodin: a systematic review." *J Clin Transl Res.* 2018.

Figure Legends

Figure-1: Figure shows the FASTA sequence and crystal structure of COVID-19 main protease (6LU7). 6LU7 is the main protease found in COVID-19, which been repositioned and structured in PDB and is accessible to public, as of early February 2020.

Figure-2: Represents the active sites (Ligand binding) in 6LU7. 6LU7 has 2 interacting chains. (a) Small molecule- 02J (5methylisoxazole-3-carboxylic acid) Interacting chain A (b) PJE (composite ligand) PJE-C-5 Interacting chains A, C.

Figure-3: Chemical structures of lead compounds: (A) Nelfinavir; (B) Rhein; (C) Aloe-emodin; (D) Enoxacin; (E) Withanolide D; (F) Withaferin A

Figure-4: Docking analysis and visualisation of 6LU7 binding with Enoxacin (A), Withaferin A (B), Nelfinavir (C), Withanolide D (D), Rhein (E), Aloe-emodin
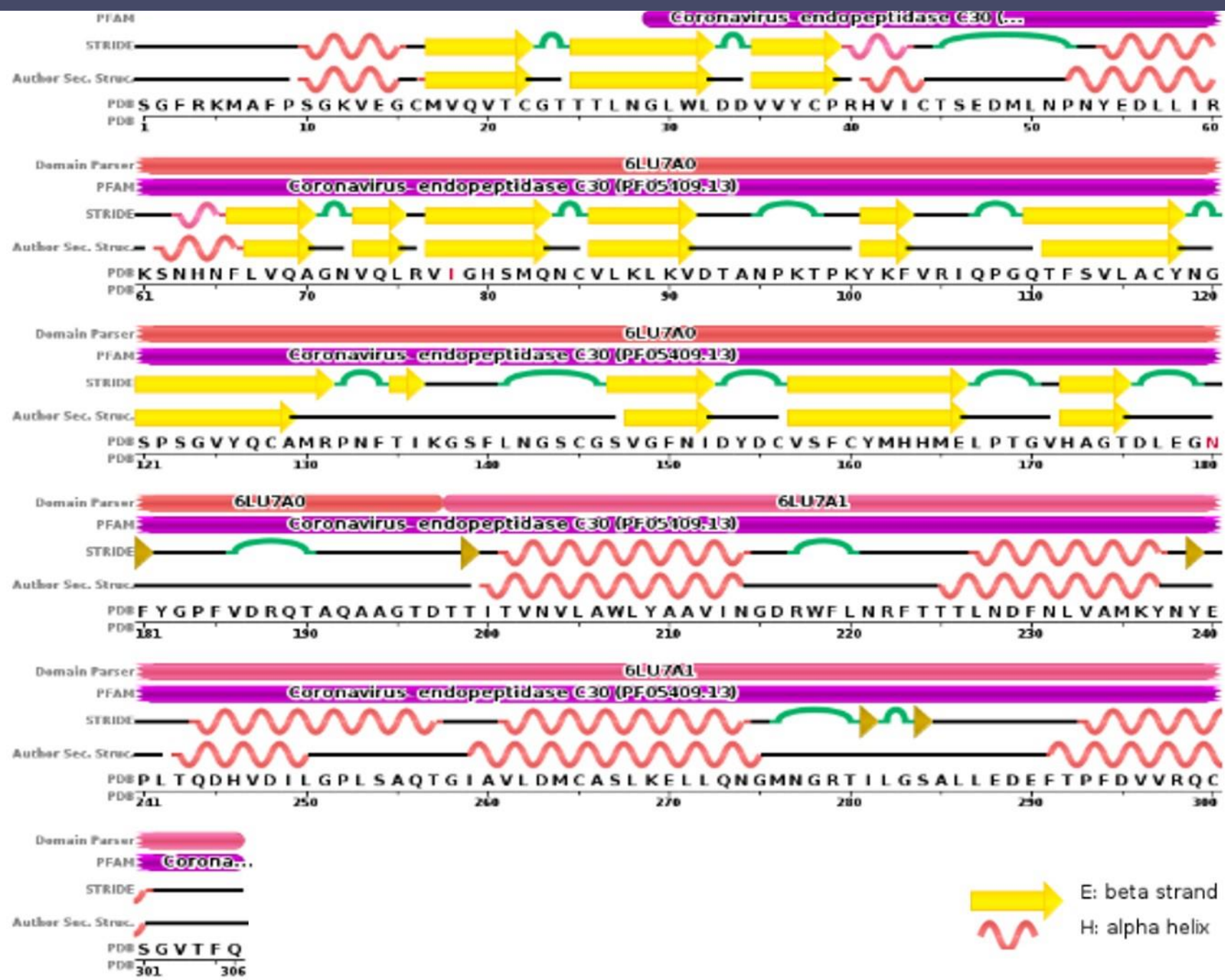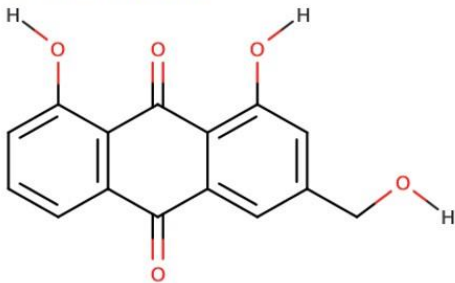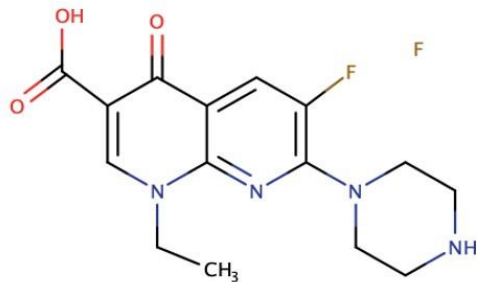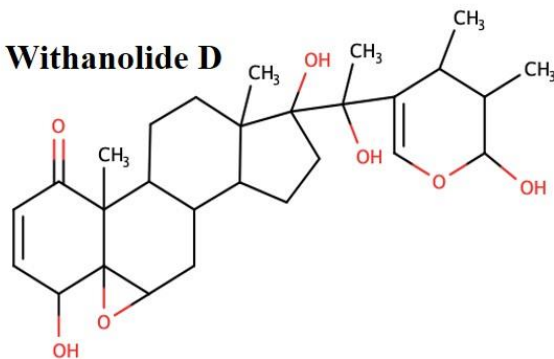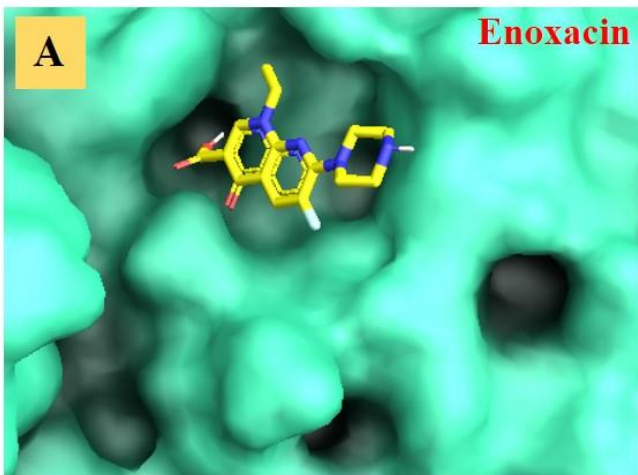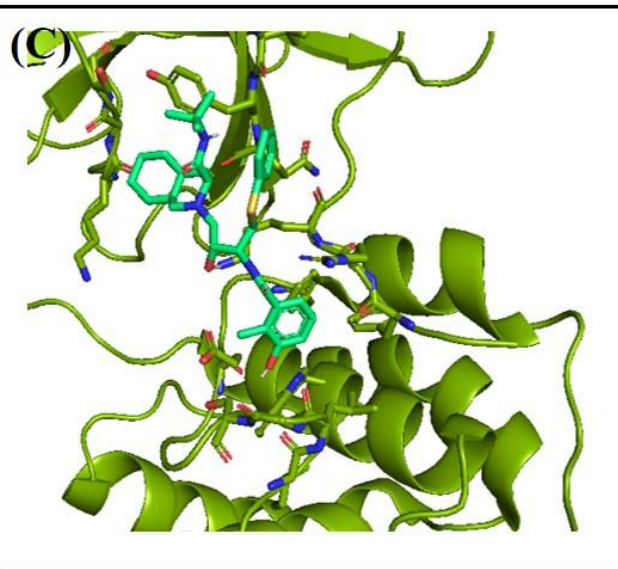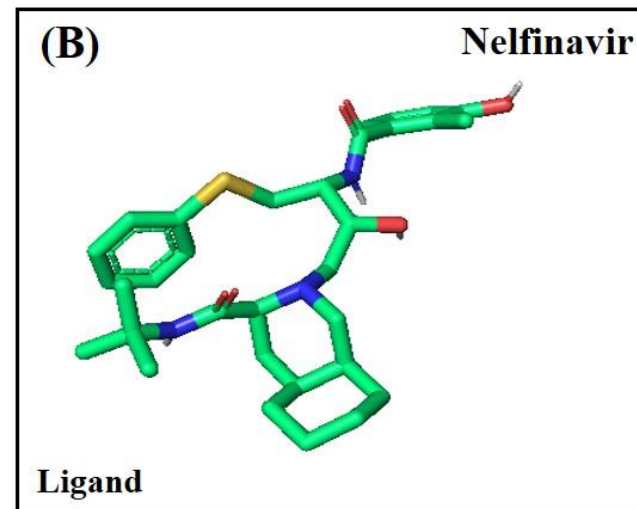
Figure-5: Molecular docking analysis between 6LU7 and Nelfinavir. (A) 3D ribbon structure of Receptor/Protein (6LU7). (B) 3D ligand structure (Nelfinavir). (C) Protein-ligand Interaction. (D) Interaction between the active site residues of the protein and ligand.

Figure-6: Molecular docking analysis between 6LU7 and Withanolide D. (A) 3D ribbon structure of Receptor/Protein (6LU7). (B) 3D ligand structure (Withanolide D). (C) Protein-ligand Interaction. (D) Interaction between the active site residues of the protein and ligand.

Figure-7: Molecular docking analysis between 6LU7 and Nelfinavir. (A) 3D ribbon structure of Receptor/Protein (6LU7). (B) 3D ligand structure (Nelfinavir). (C) Protein-ligand Interaction. (D) Interaction between the active site residues of the protein and ligand.

Figure-8: Molecular docking analysis between 6LU7 and Withaferin A. (A) 3D ribbon structure of Receptor/Protein (6LU7). (B) 3D ligand structure (Withaferin A). (C) Protein-ligand Interaction. (D) Interaction between the active site residues of the protein and ligand.

Figure-9: Molecular docking analysis between 6LU7 and Enoxacin. (A) 3D ribbon structure of Receptor/Protein (6LU7). (B) 3D ligand structure (Enoxacin). (C) Protein-ligand Interaction. (D) Interaction between the active site residues of the protein and ligand.

Figure-10: Molecular docking analysis between 6LU7 and Aloe-emodin. (A) 3D ribbon structure of Receptor/Protein (6LU7). (B) 3D ligand structure (Aloe-emodin). (C) Protein-ligand Interaction. (D) Interaction between the active site residues of the protein and ligand.

Amino Acid Sequence (FASTA) of COVID-19 Main Protease PDB-ID: (6LU7)

SGFRKMAFPSGKVEGCMVQVTCGTTTLNGLWLDDVVYCPRHVICTSEDMLNPNYEDLLIRKSNHNFLV
QAGNVQLRVIGHSMQNCVLKLKVDTANPKTPKYKFVRIQPGQTFSVLACYNGSPSGVYQCAMRPNFTI
KGSFLNGSCGSVGFNIDYDCVSFCYMHHMELPTGVHAGTDLEGNFYGPFVDRQTAQAAGTDTTITVNV
LAWLYAAVINGDRWFLNRFTTTLNDFNLVAMKYNYEPLTQDHVDILGPLSAQTGIAVLDMCASLKELLQ
NGMNGRTILGSALLEDEFTPFDVVRQCSGVTFQ

COVID-19 Main Protease
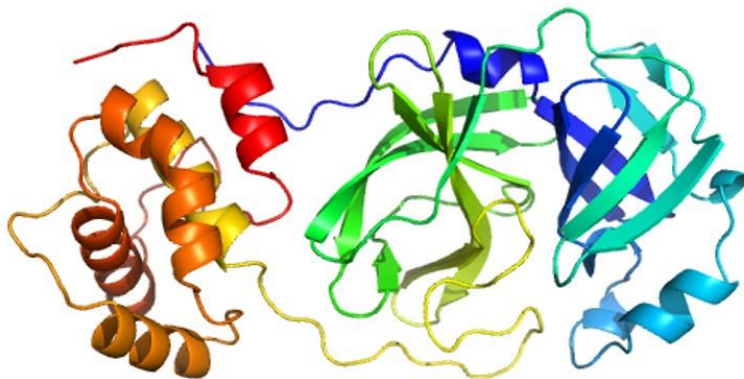
Ribbon Structure Representation

COVID-19 Main Protease
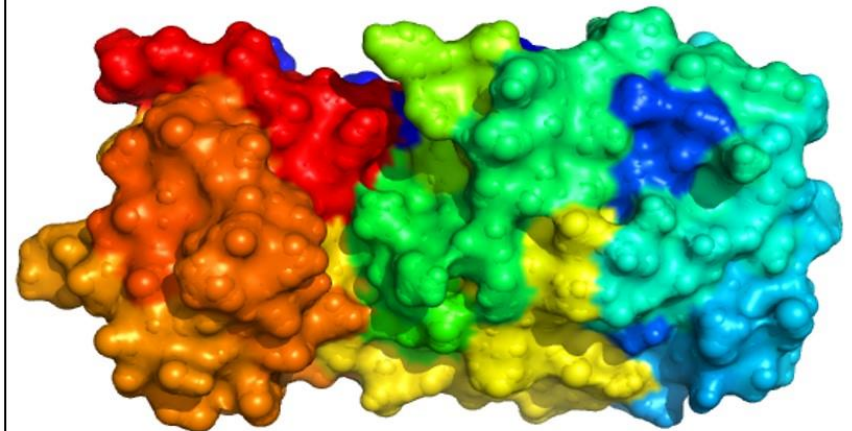
Surface Representation

Small Molecule:02J (5-Methylisoxazole-3-carboxylic acid) Interacting chain A

PRO-168

PJE (composite ligand) PJE-C-5 Interacting chains A,C

THR-26
SER-144
HIS-163
GLY-143
HIS-164
THR-25
GLU-166
VAL-3
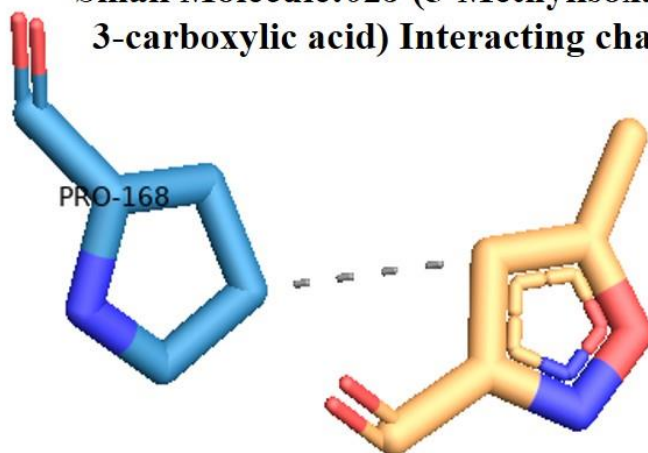
Protein
Ligand
Water
Charge Center
Aromatic Ring Center
Metal Ion
Hydrophobic Interaction
Hydrogen Bond
Water Bridge
π-Stacking (parallel)
π-Stacking (perpendicular)
π-Cation Interaction
Halogen Bond
Salt Bridge
Metal Complexation

(A) **Nelfinavir**

(B) **Rhein**

(C) **Aloe-emodin**

(D) **Enoxacin**

(E) **Withanolide D**

(F) **Withaferin A**

(A) COVID-19 Main Protease — Receptor/Protein

(B) Nelfinavir — Ligand

(C) Protein-ligand interaction at active site

(D) Ligand participating in amino-acid interaction

**(A)** COVID-19 Main Protease — Receptor/Protein

(B) Withanolide D — Ligand

(C) Protein-ligand interaction at active site

(D) Ligand participating in amino-acid interaction

(A) COVID-19 Main Protease — Receptor/Protein

(B) Rhein — Ligand

(C) Protein-ligand interaction at active site

(D) Ligand participating in amino-acid interaction

(A) COVID-19 Main Protease — Receptor/Protein

**(B) Withaferin A** — Ligand

(C) Protein-ligand interaction at active site

(D) Ligand participating in amino-acid interaction

(A) COVID-19 Main Protease

Receptor/Protein

(B) Enoxacin

Ligand

(C) Protein-ligand interaction at active site

(D) Ligand participating in amino-acid interaction

**Protein-ligand interaction
at active site**
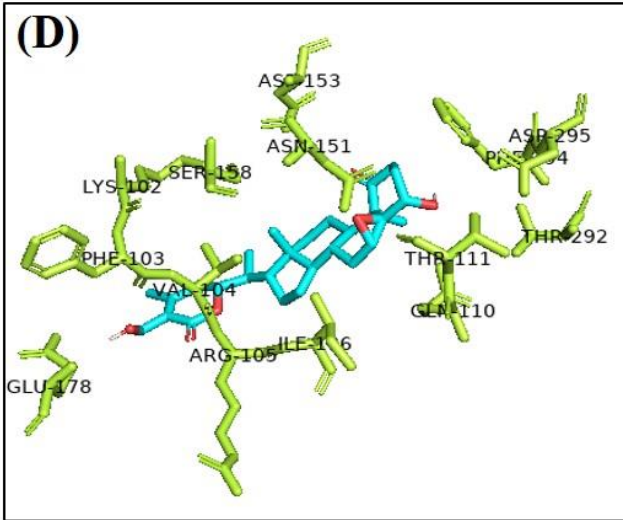
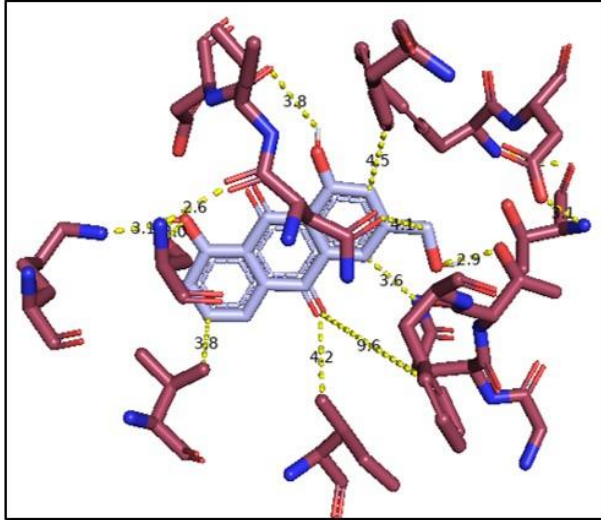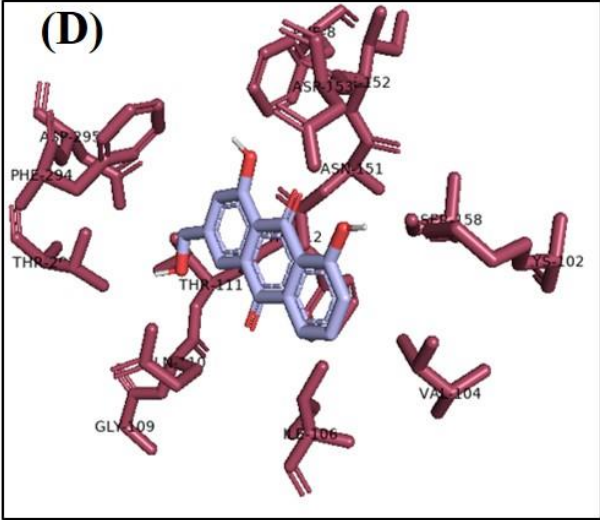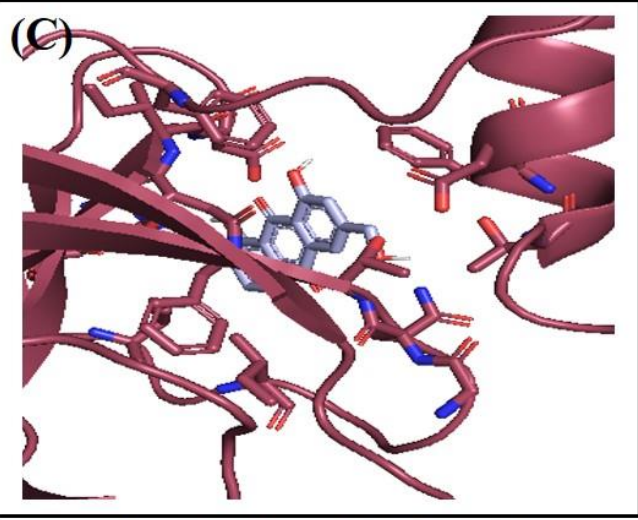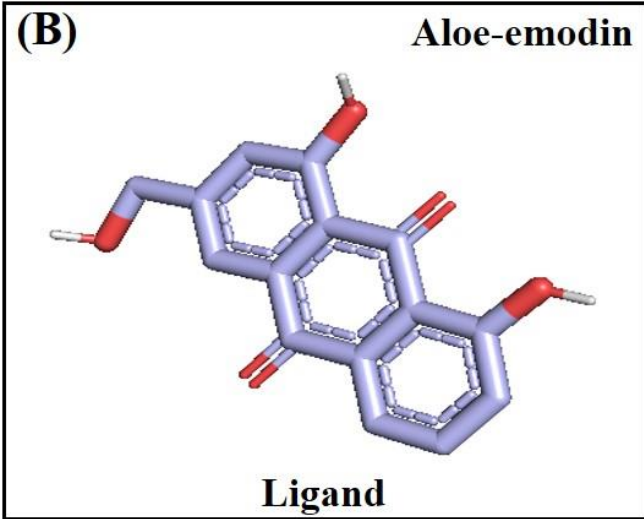**Ligand participating in amino-acid interaction**
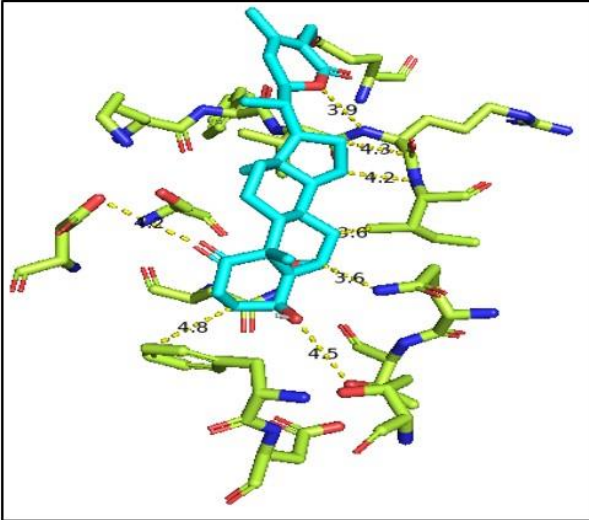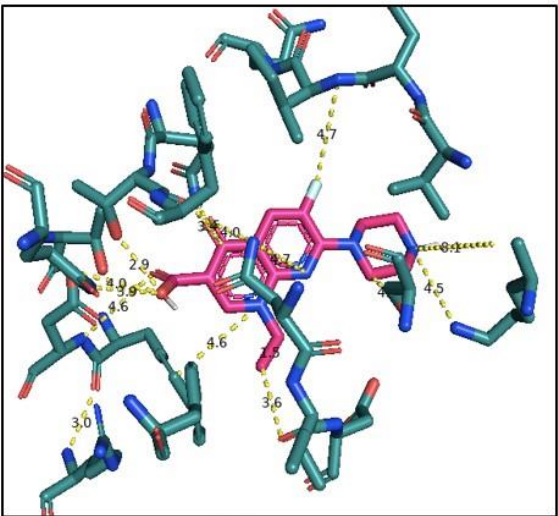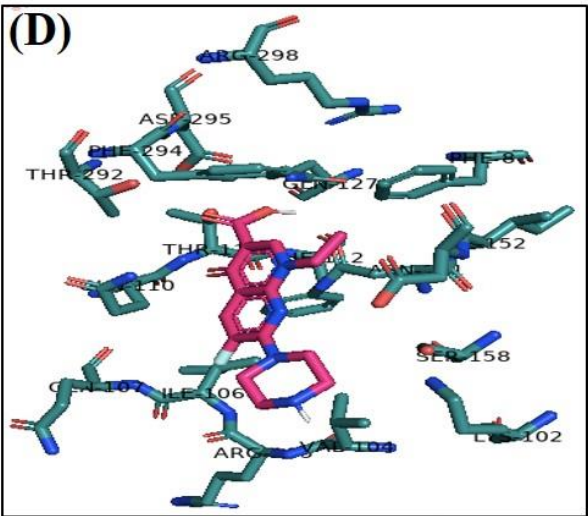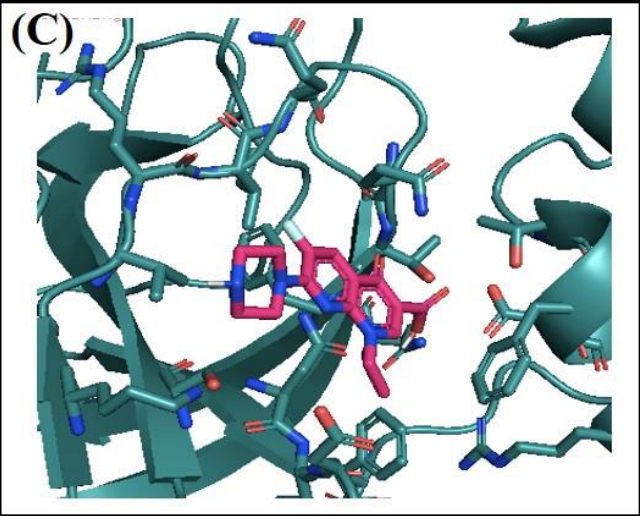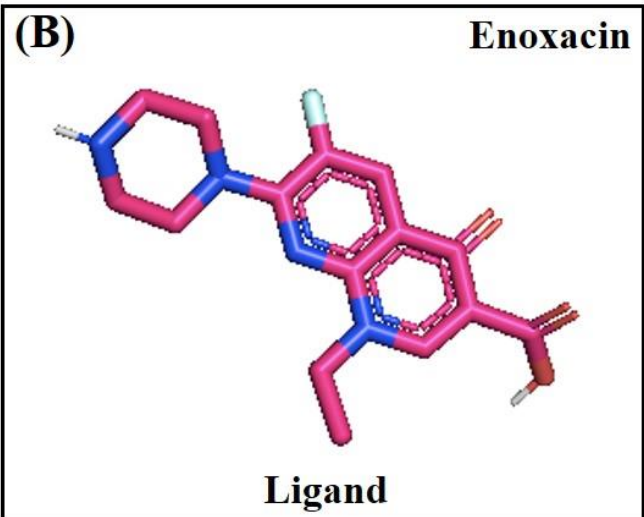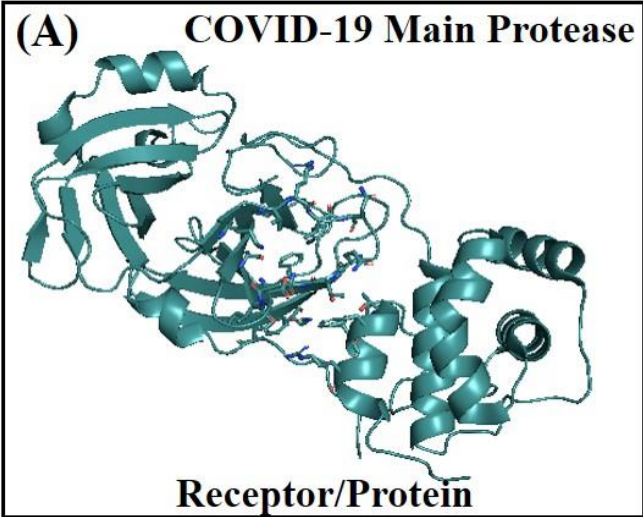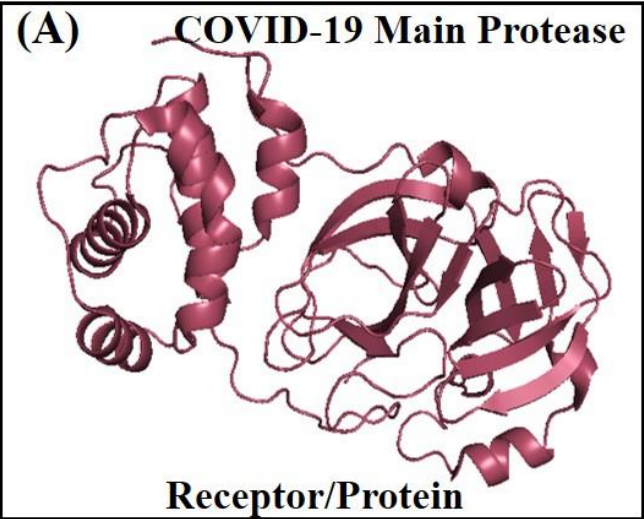
Table-1 (a): Active site properties of 6LU7 (Interacting chain A-02J)

Index	Residue	Amino Acid	Distance
1	168A	PRO	3.53

Table-1 (b): Active site properties of 6LU7 (Interacting chain A PJE-C5)

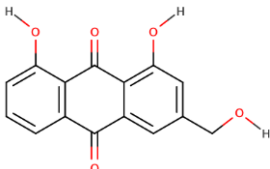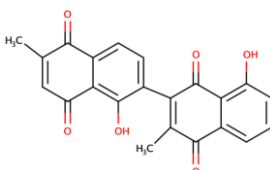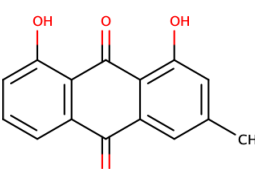| Index	Residue	Amino Acid	Distance
1	3C	VAL	3,98
2	25A	THR	3.74
3	26A	THR	3.88

Table-1 (c): Active site properties of 6LU7 (Interacting chain C PJE-C5)
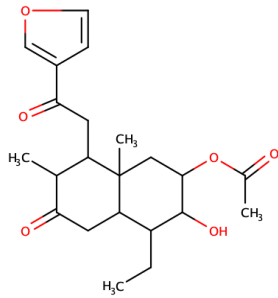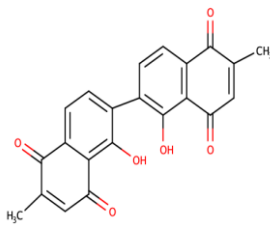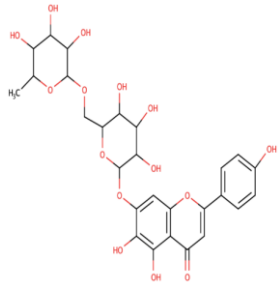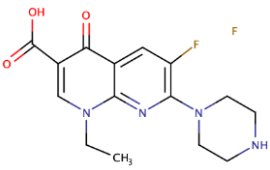
Index	Residue	Amino Acid	Distance D-A	Distance H-A	Donor angle	Protein donor	Sidechain
1	143A	GLY	2.87	2.00	145.90	+	-
2	144A	SER	3.99	3.65	104.01	+	+
3	163A	HIS	1.77	1.77	116.24	-	+
4	164A	HIS	1.85	1.85	161.75	-	-
5	166A	GLU	3.48	2.60	144.50	-	+

Table-2: Molecular docking analysis of antiviral compounds against COVID-19 Major Protease (6LU7).

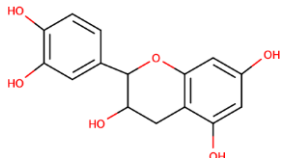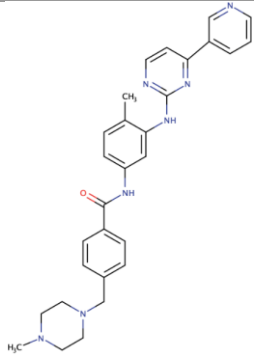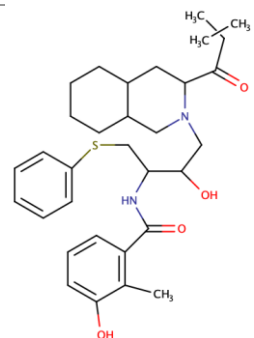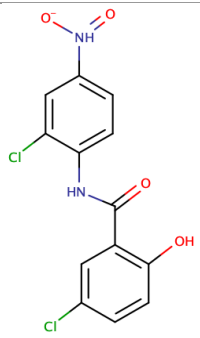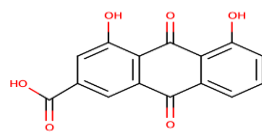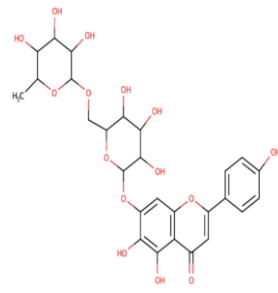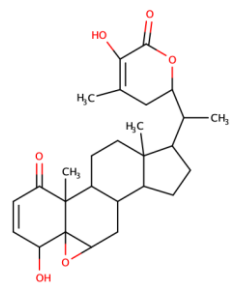| Protein	Compound	Binding Affinity (kcal/mol)	Amino acid residues
COVID-19 Major Protease (6LU7)	Aloe-emodin	-7.4	ASP295, PHE294, THR292, GLY109, THR111, ILE106, VAL104, ASN151, ASP153, GLN110, PHE112, ILE152, PHE8, PHE112
	Chitranone	-7.0	LYS5, TYR126, GLN127, SER139, GLY138, LYS137, ARG131, ASP197, LEU286, LEU287, GLU288, ASP289, GLU290, SER139
	Crysophanol	-7.0	PHE294, GLN110, THR111, GLN107, ILE106, ARG105, VAL104, LYS102, ASP153, SER158, ASN151, PHE8, ILE152
	Diterpene	-7.1	ASP295, PHE294, THR292, THR111, PHE8, GLN110, GLN107, ILE106, ARG105, VAL104, SER158, ASP153, ILE152, ASN151
	Elliptinone	-6.9	LYS5, GLN127, GLU290, ALA129, CYS128, TYR126, GLY124, VAL125, SER139
	Emetine	-7.0	PHE103, VAL104, SER158, ASN151, THR111, GLN110, ARG105, GLN107, ILE106, ASP153, LYS102
	Enoxacin	-7.4	THR292, ASP295, ARG298, GLN127, VAL104, LYS102, THR292, PHE8, GLN107, ILE106, PHE294
	+(-)Epicatechin	-7.6	LYS102, VAL104, ARG105, ILE106, GLN107, THR111, THR292, PHE294, ASP285, GLN110, ASN151, ILE152, ASP153, PHE8, GLN127, ASP295
	Imatinib	-7.4	ARG131, LYS137, GLN127, TYR126, LYS5, SER284, LEU286, LEU287, GLU288, PHE291, GLU290, ASP289, THR199, ASP197
	Nelfinavir	-8.4	TRP207, ILE281, LEU282, PHE3, PHE291, GLN127, ARG4, GLY283, GLU288, LYS5, LYS137, TYR126, GLY138, TYR126, SER139, VAL135
	Niclosamide	-7.1	ASP153, PHE8, GLN127, PHE112, THR111, ILE106, ARG105, VAL104, LYS102, SER158, ASN151, ASP295, PHE294, THR292, GLN110
	Rhein	-8.1	LYS102, VAL104, ILE106 GLN110, THR29, THR111, PHE294, ASP295, GLN127, PHE8, ASN151, ILE152, ASP153, SER158
	Scutellarein 7-rutinoside	-7.3	GLU178, ASP243, GLN116,VAL128, QLQ8, PHE9, TYR156, GLN128, ARG106
	Withaferin A	-7.7	PHE294, THR292, ASP295, ASP153, SER158, LYS102, PHE103, GLU178, ARG105, ILE106, GLN110, THR111, GLN178, VAL108
	Withanolide D	-7.8	LYS102, PHE103, VAL104, ARG105, ILE106, GLN107, GLN110, PHE294, PHE8, ASN151, TYR154, ASP153
	27-HYDROXY	-7.3	GLU288, LYS137, TYR126, GLN127, VAL125, ALA7, MET6, LYS5, GLU290
	24-Methylcholesta-	-7.3	LYS104, THR294, ASP153, ARG105, VAL108, MET7, LYS72, LEU283, TYR124, VAL127, GLU242
	17-hydroxywithanolide	-7.0	GLU288, GLU290, LYS137, GLY138, SER139, VAL125, TYR126, GLN127, CYS128, LYS5
	Aswagandhanolide	-8.1	SER139, GLY138, LYS137, ARG131, GLN290, ASP289, GLU288, LEU286, LYS5, TYR126, VAL125, HIS172, GLY179
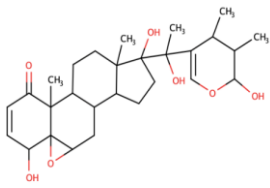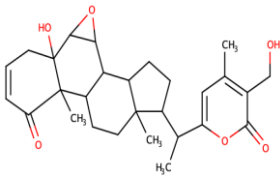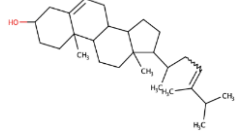
Table-3: ADME Properties of selected COVID-19 major protease inhibitors
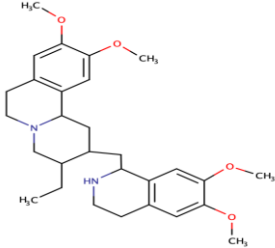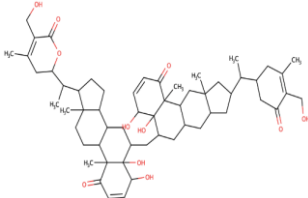
S. No	Compound	Molecular formula	ADME Properties (Lipinki's Rule of Five)		Structure	Drug likeliness
			Properties	**Value**		
1.	Aloe-emodin	C15H10O5	Molecular weight (<500Da)	270.24		Yes
			LogP (<5)	1.5		
			H-Bond donor (5)	3		
			H-bond acceptor (<10)	5		
			Violations	0		
2.	Chitranone	$C_{22}H_{14}O_6$	Molecular weight (<500Da)	374.34		Yes
			LogP (<5)	2.8		
			H-Bond donor (5)	2		
			H-bond acceptor (<10)	6		
			Violations	0		
3.	Chrysophanol	$C_5H_{10}O_4$	Molecular weight (<500Da)	254.24		Yes
			LogP (<5)	2.3		
			H-Bond donor (5)	2		
			H-bond acceptor (<10)	4		
			Violations	0		
			Molecular weight (<500Da)	392.44		Yes

4.	Diterpene	$C_{21}H_{28}O_7$				
			LogP (<5)	1.6		
			H-Bond donor (5)	2		
			H-bond acceptor (<10)	7		
			Violations	0		
5.	Elliptinone	$C_{22}H_{14}O_6$	Molecular weight (<500Da)	374.34		Yes
			LogP (<5)	2.9		
			H-Bond donor (5)	2		
			H-bond acceptor (<10)	6		
			Violations	0		
6.	Emetine	$C_{29}H_{40}N_2O_4$	Molecular weight (<500Da)	480.64		Yes
			LogP (<5)	4.1		
			H-Bond donor (5)	1		
			H-bond acceptor (<10)	6		
			Violations	0		
7.	Enoxacin	$C_{15}H_{17}FN_4O_3$	Molecular weight (<500Da)	320.32		Yes
			LogP (<5)	0.9		
			H-Bond donor (5)	2		
			H-bond acceptor (<10)	6		

			Violations	0		
8	+(-)Epicatechin	$C_{15}H_{14}O_6$	Molecular weight (<500Da)	290.27		Yes
			LogP (<5)	0.8		
			H-Bond donor (5)	5		
			H-bond acceptor (<10)	6		
			Violations	0		
9.	Imatinib	$C_{29}H_{31}N_7O$	Molecular weight (<500Da)	493.60		Yes
			LogP (<5)	3.38		
			H-Bond donor (5)	2		
			H-bond acceptor (<10)	8		
			Violations	0		
10.	Nelfinavir	$C_{32}H_{45}N_3O_4S$	Molecular weight (<500Da)	567.78		Yes
			LogP (<5)	4.3		
			H-Bond donor (5)	4		
			H-bond acceptor (<10)	5		
			Violations	1		
11.	Niclosamide	$C_{13}H_8C_{l2}N_2O_4$	Molecular weight (<500Da)	327.12		Yes
			LogP (<5)	2.9		

			H-Bond donor (5)	2		
			H-bond acceptor (<10)	4		
			Violations	0		
12	Rhein	$C_{15}H_8O_6$	Molecular weight (<500Da)	284.22		Yes
			LogP (<5)	1.4		
			H-Bond donor (5)	3		
			H-bond acceptor (<10)	6		
			Violations	0		
13	Scutellarein 7-rutinoside	$C_{27}H_{30}O_{15}$	Molecular weight (<500Da)	594.52		No
			LogP (<5)	-1.10		
			H-Bond donor (5)	9		
			H-bond acceptor (<10)	15		
			Violations	3		
14	Withaferin A	$C_{28}H_{38}O_6$	Molecular weight (<500Da)	470.60		Yes
			LogP (<5)	3.4		
			H-Bond donor (5)	2		
			H-bond acceptor (<10)	6		
			Violations	0		

15.	Withanolide D	$C_{28}H_4O_7$	Molecular weight (<500Da)	488.61		Yes
			LogP (<5)	2.1		
			H-Bond donor (5)	4		
			H-bond acceptor (<10)	7		
			Violations	0		
16.	27-hydroxywithanolide	C28H38O6	Molecular weight (<500Da)	470.60		Yes
			LogP (<5)	3.4		
			H-Bond donor (5)	2		
			H-bond acceptor (<10)	6		
			Violations	0		
17	24-Methylcholesta-5,23E-dien-3beta-ol	$C_{28}H_{46}O$	Molecular weight (<500Da)	398.66		Yes
			LogP (<5)	6.7		
			H-Bond donor (5)	1		
			H-bond acceptor (<10)	1		
			Violations	1		
18.	Ashwagandhanolide		Molecular weight (<500Da)	480.64		Yes
			LogP (<5)	4.1		
			H-Bond donor (5)	1		

			H-bond acceptor (<10)	6		
			Violations	0		
19.	17-alpha-hydroxywithanolide D		Molecular weight (<500Da)	975.28		No
			LogP (<5)	5.8		
			H-Bond donor (5)	6		
			H-bond acceptor (<10)	12		
			Violations	3		