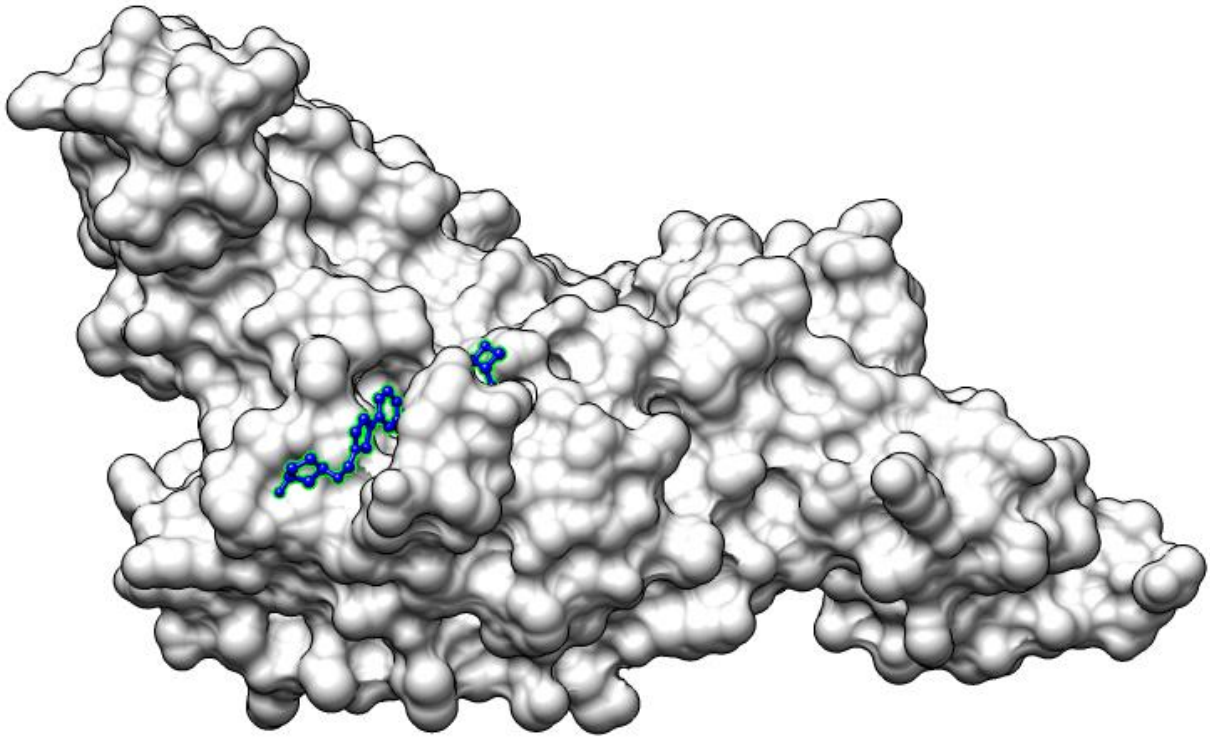
UNIVERSITAT ROVIRA i VIRGILI

ARE DOCKING SCORES AND IC_{50} VALUES OF PL^{PRO} INHIBITORS CORRELATED?

Camilla Pasquali

TREBALL FINAL DE GRAU BIOTECNOLOGIA



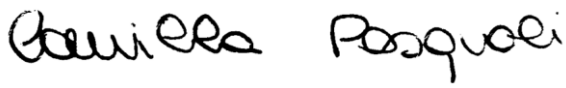
Academic Tutor: Dr. Santiago Garcia-Vallvé, Bachelor's degree in Chemistry, Biochemistry and Biotechnology Department, santi.garcia-vallve@urv.cat

In cooperation with: Cheminformatics and Nutrition Research Group, Universitat Rovira i Virgili

Supervisor: Dr. Santiago Garcia-Vallvé, Bachelor's degree in Chemistry, Biochemistry and Biotechnology Department, santi.garcia-vallve@urv.cat

Jo, Camilla Pasquali, amb DNI CA70325BH , sóc coneixedor de la guia de prevenció del plagi a la URV Prevenció, detecció i tractament del plagi en la docència: guia per a estudiants (aprovada el juliol 2017) (http://www.urv.cat/ca/vidacampus/serveis/crai/que-us-oferim/formacio-competencies-nuclears/plagi/) i afirmo que aquet TFG no constitueixen cap de les conductes considerades com a plagi per la URV.

Tarragona, 7 de juny de 2023



Index

1 Data of the centre

The University Rovira i Virgili (URV) was created in 1991 by the Parliament of Catalonia and it offers more than 50 official undergraduate degrees that meet the standards established by the European Higher Education Area.

The Department of Biochemistry and Biotechnology (DBB) teaches in several undergraduate and post graduate courses such as Biotechnology degree, Biochemistry and Molecular Biology degree, Oenology degree, Chemistry master degree and lots more. The DBB develops its research activity in eight research groups that participate in numerous European project. The research group that hosted me during the development of this project is the Chemoinformatics and Nutrition research group. Their expertise consist in the use of computational drug design tools (same tools used by the pharmaceutical industry) such as virtual screening, protein-ligand docking, pharmacophores, electrostatic or shape comparisons applied to database of natural products, to discover new and advanced bioactive ingredients. Since the COVID-19 epidemic the research group has focused their research activity on the development of new SARS-CoV-2 M-pro inhibitors and the analysis of the SARS-CoV-2 mutations.

2 Abstract and keywords

The pandemic, called COVID-19, that has hardly hit the world in the recent years was caused by the SARS-CoV-2 virus. In the virus's genome there are numerous important non-structural proteins including two proteases (the main protease M^{pro} and the papain-like protease PL^{pro}) that play fundamental roles during its viral replication working as molecular scissors. During the pandemic, therefore, different pharmacological targets have been studied and papain-like protease and its inhibitors have been greatly characterized in order to develop new drugs. Thus, this work aims to collect and analyse (by protein-ligand docking) some PL^{pro} inhibitors whose enzymatic activity has been experimentally evaluated (IC_{50} values). So the primary objective of the analysis is to evaluate whether the docking scores obtained from protein-ligand docking and the experimentally measured IC_{50} values correlated. My research group had previously reported that for M^{pro} inhibitors there is no concordance between the docking scores and IC_{50} values. In this study, we have analysed 216 SARS-CoV-2 PL^{pro} inhibitors and show that for PL^{pro} there is also no such correlation.

Keywords: SARS-CoV-2; papain-like protease, docking, 7LLF, 6WX4

3 Glossary

Abbreviations	Description
SARS-CoV-2	Severe Acute Respiratory Syndrome Coronavirus 2
SARS-CoV	Severe Acute Respiratory Syndrome Coronavirus
MERS-CoV	Middle East Respiratory Syndrome Coronavirus
ORF	open reading frames
NSP	Non-structural proteins
S	Spike protein
E	Envelope protein
N	Nucleocapsid protein
M	Membrane protein
RdRp	RNA-dependent RNA polymerase
M^{pro}	Main protease
PL^{pro}	Papain-like protease
ACE 2	Angiotensin-converting enzyme 2
USP	ubiquitin-specific proteases
SMILES	Simplified Molecular Input Line Entry System
VS	Virtual Screening
IC_{50}	Half maximal inhibitory concentration
PDB	Protein Data Bank
PRR	Protein Reliability Report
PPW	Protein Preparation Workflow
PMID	PubMed unique identifier number for each article
PCID	PubChem compound identifier

4 Introduction

4.1 COVID-19 pandemic

The Worldwide pandemic, also announced as COVID-19 by the World Health Organization (WHO), in December 2019 was rapidly spread in Wuhan (China) and was caused by the Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) (Osipiuk et al., 2021). This virus belongs to the β-coronavirus family and has led to millions of infections and to the death of approximately one million of people (Kashyap et al., 2022). To date, COVID-19 is the most severe pandemic in the recent history (Tan et al., 2022) and it has caused problems not only on people's lives but also on the global economy (Banerjee et al., 2021). As shown in Figure 1 the pandemic has affected and had a strong impact in all the countries in the world.

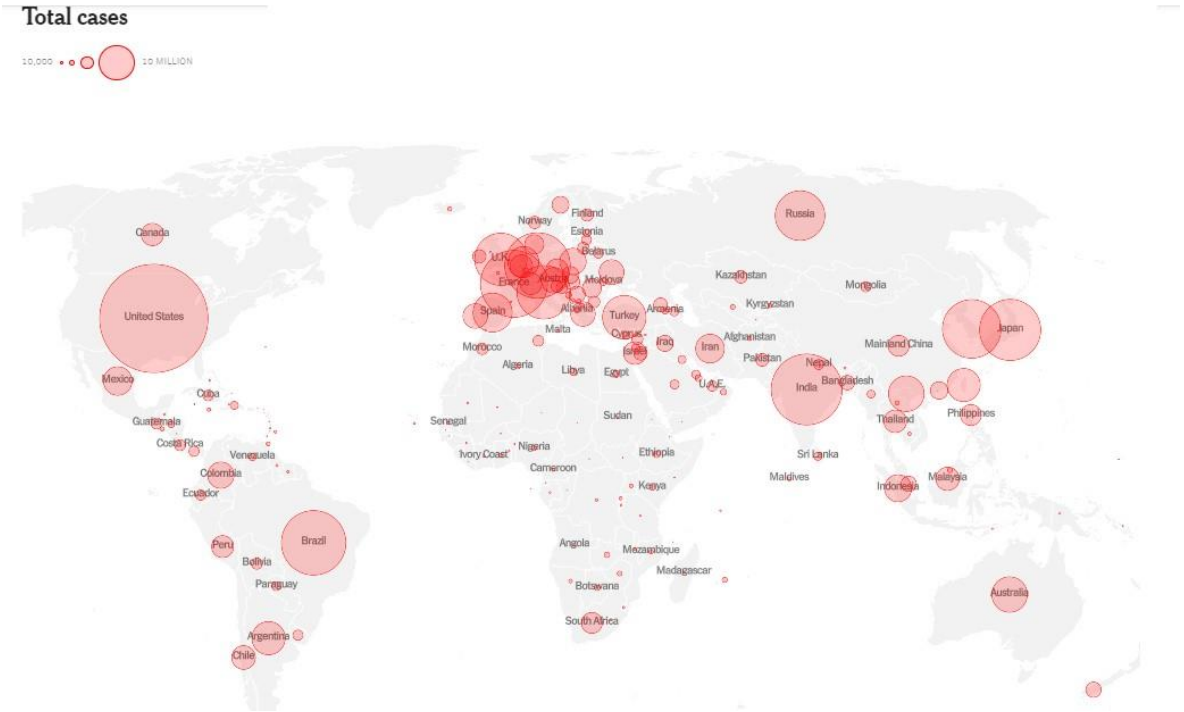


Figure 1. The map shows the known locations of cumulative reported coronavirus cases (*Covid-19 World Map: Cases, Deaths and Global Trends - The New York Times*, n.d.).

4.1.1 The disease, the symptoms and the *Coronaviridae* family

The COVID-19 disease causes the same symptoms that cause other coronavirus, such as cough, fever, vomiting, fatigue and diarrhoea (Singh et al., 2023) and in the

worse cases it can also cause septic shock, metabolic acidosis, acute respiratory distress syndrome (ARDS) and multiple organ dysfunction. Because it had such a huge and quick spread, the scientific community focused their research on the structure and mechanism of infection of the SARS-CoV-2 in order to produce a drug that could stop its spread. In the interest of doing that the knowledge of the molecular interaction between the SARS-Cov-2 and the host immune system it's really important (Kombe Kombe & Jin, 2023).

The *Coronaviridae* family is divided into four viral subcategories, such as, beta (β), alpha (α), delta (δ) and gamma (γ), but only the first two have been confirmed as capable of infecting humans. SARS-CoV-2 belongs to the family of beta coronaviruses, as do SARS-CoV and MERS-CoV, and this was recognized through a sequence analysis. Several SARS-CoV-2 variants evolved during the outbreak, caused by sequence changes. Among the variants, the Omicron variant and in particular the XBB 1.5 subvariant were the most transmissible. Other variants included the alpha, beta, gamma, epsilon, kappa and lambda variants (Polatoğlu et al., 2023).

4.1.2 Structure, encoded proteins and infection

The structure of the virus is characterized by the presence of a single-strand RNA (ssRNA) with a positive strand, consisting of about 30,000 nucleotides and encoded by fourteen functional open reading frames (ORFs) (Razali et al., 2021). This RNA is wrapped inside a spherical envelop associated with a nucleoprotein within a capsid that consists of a matrix protein (Polatoğlu et al., 2023). At both ends of the genome there are two noncoding regions and there are also multiple regions that encode for non-structural proteins (NSPs), structural proteins and accessory proteins (Bai et al., 2021).The viral genome contains two ORFs called ORF1a and ORF1b, which are directly translated into two polyproteins named pp1a and pp1ab required for RNA viral synthesis (Tan et al., 2022). As seen in the Figure 2 ORF1a and ORF1b also encode for 16 NSPs (Shereen et al., 2020). The outer surface of the envelope is covered by proteins called Spikes (S) with a glycoprotein structure, which were found to be important for mediated host cell invasion and viral replication (Polatoğlu et al., 2023). Other structural protein expressed by the SARS-Cov-2 genome (Figure 3) are a membrane protein (M), nucleocapsid protein (N), envelope protein (E) which are all

encoded in the 3' terminal region of the SARS-CoV-2 genome and are essential for viral assembly (Bai et al., 2021; Banerjee et al., 2021). As far as the aforementioned proteins are concerned, the E protein plays a very important role during infection, as well as in pathogenesis and in viral morphogenesis; the M protein binds the nucleocapsid to the host membrane and increases viral assembly while the S protein is a homotrimer that recognize the host cellular receptor and therefore is essential for infection (Banerjee et al., 2021). The SARS-CoV-2 genome also codes for NSPs such as an RNA-dependent RNA polymerase (RdRp, nsp12), the main protease (M^{pro}, also know as 3-chymotrypsin-like protease ($3CL^{pro}$), nsp5) and the papain-like protease (PL^{pro}, nsp3) (Table 1). The two proteases M^{pro} and PL^{pro} during viral replication act as molecular scissor and process the viral polyproteins pp1a and pp1ab in order to generate a functionally active viral replication complex for packaging within host cells (Banerjee et al., 2021; *PDB-101: Learn: Flyers, Posters, & Calendars: Flyers: SARS-CoV-2 Genome and Proteins*, 2023).

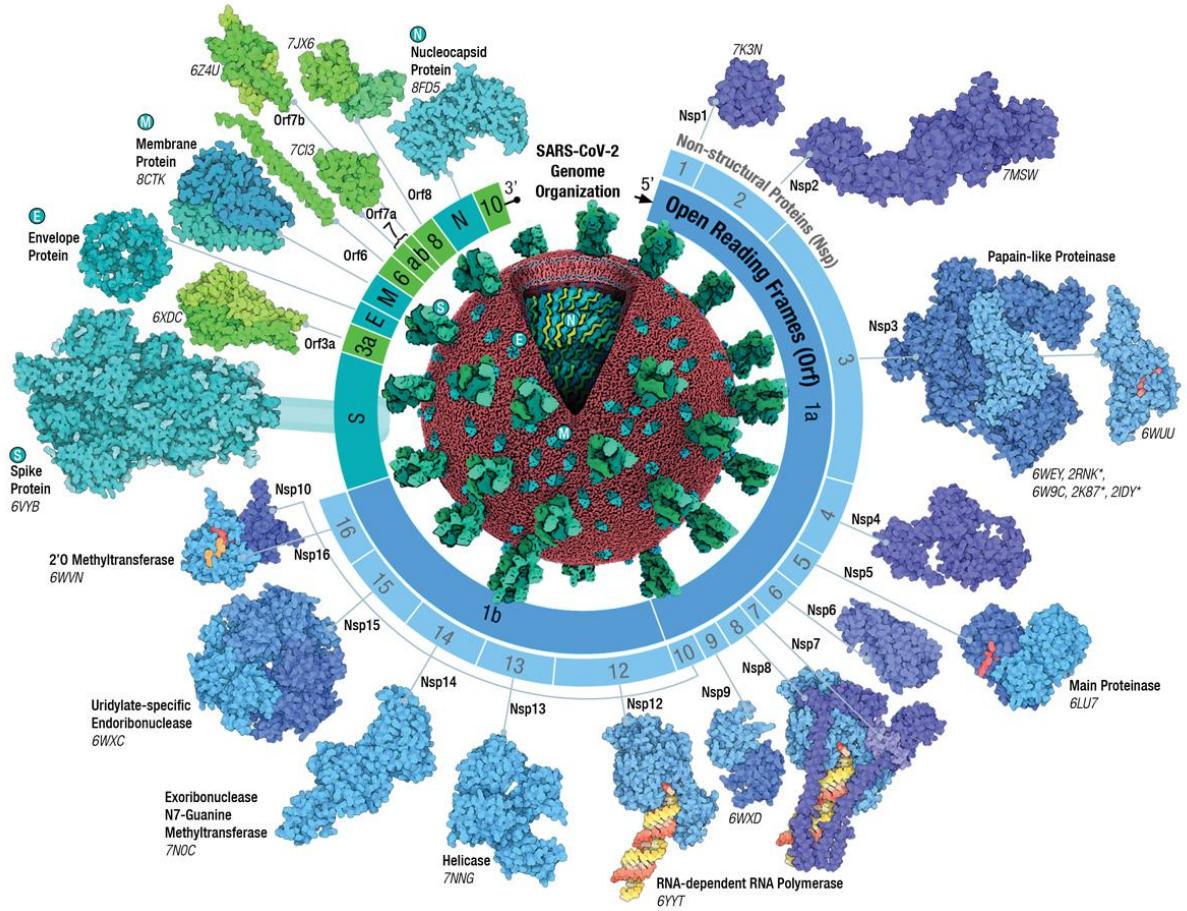


Figure 2. SARS-CoV-2 genome and its genome organization. (*PDB-101: Learn: Flyers, Posters, & Calendars: Flyers: SARS-CoV-2 Genome and Proteins*, 2023).

Table 1. Function and location of the SARS-Cov-2 genome-encoded proteins (Bai et al., 2021).

Protein name	Protein size (AA)	Location in the genome sequence	Function
Non-structural proteins			
nsp1	180	266–805	Accelerates mRNA degradation, blocks innate immune responses
nsp2	638	806-2,719	Undergoes positive selective pressure
nsp3 (PLpro)	1,945	2,729-8,554	Binds to host proteins, mediates viral replication
nsp4	500	8,555-10,054	Mediates the interaction with cellular membranes
nsp5 (Mpro)	306	10,055-10,972	The main protease with coronaviral polyprotein processing activity
nsp6	290	10,973-11,842	Host antiviral defenses, restricts autophagosome expansion
nsp7	83	11,843-12,091	Forms complex with nsp8, stabilizes nsp8
nsp8	198	12,092-12,685	Forms complex with nsp7, catalyzes the synthesis of RNA primers
nsp9	113	12,686-13,024	Nucleic acid-binding protein
nsp 10 (2'0 Methyltransfera se)	139	13,025-13,441	Interacts with nsp14 and nsp16, regulates the function of viral replicase
nsp11	13	13,442-13,480	Function unknown
nsp12 (RdRp)	932	13,442-13,468,13,468-16,236	Catalyzes the synthesis of viral RNA
nsp13 (Helicase)	601	16,237-18,039	Interacts with nsp8 and nsp12, regulates helicase activity
nsp14 (Exoribonucleas e N7-Guanine) nsp15 (Methyltransfera se	527	18,040-19,620	Decreases the incidence of mismatched nucleotides, ExoN
Uridylate-specific endoribonucleas e)	346	19,621-20,658	NendoU activity, degrades viral RNA
nsp16	298	20,659-21,552	2′-O-MTase activity, evades innate immunity
Structural proteins			
E protein	75	26,245–26,472	Interacts with M-protein to form viral membrane
M protein	222	26,523–27,191	Determines viral shape, is the central organizer of coronavirus assembly
S protein	1,282	21,563–25,384	Binds to host cell receptor, mediates viral and host membrane attachment and fusion
N protein	419	28,274–29,533	Binds to viral genomic RNA, forms the nucleocapsid
Accessory proteins			
Orf3a	275	25,393–26,220	Induces apoptosis, pathogenicity and virus release, mediates activation of inflammasome
Orf3b	22	25,814–25,882	Relates to the activation of AP-1 via the ERK and JNK pathways
Orf6	61	27,202–27,387	Inhibits cellular translation
Orf7a	121	27,394–27,759	Relates to virus-host interaction

Orf7b	43	27,756–27,887	
Orf8b	121	27,894–28,259	Mediates immune suppression and evasion activities
Orf9b	97	28,284–28,942	Mediates immune responses
Orf9c	70	28,733–28,577	Modifies the mitochondrial activity of host cells
Orf10	38	29,558–29,674	

Angiotensin-converting enzyme 2 (ACE2) is the receptor located on the surface membrane of host cells and required for the virus to enter in it. The binding happens via the S protein and its receptor binding domain (RBD) and initiate the infection process (Liu et al., 2020). After that, the virus reveals its RNA, translates its RpRd, and then constructs the RNA replicase–transcriptase complex (Polatoğlu et al., 2023). The presence of this complex allows the formation of RNA-negative strands which will then result in the production of viral proteins with replication and transcription. The proteins produced are then transported to the rough endoplasmic reticulum and Golgi where they associate with vesicles that can then release the infected cell by exocytosis in order to attack other cells (Figure 4). Each infected cell releases hundreds of new viral particles that spread to the bronchi and in some cases reach the alveoli causing pneumonia (Cao et al., 2020). Furthermore, Spike proteins were really significant during vaccine development, in fact, using Spike proteins as targets of vaccines had the production of antibodies able to block viral entry by inhibiting the interactions between the viral protein and the ACE2 receptor (Polatoğlu et al., 2023). Three vaccines against Spike protein have received approval from the U.S. Food and Drug Administration (FDA) and they are two mRNA vaccines: from Pfizer and Moderna, and one adenovirus vaccine from Jhonson and Jhonson; additional vaccines from Russia and China have been approved by the WHO (Tan et al., 2022).

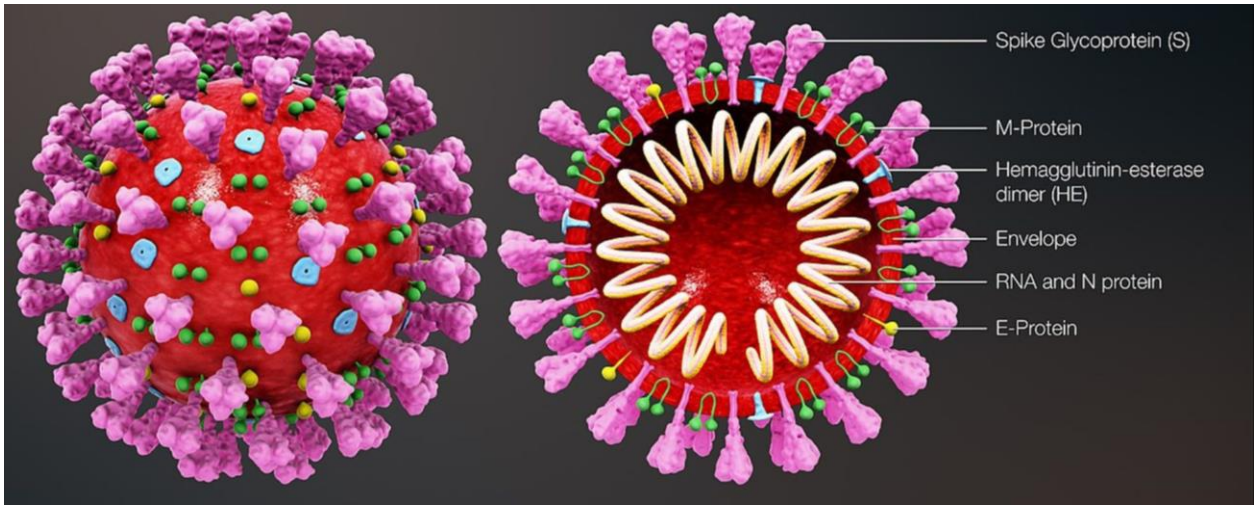


Figure 3. Schematization of the viral structure of SARS-CoV-2(Wiki Images - Scientific Animations, n.d.).

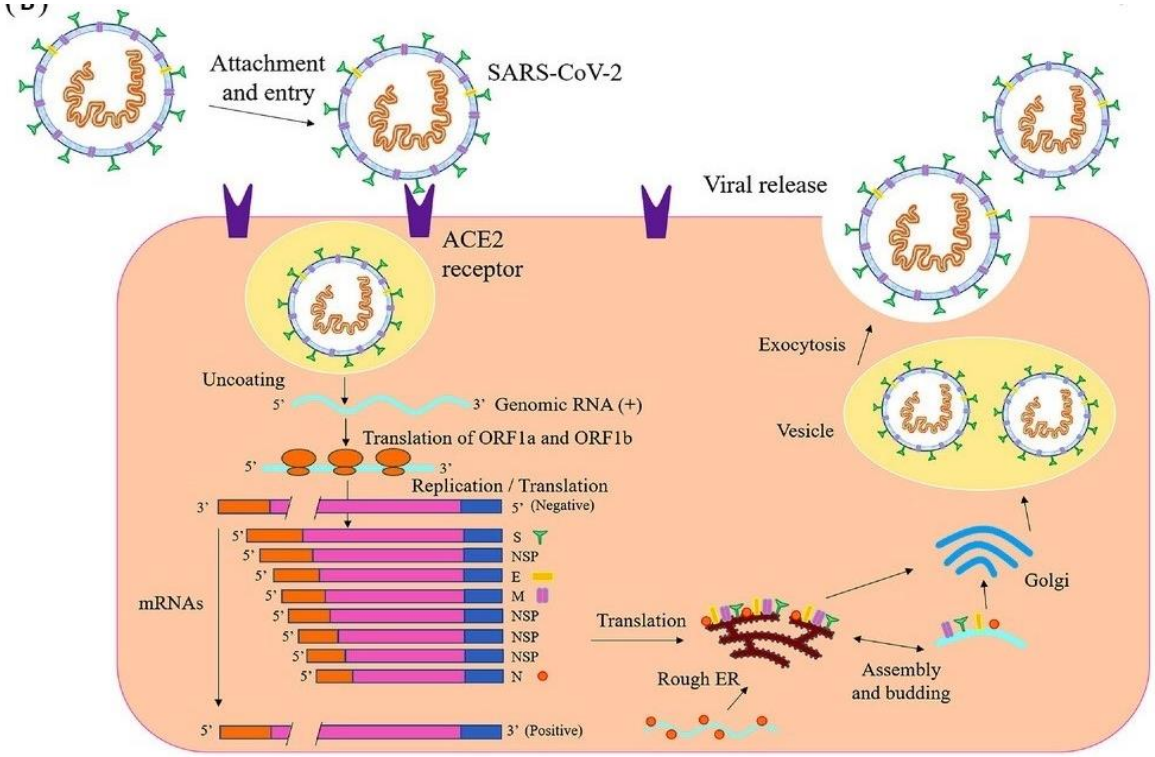


Figure 4. Representation of the mechanism of viral infection that through the glycoprotein S present on the virus and the binding with the ACE2 receptor allows the development of additional molecules that by exocytosis will infect other cells (Polatoğlu et al., 2023).

As told before the vaccines design is important to block the entry of the virus in the host cells but until now only 4 drugs are available for SARS-CoV-2, thus new drug candidates targeting the viral replication cycle are being explored. A prime target of

drug-discovery efforts is the SARS-CoV-2 M^{pro}, as well as PL^{pro} (Banerjee et al., 2021).

4.2 Papain-like protease

4.2.1 PL^{pro} sequence and domains

Protease PL^{pro} derives from a group of cysteine proteases which contains polyprotein endopeptidase activity. The fold it takes is typical for ubiquitin-specific proteases (USPs) and in fact is referred to as USP fold, topologically organised into four domains: UBL, thumb, palm and fingers (Bosken et al., 2020). The sequence alignment between the PL^{pro} of Severe Acute Respiratory Syndrome Coronavirus (SARS-CoV) and that of SARS-CoV-2 (Figure 5) allowed to establish a sequence identity of 82.9%, that at the level of the catalytic site increases to reach 100% (Razali et al., 2021). The comparison with the PL^{pro} of Middle East respiratory syndrome Conoravirus (MERS-CoV) instead shows only a 32.9% of sequence identity, but in any case, this prove that PL^{pro} can be a pharmaceutical target despite the different variants of CoV (Tan et al., 2022).

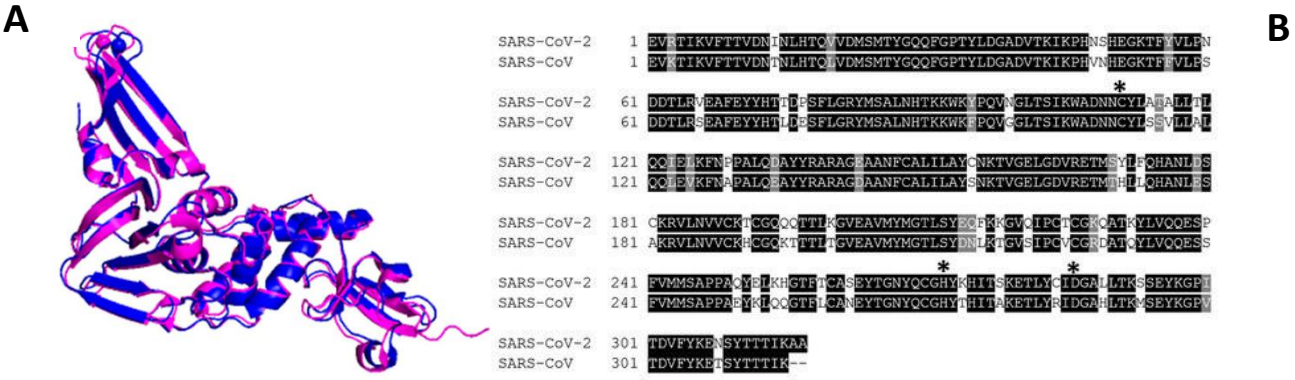


Figure 5. Image A, shows the overlapping of the crystallized structures of the PL^{pro} of SARS-CoV-2 (PDB: 7D7K, in blue) and of SARS-CoV in purple (PDB: 2FE8), while image B shows the pair-wise sequence alignment between SARS-CoV-2 and SARS-CoV PL^{pro} (Razali et al., 2021).

The four domains are (Figure 6): the thumb (residues from 61 to 180), the palm (residues from 239 to 315), the finger (residues from 181 to 238) and the ubiquitin-like domain (Ubl) which includes the residues from 1 to 60, has 5 beta strand, an

alpha helix and 310 helices and is well separated from the remaining three domains that together form the C-terminal ubiquitin-specific protease domain (Usp) (Osipiuk et al., 2021; Razali et al., 2021). Cysteine at position 111 (Cys111), located at the N-terminal of the thumb domain, histidine at position 272 (His272), located at the foot of the palm domain and adjacent to the flexible B-harpin loop called BL2, and aspartate 286 (Asp286) form the catalytic site that catalyses the hydrolysis of the peptide bond (Razali et al., 2021). The catalytic cycle therefore involves Cys111 as a nucleophile, His272 as a general acid-base, and Asp286 coupled with His272 to align and promote Cysteine 111 deprotonation (Su et al., 2021).

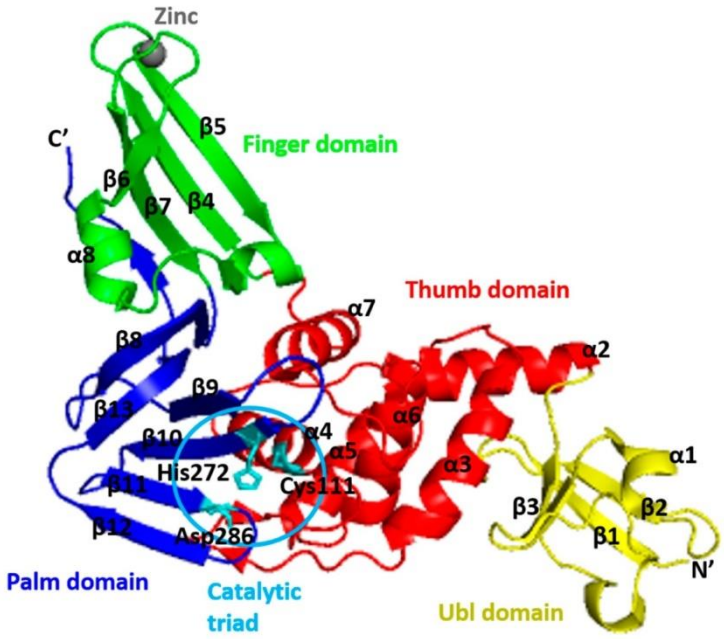


Figure 6. Representation of the SARS-CoV-2 domains (Razali et al., 2021).

Basically, PL^{pro} is located within of NSP3 (Figure 7). The enzyme recognizes a consensus tetrapeptide cleavage motif LXGG (X=any amino acid; L=leucine; G=glycine) that is present between NSP1/2, NSP2/3, NSP3/4 of the viral protein (Tan et al., 2022). Peptide hydrolysis allow the release of Nsp1, Nsp2 and Nsp3 which are indispensable for viral replication (Bai et al., 2021); pp1a contains 11 NSPs and pp1ab contains 16 NSPs. Thus, the polyproteins are then processed in NSP functional units through cleavage by PL^{pro} and M^{pro} (Tan et al., 2022).

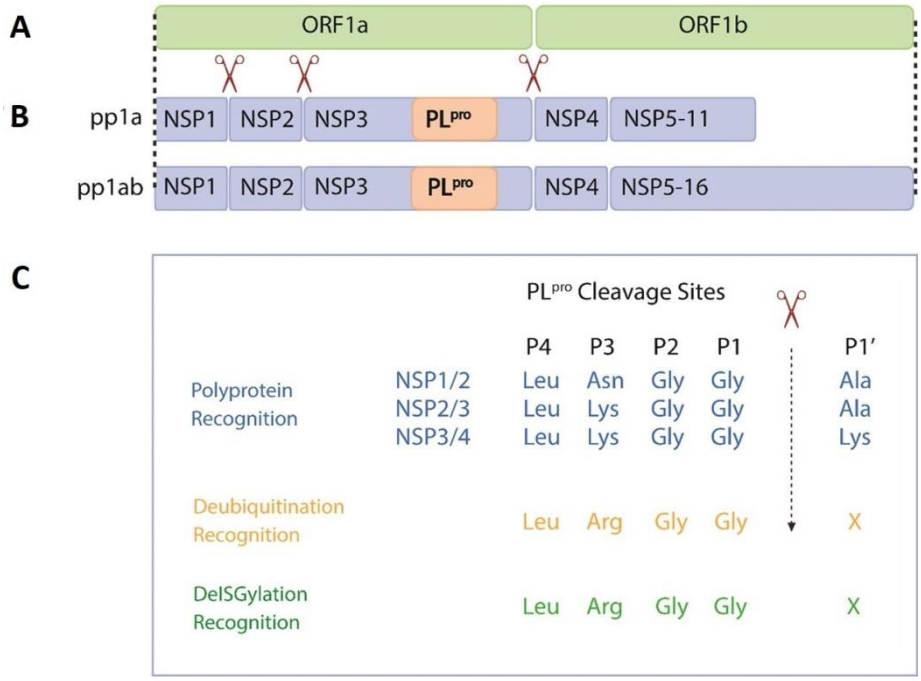


Figure 7. The image shows a schematic representation of the first two open reading frames (ORFs) of SARS-CoV-2 and SARS-CoV (A), the polyprotein replicase and the cleavage sites of PL^{pro}(B) and the motif (i.e. the amino acid sequence alignment of P4-P1') of recognition of PL^{pro} (Tan et al., 2022).

PL^{pro}, as well as the main protease of SARS-CoV-2 (M^{pro}), cleaves peptide bonds in the viral polyprotein so that NSPs necessary for viral replication and transcription are released. However, this is not the only role played by PL^{pro}, in fact it is also involved in antagonizing the host's immune response. The activities it performs are de-ubiquitination and de-ISGylatione therefore removes ubiquitin and ISG15 modification from the host protein, thus allowing the suppression of the innate immune response and promoting viral replication (Freitas et al., 2020; Tan et al., 2022). These two activities are therefore fundamental in counteracting the immediate response of the host (Freitas et al., 2020). In addition, SARS-CoV-2 infection of human macrophages allows the release of free and extracellular ISG15 through viral PL^{pro}, thus allowing the release of proinflammatory cytokines and chemokines (Munnur et al., 2021). Precisely for this reason, inhibiting the activity of PL^{pro} could be important to block the condition of hyperinflammation present in COVID patients, so that the antiviral immunity of the host is restored as well as the suppression of viral replication.

4.3 Computational biology

During the last few years there has been an exponential increase and improvement of computational biology techniques, thus making possible the functional and structural characterization of a very large number of proteins and protein complexes, as well as it has been possible to modulate and design the functional interactions between proteins. This has been allowed by the development and technological improvement that today allow the correct execution of very complex algorithms on a large scale (Mészáros et al., 2023; Tunyasuvunakool et al., 2021). Researchers also use artificial intelligence (AI) techniques to computationally predict the structure of a protein from the amino acid sequence alone (Varadi et al., 2022). In addition, efforts have been made to integrate computational approaches to experimental work. There are some computational methodologies that are used to predict biological structure or models for experimental validation. These computational techniques also help the interpretation of a large number of multi-dimensional datasets, in fact the virtual libraries can allow us to explore spaces that experimental screening methods do not let us to do. Therefore, the implementation of these techniques together with experimental approaches has permitted us to reach a level of knowledge and functional of cells that was previously unimaginable (Mészáros et al., 2023).

4.3.1 Protein-Ligand Docking

It is the technique that is most used in Virtual Screening (VS) and is based on the use of crystallized structures of the protein to predict how the compounds contained in the VS library would bind to the binding site (Gimeno et al., 2019; Macip et al., 2022). Docking is often used with the aim of identifying hit compounds while in the context of a VS, docking can enable to screen large numbers of molecules in a virtual environment. Usually, this technique is based on the following steps:

-*Protein preparation:* An X-ray crystallography protein structure may present problems such as an incomplete sidechain, undefined protonation state or missing hydrogen atoms, so these aspects need to be corrected before using the structure to perform a protein-ligand docking.

-*Binding site definition:* here it is possible to define constraints to require that the ligand occupy a certain space within the binding site or perform certain interactions with the protein

-*Conformational sampling:* A search algorithm is in charge of identifying the possible conformations (docked poses) in which each compound may fit in the binding site, thus a large number of ligand poses are generated on the receptor surface. During docking constraints can be defined to require the resulting docked poses to perform a certain interaction with the receptor or to tie to a certain region of the binding site (Ballante et al., 2021; Gimeno et al., 2019).

-*Scoring*: Each docked pose for the target is estimated with a scoring function that predicts the strength of the interaction, generating a score for each pose. Then, the docked poses are ranked according to the score provided by the docking function to obtain the docked pose that is most likely to represent the real binding mode of the compound (Gimeno et al., 2019; Macip et al., 2022).

Today, there are many docking programs with different sampling algorithms and scoring functions (Ballante et al., 2021; Gimeno et al., 2019). Programs such as Dock, AutoDock and Glide perform docking calculations efficiently (Mészáros et al., 2023).

4.3.1.1 Limits of docking

Molecular docking programs are characterized by numerous limitations although they can normally generate ligand poses that are similar to the crystallographic conformation for many targets (Macip et al., 2022). First, most docking methods use an approximation of the protein making it rigid while the ligand is treated as flexible to find possible bond conformations but at the same time we have the assignment of a score to the different binding poses depending on the binding strength (Gimeno et al., 2019; Mészáros et al., 2023). Even if we know that proteins are not static in real life, we make this concession for the sake of computational speed (this allows the docking calculation to be able to rapidly rate large ligand library). A further limitation of the scoring function is that sometimes they are unable to predict ligand binding affinity (Macip et al., 2022). Finally, contrary to what is expected, the docking score

can depend very significantly on the structure of the ligand so even small changes in it can lead to large differences in docked poses. Taking into account all these conditions, therefore, it is good to see how the docking scores should not be used to predict the activity of a compound but merely to select the compounds that adapt to the binding site (Macip et al., 2022).

4.3.1.2 How docking results can be validated

The best way to validate the results of docking is to experimentally demonstrate the bioactivity of some hits, but this is not always possible. For that reason, the methodology should undergo a computational validation. Common strategies to validate the docking are Re-docking and Cross-docking (Macip et al., 2022).

4.3.1.2.1 Re-docking

Re-docking (Figure 8) consists of removing the ligand from the complex with the protein, transforming the ligand to a 2D representation using Simplified Molecular Input Line Entry System (SMILES) and dock it back in order to evaluate whether the docking program is able to predict the initial binding pose. Subsequently, the RMSD value is calculated by superimposing the crystallized structure and the best docking pose, so this provides an estimate of the ability of the docking program to reproduce the co-crystallized ligand binding pose. Values of RMSD closest to zero are preferable but generally values below 2 ångström (Å) are considerate aceptable.

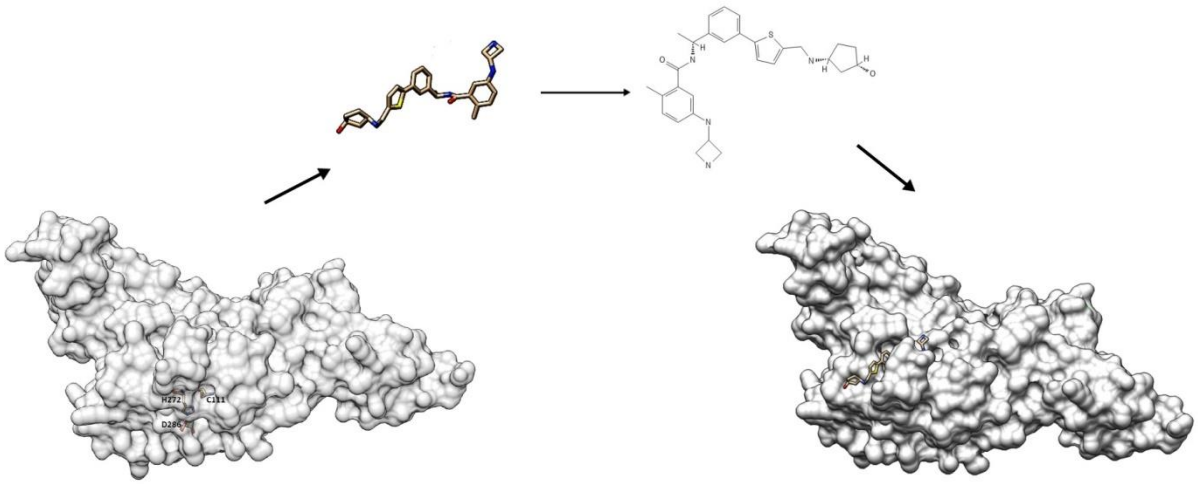


Figure 8. Example of re-docking with PDB:7LLF. The images of the proteins are generated with Chimera 1.17.1

4.3.1.2.2 Cross-docking

Cross-docking (Figure 9) uses the ligand of other complexes to perform docking, so it try to predict ligand binding pose using different structures. Also in this case, RMSD value between the crystallized ligand binding pose and the best docking pose is calculated. The closer the RMSD values are to zero, the better they are. In this way, it's possible to evaluate the effect of protein flexibility or to identify the best protein structure when various are available.

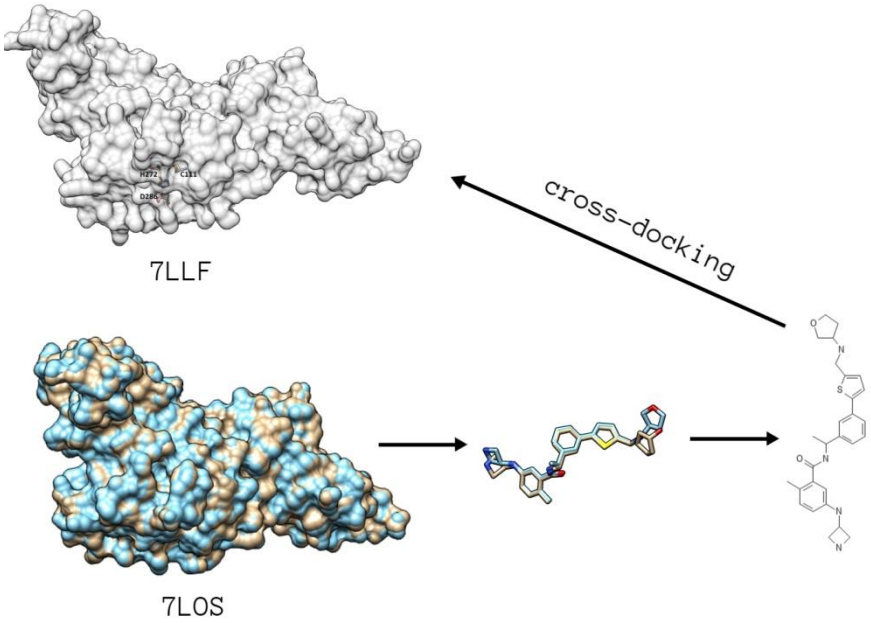


Figure 9. Example of cross-docking between the ligand from 7LOS and the protein 7LLF. The images of the proteins were generated with Chimera 1.17.1.

5 Hypothesis and objectives

5.1 Hypothesis

The pandemic had a profound impact on the scientific community, resulting in the publication of hundreds of studies, particularly in the early months of the pandemic. However, this has inevitably led to the publication of some badly performed articles with low scientific quality. In fact, the value obtained from molecular docking is frequently correlated with the enzymatic activity of an inhibitor, resulting in an inaccuracy. Our research group had previously reported that docking scores could

not be used to predict the potency of SARS-CoV-2 M^{pro} inhibitors (Macip et al., 2022).

Considering this, my work assumes as hypothesis that there is no correlation between the experimentally measured half maximal inhibitory concentration (IC_{50}) values and the docking scores of a group of PL^{pro} inhibitors.

5.2 Objectives

I. To collect from the literature the experimentally measured IC_{50} values of several SARS-CoV-2 PL^{pro} inhibitors.
II. To perform a covalent and non-covalent docking of the previously collected SARS-CoV-2 PL^{pro} inhibitors
III. To evaluate whether there is a correlation between the IC_{50} values and the docking scores for the previously collected SARS-CoV-2 PL^{pro} inhibitors.

6 Materials and methods

6.1 Recompilation of PL^{pro} inhibitors

First of all, to achieve my goal I first had to collect a library of SARS-CoV-2 PLpro inhibitors. It was created after reading a significant number of scientific articles on **PubMed** where inhibitors of SARS-CoV-2 PL^{pro} were described and their half maximal inhibitory concentration (IC_{50}) values experimentally calculated (it represents the concentration of a molecule necessary to inhibit the enzyme response by 50%). PubMed is a free resource that facilitates the search and retrieval of biomedical and life sciences literature for the purpose of improving health, both globally and personally. Available to the public online since 1996 the PubMed database contains more than 35 million citations and abstracts of biomedical literature (Fiorini et al., 2017). To find articles of interest in the database I used some keywords including: PL^{pro} inhibitors, Papain like protease, SARS-CoV-2 inhibitors. Therefore, within the libraries the PMID (unique identifier of the articles in the PubMed database), the name of the compound and its IC_{50} (µM) value have been reported.

From the IC_{50} value was then derived the pIC_{50} ($-logIC_{50}$) which allows to make the final correlation with the docking score. In some articles, the compounds are also reported with their 2D structure and their Simplified Molecular Input Line Entry System (SMILES); while, in other articles, it is only shown the 2D structure without the SMILES. In these cases it was necessary to rebuild the 2D molecule structure through the use of **PubChem** and its application Chem Draw in order to obtain the SMILES. Once the compound has been drown, Chem Draw allow you to get the SMILES. These formulas have been reported in the library as well as the CID Compound (individual for the molecule) encountered in PubChem. PubChem is the world's largest open-access collection of chemical information, so it is an open database at the National Insitute of Health (NIH). It does contain several small molecules, but also larger molecules such as nucleotides, lipids, peptides, carbohydrates and chemically modified macromolecules. It assembles information on chemical structures, physical and chemical properties, biological activities, health and much more (Kim et al., 2021).

In some cases, however, no correspondence has been identified between the SMILES and the molecules contained in PubChem. In the library there is a further column where the **Protein Data Bank** (PDB) code is indicated which represent the crystallized structure of the PL^{pro} with that particular ligand. Protein Data Bank furnish access to 3D structure data for the molecules of life found in all organisms on the planet (Burley et al., 2017).The PDB is universally deemed as a core data resource crucial for understanding the functional roles that macromolecules play in medicine and biology (Burley et al., 2019). Its use in my study has allowed to derive the crystallized structures of PL^{pro} later adopt in the docking.

This library of inhibitors was then divided into covalent and non-covalent compound depending on the interaction they develop with PL^{pro}.

6.2 Protein preparation

Before starting both covalent and non-covalent docking it is necessary to choose the most suitable protein to perform it. The research group that hosted me for this project developed a program that allowed to select the covalent and non-covalent PL^{pro} complexes. Following are reported the PDB ID found for the complexes:

- ***NO-COVALENT***: 7CMD, 7JIW, 7JN2, 7JRN, 7KOL, 7LBR, 7LBS, 7LLF, 7LLZ, 7LOS, 7OFS, 7OFT, 7RZC, 7SDR, 7SGW ,8G62
- ***COVALENT***: 8EUA, 6WX4, 6WUU, 6XA9, 6XAA

In order to select the proteins with the best root mean square deviation (RMSD) value, is fundamental to prepare all of the proteins for Re-docking and Cross-docking before starting the ligand-docking.

With the aid of the ***Protein Reliability Report*** (PRR) tool, the issues with the protein are shown so they may be fixed during the preparation of the protein. ***Protein Preparation Workflow*** (PPW) is used to prepare the proteins in order to solve many of the problems associated with a PDB structure. In these cases water molecules were removed and the pH was established (7.0 +/- 1). Through PPW it is also shown the presence of alternates. Often, a poor electronic density or incomplete electron density means that crystallography cannot accurately place or predict the positions of parts of the protein. The using of ***PrimeX*** let create electron density map to see which position of residues that have alternates is better. PrimeX allows the creation of two maps: 2Fo-Fc and Fo-Fc. For example, in no-covalent docking, the residue 133 (Gln133) was analysed because it is close to the biding site and its default position was chosen.

Later is necessary to reused the PRR to have an idea of the quality of the prepared proteins structures and to compare them with the same proteins before the preparation in order to see the improvements made. Then, the grid was created (via ***Receptor Grid Generation***) in a way that the ligand was in the centre of the grid box (without constraints and default parameters). The grid representation is used for the binding site in order to make the process a lot faster.

The ligands were prepared with ***LigPrep*** (with default settings and pH=7.0 +/- 1). which generates accurate, energy minimized 3D molecular structures. The main objective of LigPrep is to take 2D SMILES and produce the corresponding low-energy 3D structures. It can produce multiple output structures for each input structure by generating different protonation states, stereochemistry, tautomers, and ring conformations.

6.3 Protein-ligand docking

The molecular modelling discuss in this task was performed using Maestro (2022-04 release) and its docking application **Glide** (Bhachoo et al., 2017). Docking methodologies of Glide are high throughput virtual screening (HTVS), standard precision (SP) and extra precision (XP); in my study I excluded the HTVS mode because it is the less trustworthy and it is usually performed with larger libraries than those I created. Instead, Glide SP mode docks reliably for hundreds of thousands of compounds with high accuracy, while XP mode employs a more sophisticated scoring function in order to weeds out false positive that SP lets through. With Glide, the more negative the score, the better. To confirm that the chosen proteins based on PRR are the best to execute the docking and to validate the docking steps it is necessary to perform the Re-docking (or Self-docking) and then the Cross-docking with all of them.

Once, after deciding the suitable proteins, non-covalent and covalent docking were performed. For non-covalent docking, ***Ligand Docking*** was performed having as output the generation of at least 5 poses per ligand. From it, the best scores obtained for each inhibitor were then evaluated.

In covalent docking, on the other hand, following the preparation of the protein, it is necessary to break the bond between Cysteine 111 (Cys111) and the electrophilic warhead of the ligand, so that cysteine can return to function as nucleophile. In the same way that the nucleophilicity of the cysteine residue of the protein was restored, also the electrophilic warhead of the ligand was restored using the tools ***3D Builder***. Using the tool ***Merge*** between the protein and the ligand the complex required to perform the docking is repristinated. The ligands were also prepared with LigPrep but to correctly perform covalent docking (with the tool ***Covalent Docking***) it is necessary to recognize what type of reaction each inhibitor performs with the protein. Covalent docking has two modes: Virtual Screening (Fast) and Pose Prediction (Thorough) and I ran both. Different scores were obtained and the best were selected.

7 Results and discussion

7.1 Non-covalent PL^{pro} inhibitors

7.1.1 Non-covalent PL^{pro} inhibitors

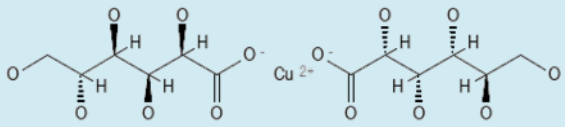
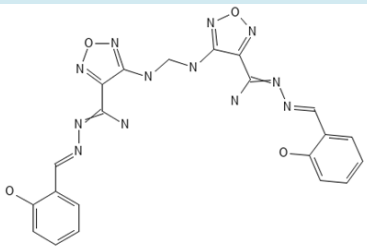
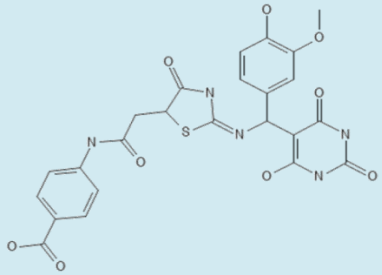
From seventeen PubMed articles have been found 206 non-covalent inhibitors of SARS-CoV-2 PL^{pro}. In the table S2 it is represented the whole library with the following information: Compound name, IC_{50}, pIC_{50}, PMID, PDB, Isomeric smiles, PCID. The range of pIC_{50} extends from 4 to 7.48. Some of the found compounds are inactive, while others have unknown values (Table 2).

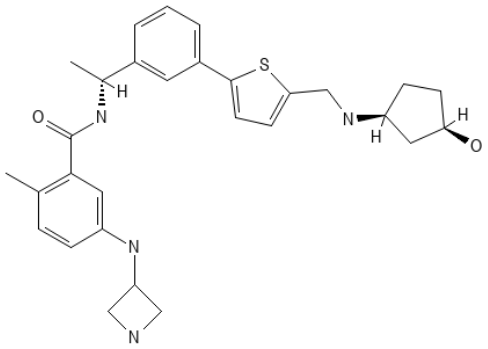
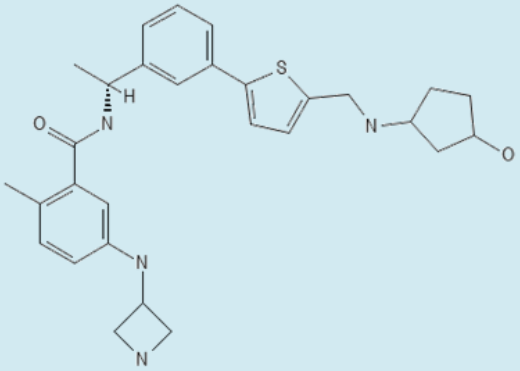
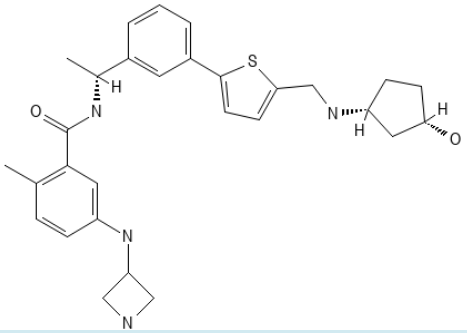
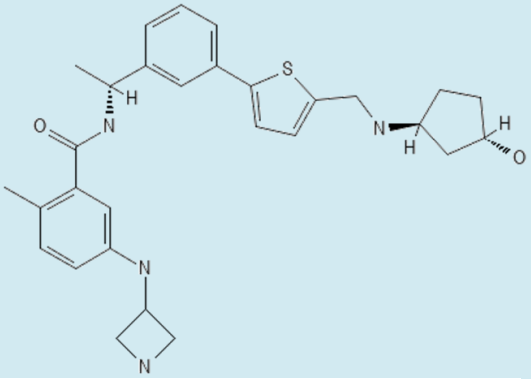
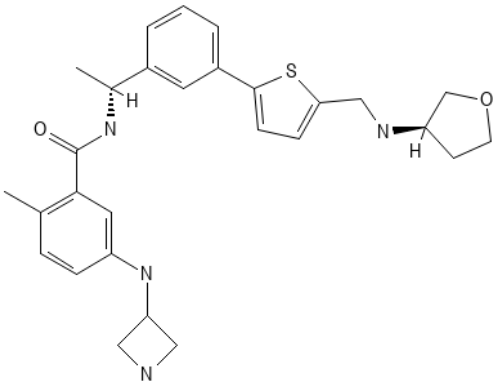
Table 2. Number of active and inactive non-covalent PL^{pro} inhibitors.

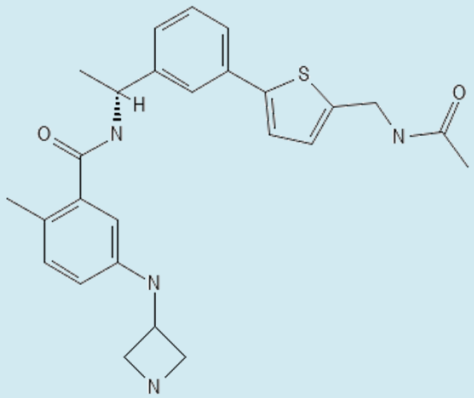
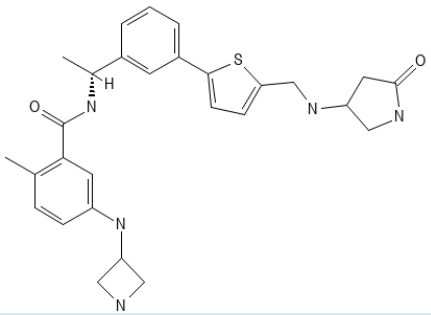
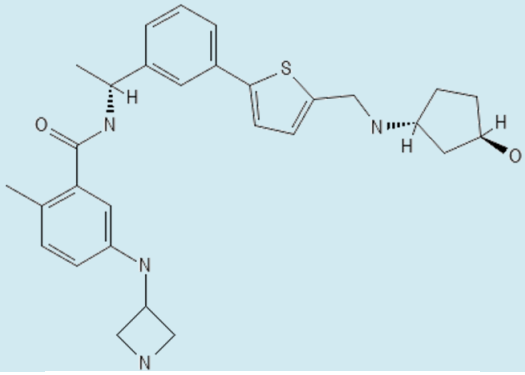
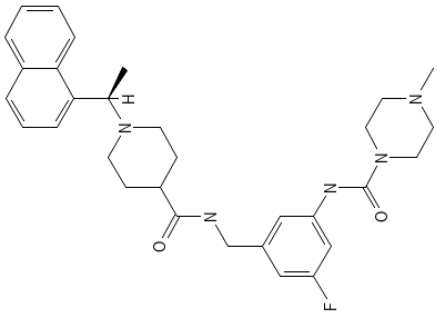
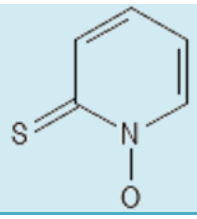
	Active	Inactive	Unknowns values
Compounds	160	7	38

Whilst, Table 3 shows the most potent non-covalent PL^{pro} inhibitors.

Table 3. Representation of non-covalent PL^{pro} inhibitors with the best pIC_{50} value (greater than 6.30).

2D structure	Compound	pIC_{50}
	Copper gluconate	7.48
	Compound 13	7.20
	BDBM50571192	7.07

	XR8-89	6.95
	XR8-67	6.77
	XR8-83	6.66
	XR8-96	6.60
	XR8-65	6.48

	XR8-69	6.43
	XR8-79	6.39
	XR8-84	6.37
	Compound 19	6.36
	Pyrithione	6.30

7.1.2 Non-covalent docking

7.1.2.1 Choosing the best PDB structure

To choose the best protein structure I based on three criteria: RMSD values derived from re-docking, from cross-docking and protein reliability.
Not all the structures mentioned above were used because for some of them were not possible to have an IC_{50} value while others had allosteric inhibitors. We chose the PDB structures: 7LBR, 7LLF, 7LOS, 7LLZ, 7LBS, 7JIW, 7JRN. Not all the structures present two chains A and B, in fact 7JIW shows only chain J. Analysing the PRR of the total protein and of each of the chains it was shown that chain B has far better PRR than the protein and than the chain A (Figure 10).

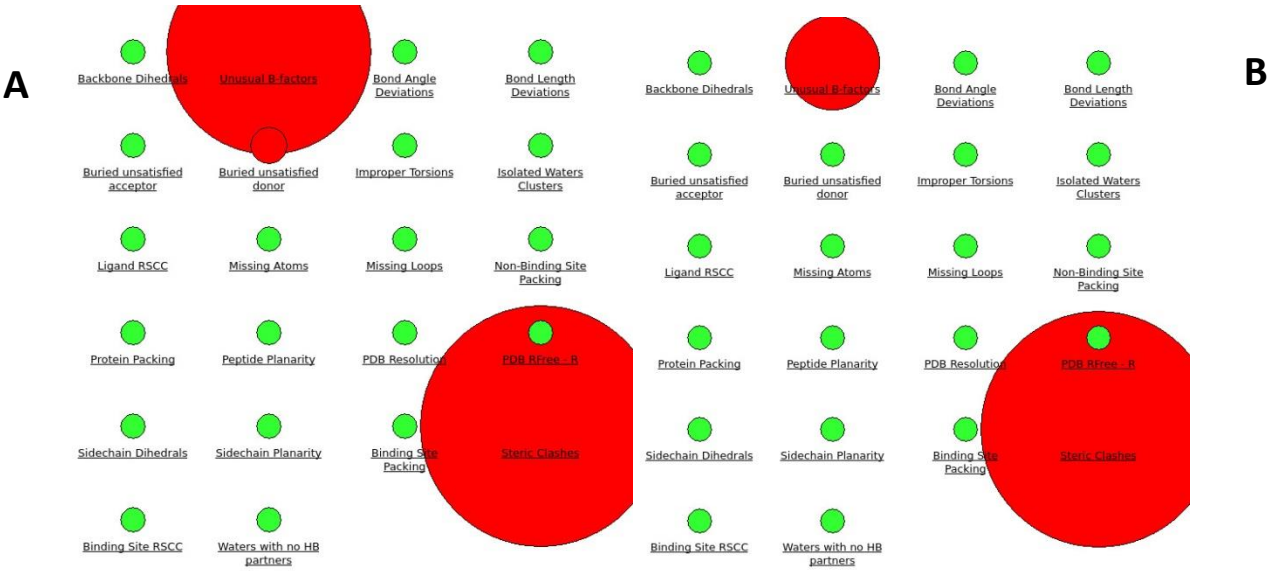


Figure 10. On image A there is the PRR of 7LLF prepared while on image B there is the PRR of Chain B of 7LLF prepared as well (which it's better).

Based on the values obtained with re-docking and cross-docking I created, the tables S2 and S3 respectably for Glide SP and Glide XP.
The protein 7LLF was chosen for doing the non-covalent docking based on both values of re-docking and cross-docking and PRR. Its RMSD values for Glide SP are far better than the other proteins, while for Glide XP 5 proteins (7LLF, 7LBR,7JRN,7LLZ and 7LBS) have similar RMSD values. For that reason the comparison of the PRR analysis between these five proteins was necessary.

In table 4 are shown the RMSD values and the docking scores obtained for chain B of 7LLF with Glide SP while in table 5 are shown the RMSD values and the docking scores obtained also for chain B of 7LLF with Glide XP.

Table 4. Docking score and RMSD values obtained with Glide SP for 7LLF structure.

SP

	7LLF	
	Docking Score	**RMSD**
7LBR	-8.41	1.18
7JRN	-7.78	0.84
7LLF	-8.41	1.15
7LOS	-6.83	1.57
7LLZ	-7.89	1.95
7JIW	-7.82	1.06
7LBS	-7.18	1.16

Table 5. Docking score and RMSD values obtained with Glide XP for 7LLF structure.

XP

	7LLF	
	Docking Score	**RMSD**
7LBR	-7.86	1.26
7JRN	-7.18	0.76
7LLF	-7.86	1.21
7LOS	-6.96	2.98
7LLZ	-7.55	2.14
7JIW	-6.54	1.36
7LBS	-6.52	1.92

7.1.2.2 Non-covalent docking scores

The protein used for non-covalent docking is 7LLF. Glide SP and Glide XP docking were performed. The best results of docking score for each compound are shown as well in the table S2.

In table 6, I selected the compound with the best SP and XP docking score.

Table 6. Compounds with best SP docking score and compounds with best XP docking score

Compound	SP Docking score	Compound	XP Docking score
DY-2-153	-8.14	***rutin***	-9.13
DY2-137	-8.78	***ZN-3-35***	-8.66
DY2-97	-8.80	***DY2-97***	-8.49
XR8-104	-8.34	***DY2-144***	-8.42
XR8-23	-8.13	***ZN-3-61***	-8.35
XR8-79	-8.14	***ZN-2-186***	-8.31
XR8-89	-8.21	***ZN-3-66***	-8.08
XR8-96	-8.30	***DY-3-70***	-7.94
ZN-2-186	-8.45	***XR8-40***	-7.90
ZN-2-187	-8.21	***XR8-9***	-7.90
ZN-3-55	-8.21	***ZN-3-55***	-7.65
ZN-3-56	-8.35	***ZN-2-184***	-7.64
ZN-3-66	-8.03	***ZN-3-32***	-7.55
ZN-3-67	-8.05	***ZN-3-70***	-7.55

7.1.3 Correlation with pIC_{50}

Figures 11 and 12 shown the lack of correlation between docking scores and IC_{50} values. The R^2 for SP non-covalent docking is 0.067 while for XP non-covalent docking is 0.0096.

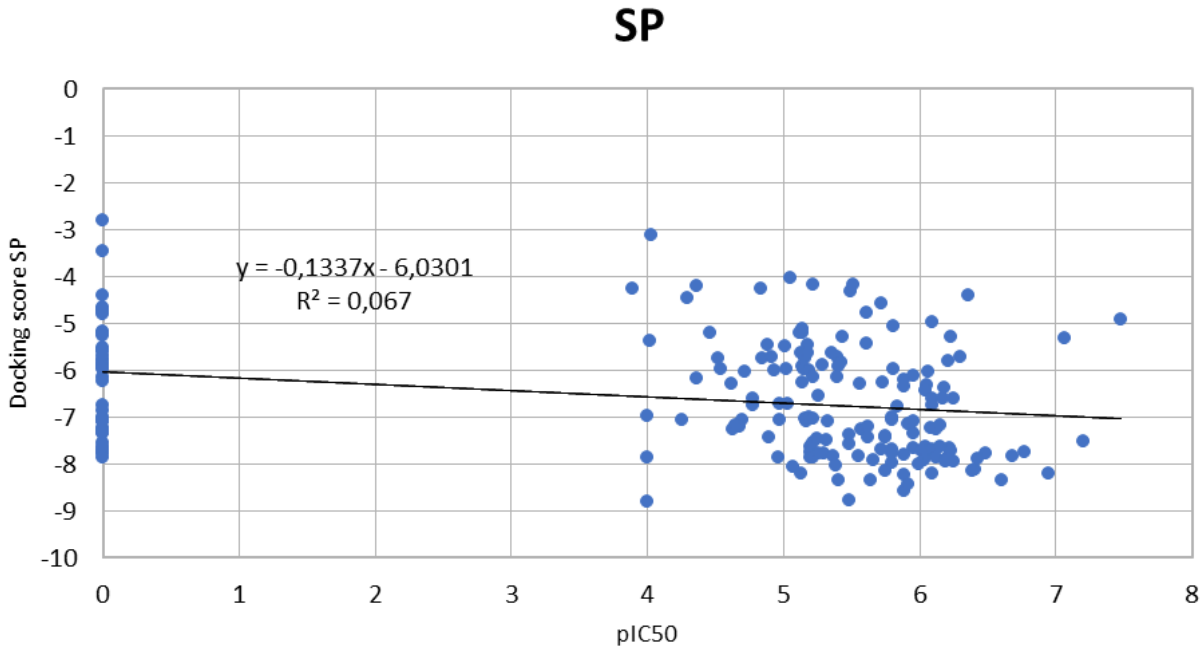


Figure 11. Correlation between the pIC_{50} values and the scores of SP non-covalent docking.

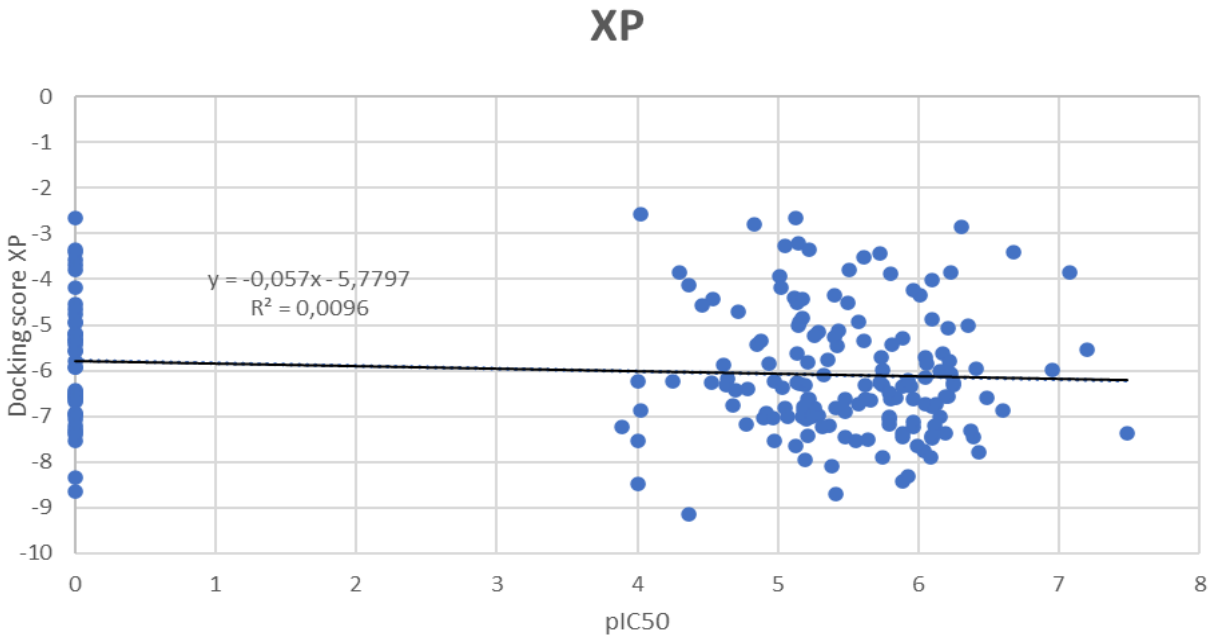


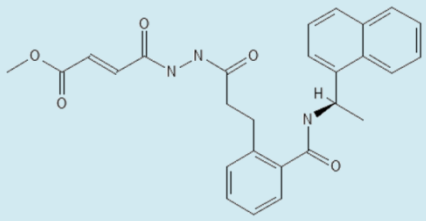
Figure 12. Correlation between the pIC_{50} values and the scores of XP non-covalent docking.

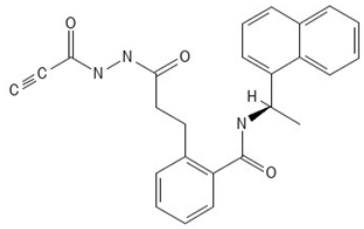
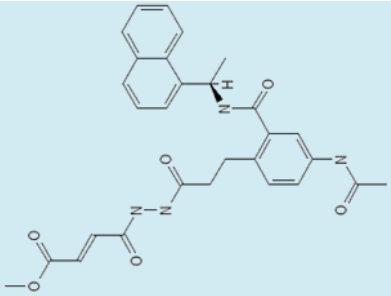
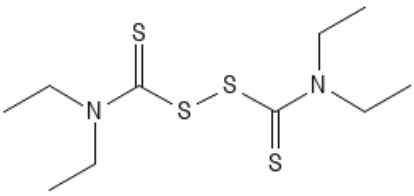
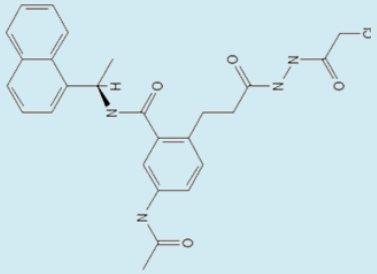
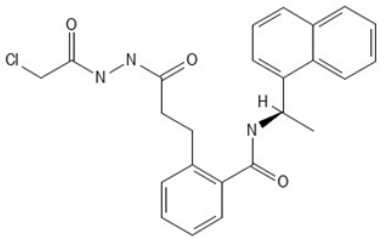
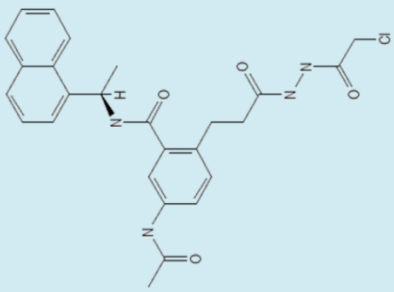
7.2 Covalent PL^{pro} inhibitors

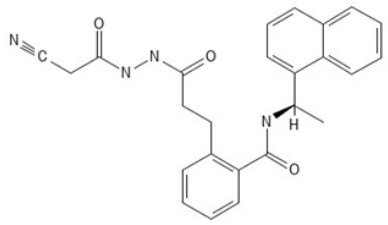
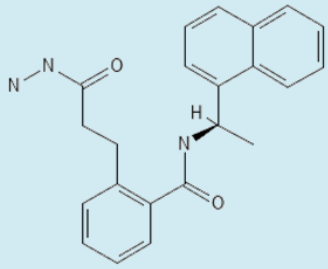
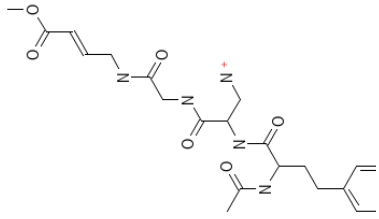
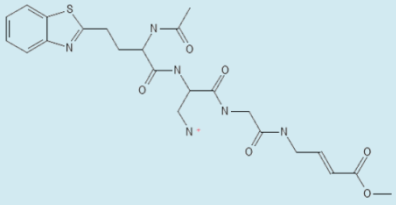
7.2.1 Covalent PL^{pro} inhibitors

Finding covalent compounds in the literature was really difficult, in fact only eleven inhibitors from three articles were found. The information collected also in this case were: Name of the compound, alternative name (if present), internal name (which I gave to them), IC_{50} value, pIC_{50} PMID, PDB, SMILES and PCID (table S4). In table 7 there are all the covalent compounds and their experimentally measured pIC_{50} values (which range is between 7.03-4.30).

Table 7. Covalent PL^{pro} inhibitors and their pIC50 values.

2D Structure	Compound	pIC_{50}
	**Compound 4**	7.03

	**Compound 9**	7.01
	**Compound 5**	6.64
	**Compound 11**	6.32
	**Compound 7**	5.36
	**Compound 6**	5.27
	**Compound 10**	5.21

	**Compound 8**	5.10
	**Compound 3**	4.62
	**Compound 2**	4.30
	**Compound 1**	4.30

7.2.2 Covalent-docking

7.2.2.1 Choosing the best PDB structure

Also, in this case I had to choose the most appropriated protein to perform the docking. PDB ID: 6XA9 and 6XAA could not be used because the first one it is complexed with ISG15 C-terminal domain propargylamide while the second one it is complexed with ubiquitin propargylamide. Also 6WUU was not used because Maestro did not recognize the ligand.

I prepared the protein 8EUA and 6WX4. As it possible to see in the Figure 13 the protein reliability report shows that after preparation 6WX4 is far better than 8EUA.

A

B

Figure 13. On the image A there is the PRR of 8EUA prepared while on the image B: PRR of 6WX4 prepared.

In any case, I performed re-docking of both proteins to validate the protocol (Table 8).

Table 8. RMSD values of re-docking

	6WX4	8EUA
RMSD	0.3	1.9

As predicted with PRR, and confirm with re-docking, the best protein to perform the covalent docking is 6WX4.

7.2.2.2 Covalent docking scores

6WX4 is the PDB ID for the protein used to performed covalent docking. In order to execute a covalent docking is necessary to recognize the reaction between the ligand and the protein; the reactions that I used are represented in the following table (Table 9).

Table 9.Reaction type used for each ligand for covalent docking.

Reaction Type	Ligands
Nucleophilic Substitution	Compound 6 Compound 7 Compound 10
Nucleophilic addition to a Double Bond	Compound 3 Compound 11
Nucleophilic addition to a Triple Bond	Compound 8
Conjugate Addition to Alkene (carbonyl activated)	Compound 9

Michael Addition	Compound 1 Compound 2 Compound 4 Compound 5

The best results obtained are represented in Table 10. For compounds 1 and 11 the protocol made with Virtual Screening mode was not appropriate while changing the accuracy of the protocol (Thorough mode) the work could be completed.

Table 10. Results of covalent docking virtual screening (VS) and thorough (TH)

Compound	Docking score VS	Docking score TH
Compound 1		-7.64
Compound 2	-8.36	-7.99
Compound 3	-4.18	-6.54
Compound 4	-8.31	-7.80
Compound 5	-7.32	-8.07
Compound 6	-7.93	-7.27
Compound 7	-6.93	-7.78
Compound 8	-8.08	-7.88
Compound 9	-8.11	-7.98
Compound 10	-6.93	-7.78
Compound 11		-2.79

7.2.3 Correlation with pIC_{50}

Figure 14 show an R^2 of 0.1346 for virtual screening mode while figure 15 show an R^2 0.0159 for Thorough mode.

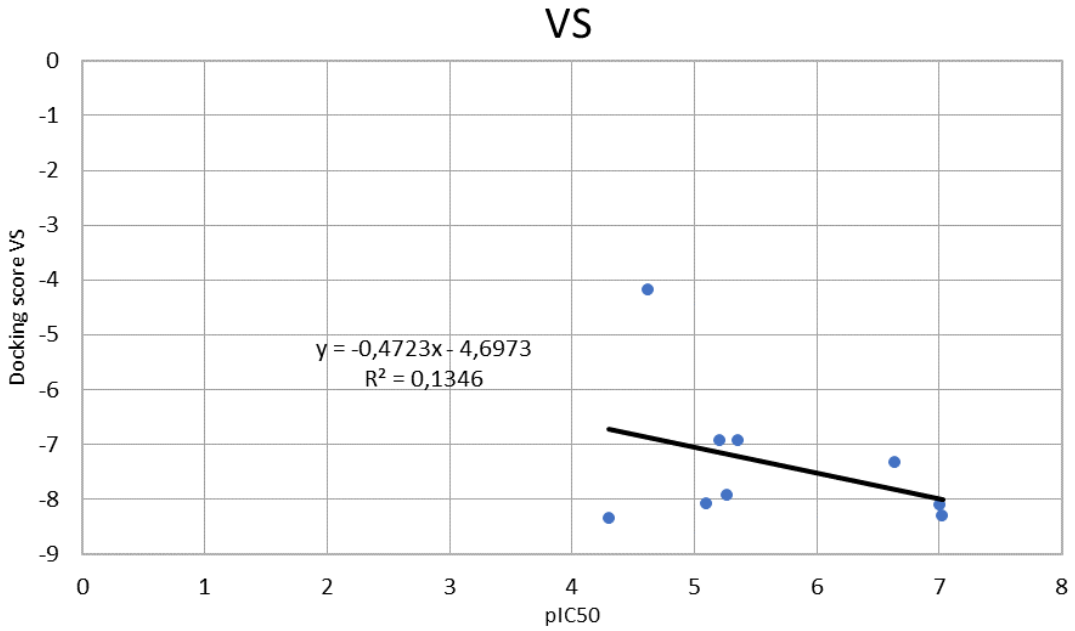


Figure 14. Correlation between pIC_{50} values and scores of VS covalent docking.

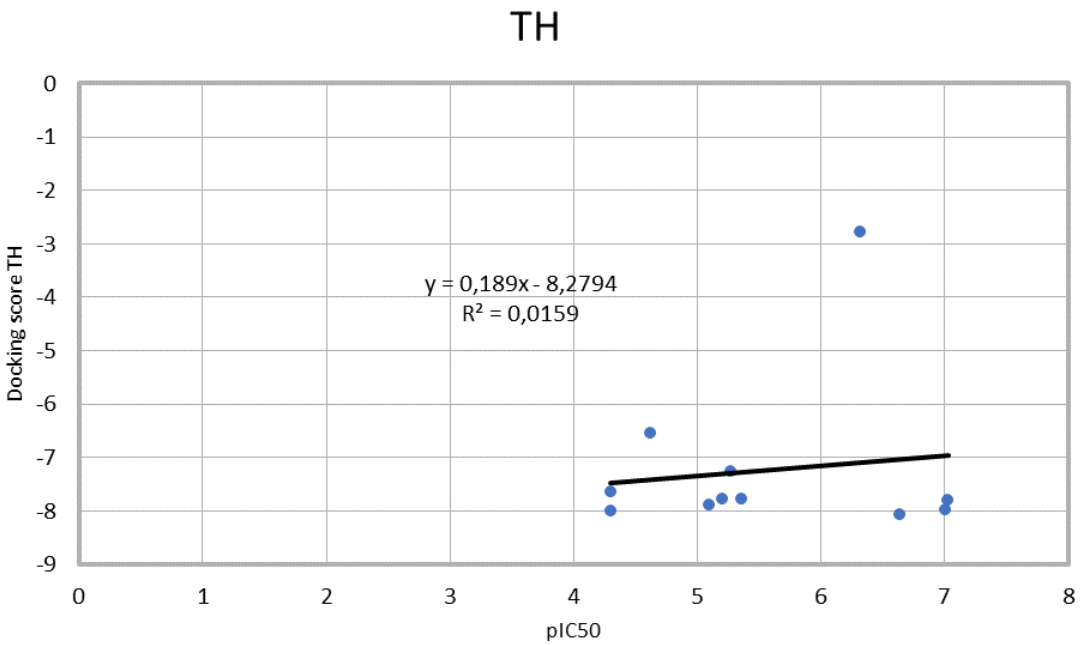


Figure 15. Correlation between the pIC_{50} values and the scores of TH covalent docking.

8 Conclusion

Starting from 20 articles I collected 216 inhibitors of SARS-CoV-2 PL^{pro}. 11 of them are covalent inhibitors while the other 205 are no-covalent inhibitors.

Through covalent and non-covalent docking we obtained the best docking score for each compound. The docking scores, however, were not correlated with the pIC_{50} values. In conclusion, this work confirms our hypothesis, proving that ligand-protein docking cannot be used to predict the potency of a PL^{pro} inhibitor.

9 Bibliography

Bai, C., Zhong, Q., & Gao, G. F. (2021). Overview of SARS-CoV-2 genome-encoded proteins. *Science China Life Sciences 2022 65:2*, *65*(2), 280–294. https://doi.org/10.1007/S11427-021-1964-4

Ballante, F., Kooistra, A. J., Kampen, S., de Graaf, C., & Carlsson, J. (2021). Structure-Based Virtual Screening for Ligands of G Protein–Coupled Receptors: What Can Molecular Docking Do for You? *Pharmacological Reviews*, *73*(4), 1698–1736. https://doi.org/10.1124/PHARMREV.120.000246

Banerjee, R., Perera, L., & Tillekeratne, L. M. V. (2021). Potential SARS-CoV-2 main protease inhibitors. *Drug Discovery Today*, *26*(3), 804–816. https://doi.org/10.1016/J.DRUDIS.2020.12.005

Bhachoo J., & Beuming T. (2017). Investigating Protein–Peptide Interactions Using the Schrödinger Computational Suite. *Methods in Molecular Biology (Clifton, N.J.),* 1561, 235–254. https://doi.org/10.1007/978-1-4939-6798-8_14

Bosken, Y. K., Cholko, T., Lou, Y.-C., Wu, K.-P., & Chang, C. A. (2020). Insights Into Dynamics of Inhibitor and Ubiquitin-Like Protein Binding in SARS-CoV-2 Papain-Like Protease. *Frontiers in Molecular Biosciences*, *7*, 174. https://doi.org/10.3389/FMOLB.2020.00174

Burley, S. K., Berman, H. M., Bhikadiya, C., Bi, C., Chen, L., Di Costanzo, L., Christie, C., Duarte, J. M., Dutta, S., Feng, Z., Ghosh, S., Goodsell, D. S., Green, R. K., Guranovic, V., Guzenko, D., Hudson, B. P., Liang, Y., Lowe, R., Peisach, E., … Ioannidis, Y. E. (2019). Protein Data Bank: the single global archive for 3D macromolecular structure data. *Nucleic Acids Research*, *47*(D1), D520–D528. https://doi.org/10.1093/NAR/GKY949

Burley, S. K., Berman, H. M., Kleywegt, G. J., Markley, J. L., Nakamura, H., & Velankar, S. (2017). Protein Data Bank (PDB): The single global macromolecular structure archive. *Methods in Molecular Biology*, *1607*, 627–641. https://doi.org/10.1007/978-1-4939-7000-1_26/COVER

Cao, Y. chen, Deng, Q. xin, & Dai, S. xue. (2020). Remdesivir for severe acute respiratory syndrome coronavirus 2 causing COVID-19: An evaluation of the evidence. *Travel Medicine and Infectious Disease*, *35*, 101647. https://doi.org/10.1016/J.TMAID.2020.101647

Covid-19 World Map: Cases, Deaths and Global Trends - The New York Times. (n.d.). Retrieved May 29, 2023, from https://www.nytimes.com/interactive/2021/world/covid-cases.html

Fiorini, N., Lipman, D. J., & Lu, Z. (2017). Towards PubMed 2.0. *ELife*, *6*. https://doi.org/10.7554/ELIFE.28801

Freitas, B. T., Durie, I. A., Murray, J., Longo, J. E., Miller, H. C., Crich, D., Hogan, R. J., Tripp, R. A., & Pegan, S. D. (2020). Characterization and Noncovalent Inhibition of the Deubiquitinase and deISGylase Activity of SARS-CoV-2 Papain-Like Protease. *ACS Infectious Diseases*, *6*(8), 2099–2109. https://doi.org/10.1021/ACSINFECDIS.0C00168/ASSET/IMAGES/LARGE/ID0C00168_0005.JPEG

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. https://doi.org/10.1021/JM0306430

Gimeno, A., Ojeda-Montes, M. J., Tomás-Hernández, S., Cereto-Massagué, A., Beltrán-Debón, R., Mulero, M., Pujadas, G., & Garcia-Vallvé, S. (2019). The Light and Dark Sides of Virtual Screening: What Is There to Know? *International Journal of Molecular Sciences*, *20*(6), 1375. https://doi.org/10.3390/IJMS20061375

Kashyap, P., Thakur, M., Singh, N., Shikha, D., Kumar, S., Baniwal, P., Yadav, Y. S., Sharma, M., Sridhar, K., & Inbaraj, B. S. (2022). In Silico Evaluation of Natural Flavonoids as a Potential Inhibitor of Coronavirus Disease. *Molecules (Basel, Switzerland)*, *27*(19), 6374. https://doi.org/10.3390/MOLECULES27196374

Kim, S., Chen, J., Cheng, T., Gindulyte, A., He, J., He, S., Li, Q., Shoemaker, B. A., Thiessen, P. A., Yu, B., Zaslavsky, L., Zhang, J., & Bolton, E. E. (2021). PubChem in 2021: new data content and improved web interfaces. *Nucleic Acids Research*, *49*(D1), D1388–D1395. https://doi.org/10.1093/NAR/GKAA971

Kombe Kombe, A. J., & Jin, T. (2023). Editorial: Host-microbe interaction in SARS-CoV-2 infection: mechanism and intervention. *Frontiers in Immunology*, *14*, 2094. https://doi.org/10.3389/FIMMU.2023.1198868

Liu, C., Zhou, Q., Li, Y., Garner, L., Watkins, S., Carter, L., Smoot, J., Gregg, A., Daniels, A., Jevrey, S., & Albaiu, D. (2020). Research and Development on

Therapeutic Agents and Vaccines for COVID-19 and Related Human Coronavirus Diseases. *ACS Cent. Sci.* 6(3), 315-331. https://doi.org/10.1021/acscentsci.0c00272

Macip, G., Garcia-Segura, P., Mestres-Truyol, J., Saldivar-Espinoza, B., Ojeda-Montes, M. J., Gimeno, A., Cereto-Massagué, A., Garcia-Vallvé, S., & Pujadas, G. (2022). Haste makes waste: A critical review of docking-based virtual screening in drug repurposing for SARS-CoV-2 main protease (M-pro) inhibition. *Medicinal Research Reviews*, *42*(2), 744–769. https://doi.org/10.1002/MED.21862

Madhavi Sastry, G., 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. https://doi.org/10.1007/s10822-013-9644-8

Mészáros, B., Park, E., Malinverni, D., Sejdiu, B. I., Immadisetty, K., Sandhu, M., Lang, B., & Babu, M. M. (2023). Recent breakthroughs in computational structural biology harnessing the power of sequences and structures. *Current Opinion in Structural Biology*, *80*, 102608. https://doi.org/10.1016/J.SBI.2023.102608

Munnur, D., Teo, Q., Eggermont, D., Lee, H. H. Y., Thery, F., Ho, J., van Leur, S. W., Ng, W. W. S., Siu, L. Y. L., Beling, A., Ploegh, H., Pinto-Fernandez, A., Damianou, A., Kessler, B., Impens, F., Mok, C. K. P., & Sanyal, S. (2021). Altered ISGylation drives aberrant macrophage-dependent immune responses during SARS-CoV-2 infection. *Nature Immunology 2021 22:11*, *22*(11), 1416–1427. https://doi.org/10.1038/s41590-021-01035-8

Osipiuk, J., Azizi, S. A., Dvorkin, S., Endres, M., Jedrzejczak, R., Jones, K. A., Kang, S., Kathayat, R. S., Kim, Y., Lisnyak, V. G., Maki, S. L., Nicolaescu, V., Taylor, C. A., Tesar, C., Zhang, Y. A., Zhou, Z., Randall, G., Michalska, K., Snyder, S. A., … Joachimiak, A. (2021). Structure of papain-like protease from SARS-CoV-2 and its complexes with non-covalent inhibitors. *Nature Communications*, *12*(1), 743. https://doi.org/10.1038/S41467-021-21060-3

PDB-101: Learn: Flyers, Posters, & Calendars: Flyers: SARS-CoV-2 Genome and Proteins. (2023, April 1). https://pdb101.rcsb.org/learn/flyers-posters-and-calendars/flyer/sars-cov-2-genome-and-proteins

Polatoğlu, I., Oncu-Oner, T., Dalman, I., & Ozdogan, S. (2023). COVID-19 in early 2023: Structure, replication mechanism, variants of SARS-CoV-2, diagnostic tests, and vaccine & drug development studies. *MedComm*, *4*(2), e228. https://doi.org/10.1002/MCO2.228

Razali, R., Asis, H., & Budiman, C. (2021). Structure-Function Characteristics of SARS-CoV-2 Proteases and Their Potential Inhibitors from Microbial Sources. *Microorganisms 2021, Vol. 9, Page 2481*, *9*(12), 2481. https://doi.org/10.3390/MICROORGANISMS9122481

Shereen, M. A., Khan, S., Kazmi, A., Bashir, N., & Siddique, R. (2020). COVID-19 infection: Emergence, transmission, and characteristics of human coronaviruses. *Journal of Advanced Research*, *24*, 91–98. https://doi.org/10.1016/J.JARE.2020.03.005

Singh, D. D., Han, I., Choi, E. H., & Yadav, D. K. (2023). A Clinical Update on SARS-CoV-2: Pathology and Development of Potential Inhibitors. *Current Issues in Molecular Biology 2023, Vol. 45, Pages 400-433*, *45*(1), 400–433. https://doi.org/10.3390/CIMB45010028

Structure-Based Virtual Screening High-Throughput Virtual Screening for Hit Finding and Evaluation A Schrödinger Online Course. (n.d.).

Su, H., Zhou, F., Huang, Z., Ma, X., Natarajan, K., Zhang, M., Huang, Y., & Su, H. (2021). Molecular Insights into Small-Molecule Drug Discovery for SARS-CoV-2. In *Angewandte Chemie - International Edition* (Vol. 60, Issue 18, pp. 9789–9802). John Wiley and Sons Inc. https://doi.org/10.1002/anie.202008835

Tan, H., Hu, Y., Jadhav, P., Tan, B., & Wang, J. (2022). Progress and Challenges in Targeting the SARS-CoV-2 Papain-like Protease. *Journal of Medicinal Chemistry*, *65*(11), 7561–7580. https://doi.org/10.1021/ACS.JMEDCHEM.2C00303

Tunyasuvunakool, K., Adler, J., Wu, Z., Green, T., Zielinski, M., Žídek, A., Bridgland, A., Cowie, A., Meyer, C., Laydon, A., Velankar, S., Kleywegt, G. J., Bateman, A., Evans, R., Pritzel, A., Figurnov, M., Ronneberger, O., Bates, R., Kohl, S. A. A., … Hassabis, D. (2021). Highly accurate protein structure prediction for the human proteome. *Nature 2021 596:7873*, *596*(7873), 590–596. https://doi.org/10.1038/s41586-021-03828-1

Varadi, M., Anyango, S., Deshpande, M., Nair, S., Natassia, C., Yordanova, G., Yuan, D., Stroe, O., Wood, G., Laydon, A., Zídek, A., Green, T.,

Tunyasuvunakool, K., Petersen, S., Jumper, J., Clancy, E., Green, R., Vora, A., Lutfi, M., … Velankar, S. (2022). AlphaFold Protein Structure Database: massively expanding the structural coverage of protein-sequence space with high-accuracy models. *Nucleic Acids Research*, *50*(D1), D439–D444. https://doi.org/10.1093/NAR/GKAB1061

Wiki Images - Scientific Animations. (n.d.). Retrieved May 29, 2023, from https://www.scientificanimations.com/wiki-images/

10 Self evaluation

This Erasmus experience has enabled me to integrate myself in a new environment, to greatly improve my computational skills, and to get to know different realities compared to my Italian university. This incredible experience also allowed me to grow a lot as a person, to learn how to work in a team and to interface with highly knowledgeable professors and PhD students. Coming into contact with such passionate people made me realise how incredible the world of science and research is. I would have loved to have had more time to continue these analyses in this wonderful research group but since I am definitely passionate about this area of research, once back in Italy, I will continue my studies focusing on bioinformatics and computational biology.

11 Acknowledgements

I would like to sincerely thank Professors Santiago and Gerard who included me in their research group and followed me during this path, as their constant presence and help was fundamental for me, especially since my bioinformatics background was not the best. For the same reason I would thanks my research group and my colleagues that were of great help. They always supported me, giving me the necessary knowledge whenever I did not have it.

12 Annexes

Table S1. No-covalent compounds. (DS= docking score).

Compound	IC_{50}	pIC_{50}	PMID	PDB	Isomeric SMILES	PCID	DS SP	DS XP
7724772	20	4.70	36500659		CC1=CC=CC=C1C(=O)NC(C)C2=CC3=CC=CC=C3C=C2	2949898	-7.1	-6.4
(2S)-1',3',3',5,7,8-hexamethylspiro[1,4-benzoxazine-2,2'-indole]	93.74±29.97	4.03	36591160		CC1=CC(=C2C(=C1C)O[C@]3(C=N2)C(C4=CC=CC=C4N3C)(C)C)C	1399150	-3.1	-2.6
2-[5-(pyridin-4-yl)-1,2,4-oxadiazol-3-yl]aniline	7.197	5.14	36966652		C1=CC=C(C(=C1)C2=NOC(=N2)C3=CC=NC=C3)N	61832598	-5.2	-3.2
2-Methyl-N-(1S-naphthalen-2-yl-ethyl)-benzamide	23.5	4.63	36689835		CC1=CC=CC=C1C(=O)N[C@@H](C)C2=CC3=CC=CC=C3C=C2	1287747	-7.3	-6.3
2,2-Dimethyl-3-(4-phenyl-1H-1,2,3-triazole-1-yl)-2,3-dihydronaphtho[1,2-b]furan-4,5-dione	1.9	5.72	35960688		CC1(C(C2=C(O1)C3=CC=CC=C3C(=O)C2=O)N4C=C(N=N4)C5=CC=CC=C5)C	25105621	-4.6	-3.4
3-[4-[[(1,4-dioxonaphthalen-2-yl)amino]methyl]triazol-1-yl]-2,2-dimethyl-3H-benzo[g][1]benzofuran-4,5-dione	3.2	5.49	35960688		CC1(C(C2=C(O1)C3=CC=CC=C3C(=O)C2=O)N4C=C(N=N4)CNC5=CC(=O)C6=CC=CC=C6C5=O)C	72203992	-4.3	-4.5
3-[4-[[(3-bromo-1,4-dioxonaphthalen-2-yl)amino]methyl]triazol-1-yl]-2,2-dimethyl-3H-benzo[g][1]benzofuran-4,5-dione	3.1	5.51	35960688		CC1(C(C2=C(O1)C3=CC=CC=C3C(=O)C2=O)N4C=C(N=N4)CNC5=C(C(=O)C6=CC=CC=C6C5=O)Br)C	72204157	-4.2	-3.8
3-amino-N-(naphthalen-1-yl)-5-(trifluoromethyl)benzamide	43.2±9.4	4.36	33531496		C1=CC=C2C(=C1)C=CC=C2NC(=O)C3=CC(=CC(=C3)N)C(F)(F)F	155817515	-6.2	-4.1
4-{[(2-{[(6-Hydroxy-2,4-dioxo-1,2,3,4-tetrahydropyrimidin-5-yl)(4-hydroxy-3-methoxyphenyl)methyl]amino}-4-oxo-4,5-dihydro-1,3-thiazol-5-yl)acetyl]amino}benzoic acid	0.085	7.07	36114026		COC1=C(C=CC(=C1)C(C2=C(NC(=O)NC2=O)O)N=C3NC(=O)C(S3)CC(=O)NC4=CC=C(C=C4)C(=O)O)O	135936503	-5.3	-3.9
4-Chloroquinolin-8-ol	9.5± 3.44	5.02	36591160		C1=CC2=C(C=CN=C2C(=C1)O)Cl	384164	-6.0	-4.2
4-methyl-N-(naphthalen-2-ylmethyl)benzamide	29.1	4.54	36689835		CC1=CC=C(C=C1)C(=O)NCC2=CC3=CC=CC=C3C=C2	60427498	-6.0	-4.4
5-(((4-nitrophenoxy)carbonyl)amino)-2-methyl-N-[(1R)-1-naphthalen-1-ylethyl]benzamide	7.0±0.6	5.15	33531496		[N+](=O)([O-])C1=CC=C(OC(=O)NC=2C=CC(=C(C(=O)N[C@H](C)C3=CC=CC4=CC=CC=C34)C2)C)C=C1	-	-5.7	-4.9
5-(acryloylamino)-2-methyl-N-[(1R)-1-(naphthalen-1-yl)ethyl]benzamide	6.4±0.6	5.19	33531496	7JIW	CC1=C(C=C(C=C1)NC(=O)C=C)C(=O)N[C@H](C)C2=CC=CC3=CC=CC=C32	146681430	-7.6	-6.3
5-(butylcarbamoylamino)-2-methyl-N-[(1R)-1-naphthalen-1-ylethyl]benzamide	16.8±2.9	4.77	33531496		C(CCC)NC(=O)NC=1C=CC(=C(C(=O)N[C@H](C)C2=CC=CC3=CC=CC=C23)C1)C	-	-6.6	-7.2
5-[(carbamoylcarbamoyl)amino]-2-methyl-N-[(1R)-1-(naphthalen-1-yl)ethyl]benzamide	5.1±0.7	5.29	33531496		CC1=C(C=C(C=C1)NC(=O)NC(=O)N)C(=O)N[C@H](C)C2=CC=CC3=CC=CC=C32	146681429	-7.8	-7.0
5-Chloro-8-hydroxyquinoline	9.7± 1.66	5.01	36591160		C1=CC2=C(C=CC(=C2N=C1)O)Cl	2817	-5.5	-3.9
5-pentanoylamino-2-methyl-N-[(1R)-1-naphthalen-1-ylethyl]benzamide	12.7±1.3	4.90	33531496		C(CCCC)(=O)NC=1C=CC(=C(C(=O)N[C@H](C)C2=CC=CC3=CC=CC=C23)C1)C	-	-7.4	-7.0
8-Quinolinol, 5,7-dichloro-2-((dimethylamino)methyl)-	7.6±3.49	5.12	36591160		CN(C)CC1=NC2=C(C=C1)C(=CC(=C2O)Cl)Cl	10016012	-5.2	-4.4
acetylshikonin	24.3	4.61	36651644		CC(=CC[C@H](C1=CC(=O)C2=C(C=CC(=C2C1=O)O)O)OC(=O)C)C	479501	-6.3	-5.9
aurintricarboxylic acid (ATA)	30	4.52	36682650		C1=CC(=C(C=C1C(=C2C=CC(=O)C(=C2)C(=O)O)C3=CC(=C(C=C3)O)C(=O)O)C(=O)O)O	2259	-5.8	-6.3
avasimibe	>20		33594371		c1(cc(c(c(c1)C(C)(C)C)CC(=O)N1C(C)c2c(c(ccc2)C(C)C)OS1(=O)=O)C(C)C)C(C)(C)C	-	-2.8	-2.7
Broxyquinoline	6.0±1.2	5.22	36591160		C1=CC2=C(C(=C(C=C2Br)Br)O)N=C1	2453	-4.2	-3.3
candesartan cilexitil	>20		33594371		CCOC1=NC2=CC=CC(C(OC(OC(OC3CCCCC3)=O)C)=O)=C2N1C(C4=CC=C(C5=CC=CC=C5C6=NNN=N6)C=C4)C	-	-3.4	-3.6
catechin	14.2	4.85	36712981		C1[C@@H]([C@H](OC2=CC(=CC(=C21)O)O)C3=CC(=C(C=C3)O)O)O	9064	-5.7	-5.4
Chloroquine	None		36591160		CCN(CC)CCCC(C)NC1=C2C=CC(=CC2=NC=C1)Cl	2719	-4.7	-3.4
Clioquinol	8.9±2.03	5.05	36591160		C1=CC2=C(C(=C(C=C2Cl)I)O)N=C1	2788	-4.0	-3.3
Copper Gluconate	0.033	7.48	36689835		C([C@H]([C@H]([C@@H]([C@H](C(=O)[O-])O)O)O)O)O.C([C@H]([C@H]([C@@H]([C@H](C(=O)[O-])O)O)O)O)O.[Cu+2]	10692	-4.9	-7.4

CPI-169	7.3	5.14	33594371		CCS(=O)(=O)N1CCC(CC1)[C@@H](C)N2C(=C(C3=CC=CC=C32)C(=O)NCC4=C(C=C(NC4=O)C)OC)C	71712226	-5.1	-4.5
Cryptotanshinone	1.34;>100	4.30	36689835;36591160		C[C@H]1COC2=C1C(=O)C(=O)C3=C2C=CC4=C3CCCC4(C)C	160254	-4.5	-3.8
Danshensu	>100		36591160		C1=CC(=C(C=C1C[C@H](C(=O)O)O)O)O	11600642	-5.9	-5.3
Dihydrotanshinone	0.586	6.23	36689835		CC1COC2=C1C(=O)C(=O)C3=C2C=CC4=C(C=CC=C43)C	5316743	-5.3	-3.8
Dihydrotanshinone I	68.3±18	4.17	36591160		C[C@H]1COC2=C1C(=O)C(=O)C3=C2C=CC4=C(C=CC=C43)C	11425923		
DY-2-153	1.8	5.74	34665619		C[C@@H](NC(C1=C(C)C=CC(NC2CNC2)=C1)=O)C3=C4C(NC5=C4C=CC=C5)=CC=C3	156907306	-8.1	-6.0
DY-3-14	>10		34665619		C[C@@H](NC(C1=CC=CC2=C1C=CN2)=O)C3=CC=CC4=C3C=CC=C4	156907275	-7.0	-7.1
DY-3-15	0.8	6.10	34665619		C[C@@H](NC(C1=CC=C(C=CN2)C2=C1)=O)C3=CC=CC4=C3C=CC=C4	39704311	-7.8	-6.8
DY-3-49	NI		34665619		C[C@@H](NC(C1=C(C)C=CC(N)=C1)=O)C2=CC=CC3=CC=CN=C32	156907297	-7.3	-6.4
DY-3-59	6.7	5.17	34665619		CC1=C(C(N[C@H](C)C2=CC(C=CS3)=C3C=C2)=O)C=C(N(C)C4CNC4)C=C1	156907299	-7.1	-7.0
DY-3-65	6.3	5.20	34665619		C[C@@H](NC(C1=C(C)C=CC2=C1C=CN2)=O)C3=CC=CC4=C3C=CC=C4	156907276	-7.8	-7.1
DY-3-66	3.3	5.48	34665619		C[C@@H](NC(C1=C(C)C=CC(N(C)C2CNC2)=C1)=O)C3=CC=CC4=C3C=CN4CC5CCC5	156907300	-7.4	-6.9
DY-3-70	6.4	5.19	34665619		C[C@@H](NC(C1=C(C)C=CC2=C1C=CN2CC3CNC3)=O)C4=CC=CC5=C4C=CC=C5	156907277	-7.9	-7.9
DY-3-8	>100		34665619		C[C@@H](NC(C1=CNC2=C1C=CC=C2)=O)C3=CC=CC4=C3C=CC=C4	156907274	-7.7	-6.5
DY2-109	21	4.68	34665619		C[C@@H](NC(C1=C(Br)C=CC(N)=C1)=O)C2=CC=CC3=C2C=CC=C3	156907272	-7.2	-6.8
DY2-115	7	5.15	34665619		C[C@@H](NC(C1=C(Br)C=NC(N)=C1)=O)C2=CC=CC3=C2C=CC=C3	156907273	-7.0	-6.3
DY2-116	NI		34665619		O=C(C1=CC=CC(N(C)C2CNC2)=C1)N[C@H](CC(N(C)C)=O)C3=CC=CC4=C3C=CC=C4	156907290	-4.8	-6.6
DY2-117	NI		34665619		O=C(C1=CC=CC(N(C)C2CNC2)=C1)N[C@H](CC(NC)=O)C3=CC=CC4=C3C=CC=C4	164866171	-5.9	-6.5
DY2-137	3.3	5.48	34665619		C[C@@H](NC(C1=C(C)C=CC(CN2CC(C(O)=O)C2)=C1)=O)C3=CC=CC4=C3C=CC=C4	156907236	-8.8	-7.5
DY2-138-2	6.1	5.21	34665619		C[C@@H](NC(C1=C(C)C=CC(CN2CCC(C(O)=O)CC2)=C1)=O)C3=CC=CC4=C3C=CC=C4	156907245	-7.9	-7.4
DY2-139	>40		34665619		C[C@@H](NC(C1=C(C)C=CC(N(C)C2CNC2)=C1)=O)C3=CC=CC=C3C4=CC=CS4	156907308	-6.7	-5.9
DY2-144	1.3	5.89	34665619		C[C@@H](NC(C1=C(C)C=CC(CNC2CNC2)=C1)=O)C3=CC=CC4=CC=CC=C43	156907198	-8.6	-8.4
DY2-149	1.6	5.80	34665619		C[C@@H](NC(C1=C(C)C=CC(N(C)C2CNC2)=C1)=O)C3=C4C(NC5=C4C=CC=C5)=CC=C3	156907304	-8.0	-7.2
DY2-97	100	4.00	34665619		O=C(C1=C(C)C=CC(N)=C1)N[C@H](CCO)C2=CC=CC3=CC=CC=C32	156907285	-8.8	-8.5
EGCG	128.4	3.89	36712981		C1[C@H]([C@H](OC2=CC(=CC(=C21)O)O)C3=CC(=C(C(=C3)O)O)O)OC(=O)C4=CC(=C(C(=C4)O)O)O	65064	-4.3	-7.2
GRL0617	1.61±0.09	5.79	34665619	3E9S,7JRN,7CMD,7CJM.	C[C@@H](NC(C1=C(C)C=CC(=C1)N)=O)C2=CC=CC3=C2C=CC=C3	24941262	-7.8	-7.0
HBA	3.99 ± 1.33	5.40	35953531		OC1=CC=C(C=O)C=C1	126	-5.7	-4.3
HE9	3.76 ± 1.13	5.42	35953531		COC(=O)C1=CC(=C(C=C1)O)O	287064	-5.8	-5.5

Jun9-13-4	>20		34341772		CN([C@H](C)C1=CC=C(Cl)C=C1)CC2=C(Cl)C=CC(O)=C2	139014904	-5.8	-4.9
Jun9-13-5	>20	#VALORE!	34341772		ClC1=C(CN(CC)C2=CC=CC(C)=C2C)C=C(O)C=C1	154866238	-6.0	-4.8
Jun9-13-6	>20		34341772		CN(C(C)C1CCCCC1)CC2=C(Cl)C=CC(O)=C2	139014884	-4.4	-4.6
Jun9-13-7	7.29 ± 1.03	5.14	34341772		ClC1=CC=CC=C1C(C)N(C)CC2=C(Cl)C=CC(O)=C2	-	-5.9	-5.6
Jun9-13-8	>20		34341772		CN(C(C)C1=CC=C(C)C=C1)CC2=C(Cl)C=CC(O)=C2	139014851	-5.2	-5.2
Jun9-13-9	6.67 ± 0.05	5.18	34341772		ClC1=CC=CC=C1[C@@H](C)N(C)CC2=C(Cl)C=CC(O)=C2	-	-5.5	-4.8
Jun9-25-4	>20		34341772		CC1=CC=CC=C1[C@@H](C)N(C)CC2=C(Cl)C=CC(OC)=C2	-	-5.6	-5.6
Jun9-26-2	>20		34341772		CC1=CC=CC=C1CN(C)CC2=C(Cl)C=CC(O)=C2	-	-5.8	-4.2
Jun9-29-1	>20		34341772		CN(C1C(C)CCCC1)CC2=C(Cl)C=CC(O)=C2	139014882	-4.7	-4.7
Jun9-29-3	>20		34341772		CN(C1=CC=CC=C1C(C)C)CC2=C(Cl)C=CC(O)=C2	139021260	-4.7	-3.7
Jun9-29-4	>20		34341772		ClC1=CC=CC(Cl)=C1CN(C)CC2=C(Cl)C=CC(O)=C2	139014859	-5.3	-5.0
Jun9-29-5	>20		34341772		CN(C(C)C1=C(F)C=CC(F)=C1)CC2=C(Cl)C=CC(O)=C2	139014854	-6.2	-5.3
Jun9-29-6	>20		34341772		CN(C(C)C1=CC(O)=CC=C1)CC2=C(Cl)C=CC(O)=C2	139020943	-5.6	-5.2
Jun9-29-7	>20		34341772		CN(C(C)C1=C(F)C=CC=C1)CC2=C(Cl)C=CC(O)=C2	139014855	-5.9	-5.4
Jun9-44-5	>20		34341772		ClC1=C(CN2CCC(C=CC=C3)=C3C2)C=C(O)C=C1	-	-6.0	-5.8
Jun9-53-2	0.89±0.12	6.05	34341772		CN([C@@H](C1=CC=CC2=C1C=CC=C2)C)CC3=C(C)C=CC(N)=C3	165122744	-6.3	-5.7
Jun9-67-1	2.72±0.28	5.57	34341772		CN([C@@H](C1=CC=CC2=C1C=CC=C2)C)CC3=CC=C(O)C(O)=C3	-	-6.3	
Jun9-67-2	7.10±2.47	5.15	34341772		CN([C@H](C)C1=CC=CC2=C1C=CC=C2)CC3=CC=CN3	-	-6.0	-5.0
Jun9-67-5	>20		34341772		CN([C@@H](C1=CC=CC2=C1C=CC=C2)C)CC3=CC=C(N4C=CC=C4)C=C3	-	-5.5	-3.4
Jun9-68-1	11.71±2.05	4.93	34341772		CN([C@H](C)C1=CC=CC2=C1C=CC=C2)CC3=CN=C(OC)C=C3	-	-6.0	-5.8
Jun9-68-2	1.30±0.14	5.89	34341772		CN([C@H](C)C1=CC=CC2=C1C=CC=C2)CC3=CC4=C(C=C3)C=CN4	-	-6.3	-5.3
Jun9-68-3	5.13±0.74	5.29	34341772		CN([C@H](C)C1=CC=CC2=C1C=CC=C2)CC3=CC=C4C(C=CC=N4)=C3	-	-5.9	-5.2
Jun9-68-4	>20		34341772		CN([C@H](C)C1=CC=CC2=C1C=CC=C2)CC3=CN=CC=N3	-	-5.7	-5.2
Jun9-72-2	0.67±0.08	4.97	34341772	7SDR	CN([C@@H](C1=CC=CC2=C1C=CC=C2)C)CC3=CC=C(O)C=C3	156612931	-6.7	-6.2
Jun9-75-2	1.54±0.30	5.81	34341772		CN([C@@H](C1=CC=CC2=C1C=CC=C2)C)CC3=CC=CC(O)=C3	-	-6.0	-5.4
Jun9-75-3	8.89±2.23	5.05	34341772		CN([C@@H](C1=CC=CC2=C1C=CC=C2)C)CC3=CC=C(O)C(O)=C3	-		-6.8
Jun9-75-4	0.62±0.06	6.21	34341772		CN([C@H](C)C1=CC=CC2=C1C=CC=C2)CC3=CC=CC4=C3C=CN4	-	-5.8	-5.1
Jun9-75-5	0.56±0.03	6.25	34341772		CN([C@@H](C1=CC=CC2=C1C=CC=C2)C)CC3=C(OC)C=C(O)C=C3	-	-6.6	-6.3
Jun9-80-4	1.58±0.26	5.80	34341772		CN([C@H](C)C1=CC=CC2=C1C=CC=C2)CC(C=C3)=CC4=C3NC=C4	-	-7.0	-6.6
Jun9-81-1	19.23±1.70	4.72	34341772		O=C1C(CN([C@H](C)C2=CC=CC3=C2C=CC=C3)C)=CNC(N1)=O	-	-6.0	-4.7
Jun9-81-2	1.44±0.22	5.84	34341772		CN([C@@H](C1=CC=CC2=C1C=CC=C2)C)CC3=C(F)C=C(N)C=C3F	-	-6.8	-6.6
Jun9-81-3	1.85±0.15	5.73	34341772		CN([C@@H](C1=CC=CC2=C1C=CC=C2)C)CC3=C(Cl)C=CC(N)=C3	165122754	-6.3	-5.7
Jun9-84-2	2.47±0.44	5.61	34341772		CN([C@@H](C1=CC=CC2=C1C=CC=C2)C)CC3=CC=C(NN=N4)C4=C3	-	-5.4	-5.4
Jun9-84-3	0.67±0.14	6.17	34341772	7RZC	CN([C@@H](C1=CC=CC2=C1C=CC=C2)C)CC3=CNC4=C3C=CC=C4	156595870	-6.6	-5.6
Jun9-84-4	1.30±0.10	5.89	34341772		CN([C@@H](C1=CC=CC2=C1C=CC=C2)C)CC3=CC=C(B(O)O)C=C3	-	-6.2	-6.3

Jun9-84-5	>20		34341772		CN([C@H](C1=CC=CC2=C1C=CC=C2)C)CC3=CC=C(O)C=C3	165122781	-7.1	-7.3
Jun9-84-6	>20		34341772		CN([C@@H](C1=CC=CC2=C1C=CC=C2)C)CC3=CC=NC(O)=C3	-	-6.2	-5.6
Jun9-85-1	0.66±0.03	6.18	34341772		CN([C@@H](C1=CC=CC2=C1C=CC=C2)C)CC3=CC=C(NN=C4)C4=C3	-	-6.4	-6.0
Jun9-85-2	1.15±0.10	5.94	34341772		CN([C@@H](C1=CC=CC2=C1C=CC=C2)C)CC3=CN(CC(N)=O)C4=C3C=CC=C4	-	-7.2	-6.3
Jun9-85-5	16.59±2.60	4.78	34341772		CN([C@@H](C1=CC=CC2=C1C=CC=C2)C)CC3=CN=C(N)N=C3	-	-6.7	-6.4
Jun9-85-6	4.42±1.22	5.35	34341772		CN([C@@H](C1=CC=CC2=C1C=CC=C2)C)CC3=CC=C(NC(C)=O)C=C3	-	-5.6	-5.8
Jun9-86-1	3.99±0.56	5.40	34341772		CN([C@@H](C1=CC=CC2=C1C=CC=C2)C)CC3=NNC4=C3C=C(O)C=C4	-	-6.2	-5.3
Jun9-86-2	6.49±1.14	5.19	34341772		CN([C@@H](C1=CC=CC2=C1C=CC=C2)C)CC3=CC(F)=C(O)C(F)=C3	165122771	-6.0	-6.9
Jun9-86-4	9.22±3.31	5.04	34341772		CN([C@@H](C1=CC=CC2=C1C=CC=C2)C)CC3=CC=C(N(C)CCO)C=C3	165122788	-6.7	-6.4
Jun9-86-5	6.12±0.51	5.21	34341772		CN([C@@H](C1=CC=CC2=C1C=CC=C2)C)CC3=CC=CC(S(=O)(N)=O)=C3	-	-6.2	-5.8
Jun9-86-8	5.50±0.27	5.26	34341772		CN([C@@H](C1=CC=CC2=C1C=CC=C2)C)CC3=NC=C(O)C=C3	165122782	-6.5	-5.2
Jun9-87-1	0.87±0.10	6.06	34341772		CN([C@@H](C1=CC=CC2=C1C=CC=C2)C)CC3=C(C)C=CC(O)=C3	-	-6.0	-5.8
Jun9-87-2	0.90±0.24	6.05	34341772		CN([C@@H](C1=CC=CC2=C1C=CC=C2)C)CC3=C(C)C=C(O)C=C3	-	-6.4	-6.1
Jun9-87-3	0.80±0.10	6.10	34341772		CN([C@@H](C1=CC=CC2=C1C=CC=C2)C)CC3=C(Cl)C=C(O)C=C3	-	-6.6	-4.0
mangiferin	95.3	4.02	36712981		C1=C2C(=CC(=C1O)O)OC3=C(C2=O)C(=C(C(=C3)O)[C@H]4[C@@H]([C@H]([C@@H]([C@H](O4)CO)O)O)O)O	5281647	-5.4	-6.9
MK-3903	>20		33594371		CC1=C(C=C(C=C1)OC2=NC3=C(N2)C=C(C(=C3)C4=CC=C(C=C4)C5=CC=CC=C5)Cl)C(=O)O	45256689	-4.6	-3.8
myricetin	12,1	4.92	36712981		O=c2c(O)c(c1cc(O)c(O)c(O)c1)oc3cc(O)cc(O)c23	5281672	-5.7	-6.9
N-(2-chlorophenyl)-5-oxo-1-thioxo-4,5-dihydro[1,3]thiazolo[3,4-a]quinazoline-3-carboxamide	7.4	5.13	36114026		C1=CC=C2C(=C1)C(=O)NC3=C(SC(=S)N23)C(=O)NC4=CC=CC=C4Cl	50760592	-5.6	-2.7
N-[(2-methoxyphenyl)methyl]-1-[(1R)-1-naphthalen-1-ylethyl]piperidine-4-carboxamide	34.8	4.46	36689835		C[C@H](C1=CC=CC2=CC=CC=C21)N3CCC(CC3)C(=O)NCC4=CC=CC=C4OC	46854593	-5.2	-4.6
N-[(3-methoxyphenyl)methyl]-1-[(1R)-1-naphthalen-1-ylethyl]piperidine-4-carboxamide	13.2	4.88	36689835		C[C@H](C1=CC=CC2=CC=CC=C21)N3CCC(CC3)C(=O)NCC4=CC(=CC=C4)OC	46854594	-5.5	-5.3
N-[3-(Acetylamino)benzyl]-1-[(R)-1-(1-naphthyl)ethyl]piperidine-4-carboxamide	2.69	5.57	36114026		C[C@H](C1=CC=CC2=CC=CC=C21)N3CCC(CC3)C(=O)NCC4=CC(=CC=C4)NC(=O)C	86276129		-4.9
N-{3-fluoro-5-[({1-[(1R)-1-(naphthalen-1-yl)ethyl]piperidin-4-yl}formamido)methyl]phenyl}-4-methylpiperazine-1-carboxamide	0.44	6.36	36114026		C[C@H](C1=CC=CC2=CC=CC=C21)N3CCC(CC3)C(=O)NCC4=CC(=CC(=C4)F)NC(=O)N5CCN(CC5)C	155925847	-4.4	-5.0
N,2-dimethyl-N-(naphthalen-1-ylmethyl)benzamide	22.6	4.65	36689835		CC1=CC=CC=C1C(=O)N(C)CC2=CC=CC3=CC=CC=C32	60428070	-7.2	-6.2
N'-[(Z)-(2-hydroxyphenyl)methylideneamino]-4-[[[4-[N'-[(Z)-(2-hydroxyphenyl)methylideneamino]carbamimidoyl]-1,2,5-oxadiazol-3-yl]amino]methylamino]-1,2,5-oxadiazole-3-carboximidamide	0.063	7.20	36114026		C1=CC=C(C(=C1)C=NN=C(C2=NON=C2NCNC3=NON=C3C(=NN=CC4=CC=CC=C4O)N)N)O	135868178	-7.5	-5.5
Papyriflavonol	3.7	5.43	36234910		CC(=CCC1=C(C(=CC(=C1)C2=C(C(=O)C3=C(O2)C=C(C(=C3O)CC=C(C)C)O)O)O)O)C	10343070	-5.3	-5.1
pyrantel pamoate	>20		33594371		CN1CCCN=C1/C=C/C2=CC=CS2.C1=CC=C2C(=C1)C=C(C(=C2CC3=C(C(=CC4=CC=CC=C43)C(=O)O)O)O)C(=O)O	5281033	-5.2	-5.4
pyrithione	0.50 ± 0.07	6.30	35943189		C1=CC(=S)N(C=C1)O	1570	-5.7	-2.8
Rac5c	0.81	6.09	36966652		CC(C1=CC=CC2=CC=CC=C21)N3CCC(CC3)C(=O)NCC4=CC(=NC=C4)OC	76853649	-5.0	-4.9
rutin	43.4	4.36	36712981		C[C@@H]5O[C@@H](OC[C@H]4O[C@@H](Oc3c(c1ccc(O)c(O)c1)oc2cc(O)cc(O)c2c3=O)[C@H](O)[C@@H](O)[C@@H]4O)[C@H](O)[C@H](O)[C@H]5O	5280805	-4.2	-9.1
schaftoside	3.91 ± 0.19	5.41	35968270		C1[C@@H]([C@@H]([C@H]([C@@H](O1)C2=C3C(=C(C(=C2O)[C@H]4[C@@H]([C@H]([C@@H]([C@H](O4)CO)O)O)O)O)C(=O)C=C(O3)C5=CC=C(C=C5)O)O)O)O	442658	-5.9	-6.8
Tanshinone I	none		36591160		CC1=C2C=CC3=C(C2=CC=C1)C(=O)C(=O)C4=C3OC=C4C	114917	-5.7	-3.8
Tanshinone IIA	1.57	5.80	36689835		CC1=COC2=C1C(=O)C(=O)C3=C2C=CC4=C3CCCC4(C)C	164676	-5.0	-3.9

theaflavin	7.3	5.14	36712981		O=c6cc(C2Oc1cc(O)cc(O)c1CC2O)cc5c(C4Oc3cc(O)cc(O)c3CC4O)cc(O)c(O)c5c6O	135438671	-6.3	-6.3
thiophene	14.7	4.83	36213008		C1=CSC(=C1)C2=CSC(=C2C(=O)O)NC(=O)C3=CC=C(C=C3)Br	1223274	-4.3	-2.8
XDY2-58	<10		34665619		C[C@@H](NC(C1=C(C)C=CC(N)=C1)=O)C2=CN(C)C3=C2C=CC=C3	156907292	-7.3	-6.5
XDY2-62	3.3	5.48	34665619		C[C@@H](NC(C1=C(C)C=CC(N)=C1)=O)C2=CSC3=C2C=CC=C3	156907291	-7.6	-6.6
XR8-101	1.8	5.74	34665619		C[C@@H](NC(C1=C(C)C=CC(N)=C1)=O)C2=CC(C3=CC=C(CN)S3)=CC=C2	156907336	-7.4	-6.3
XR8-103	1.1	5.96	34665619		C[C@@H](NC(C1=C(C)C=CC(N)=C1)=O)C2=CC(C3=CC=C(CNC(C4CCCC4)=O)S3)=CC=C2	156907337	-7.7	-7.2
XR8-104	2.3	5.64	34665619		C[C@@H](NC(C1=C(C)C=CC(NC(C)=O)=C1)=O)C2=CC(C3=CC=C(CNC(C)=O)S3)=CC=C2	156907338	-8.3	-7.5
XR8-106	1.4	5.89	34665619		C[C@@H](NC(C1=C(C)C=CC(NC(C)=O)=C1)=O)C2=CC(C3=CC=C(CN[C@H]4CC[C@@H](O)C4)S3)=CC=C2	156907339	-8.2	-7.4
XR8-14	1.2	5.92	34665619		C[C@@H](NC(C1=C(C)C=CC(NC2CNC2)=C1)=O)C3=CC(C4=CC=C(CN5CCNCC5)S4)=CC=C3	156907313	-7.2	-6.2
XR8-15	0.9	6.05	34665619		C[C@@H](NC(C1=C(C)C=CC(NC2CNC2)=C1)=O)C3=CC(C4=CC=C(CN5CCOCC5)S4)=CC=C3	156907314	-7.6	-6.7
XR8-16	1.6	5.80	34665619		C[C@@H](NC(C1=C(C)C=CC(NC2CNC2)=C1)=O)C3=CC(C4=CC=C(CNC5CCN(C)CC5)S4)=CC=C3	156907315	-7.1	-6.5
XR8-17	2.7	5.57	34665619		C[C@@H](NC(C1=C(C)C=CC(NC2CNC2)=C1)=O)C3=CC(C4=CC=C(CNCC5CCNCC5)S4)=CC=C3	156907316	-7.3	-6.7
XR8-23	0.39	6.41	34665619		C[C@@H](NC(C1=C(C)C=CC(NC2CNC2)=C1)=O)C3=CC(C4=CC=C(CNC5CCCC5)S4)=CC=C3	156907317	-8.1	-6.0
XR8-24	0.56	6.25	34665619	7LBS	C[C@@H](NC(C1=C(C)C=CC(NC2CNC2)=C1)=O)C3=CC(C4=CC=C(CN5CCCC5)S4)=CC=C3	155801618	-7.9	-6.3
XR8-30	0.75	6.12	34665619		C[C@@H](NC(C1=C(C)C=CC(NC2CNC2)=C1)=O)C3=CC(C4=CC=C(C)S4)=CC=C3	156907318	-7.3	-6.7
XR8-32-1	0.97	6.01	34665619		C[C@@H](NC(C1=C(C)C=CC(NC2CNC2)=C1)=O)C3=CC(C4=CC=C(C(O)=O)S4)=CC=C3	156907319	-7.7	-4.4
XR8-32-2	0.81	6.09	34665619		C[C@@H](NC(C1=C(C)C=CC(NC2CNC2)=C1)=O)C3=CC(C4=CC=C(C(OC)=O)S4)=CC=C3	156907320	-7.7	-7.5
XR8-35	0.92	6.04	34665619		C[C@@H](NC(C1=C(C)C=CC(NC2CNC2)=C1)=O)C3=CC(C4=CC=C(C(NC[C@H]5CCCO5)=O)S4)=CC=C3	156907321	-7.9	-7.8
XR8-38	0.76	6.12	34665619		C[C@@H](NC(C1=C(C)C=CC(NC2CNC2)=C1)=O)C3=CC(C4=CC=C(C(NC[C@@H]5CCCO5)=O)S4)=CC=C3	-	-7.9	-7.2
XR8-39	1.1	5.96	34665619		C[C@@H](NC(C1=C(C)C=CC(NC2CNC2)=C1)=O)C3=CC(C4=CC=C(C(NC)=O)S4)=CC=C3	156907322	-7.3	-7.1
XR8-40	0.82	6.09	34665619		C[C@@H](NC(C1=C(C)C=CC(NC2CNC2)=C1)=O)C3=CC(C4=CC=C(C(NCC5CCO5)=O)S4)=CC=C3	156907323	-7.2	-7.9
XR8-49	0.64	6.19	34665619		C[C@@H](NC(C1=C(C)C=CC(NC2CNC2)=C1)=O)C3=CC(C4=CC=C(CN5CC[C@@H](O)C5)S4)=CC=C3	156907324	-7.9	-7.4
XR8-51	1.1	5.96	34665619		C[C@@H](NC(C1=C(C)C=CC(NC2CNC2)=C1)=O)C3=CC(C4=CC=C(CN5CC[C@@H](C)C5)S4)=CC=C3	156907325	-7.1	-6.6
XR8-56	2.2	5.66	34665619		C[C@@H](NC(C1=C(C)C=CC(NC2CNC2)=C1)=O)C3=CC(C4=CC=C(CN5CCC5)S4)=CC=C3	156907326	-7.9	-6.7
XR8-57	0.7	6.15	34665619		C[C@@H](NC(C1=C(C)C=CC(NC2CNC2)=C1)=O)C3=CC(C4=CC=C(CN5CC[C@H](O)C5)S4)=CC=C3	156907327	-7.6	-6.0
XR8-61	6.5	5.19	34665619		C[C@@H](NC(C1=C(C)C=CC(NC2CNC2)=C1)=O)C3=CC(Br)=CC=C3	156907328	-7.0	-6.8
XR8-65	0.33	6.48	34665619	7LOS	C[C@@H](NC(C1=C(C)C=CC(NC2CNC2)=C1)=O)C3=CC(C4=CC=C(CN[C@@H]5CCOC5)S4)=CC=C3	155801623	-7.8	-6.6

XR8-66	0.62	6.21	34665619		C[C@@H](NC(C1=C(C)C=CC(NC2CNC2)=C1)=O)C3=CC(C4=CC=C(CN[C@H]5CCOC5)S4)=CC=C3	156907329	-7.9	-6.6
XR8-67	0.17	6.77	34665619		C[C@@H](NC(C1=C(C)C=CC(NC2CNC2)=C1)=O)C3=CC(C4=CC=C(CNC5CC(O)CC5)S4)=CC=C3	156907330	-7.8	
XR8-69	0.37	6.43	34665619	7LLZ	C[C@@H](NC(C1=C(C)C=CC(NC2CNC2)=C1)=O)C3=CC(C4=CC=C(CNC(C)=O)S4)=CC=C3	155801622	-7.9	-7.8
XR8-77	0.64	6.19	34665619		C[C@@H](NC(C1=C(C)C=CC(NC2CNC2)=C1)=O)C3=CC(C4=CC=C(CNC5CCNC5=O)S4)=CC=C3	156907331	-7.9	-6.6
XR8-79	0.41	6.39	34665619		C[C@@H](NC(C1=C(C)C=CC(NC2CNC2)=C1)=O)C3=CC(C4=CC=C(CNC5CC(NC5)=O)S4)=CC=C3	156907332	-8.1	-7.5
XR8-8	1.3	5,89	34665619		C[C@@H](NC(C1=C(C)C=CC(NC2CNC2)=C1)=O)C3=CC(C4=CSC=C4)=CC=C3	156907311	-7.8	-7.5
XR8-83	0.21	6.68	34665619	7LLF	C[C@@H](NC(C1=C(C)C=CC(NC2CNC2)=C1)=O)C3=CC(C4=CC=C(CN[C@@H]5CC[C@H](O)C5)S4)=CC=C3	155801621	-7.8	-3.4
XR8-84	0.43	6.37	34665619		C[C@@H](NC(C1=C(C)C=CC(NC2CNC2)=C1)=O)C3=CC(C4=CC=C(CN[C@@H]5CC[C@@H](O)C5)S4)=CC=C3	156907333		-7.3
XR8-89	0.113	6.95	34665619	7LBR	C[C@@H](NC(C1=C(C)C=CC(NC2CNC2)=C1)=O)C3=CC(C4=CC=C(CN[C@H]5CC[C@@H](O)C5)S4)=CC=C3	155801617	-8.2	-6.0
XR8-9	1.8	5.74	34665619		C[C@@H](NC(C1=C(C)C=CC(NC2CNC2)=C1)=O)C3=CC(C4=CNC=C4)=CC=C3	156907312	-7.4	-7.9
XR8-96	0.25	6.60	34665619		C[C@@H](NC(C1=C(C)C=CC(NC2CNC2)=C1)=O)C3=CC(C4=CC=C(CN[C@H]5CC[C@H](O)C5)S4)=CC=C3	156907334	-8.3	-6.9
XR8-98	0.81	6.09	34665619		C[C@@H](NC(C1=C(C)C=CC(N)=C1)=O)C2=CC(C3=CC=C(CNC(OC(C)(C)C)=O)S3)=CC=C2	156907335	-6.7	-7.4
YF4-134	>10		34665619		CC(NC(C1=C(C)C=CC([N+]([O-])=O)=C1)=O)C2=CC=CC3=C2NC=C3	156907293	-7.1	-5.9
YF4-136	4.7	5.33	34665619		CC(NC(C1=C(C)C=CC(N)=C1)=O)C2=CC=CC3=C2NC=C3	156907294	-7.1	-6.1
YF4-137	100	4.00	34665619		CC(NC(C1=C(C)C=CC(N)=C1)=O)C2=CNC3=C2C=CC=C3	156907295	-7.0	-6.2
YF4-145	>>100		34665619		CC(NC(C1=C(C)C=CC(N)=C1)=O)C2=CC=CC3=C2N(C)C=C3	156907296	-7.0	-6.6
YM155	2.47	5.61	36689835		CC1=[N+](C2=C(N1CCOC)C(=O)C3=CC=CC=C3C2=O)CC4=NC=CN=C4.[Br-]	11178236	-4.8	-3.5
YRL	6.68 ± 1.20	5.18	35953531	7OFS	OCCC1=CC=C(C=C1)O	10393	-5.6	-4.4
ZN-2-180	6.08	5.22	34665619		C[C@@H](NC(C1=C(C)C=CC(NC2CN(C2)C(OC(C)(C)C)=O)=C1)=O)C3=CC=CC4=C3C=CC=C4	156907057	-7.0	-6.6
ZN-2-181	1.1	5.96	34665619		C[C@@H](NC(C1=C(C)C=CC(NC2CCN(CC2)C(OC(C)(C)C)=O)=O)=C1)=O)C3=CC=CC4=C3C=CC=C4	156907058	-6.1	-4.2
ZN-2-182	5.5	5.26	34665619		C[C@@H](NC(C1=C(C)C=CC(NC(C2CN(C2)C(OC(C)(C)C)=O)=O)=O)=C1)=O)C3=CC=CC4=C3C=CC=C4	156907094	-7.8	-6.8
ZN-2-183	6	5.22	34665619		C[C@@H](NC(C1=C(C)C=CC(NC(C2CCN(CC2)C(OC(C)(C)C)=O)=O)=O)=C1)=O)C3=CC=CC4=C3C=CC=C4	156907119	-7.5	-6.6
ZN-2-184	1.01 ±0.15	6.00	34665619		C[C@@H](NC(C1=C(C)C=CC(NC2CNC2)=C1)=O)C3=CC=CC4=C3C=CC=C4	156907132	-8.0	-7.6
ZN-2-185	0.6	6.22	34665619		C[C@@H](NC(C1=C(C)C=CC(NC2CCNCC2)=C1)=O)C3=CC=CC4=C3C=CC=C4	156907148	-7.7	-5.8
ZN-2-186	1.2	5.92	34665619		C[C@@H](NC(C1=C(C)C=CC(NC(C2CNC2)=O)=C1)=O)C3=CC=CC4=C3C=CC=C4	156907156	-8.4	-8.3
ZN-2-187	0.8	6.10	34665619		C[C@@H](NC(C1=C(C)C=CC(NC(C2CCNCC2)=O)=C1)=O)C3=CC=CC4=C3C=CC=C4	156907157	-8.2	-7.4
ZN-2-188-1	1.6	5.80	34665619		C[C@@H](NC(C1=C(C)C=CC(N(C)C2CN(C)C2)=C1)=O)C3=CC=CC4=C3C=CC=C4	156907159	-7.7	-7.1
ZN-2-188-2	4.3	5.37	34665619		C[C@@H](NC(C1=C(C)C=CC(NC2CN(C)C2)=C1)=O)C3=CC=CC4=C3C=CC=C4	15690716	-7.8	-7.2

						0		
ZN-2-189	0.7	6.15	34665619		C[C@@H](NC(C1=C(C)C=CC(NC2CCN(C)CC2)=C1)=O)C3=CC=CC4=C3C=CC=C4	156907158	-7.2	-7.0
ZN-2-190	>100		34665619		C[C@@H](NC(C1=C(F)C=CC(N)=C1)=O)C2=CC=CC3=C2C=CC=C3	156907248	-7.8	-7.0
ZN-2-192	4.8	5.32	34665619		C[C@@H](NC(C1=C(Cl)C=CC(N)=C1)=O)C2=CC=CC3=C2C=CC=C3	156907262	-7.5	-7.2
ZN-2-193	>10		34665619		C[C@@H](NC(C1=C(C(F)(F)F)C=CC(N)=C1)=O)C2=CC=CC3=C2C=CC=C3	156907270	-7.4	-6.9
ZN-2-197	2.4	5.62	34665619		C[C@@H](NC(C1=C(C)C=CC(N(C)C2CNC2)=C1)=O)C3=CC=CC4=C3C=CC=C4	156907182	-7.4	-6.3
ZN-3-13	100	4.00	34665619		O=C(C1=C(C)C=CC(N)=C1)NC(CO)C2=CC=CC3=CC=CC=C32	156907283	-7.9	-7.5
ZN-3-19	>100		34665619		C[C@@H](NC(C1=C(C)C=CC(N)=C1)=O)C2=C(O)C=CC3=CC=CC=C32	156907284	-7.7	-7.3
ZN-3-3	>10		34665619		C[C@@H](NC(C1=C(C=C)C=CC(N)=C1)=O)C2=CC=CC3=C2C=CC=C3	156907271	-7.6	-7.0
ZN-3-32	<100		34665619		C[C@@H](NC(C1=C(C)C=CC(N(C2CNC2)C)=C1)=O)C3=C(O)C=CC4=CC=CC=C43	156907287	-7.9	-7.6
ZN-3-33	>10	!	34665619		C[C@@H](NC(C1=C(C)C=CC(N(C2CNC2)C)=C1)=O)C3=C(OC4CNC4)C=CC5=CC=CC=C53	156907288	-5.6	-7.2
ZN-3-34	NI		34665619		C[C@@H](NC(C1=C(C)C=CC(N(C2CNC2)C)=C1)=O)C3=C(OCCN(C)C)C=CC4=CC=CC=C43	164866242	-7.2	-7.4
ZN-3-35	>100		34665619		O=C(C1=C(C)C=CC(N(C2CNC2)C)=C1)NC(CO)C3=CC=CC4=CC=CC=C43	156907289	-7.3	-8.7
ZN-3-36	5.6	4.25	34665619		C[C@@H](NC(C1=C(C)C=CC(N(C2CNC2)C)=C1)=O)C3=NC=CC4=CC=CC=C43	156907307	-7.1	-6.2
ZN-3-40	NI		34665619		C[C@H](NC(C1=C(C)C=CC(N(C2CNC2)C)=C1)=O)C3=NC=CC4=CC=CC=C43	164866161	-7.3	-6.7
ZN-3-41	10-100		34665619		O=C(N[C@@H](C1=CC=CC2=C1C=CC=C2)C)C3=NC(N)=CC=C3C	156907278	-7.5	-6.6
ZN-3-45	5.7	5.24	34665619		O=C(C1=C(C)C=CC(N(C2CNC2)C)=C1)NC(C)C3=CC=CC4=C3NC=C4	156907298	-7.5	-7.0
ZN-3-55	7.4	5.13	34665619		C[C@@H](NC(C1=C(C)C=NC(NC2CNC2)=C1)=O)C3=CC=CC4=CC=CC=C43	156907279	-8.2	-7.7
ZN-3-56	3.9	5.41	34665619		C[C@@H](NC(C1=C(C)C=CC(CNCC(O)=O)=O)=C1)=O)C2=CC=CC3=CC=CC=C32	156907183	-8.4	-8.7
ZN-3-57	10-100		34665619		C[C@@H](NC(C1=C(C)C=NC(N(C)C2CNC2)=C1)=O)C3=CC=CC4=CC=CC=C43	156907280	-7.1	-6.7
ZN-3-59	2.4	5.62	34665619		C[C@@H](NC(C1=C(C)C=CC(N(C2CNC2)C)=C1)=O)C3=CSC4=C3C=CC=C4	156907301	-7.2	-6.6
ZN-3-61	>>10		34665619		O=C(C1=C(C)C=CC(N(C2CNC2)C)=C1)N[C@H](CC)C3=CC=CC4=CC=CC=C43	156907286	-6.9	-8.4
ZN-3-66	4.1	5.39	34665619		C[C@@H](NC(C1=C(Cl)C=CC(NC2CNC2)=C1)=O)C3=CC=CC4=CC=CC=C43	156907281	-8.0	-8.1
ZN-3-67	8.5	5.07	34665619		C[C@@H](NC(C1=C(Cl)C=CC(NC2CNC2)=C1)=O)C3=CSC4=C3C=CC=C4	156907302	-8.0	-7.0
ZN-3-70	10.7	4.97	34665619		C[C@@H](NC(C1=C(Cl)C=CC(N(C)C2CNC2)=C1)=O)C3=CC=CC4=CC=CC=C43	156907282	-7.1	-7.5
ZN-3-71	10.9	4.96	34665619		C[C@@H](NC(C1=C(Cl)C=CC(N(C2CNC2)C)=C1)=O)C3=CSC4=C3C=CC=C4	156907303	-7.8	-7.0
ZN-3-74	2.8	5.55	34665619		C[C@@H](NC(C1=C(Cl)C=CC(NC2CNC2)=C1)=O)C3=CC(C4=CC=CS4)=CC=C3	156907309	-7.8	-7.5
ZN-3-79	1.9	5.72	34665619		C[C@@H](NC(C1=C(C)C=CC(NC2CNC2)=C1)=O)C3=CSC4=C3C=CC=C4	15690730	-7.7	-6.3

						5		
ZN-3-80	0.59	6.23	34665619		C[C@@H](NC(C1=C(C)C=CC(NC2CNC2)=C1)=O)C3=CC(C4=CC=CS4)=CC=C3	156907310	-7.7	-6.1
7724772	20	4.70	36500659		CC1=CC=CC=C1C(=O)NC(C)C2=CC3=CC=CC=C3C=C2	2949898	-7.1	-6.4

Table S2. SP re-docking and cross-docking values .

		PROTEIN													
		7LBR		***7JRN***		***7LLF***		***7LOS***		***7LLZ***		***7JIW***		***7LBS***	
		Docking score	RMSD	Docking score	RMSD	Docking score	RMSD	Docking score	RMSD	Docking score	RMSD	Docking score	RMSD	Docking score	RMSD
LIGAND	***7LBR***	-8.105	1.182	-5.433	2.401	-8.409	1.177	-8.460	1.951	-7.527	1.323	-7.541	1.711	-8.369	1.932
	7JRN	-7.614	0.383	-7.835	0.171	-7.775	0.844	-7.728	0.402	-8.054	0.496	-7.561	0.852	-7.597	0.730
	7LLF	-8.105	1.444	-5.433	2.496	-8.409	1.149	-8.460	2.014	-7.527	1.261	-7.541	2.018	-8.369	2.334
	7LOS	-7.233	2.220	-6.657	1.640	-6.825	1.570	-7.244	2.489	-7.229	2.145	-7.110	2.742	-6.419	1.361
	7LLZ	-8.130	1.987	-8.149	1.498	-7.894	1.950	-8.368	1.200	-8.130	1.987	-7.900	2.019	-8.044	2.002
	7JIW	-6.995	1.220	-7.320	1.110	-7.824	1.062	-7.417	1.338	-7.457	1.040	-7.933	0.420	-7.188	1.372
	7LBS	-8.067	1.225	-6.394	6.500	-7.178	1.163	-7.147	1.228	-6.778	0.732	-7.225	3.065	-7.187	3.032

Table S3. XP re-docking and cross-docking values .

		PROTEIN													
		7LBR		***7JRN***		***7LLF***		***7LOS***		***7LLZ***		***7JIW***		***7LBS***	
		Docking score	RMSD	Docking score	RMSD	Docking score	RMSD	Docking score	RMSD	Docking score	RMSD	Docking score	RMSD	Docking score	RMSD
LIGAND	***7LBR***	-6.766	2.323	-5.759	1.838	-7.860	1.257	-6.950	1.367	-6.731	1.633	-6.360	1.653	-7.092	1.931
	7JRN	-6.595	0.325	-7.227	0.104	-7.184	0.755	-7.063	0.432	-6.845	0.470	-6.428	0.844	-6.714	0.541
	7LLF	-6.766	1.964	-5.759	1.600	-7.860	1.208	-6.950	1.682	-6.731	1.544	-6.360	1.930	-7.092	1.785
	7LOS	-5.109	2.281	-5.934	2.054	-6.958	2.984	-7.225	2.392	-5.708	2.147	-5.308	2.200	-6.557	2.549
	7LLZ	-8.723	1.817	-8.295	1.961	-7.552	2.143	-6.832	1.202	-7.414	1.819	-6.954	2.085	-7.378	1.788
	7JIW	-6.28	1.204	-6.923	1.245	-6.538	1.359	-6.441	2.619	-6.158	1.313	-6.559	0.451	-6.376	1.603

D														
	7LBS	-6.306	1.006	-5.784	2.732	-6.527	1.916	-3.908	2.066	-6.242	2.134	-6.152	2.093	-6.743	2.979

Table S4. Covalent compounds

Compound	Internal name	Name_alt	IC50(microM)	Pic50	PMID	PDB	Isomeric_smiles
VIR250	Compound 1		50	4.30103	34665619	6WUU	c1cccc(c12)sc(n2)CCC(NC(=O)C)C(=O)NC(C[N+])C(=O)NCC(=O)NCC=CC(=O)OC
VIR251	Compound 2		50	4.30103	34665619	6WX4	c1cc(O)ccc1CCC(NC(=O)C)C(=O)NC(C[N+])C(=O)NCC(=O)NCC=CC(=O)OC
(R)-2-(3-hydrazineyl-3-oxopropyl)-N-(1-(naphthalen-1-yl)ethyl)benzamide	Compound 3		24	4.619788 76	36977673		N(N)C(CCC1=C(C(=O)N[C@H](C)C2=CC=CC3=CC=CC=C23)C=CC=C1)=O
methyl(R,E)-4-(2-(3-(2-((1-(naphthalen-1 yl)ethyl)carbamoyl)phenyl)propanoyl)hydrazineyl)-4-oxobut-2-enoate	Compound 4	PLpro inhibitor 7	0.094	7.026872 15	36977673	8EUA	COC(\C=C\C(=O)NNC(CCC1=C(C=CC=C1)C(N[C@H](C)C1=CC=CC2=CC=CC=C12)=O)=O)=O
methyl-(R,E)-4-(2-(3-(4-acetamido-2-((1-(naphthalen-1-yl)ethyl)carbamoyl)phenyl)propanoyl)hydrazineyl)-4-oxobut-2-enoate	Compound 5		0.23	6.638272 16	36977673		COC(\C=C\C(=O)NNC(CCC1=C(C=C(C=C1)NC(C)=O)C(N[C@H](C)C1=CC=CC2=CC=CC=C12)=O)=O)=O
(R)-2-(3-(2-(2-chloroacetyl)hydrazineyl)-3-oxopropyl)-N-(1-(naphthalen-1-yl)ethyl)benzamide	Compound 6		5.4	5.267606 24	36977673		ClCC(=O)NNC(CCC1=C(C(=O)N[C@H](C)C2=CC=CC3=CC=CC=C23)C=CC=C1)=O
(R)-5-acetamido-2-(3-(2-(2-chloroacetyl)hydrazineyl)-3-oxopropyl)-N-(1-(naphthalen-1-yl)ethyl)benzamide	Compound 7		4.4	5.356547 32	36977673		C(C)(=O)NC=1C=CC(=C(C(=O)N[C@H](C)C2=CC=CC3=CC=CC=C23)C1)CCC(=O)NNC(CCl)=O
(R)-2-(3-(2-(2-cyanoacetyl)hydrazineyl)-3-oxopropyl)-N-(1-(naphthalen-1-yl)ethyl)benzamide	Compound 8		8	5.096910 01	36977673		C(#N)CC(=O)NNC(CCC1=C(C(=O)N[C@H](C)C2=CC=CC3=CC=CC=C23)C=CC=C1)=O
(R)-N-(1-(naphthalen-1-yl)ethyl)-2-(3-oxo-3-(2-propioloylhydrazineyl)propyl)benzamide	Compound 9		0.098	7.008773 92	36977673		C1(=CC=CC2=CC=CC=C12)[C@@H](C)NC(C1=C(C=CC=C1)CCC(NNC(C#C)=O)=O)=O
(R)-5-acetamido-2-methyl-N-(1-(naphthalen-1-yl)ethyl)benzamide	Compound 10	1093070 -14-4	6.2	5.207608 3	36977673		C(C)(=O)NC=1C=CC(=C(C(=O)N[C@H](C)C2=CC=CC3=CC=CC=C23)C1)CCC(=O)NNC(CCl)=O
Disulfiram	Compound 11		0.48	6.318758 76	36689835		CCN(CC)C(=S)SSC(=S)N(CC)CC