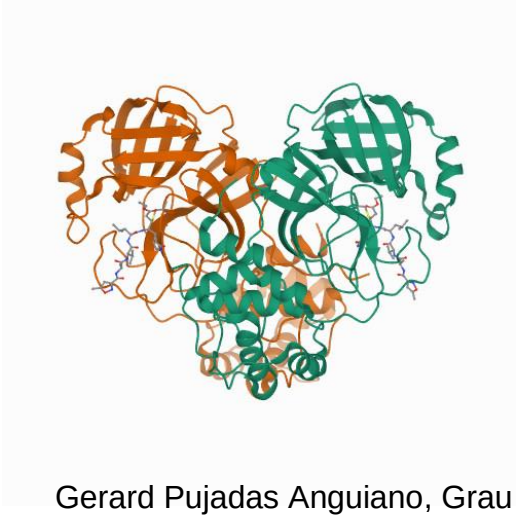
UNIVERSITAT ROVIRA i VIRGILI

BINDING AFFINITY AND POSE PREDICTION FOR NON-COVALENT M-PRO SARS-CoV-2 INHIBITORS: AN EVALUATION OF MOST POPULAR PREDICTION METHODOLOGIES

Júlia Vilalta Mor

FINAL DEGREE PROJECT
Bachelor's Degree in Biotechnology



Academic and professional tutor:	Gerard Pujadas Anguiano, Grau en Biotecnologia, Departament de Bioquímica i Biotecnologia, gerard.pujadas@urv.cat
In cooperation with:	Grup de recerca en Quimioinformàtica i nutrició Universitat Rovira i Virgili
Supervisors:	Santiago Garcia Vallvé, Grau en Biotecnologia, Departament de Bioquímica i Biotecnologia, santi.garcia-vallve@urv.cat
	Guillem Macip Sancho, PhD, Grau en Biotecnologia, Departament de Bioquímica i Biotecnologia, guillem.macip@urv.cat

June 2021

Jo, Júlia Vilalta Mor , amb DNI 39933805D, soc coneixedora 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/vida-campus/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, 1 de juny de 2021

TABLE OF CONTENTS

INFORMATION OF THE CENTRE

This research project has been developed at the Cheminformatics and Nutrition research group, which is part of the Department of Biochemistry and Biotechnology of the Rovira i Virgili University. The lab is located in the *Campus Sescelades, carrer Marcel·lí Domingo núm. 1*. Universitat Rovira i Virgili is a public institution created by the Parliament of Catalonia in 1991 to open doors to education in the *Camp de Tarragona*, and it has become a centre of excellence in training, research, and development.

The Cheminformatics and Nutrition research group focuses its research on using computational tools to develop new strategies for the identification of bioactive compounds and provide new functions to certain molecules, such as authorized drugs. Thus, it is a leading research group in the field of bioinformatics where its expertise is demonstrated through several articles published in the latest years in top international peer-reviewed journals.

Following the outbreak of the pandemic caused by the COVID-19 infection, the group has focused specially on applying its expertise to find new non-covalent inhibitors for the main protease (M-pro) of the virus that cause the infection (*i.e.,* SARS-CoV-2), and consequently contribute to this social emergency providing rapid and non-invasive studies. The methodology carried out in this group is essentially based on combining a computational part (*in silico*) with an experimental part (*in vitro and in vivo*), which will allow to correlate them in order to identify which are the most bioactive compounds.

Júlia Vilalta Mor
Universitat Rovira i Virgili

ABSTRACT AND KEY WORDS

COVID-19 disease pandemic has resulted in a change in global health and economy that has involved the scientific community to find effective treatments in order to inhibit the virus responsible for this disease, the SARS-CoV-2. One of the key targets is its main protease (M-pro) as it has an important role in the replication of the virus. Computational approaches are essential for its important role in drug discovery research. Here, we make an evaluation of the two most popular methodologies used in drug discovery studies: prediction of protein-ligand binding affinity and pose prediction for non-covalent M-pro inhibitors. The combination of both prediction methodologies could be useful to predict the bioactivity of potential M-pro inhibitors and step forward on the drug discovery used for the COVID-19 disease treatment. In this way, we take advantage of the protein-ligand complexes between M-pro and non-covalent inhibitors available thanks to the COVID Moonshot project to analyse the performance of different commonly used tools for both predictive methodologies and evaluate the state-of-art of the accuracy of their predictions.

Key words:

M-pro, SARS-CoV-2, non-covalent inhibitors, binding affinity, docking, prediction

1. INTRODUCTION

The actual pandemic of coronavirus disease (COVID-19) is caused by the virus named **s**evere **a**cute **r**espiratory **s**yndrome-**co**rona**v**irus **2**, known as SARS-CoV-2. The outbreak started in December 2019 in the city of Wuhan (China) and it has been rapidly spreading globally despite the efforts to contain the outbreak, causing a global health emergency situation declared by the World Health Organization (WHO) [1,2]. Symptoms of this disease include fever, cough, sore throat, runny nose and difficulty in breathing, but it could further aggravate the situation to a severe progressive pneumonia, multiorgan failure and death. This pandemic counts on 170,377,345 global confirmed cases, including 3,542,391 deaths as of May 31, 2021, according to the Johns Hopkins University Coronavirus tracker (https://coronavirus.jhu.edu/), and has not only changed human health but also affected adversely the global economy [3].

Although vaccines against SARS-CoV-2 infection have emerged at an enormously high speed, the development of anti-viral drugs is crucial for the treatment of this disease in infected people. The current treatment for COVID-19 mainly consist of repurposed agents, such as protease inhibitors, mainly against the main protein M-pro [4].

For this reason, drug discovery studies are essential in order to find new drugs capable to revert this disease caused by a virus and, in this way, make a contribution in this global health emergency. Thanks to *in silico* studies, drug candidates can be discovered quickly, economically and safely through computational approaches such as drug-target interaction analyses [5].

1.1. SARS-CoV-2 and its main protease (M-pro)

SARS-CoV-2 belongs to the family of *Coronaviridae* and the genus *Betacoronavirus*, which are characterised by the following features: they are enveloped virus, with positive-sense and single-stranded RNA ranging from 26 kb to 32 kb approximately [6].

SARS-CoV-2 genome size varies from 29.8 kb to 29.9 kb, its structure contains an ORF1ab region on 5' extreme that encodes orf1ab polyproteins and takes more than two-thirds of the genome, whereas the 3' extreme contains genes that encode structural proteins such as surface proteins or Spike protein (S), envelope proteins (E), membrane proteins (M), and nucleocapsid N proteins (N). Moreover, it contains 6 accessory proteins encoded by ORF3a, ORF6, ORF7a, ORF7b, and ORF8 genes [7]. SARS-COV-2 structure is represented in Figure 1a.

The main protease or M-pro (also known as 3CL-pro) has become a key drug target for antiviral treatments studies, as the protein exclusively cleaves polypeptide sequences after a glutamine residue, being a high substrate specificity protease. SARS-CoV-2 M-pro has a lytic activity, and it cleaves the overlapping pp1a and pp1ab polyproteins to convert them into functional proteins. This step is critical during viral replication. The processing of these polyproteins is made together with the papain-like protease (PL-pro), and they give rise to 16 non-structural proteins (nsp1-16), as Figure 1b shows. Therefore, M-pro inhibition will be able to stop the production of infectious viral particles and thus alleviate disease symptoms [6].

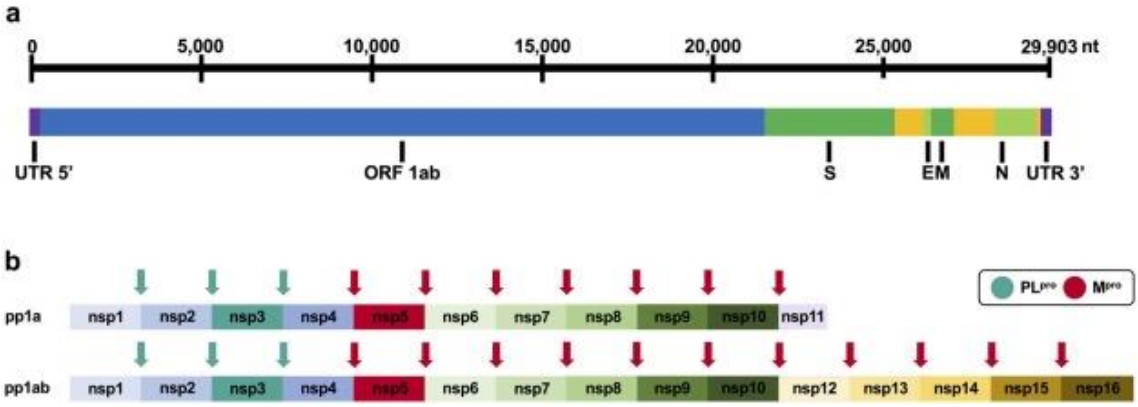


Figure 1. (a) Organization of the genome of SARS-CoV-2, where the RNA is indicated with all genes. (b) Polyproteins pp1a and pp1ab cleavage sites of SARS-CoV-2, where M-pro cleavage sites are indicated in red arrows and PLpro cleavage sites are indicated in green arrows, and non-structural proteins as a result of this lytic process are also indicated. Figure obtained from *Ullrich et al.* [6]

The structure of the SARS-CoV-2 main protease is composed of two dimers with a crystallographic two-fold axis of symmetry, known as protomers. Each protomer is composed of three domains: domain I (residues 8 - 101), domain II (residues 102 - 184), and domain III (residues 201 - 303) (Figure 2) [8,9]. Domains I and II have an antiparallel β-barrel structure, where the substrate-binding site is located between them, known as Cys-His catalytic dyad (residues His^{41} and Cys^{145}). In contrast, domain III has a largely antiparallel globular cluster composed of five α-helices, and it is connected to domain II by a long loop region (residues 185-200) [8,9]. Moreover, Domain III is involved in regulating the dimerization of the M-pro, a necessary process for catalytic activity of the enzyme [1]. The two protomers interact together with the NH_2-terminal residues ("N-finger") of one protomer and the Glu^{166} of the other one, and thereby the S1 pocket is helped to shape for the substrate binding site. To reach this interaction site, the N-finger is squeezed in between domain II of the other monomer and domains II and III of the parent one. Thanks to this event, the M-pro protein can be active and catalyse their substrates using amino acid residues of the catalytic site, Cys^{145} and His^{41} [1].

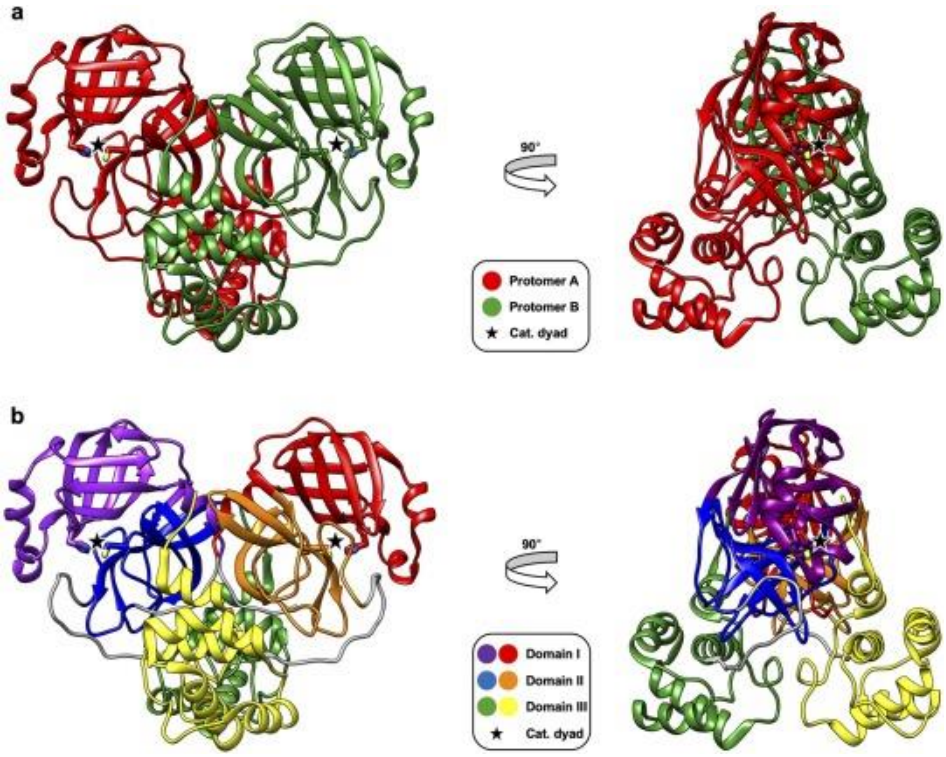


Figure 2. Three-dimensional structure of SARS-CoV-2 M-pro in two different views. Residues of the catalytic dyad (His^{41} and Cys^{145}) are indicated. (a) Protomers are indicated in red and green colours. (b) Domains are indicated in 6 different colours. Figure obtained from *Ullrich et al.* [6]

SARS-CoV-2 M-pro structures forming complexes with non-covalent inhibitors can be obtained from COVID Moonshot [10]. COVID Moonshot is a collaborative world project aimed to accelerate the development of a COVID antiviral. To this end, they have published at the Fragalysis website (https://fragalysis.diamond.ac.uk/viewer/react/preview/target/Mpro) the 3D structure of several SARS-CoV-2 M-pro complexes with small drug fragments that can be freely used by scientifics around the world as the basis for designing new putative M-pro inhibitors. Then, some of the designed drugs have been built, their bioactivity measured and, some of them, co-crystallized again with SARS-CoV-2 M-pro [11]. The enzymatic activity of different compounds has been measured using the two complementary biochemical assays fluorescence based assay and RapidFire Mass Spectrometry assay, and structural characterization has been made with crystal structures when combined with the fragment screening data [11]. Thanks to this initiative, the structural data and the activity have been taken for this study. All data can be found on the COVID Moonshot website [10], that also links to crystal structures found on Fragalysis [12].

Furthermore, we only selected non-covalent SARS-CoV-2 M-pro inhibitors from all ligands taken from these ligand-protein complexes that COVID Moonshot provides, since covalent and non-covalent inhibitors have noticeable differences such as the mode of binding with a protein.

Prediction of drug-target interactions can substantially accelerate lead generation and optimization in drug discovery to yield successful candidate compounds. Detecting interactions between drugs and targets is currently applied in virtual screening (a.k.a. *in silico* screening), a computational method used to filter a subset of compounds from a large library of chemical compounds with the aim of identifying novel active compounds [13]. M-pro non-covalent inhibitors interactions are one of the promising strategies for developing effective antivirals against SARS-CoV-2 since this protease is a key piece of the cell-replication machinery of the virus. Therefore, we provide a useful evaluation of the most popular prediction methodologies for binding free energy and for pose prediction (protein-ligand docking) for non-covalent SARS-CoV-2 M-pro inhibitors.

1.2. Protein-ligand binding affinity

Interaction between proteins and their ligands is measured by binding affinity, which is the quantification by how strong the chemical compounds, also known as a ligand, binds to its counterpart protein [14]. How stronger the readout for binding affinity is, the stronger will be the interaction between protein and ligand to be inferred [15]. Predicting the binding affinity between compounds and targets with reasonable accuracy is crucial in drug discovery, in our case between proteins and ligands, and its prediction through computational approaches enhances the probability of finding lead compounds by reducing the number of wet-lab experiments and saving time and cost in drug discovery studies [13].

Terms used to quantify the binding affinity are inhibition constant (K_i), dissociation constant (K_d), changes in free energy measures (ΔG, ΔH) and the half-maximal inhibitory concentration (IC_{50}) [15]. Moreover, the binding affinity is the single most important initial indicator of drug potency despite it is also the most challenging to predict. A number of methods exists to determine it experimentally, such as the half-maximal inhibitory concentration (IC_{50}), the isothermal titration calorimetry (ITC) or the surface plasmon resonance (SPR) [16]. In this study, the experimental binding affinity data is obtained from COVID Moonshot in terms of IC_{50} or pIC_{50} (negative logarithm of IC_{50}). However, we focus our study on the binding affinity prediction, not on the experimentally determined. We use different computational approaches in order to obtain an accurate binding affinity prediction between M-pro and different compounds, and find an accurate correlation between the predicted and the experimental data. The combination between a computational part (*in silico*) and an experimental part (*in vitro* and *in vivo*) aims to identify new active compounds and demonstrate their predicted activities.

In this study, we used web servers since they are user-friendly online tools for the prediction of binding affinity which can be used without a program installation [17]. Programs can be limited by computational cost or required skills and resources needed to properly make use of the software. Therefore, we selected the following web servers: PRODIGY [17], CSM-lig [18], Alloscore [19], GraphDelta [20] and DeltaDelta [21].

PRODIGY web server (https://bianca.science.uu.nl/prodigy/) is a simple binding affinity descriptor for protein-small ligand systems which makes use of atomic contacts instead of residue contacts. It has been trained on 200 protein-ligand complexes with known experimental binding affinity and structure from 2P2I dataset (a hand-curated structural database dedicated to protein-protein interactions with known small molecule orthosteric modulators [22]) and refined through HADDOCK (a docking software [23]), and the combination of structural and energy based terms is used to train multiple linear regression models [17]. In addition, CSM-lig web server (http://biosig.unimelb.edu.au/csm_lig/prediction) uses a class of graph-based signatures, called cutoff scanning matrix (CSM), to represent 3D environment of proteins and small ligands that are used to train predictors for receptor-based ligand prediction. It was built and evaluated using the release of PDBbind 2014 [18]. Moreover, Alloscore web server (http://mdl.shsmu.edu.cn/alloscore/) was developed using experimentally determined allosteric complexes with binding affinity from ASD v3.0 (an Allosteric Database developed to provide information of allosteric regulation [24]), and after its training and testing, the final model was deployed on the web server [19]. Furthermore, Graphdelta web server (https://mpnn.syntelly.com/) is a graph-convolutional neural network model, using as multitask learning the target variables of K_d, K_i and IC_{50}, and it has been trained on the PDBbind dataset v.2018 [20]. Finally, DeltaDelta is a web server from Playmolecule (https://www.playmolecule.com/DeltaDelta/) that predicts protein-ligand binding affinity through a deep learning-based methodology [25]. Its applicability is the prediction of delta delta free energies in congeneric series. It is a neural network capable of predict the binding affinity running two phases: (1) Training: the initial predictive model has been pre-trained on hundreds of congeneric series from BindingDB database, and this input allows to optimise and fine tune this model with the training dataset provided by the user (*i.e.,* M-pro non-covalent ligands). (2) Testing: the affinity prediction of the test dataset is calculated from the previous trained model [21,25].

However, different methodologies have been used in order to predict the protein-ligand binding affinity. Currently, scientific advances certainly opt for deep learning and machine learning models since they are an alternative to traditional physics-based free energy scoring functions

[26]. Deep learning approach benefit is that it learns binding interaction rules directly from an atomic representation without relying on hand-curated features that may not capture the mechanism of binding [26]. This method predicts the binding affinity using regression and learning intrinsic patterns in a complex plane of available data, that make an optimal prediction without compromising on accuracy [15]. There are two types of three-dimensional structure-based deep learning approaches commonly used: (1) 3D-convolutional neural network (3D-CNN) and (2) spatial graph convolutional neural network (SG-CNN) [26]. The 3D-CNN uses a 3D voxel representation of atoms that accounts for pairwise relationships among them through their relative positioning in the 3D voxel grip but does not predetermine a minimum atomic resolution, whereas the SG-CNN uses explicit distance thresholds to determine which pairs of atoms to consider in pairwise interactions [26]. Examples of deep learning 3D-CNN methodologies are DeltaDelta web server [21] previously described and KDeep [27], also from Playmolecule [28], that in this case it has been pre-trained using PDBbind v.2016 database and it was voxelized into 8 different pharmacophoric-like features (hydrophobic, aromatic, hydrogen-bond donor and acceptor, positive and negative ionizable, metallic and total excluded volume) [28]. DeepAtom [14] is another 3D-CNN also trained with PDBbind v.2016 core set, but this model improves representational capacity. In addition, other currently 3D-CNN approaches are, for instance, DEELIG [15], DeepDTA [29], AK-score [30], and DeepBindRG [31]. On the other hand, one example of spatial graph convolutional neural network (SG-CNN) methodology is graphDelta [20]. Furthermore, Jones et al. approach combine both 3D model representations (3D-CCN and SG-CNN) for protein-ligand binding affinity prediction, and they successfully demonstrated an improved accuracy [26]. Finally, other current methodologies are based on free energy perturbations (FEP), such as CHARMM-GUI free energy calculator [32] and gREST + FEP method [33] that both combines FEP and MD (molecular dynamics) methodology, as well as there are the RASPD+ method [34] that combines a rapid screening with physiochemical descriptors and machine learning.

1.3. Docking

Protein-ligand docking methods predict multiple binding poses and calculate their affinities through scoring functions. Molecular docking methods have been widely used to virtually screen potential drug candidates since they are considered as fast and efficient computational approaches. The determination of reliable poses is a prerequisite for high-precision drug design. For this reason, a variety of computational docking methods have been proposed in order to predict an accurate pose [33].

Molecular docking generally involves two stages: (1) sampling the pose of the ligand in the protein binding pocket and (2) scoring the binding energy of the ligand to the receptor [35]. These two stages are shown at the top of Figure 3. At first step, docking programs like Glide [36] and FRED [37] use an anchor and growth strategy to gradually build up the ligand in the binding site whereas docking program AutoDock [31] use a generic algorithm (GA) approach that address the high computational cost associated with stochastic methods. At second step, a scoring function (SF) is required in order to predict the most likely pose within the set of all generated poses and then, use it to rank-order the top predicted poses for each compound and compare energies between compounds [35].

Scoring functions (SF) are the most important component of molecular docking as they have three major functions: (1) determine the binding mode and site of a ligand binding to a protein, (2) predict the absolute protein-ligand binding affinity in lead optimization, and (3) virtual screening in order to identify the potential drug leads for a given protein target [38]. Since current research, scoring functions are classified into four categories, described in Figure 3: (1) physics-based, (2) empirical, (3) knowledge-based, and (4) machine-learning method [38]. To begin with, physics-based scoring functions are based on force field, solvation models and quantum mechanics methods. The prediction of these scoring functions is greater accurate due to the consideration of the enthalpy, solvation and entropy although it has an expensive computational cost [38]. Otherwise, empirical scoring functions approximate protein-ligand interactions using equations of several physics-based terms, mimicking van de Waals interactions, solvation free energy, electrostatic interactions, etc. These empirical scoring functions have been implemented to various docking programs, although they are still actively developed because of their simple calculations and their close relationship with physic-based interactions [30]. In contrast, knowledge-based scoring functions use the three-dimensional structures of a large set of protein-ligand complexes in order to derive the desired pairwise potentials, assuming that the frequency of different atom pairs in different distances is related to the interaction of two atoms. Then, this frequency is converted into the distance-dependent potential of mean force. It has lower computing cost and higher predictive accuracy compared with the physics-based and empirical SFs, but it could have problems to locate the reference state [38]. Finally, machine-learning-based scoring functions use a training dataset docked by classical docking software, and then the docked structure is rescored using machine learning algorithms such as random forest, neural network or deep-learning. In this way, the accuracy can be improved [38].

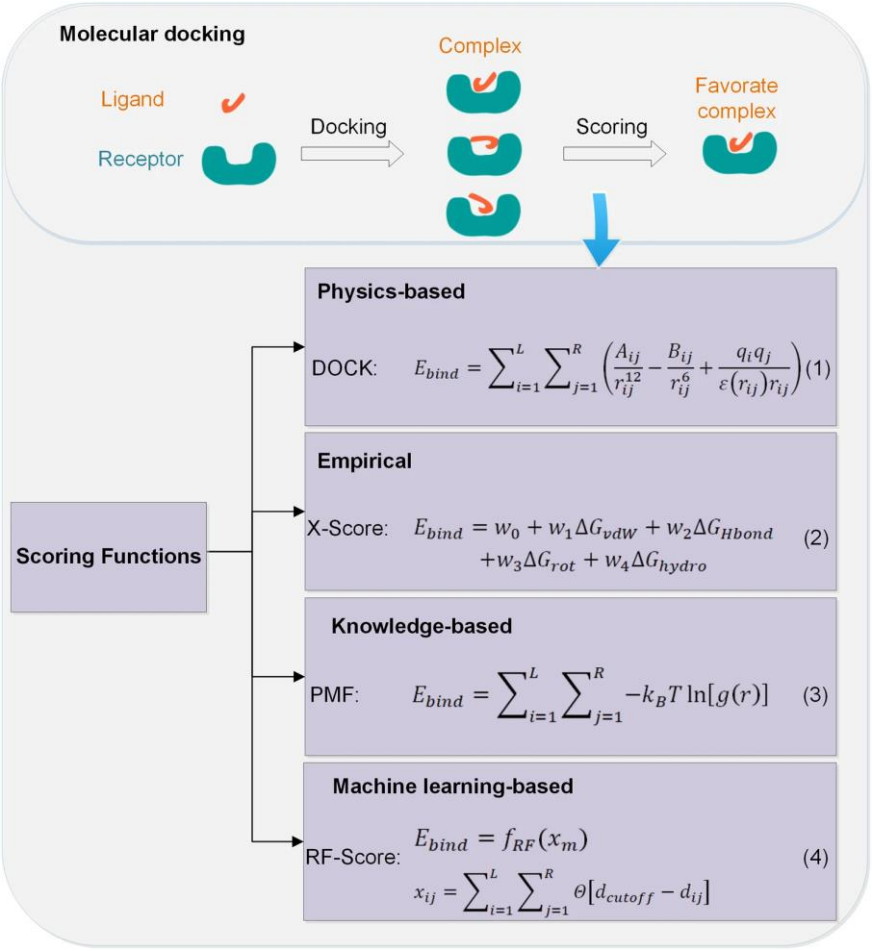


Figure 3. Two steps of molecular docking and four categories of the scoring functions for protein-ligand docking. Figure obtained from *Li J et al.* [38]

Assessing a docking program's ability to successfully predict tan accurate pose is usually done by redocking (*i.e.,* rebuilding the 3D complex from its individual components) and comparing the conformations of crystalized (*i.e.,* experimental) and docked ligands poses [39]. The quantitative measure of comparison is determined by the root mean square deviation (RMSD). The lower value, the similar conformation to the crystal structure. Molecular docking scoring functions are considered to predict a correct pose if the RMSD values are less than the accepted threshold [39].

In this study we perform a pose prediction using 5 different docking programs/methodologies: Glide SP, Glide XP, Glide HTVS [36], FRED [37] and AutoDock Vina [40]. Glide is a program of Schrödinger [41] that uses Glide Score as a scoring function between protein-ligand complexes and have 3 different docking modes: SP, XP and HTVS. Glide HTVS and SP search for possible locations of the ligand in the binding-site region of a receptor using a series of hierarchical filters and they follow this process: first, the shape and properties of the receptor are represented on a grid by different sets of fields that provide progressively more accurate scoring of the ligand

pose; then, an enumeration of ligand torsions generates a collection of ligand conformations to examinate during the docking process; after these ligand conformations, initial screens are performed over the entire phase space available to the ligand to locate promising ligand poses; then, the ligand is refined in torsional space in the field of the receptor with a distance-dependent dielectric model (this step differs in Glide SP and XP that use OPLS3, and Glide HTVS uses OPLS2005 refine methods); finally, a small number of poses are minimized within the field of the receptor with full ligand flexibility [42]. The main distinction within these 3 programs depends on their speed and accuracy: Glide HTVS can dock and trades sampling breath for higher speeds; Glide SP performs exhaustive sampling and is recommended for its balance between speed and accuracy; and Glide XP employs and anchor-and grow sampling approach and a different functional form for GlideScore, being also the slower mode [42]. In contrast, FRED is a Fast Exhaustive Docking from Openeye scientific [37] that performs a systematic and non-stochastic examination of all possible protein-ligand poses, filters for shape complementary and chemical features alignment before selecting and optimizing poses. It uses the Chemgauss4 scoring function and the best scoring pose is then used to rank the ligand against other ligands in the screening database [43]. Finally, AutoDock Vina automatically calculates the grid maps and clusters the results in a transparent way to the user. The program functions as a machine learning method and it have a refined scoring function [40].

Nevertheless, other protein-ligand docking programs have been particularly used in other virtual screening studies. Some of these docking programs are the following: DOCK 5 [44], AutoDock 4 [45] (previous to AutoDock Vina [40]), GOLD [46], FlexX [47], Surflex-DOCK [48], and SwissDock [49]. As well as that, there are scoring functions to re-score docked poses in order to get better results, such as X-Score [50], ChemScore [51], and ChemPLP [52] [30].

2. HYPOTHESIS AND AIM OF THIS STUDY

The hypothesis of this research project is that it is possible to find: (1) a methodology capable of quantitatively predict the bioactivity of a drug from its 3D non-covalent complex with the M-pro from SARS-CoV-2, and (2) a protein-ligand docking methodology capable of predicting the 3D structure of the complex between the SARS-CoV-2 M-pro and non-covalent inhibitors. This combination of prediction will be useful in order to predict the bioactivity of potential inhibitors that are not co-crystallised with the SARS-CoV-2 M-pro and, therefore find new drugs that can be used for the treatment of the COVID-19 infection.

Therefore, the main aim of this study is to benefit from the large number of available protein-ligand 3D complexes between the SARS-CoV-2 M-pro and non-covalent inhibitors with known IC_{50} value in order to evaluate:

1. Which is the drug-target binding affinity program or web server predictor capable of perform a better correlation between predicted binding affinity values and experimentally known IC_{50} values for these non-covalently bound M-pro inhibitors.
2. Which is the protein-ligand docking program capable to better predict the structure of these complexes from their individual elements (*i.e.,* the SARS-CoV-2 M-pro and their corresponding non-covalent inhibitors).

3. MATERIALS AND METHODS

3.1. Library preparation

The library of compounds used in this study was obtained from the COVID Moonshot initiative, that has the activity data in its web page available to download (https://covid.postera.ai/covid) and the crystal structures posted on FRAGALYSIS database (https://fragalysis.diamond.ac.uk/viewer/react/preview/target/Mpro).

The M-pro complexes were visualized using the ligand interaction option in Maestro [53] in order to select only the structures that contain a non-covalent bonded ligand. Therefore, the library was created with these non-covalent M-pro complexes and then was updated with the new M-pro structures as the content of the project is constantly growing. Thus, depending on the date, the libraries are: (1) Old library: library of 40 non-covalent compounds created on 2020-12-02, and (2) Updated library: library of 111 non-covalent compounds created on 2021-02-26 (libraries shown in *Annex 1*).

Furthermore, these complexes were processed in Maestro in order to obtain separately the protein structure and the ligand structure in the file format needed for further predictions. It was done by the 'split' option and exported in different formats. Wherefore, the libraries contained the following files: (1) Complexes (M-pro - non-covalent ligand) in PDB format, (2) proteins (M-pro) in PBD format, (3) proteins (M-pro) in MOL2 format, (4) ligands in MOL2 format, (5) ligands in SDF format, and (6) a compound tracker CSV containing the compound SMILES and the data from COVID Moonshot.

3.2. Binding affinity prediction

3.2.1. BINDING AFFINITY PREDICTION PROGRAM OR WEB SERVER

The research for an appropriate program or web served capable to predict the protein-ligand binding affinity was done through bibliographic searches. After this research we decided to use web servers as they are easier to handle. Therefore, we select five different web servers: PRODIGY [17], CSM-lig [18], Alloscore [19], GraphDelta [20] and DeltaDelta[21].

3.2.2. WEB SERVER PREDICTORS SETUP

Compounds to be predicted were upload as an input on a different way depending on the web server used. PRODIGY [17] web server inputs were (1) a PDB file containing the structure of M-pro complex, (2) selection of the protein chain ID where the ligand is located (ex., A) and (3) ligand ID as in PDB file. CSM-lig [18] inputs were similar: (1) A protein/small-molecule complex

in PDB format, (2) the small-molecule ID as in PDB and (3) the canonical SMILES string of the small-molecule. In contrast, Alloscore [19] inputs were the following: (1) a PDB file containing only the protein structure and (2) a ligand file in MOL2 format, as well as both protein and ligand should be hydrogen-added and pre-docked. Furthermore, GraphDelta [20] needed the following inputs: (1) a PDB file containing the protein structure without hydrogens and only with standard amino acids, and (2) a SDF file with the ligand molecule.

On the other hand, the operation of DeltaDelta web server was different as it is a neural network that allows to put as an input a ligand set for training. Thus, for running this web server it was necessary to provide (1) a file of the receptor in MOL2 format (*i.e.,* a file containing the M-pro structure), (2) a SDF file containing a set of training ligands containing a field called pIC50, (3) a SDF file containing a set of test ligands, and (4) select the validation field for assess the accuracy of the prediction. Figure 3 shows the web server with all fields commented. In this case, we run 40 protein structures from the initial library with 4 different ligand distributions between test and training. A 75% of the ligands were set on the training SDF file whereas a 25% of the ligands were set on the test SDF file, that this corresponds to train 30 ligands and test 10 ligands. The distributions were made randomly according to their experimental pIC_{50} values. The reason of run 4 different distributions is for validating if the neural network works properly (cross-validation method).

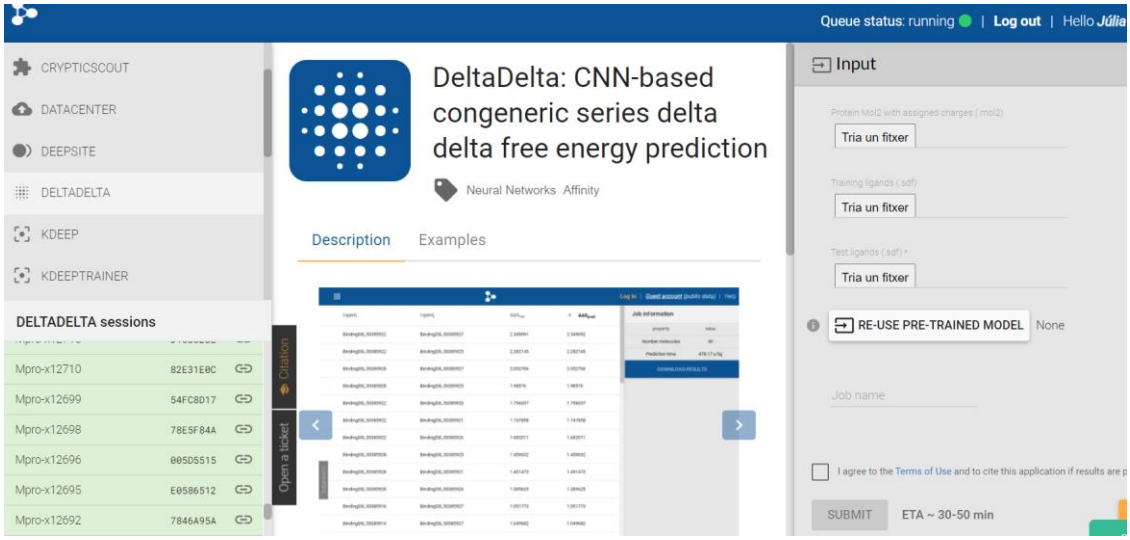


Figure 4. DeltaDelta web server from Playmolecule [21].

3.2.3. PREDICTED AND EXPERIMENTAL BINDING AFFINITY CORRELATION

The outputs obtained in the four web servers were downloaded and added into a spreadsheet software. The spreadsheet was built to have the following information: (1) ID of each M-pro-ligand complex, (2) experimental pIC_{50} obtained from COVID Moonshot (*i.e.,* we used Fluorescence pIC_{50} values), and (3) predicted binding affinity obtained from the web servers used in this study.

Afterward, the data of experimental and predicted binding affinity from this spreadsheet was selected and the Pearson correlation coefficient was calculated and a correlation plot was performed for each web server.

3.2.4. BETTER WEB SERVER AND M-PRO STRUCTURE

Finally, we analysed which web server was able to better predict the binding affinity looking at the correlation results. The higher Pearson correlation coefficient, the better web server predictor. After this web server was selected, we continued using it in order to figure out which M-pro complex was better for the following step of docking.

Then, as the initial library was updated, we need to run all the 111 compounds to obtain more accurate results. Thus, DeltaDelta web server was carried out by M-pro new structures obtained from the library update using the first ligand distribution. At the end of the process, the 10 best M-pro structures were selected in order to figure out which one was better for the following step of docking. Therefore, these 10 selected structures were run with 5 different distributions containing all 111 ligands from the updated library, as in order to optimize the initial model of the neural network it is necessary to further train it with as many examples as possible of our compounds of interest.

Finally, the results of the predicted affinity were correlated with the experimental affinity and the best M-pro structure was selected for the next step according to their Pearson correlation coefficient value.

3.3. Docking

3.3.1. M-PRO PREPARATION, LIGANDS PREPARATION AND GRID GENERATION

The compounds were prepared to be docked using 5 different programs/methods. On the one hand, compounds to be docked using Glide SP [36], Glide XP [36], Glide HTVS [36] and AutoDock Vina [40] were prepared as follows: (1) one 3D conformation was generated per compound and

compounds with unspecified chiralities were discarded by Omega [54], and (2) LigPrep [55] was used to prepare the compound for docking generating all the protonation states for each compound. On the other hand, compounds to be docked using FRED [37] were prepared as follows: (1) the ionization states of the compounds were set by fixpa [56], (2) tautomeric forms were enumerated by tautomer [56], (3) atomic partial charges were assigned by molcharge [56], and (4) one conformation per compound was generated and compounds with unspecified chiralities were discarded by Omega [54].

The selected M-pro structure was the complex with ID Mpro-x12419. We took this protein from our library and it was prepared using different tools depending on the docking program/methodology used.

For Glide SP, XP and HTVS [36], the protein structure was prepared using the Protein Preparation Wizard software [57], where the following modifications relative to default values were applied: (1) N and C termini were capped, (2) original hydrogens were removed to add new hydrogens, (3) water molecules were removed, (4) tautomers and protonation states with neutral pH were generated using Epik, (5) H-bond assignment was optimized using PROPKA, and (5) the structure was minimized with the default force field. Then, the grid generation was done around the binding site of the protein where the compounds were supposed to bind, and the mode of docking was selected: Glide SP (standard precision), Glide XP (extra precision) or Glide HTVS (high-throughput virtual screening). In contrast, for AutoDock Vina [40] dockings we used AutoDockTools [45] in order to prepare the protein structure as follows: (1) all water molecules were removed, and (2) polar hydrogens were added. Then, the grid was defined with the size and coordinated in order to perform the docking. Finally, for FRED [37] dockings the protein was prepared using MakeReceptor [58], that carried out the following procedure: (1) a box was defined according to the centre coordinates and dimensions of the grid previously defined with Glide, (2) shape potential was set, and (3) the inner and outer contours of the receptor was defined with the default options.

3.3.2. DOCKING SETUP

The docking setup in all 5 methodologies used in this study was to generate 20 binding poses per compound (*i.e.,* 111 non-covalent ligands). Thus, we obtained a total of 2220 docked poses for each program.

3.3.3. IDENTIFICATION OF ACCURATE DOCKED POSES: RMSD AND THRESHOLD

The comparison used to assess the docking programs' ability to successfully predict accurate poses was done by comparing the root mean square deviation (RMSD) between the docked poses and the crystalized poses.

After assigning a quantitative comparison measure (RMSD) in each docked pose for each program, we identified the similarity among all results using a visual inspection with Maestro [53]. The lower RMSD value, the more similar the docked pose against the crystalized pose. Thus, we stablished a threshold value of 2,7 Å in order to determine which poses had highest affinity.

Then, this threshold was useful to approve or discard compounds depending on their RMSD value: (1) the compounds with a RMSD value lower than 2,7 Å were considered as true docked poses, and (2) the compounds with a RMSD value higher than 2,7 Å were considered as false docked poses.

3.3.4. ANALYSIS OF DOCKING RESULTS

Finally, the evaluation of the 5 docking programs/methodologies used in this study was done by the creation of statistic figures: (1) histograms and (2) violin plots. These figures were created using R Studio [59] considering the RMSD values for each pose and for each program/methodology. Thanks to these figures, we analysed the docking results and determined which program/methodology was better in order to predict more accurate poses.

4. RESULTS AND DISCUSSION

4.1. Protein-ligand binding-affinity prediction

4.1.1. BEST WEB SERVER PREDICTOR

Nowadays, most computational approaches in the field of drug discovery have been focused on the calculation of protein-ligand binding affinity using different methods of prediction. Nonetheless, it remains a question which is the most accurate and precise program or web server to use in some specific cases, such as the affinity of different drugs targeted on the main protease (M-pro) of SARS-CoV-2. Here we used 4 different web servers as informatic tool predictors since they are easier to manage and set up than a downloadable program.

The protein-ligand binding affinity prediction was performed using the following web servers: PRODIGY [17], CSM-lig [18], Alloscore [19], GraphDelta [20], and DeltaDelta [25]. 40 non-covalent protein-ligand SARS-CoV-2 M-pro complexes obtained from the M-pro FRAGALYSIS database (https://fragalysis.diamond.ac.uk/viewer/react/preview/target/Mpro) were run in these web servers. Figure 5 shows correlation plots between experimental pIC_{50} and binding-affinity predicted by PRODIGY (Fig. 5a), CSM-lig (Fig. 5b), Alloscore (Fig. 5c), and Graphdelta (Fig. 5d). The results obtained using these web servers were comparable. It shows that the correlation is not good, as Pearson correlation coefficient were -0.4945, 0.3394, -0.0624 and -0.0732 in PRODIGY, CSM-lig, Alloscore, and GraphDelta, respectively. The binding affinity is predicted in terms of free energy (ΔG) in PRODIGY results, thus the higher experimental pIC_{50}, the lower ΔG as the negative correlation coefficient obtained in Figure 5a. CSM-lig results are expressed in terms of $-\log10(K_D/K_i)$ so the correlation coefficient is positive as our results in Figure 5b. In contrary, Alloscore uses its own measure unit, Alloscore. The higher Alloscore, the higher pIC_{50}, but our results do not show a positive correlation coefficient (Figure 5c). Finally, results obtained using GraphDelta are expressed in terms of pK_D, and it shows a negative correlation coefficient (Figure 5d) although the highest pK_D value, the greater binding affinity of the ligand for its target.

This study showed that web servers PRODIGY [17], CSM-lig [18], Alloscore [19] and GraphDelta [20] are not able to predict accurately the binding affinity of M-pro with different ligands, as the correlation between experimental pIC_{50} and predicted affinity was low in all cases (see above). These web servers were trained by using datasets that are different than the ones from our case of study (*i.e.,* M-pro complexes with non-covalent inhibitors), as well as they were published in 2016 (CSM-lig [18] and Alloscore [19]) and 2018 (PRODIGY [17]) when further approaches were

not launched. For these reasons, they are not able to predict with the same reliability our study molecules, and we should try other methods instead of using these binding affinity web server predictors.

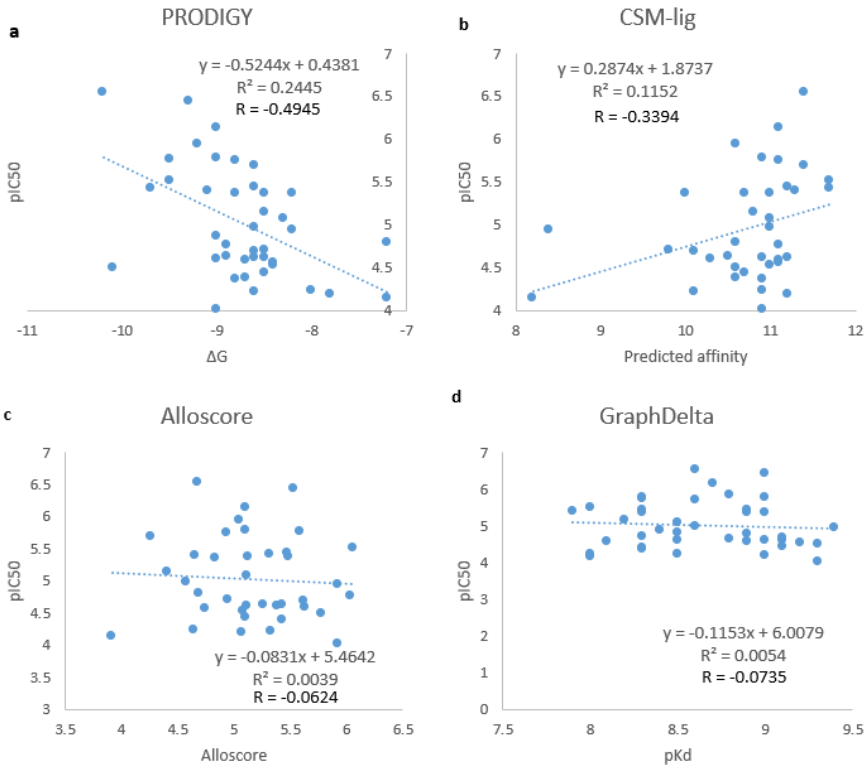


Figure 5. Correlation plots between experimental pIC_{50} and predicted binding affinity using the following web servers: (a) PRODIGY, (b) CSM-lig, (c) Alloscore, and (d) GraphDelta.

On the other hand, DeltaDelta web server [21] needed as an input a set of training ligands and a set of test ligands as it is a conventional neural network that allows us to train it with our study case, M-pro inhibitors. This web server was first run using the initial library of 40 non-covalent ligands, where 10 of them were tested. For this reason, we show the results separately from Figure 5. The neural network was validated using 4 different ligand distributions as a cross-validation method. After the library update dated on 26th February 2021, 111 protein structures were run with these 4 distributions of 40 ligands, putting as an input a set of 30 ligands to the training and a set of 10 ligands to the test. Correlation between experimental pIC_{50} and DeltaDelta predicted pIC_{50} was good in most cases, showing high values with a Pearson correlation coefficient over 0.5 in most cases, even going to a value of 0.94 when the ligands are docked at the M-pro structure with ID Mpro-x11488 (Figure 6). We saw in the previous study that KDeep [28] also performed high Pearson correlation coefficient between experimental pIC_{50} and predicted binding affinity (Macip *et al.* manuscript in revision: *Haste makes waste: a critical review of docking-based virtual screening in drug repurposing for SARS-CoV-2 main protease (M-pro) inhibition*) but here we demonstrate that results obtained using DeltaDelta web server [21]

are even better, because with this methodology we have the possibility to train the neural network with our set of interest, in our study, with non-covalent ligands that have been co-crystallized with M-pro.

DeltaDelta results are promising, as the combination between the 3D structures of the 40 different complexes with the experimental 40 ligands co-crystallized with the M-pro showed a high Pearson correlation coefficient, showed in Figure 6. Thanks to the prediction using different distributions we have validated the neural network, in other words, we have tested that they are not random results, and this neural network works properly. Notice that Figure 4 only shows structures obtained by using distribution 1 and 3. This is because ligands contained in the training set have different characteristics when they bind to the binding site of the protein, depending on the ligand there will be more or less affinity. As ligand distribution between training set and test set has been done randomly, our results indicate that distribution 1 and 3 correlates better than distributions 2 and 4.

Noting that the results showed a good correlation in only one of the 5 web servers used in this study, even with a huge difference in their correlation coefficient, we proceeded using DeltaDelta.

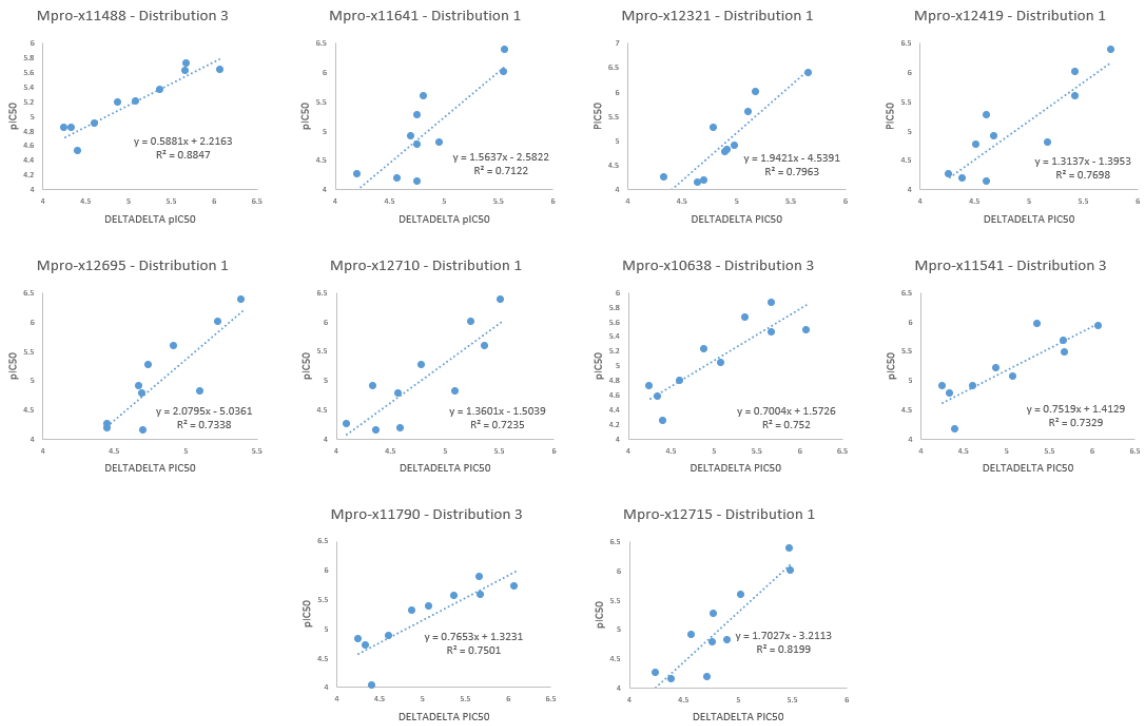


Figure 6. M-pro structures with higher correlation coefficient between experimental pIC_{50} and predicted pIC_{50} by DeltaDelta web server.

4.1.2. BEST M-PRO STRUCTURE

After the correlation between experimental pIC_{50} and DeltaDelta predicted pIC_{50}, structures that correlated better were selected, as they show a correlation coefficient higher than 0.7 ($R^2 > 0,7$) (Figure 6). These structures are the following: (1) M-pro-x11488, (2) M-pro-x11641, (3) M-pro-x12321, (4) M-pro-x12419, (5) M-pro-x12695, (6) M-pro-x12710, (7) M-pro-x10638, (8) M-pro-x11541, (9) M-pro-x11790 and (10) M-pro-x12715.

In order to identify which one of the selected SARS-CoV-2 M-pro structures is the best one to perform the following part of this study (*i.e.,* protein-ligand docking), we analysed which protein complex is able to generate better correlation between predicted and experimental pIC_{50} using the best web server previously identified, DeltaDelta. Table 1 collects the results of these experimental-predicted pIC_{50} correlations, realized using 5 random different sets of ligand distributions between training and test, obtained from the updated library that has a total of 111 ligands. We have also run 5 different distributions for a cross-validation purpose.

Results in Table 1 show that there are different structures with high Person correlation coefficients, as correlation can vary depending on the ligand distribution. Thus, we selected M-pro-x12419 structure for the following protein-ligand docking, as it shows high correlation in most cases.

Table 1. Correlations between experimental pIC_{50} and predicted pIC_{50} by DeltaDelta web server of the 10 best M-pro complexes and performed 5 times using different ligand distributions.

	Distribution 1		Distribution 2		Distribution 3		Distribution 4		Distribution 5	
Dataset	**R**	**R^2**	**R**	**R^2**	**R**	**R^2**	**R**	**R^2**	**R**	**R^2**
M-pro-x10638	0.4812	0.2315	0.4001	0.16	0.6216	0.3864	0.6378	0.4068	0.6116	0.374
M-pro-x11488	0.4051	0.1641	0.3174	0.1007	0.6615	0.4376	0.6116	0.374	0.569	0.3238
M-pro-x11541	0.4948	0.2448	0.4973	0.2473	0.5146	0.2649	0.6906	0.4769	0.5128	0.263
M-pro-x11641	0.5368	0.2882	0.2269	0.0515	0.6032	0.3638	0.5573	0.3106	0.6228	0.3879
M-pro-x11790	0.4815	0.2318	0.4475	0.2003	0.588	0.3457	0.5427	0.2945	0.4878	0.2379
M-pro-x12321	0.4033	0.1626	0.4084	0.1668	0.5701	0.325	0.5804	0.3369	0.4863	0.2364
M-pro-x12419	0.4018	0.1614	0.4985	0.2485	0.6826	0.4659	0.6086	0.3704	0.6831	0.4667
M-pro-x12695	0.3461	0.1198	0.3129	0.0979	0.5699	0.3248	0.4923	0.2424	0.6277	0.394
M-pro-x12710	0.5036	0.2536	0.5799	0.3363	0.4994	0.2494	0.5643	0.3184	0.6184	0.3824
M-pro-x12715	0.3511	0.1232	0.458	0.2098	0.4813	0.2316	0.5279	0.2786	0.571	0.326

4.2. Protein-ligand docking

Most studies use protein-ligand docking methodology in order to predict the structure of a compound on a specific target, the protein. Here we applied this computational methodology in the case of the SARS-CoV-2 M-pro and different putative drugs obtained from our library of individual structures of M-pro and their non-covalent inhibitors, using 5 different docking programs/methodologies: FRED [37], AutoDock Vina [40], Glide SP [36], Glide XP [36] and, Glide HTVS [36].

Since one purpose of this study was to find which docking program is capable of better predict the structure of these complexes, this part of the study demonstrates that Glide SP perform better than the other docking programs, although there are some points that are necessary to review such as the ranking order and the number of poses that the program needs to obtain an accurate result. Docking results of the 5 different programs/protocols used in this study (FRED, Glide SP, Glide XP, Glide HTVS and AutoDock Vina) are graphically expressed as follows.

Figure 7 shows 5 histograms, each one for a different docking program, where the number of true and false docking poses are represented on y-axis for the 20 docking poses on x-axis. We considered that a pose is *true* if its RMSD value against its experimental pose is under the established threshold of 2.7 Å. At this point, we can observe which of the docking programs is capable to predict a major number of true docked poses. Figure 7 clearly shows that FRED, Glide XP, Glide HTVS and AutoDock Vina place the major number of true docked positions at first generated poses, and this number decreases as programs predict new docked poses. FRED results show a strong tendency to decrease the number of true poses as docked poses are predicted whereas in AutoDock Vina results this tendency is downward. Results obtained by Glide XP and Glide HTVS also show that number of false docked poses decrease as the true ones, since programs dismiss the possibility of generating new docked poses. Otherwise, there is no decreasing tendency at Glide SP results, as well as we can observe that in all the 20 docked predicted poses the number of true poses is higher than in the other docking programs. Thus, Glide SP is the program capable of guess more true docked poses, but all of these poses are not ranked properly. For instance, we cannot affirm that pose 19 is less valid than pose 1. Moreover, all true numbers in all cases are very low, which stands below 25% of the total. Therefore, although we distinguished Glide SP as the better docking predictor, we cannot affirm that this program is totally successful.

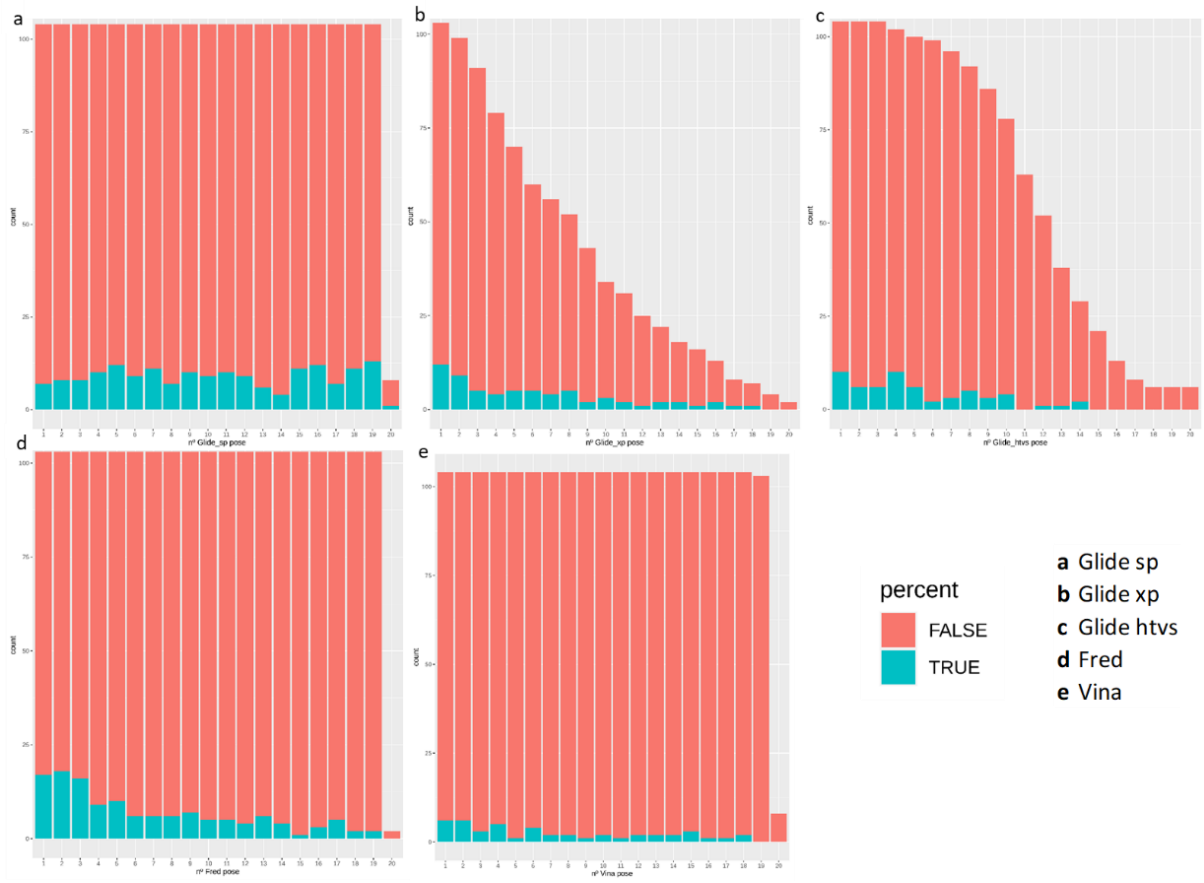


Figure 7. Histograms showing the number of true and false docking poses for (a) Glide SP, (b) Glide XP, (c) Glide HTVS, (d) FRED and (e) AutoDock Vina.

Docking programs used in this study were reviewed through the help of violin plots shown in Figures 8 and 9. In both Figures, the RMSD value relative to the experimental pose is located on the y-axis whereas the results for the 5 different programs/methodologies assayed are located on the x-axis. Thus, Figure 8a shows the RMSD value of the first pose (the one that is selected as the best one by the scoring function of each program) relative to the experimental pose. In contrast, Figure 8b shows the RMSD distribution of the docked pose with lower RMSD value relative to the experimental one (irrespective of how the corresponding scoring function has ranked it in the first twenty poses). Furthermore, Figure 9 shows 5 violin plots, one for each program, that allows to analyse the RMSD distribution of the 20 docked poses per ligand. Thus, for each program, the first violin plot shows how the poses scored first by the scoring function resembles the experimental pose, the second violin plot shows how the poses scored second by the scoring function resembles the experimental pose, and so on for the remaining eighteen violin plots.

Violin plots are very interesting as they allow us to analyse data results similarly to a box plot with the addition of have a visual density map in their sides. Figure 8a shows the first generated docked pose by all 5 docking programs/methodologies used in this study. Thanks to this plot, we demonstrate that with only one docked pose generation is not enough to predict an accurate docked pose that best approximates to their relative experimental crystalized pose. For this reason, it is essential to generate as many poses as necessary in order to find the one that best approximate to the crystalized one. As it stands, in Figure 8a the average in all programs is located higher that the established RMSD threshold and the major density of data is located around them, thus these values are set too high to be considered similar to the experimental pose. It is true that there exists a major density band around the threshold, but with only the first docked pose representation it is not enough to affirm that programs could be able to generate better poses, and for this reason we evaluate all 20 docked poses later (Figure 9).

On the other hand, Figure 8b shows a comparison among all better docked poses of the five different programs used in this study. Note that this situation is not normal, but it allows us to see how good a program is on generating poses, as it shows that they really can approximate to the crystalized pose. Glide SP has the lower average value, thus it is clearly capable of predict better docked poses that are similar to the experimental ones as they show lower RMSD values. Average point is higher than the threshold at results obtained by FRED, Glide HTVS, Glide XP, and almost AutoDock Vina. Moreover, if we analyse data distribution we can see a major density under the threshold in Glide SP, stating that all these docked poses are considered as true, not as other docking programs. AutoDock Vina also has a lower average value, but their distribution is located higher than Glide SP as we can see in the density map, and their standard deviation is shorter. FRED, Glide HTVS and Glide XP have their data uniformly distributed, with longer violin plots that present less pronounced density at a particular point and they all have a huge standard deviation.

We show one violin plot for each docking program in Figure 9, where 20 docked poses are appreciated in more detail. Here, we demonstrate that there are no better docked poses according to the scored order, as we sensed in the previous figures. Average numbers are similar in all docking programs and almost remain unchanged in all 20 docked poses, as well as there is a major density around this average point. Moreover, a density zone is located around the 2,7 Å RMSD threshold in all 20 poses predicted by Glide SP (except pose 20 that have lost some compound poses). There exists a similar tendency at poses generated by Glide XP, but this density zone around the threshold is maintained until the last docked poses. The same tendency is observed at poses predicted by Glide HTVS but less pronounced, here the density around

threshold is lost at last docked poses and all values are lower. According to FRED, the density zone is higher around the threshold at first poses and it decreases as the program generate new predicted poses. Finally, AutoDock Vina shows that from pose 11 it generates docked poses with RMSD values far from our interest. Thus, it generates a lot of variability from this pose, and it loses its reliability.

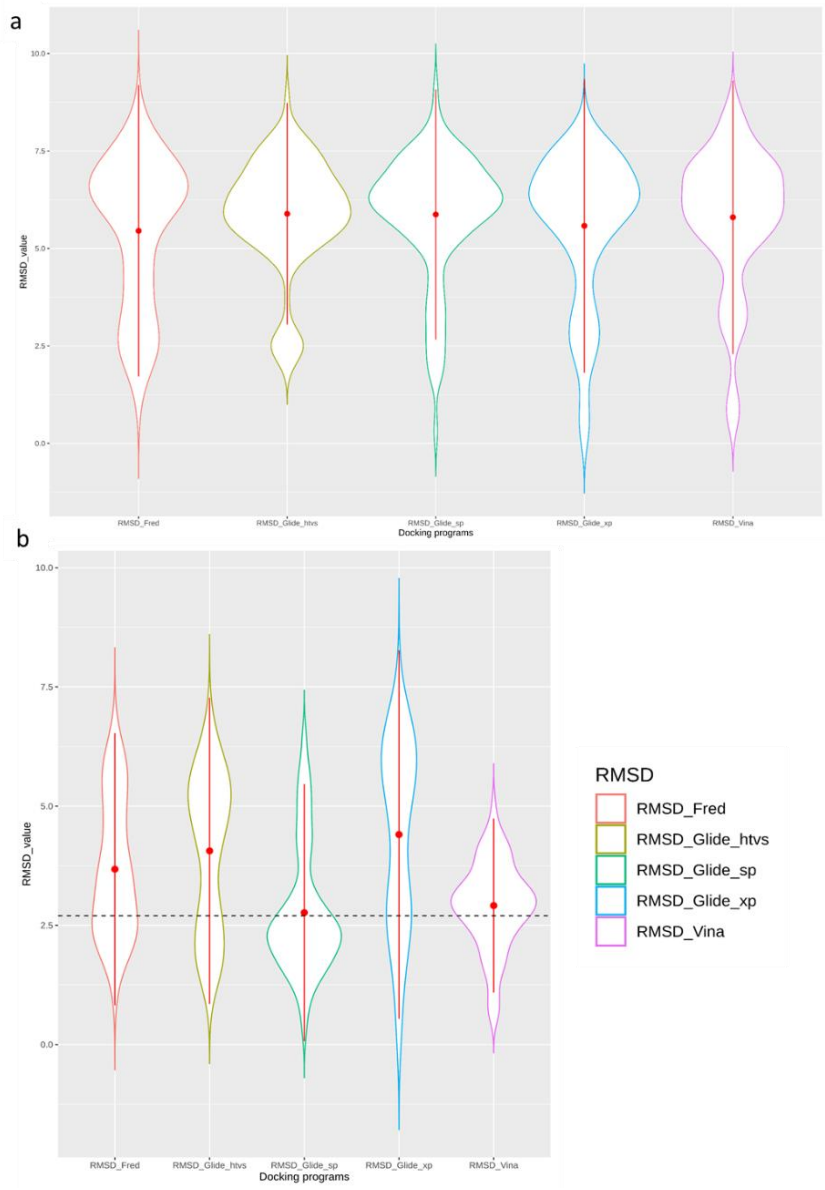


Figure 8. Violin plots of the 5 docking programs used in this study showing the RMSD value distribution at (a) first docked pose (b) best docked poses.

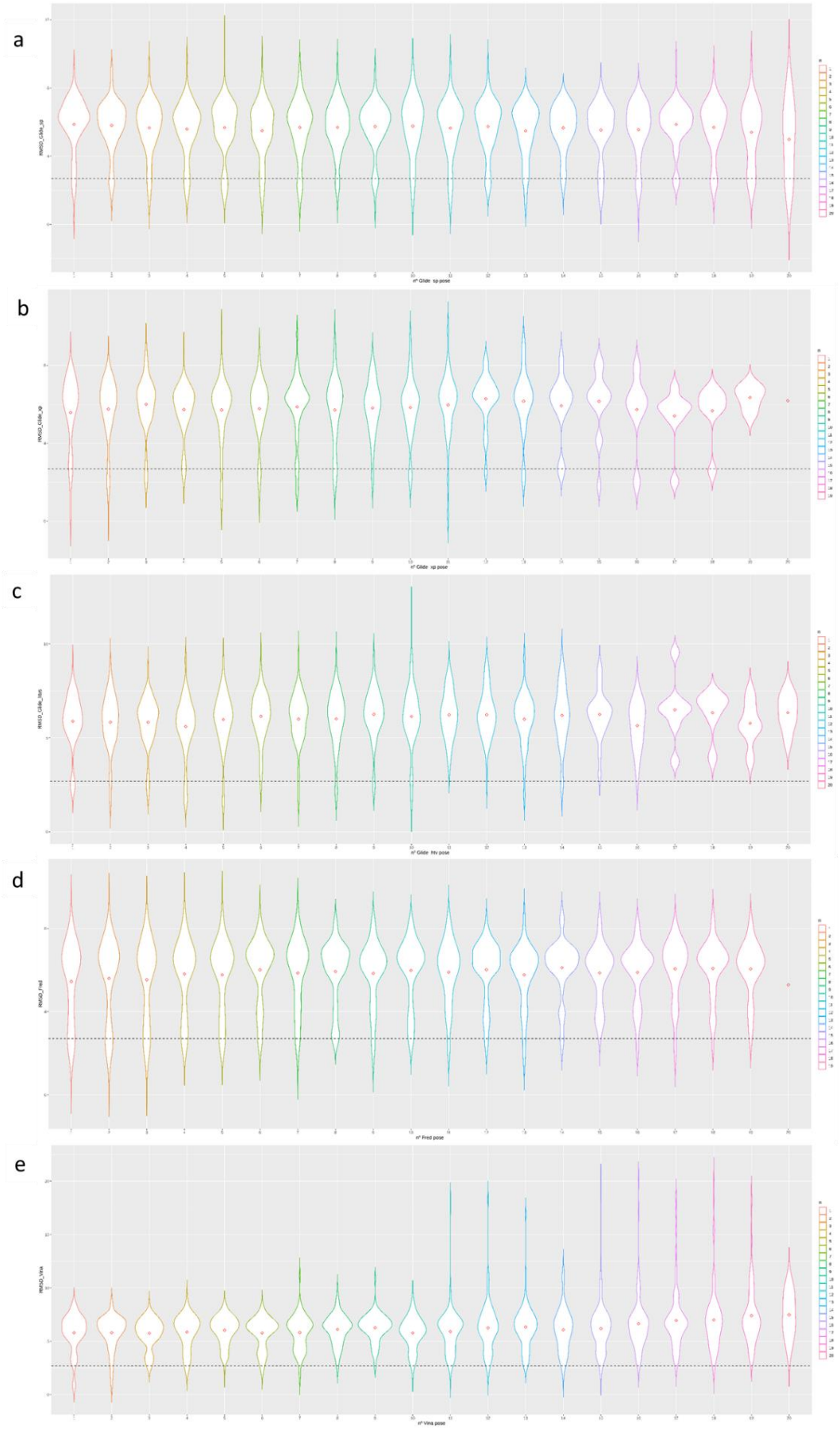


Figure 9. Violin plots depending on the 20 predicted docked poses by (a) Glide SP, (b) Glide XP, (c) Glide HTVS, (d) FRED and (e) AutoDock Vina.

To sum up, Glide SP is the best docking predictor program used in this study for all reasons mentioned above, but still has some shortcoming to improve as a result of their lack of accuracy and success. Non-existence of a computational methodology capable to correctly predict experimental poses in M-pro non-covalent inhibitors set a challenging situation to identify potential drugs against this protein essential for SARS-CoV-2 replication. Therefore, we demonstrated that this methodology is not good at predicting poses similar to crystalized structures in our study case of M-pro, and it is necessary to change and review this methodology that have been used in a lot of previous studies.

We hypothesized if the combination of prediction programs would be useful to predict the bioactivity of potential inhibitors of SARS-CoV-2 M-pro. As mentioned, although we have found a methodology that properly predict bioactivity between M-pro and non-covalent inhibitors, we cannot affirm that the docking methodology works properly, therefore, the combination of both programs will not be useful to find potential inhibitors of M-pro.

6. CONCLUSIONS

M-pro is an attractive target because it has a key role in the replication of SARS-CoV-2 virus. Thanks to computational approaches, potential inhibitors of this protease can be discovered and thus, contribute to find a new treatment for infected people with the COVID-19 disease.

Our study has served to evaluate the performance on the SARS-CoV-2 M-pro of different commonly used tools for two of the most popular prediction methodologies used in the drug discovery of non-covalent inhibitors: prediction of protein-ligand binding affinity and pose prediction (a.k.a. docking).

We demonstrate that the web server capable that best predicts the binding affinity between M-pro and their non-covalent ligands is DeltaDelta [21], as it allows to train a neural network with our set of interest (*i.e.,* with non-covalent M-pro inhibitors that have been co-crystallized with the M-pro from SARS-CoV-2).

The evaluation of different protein-ligand docking programs/methodologies showed that the best program capable to most accurately predict the poses for non-covalent M-pro inhibitors is Glide SP [36]. Here, we demonstrate too that all these methodologies might not have been as affective as they are expected in their predictions. Docking has been used in a lot of drug discovery studies but however, our results show that this methodology is not really successful in predicting poses for M-pro non-covalent inhibitors. Therefore, it should be necessary to look for other protein-ligand docking programs/methodologies that perform better with potential non-covalent SARS-CoV-2 inhibitors than the ones used here.

In conclusion, the combination of both prediction programs cannot be useful to predict the bioactivity of potential non-covalent M-pro inhibitors as the step of protein-ligand docking is not capable to predict accurate poses.

7. REFERENCES

1. Zhang L, Lin D, Sun X, et al. Crystal structure of SARS-CoV-2 main protease provides a basis for design of improved a-ketoamide inhibitors. Science (80-.). 2020; 368:409–412

2. Dai W, Zhang B, Jiang XM, et al. Structure-based design of antiviral drug candidates targeting the SARS-CoV-2 main protease. Science (80-.). 2020; 368:1331–1335

3. Ghosh R, Chakraborty A, Biswas A, et al. Evaluation of green tea polyphenols as novel corona virus (SARS CoV-2) main protease (Mpro) inhibitors–an in silico docking and molecular dynamics simulation study. J. Biomol. Struct. Dyn. 2020; 1

4. Vallianou NG, Tsilingiris D, Christodoulatos GS, et al. Anti-viral treatment for SARS-CoV-2 infection: A race against time amidst the ongoing pandemic. Metab. Open 2021; 10:100096

5. Gimeno A, Mestres-Truyol J, Ojeda-Montes MJ, et al. Prediction of novel inhibitors of the main protease (M-pro) of SARS-CoV-2 through consensus docking and drug reposition. Int. J. Mol. Sci. 2020; 21:

6. Ullrich S, Nitsche C. The SARS-CoV-2 main protease as drug target. Bioorganic Med. Chem. Lett. 2020; 30:127377

7. Khailany RA, Safdar M, Ozaslan M. Genomic characterization of a novel SARS-CoV-2. Gene Reports 2020; 19:100682

8. Banerjee R, Perera L, Tillekeratne LMV. Potential SARS-CoV-2 main protease inhibitors. Drug Discov. Today 2021; 26:804–816

9. Jin Z, Du X, Xu Y, et al. Structure of Mpro from SARS-CoV-2 and discovery of its inhibitors. Nature 2020; 582:289–293

10. . PostEra | COVID-19.

11. Achdout H, Aimon A, Bar-David E, et al. COVID moonshot: Open science discovery of SARS-CoV-2 main protease inhibitors by combining crowdsourcing, high-throughput experiments, computational simulations, and machine learning. bioRxiv 2020; 2020.10.29.339317

12. . Mpro: Fragalysis.

13. D'Souza S, Prema K V., Balaji S. Machine learning models for drug–target interactions: current knowledge and future directions. Drug Discov. Today 2020; 25:748–756

14. Rezaei MA, Li Y, Wu DO, et al. Deep Learning in Drug Design: Protein-Ligand Binding Affinity Prediction. IEEE/ACM Trans. Comput. Biol. Bioinforma. 2020;

15. Ahmed A, Mam B, Sowdhamini R. DEELIG: A Deep Learning-based approach to predict protein-ligand binding affinity.

16. Wan S, Bhati AP, Zasada SJ, et al. Rapid, accurate, precise and reproducible ligand–protein binding free energy prediction. Interface Focus 2020; 10:20200007

17. Vangone A, Schaarschmidt J, Koukos P, et al. Large-scale prediction of binding affinity in protein-small ligand complexes: The PRODIGY-LIG web server. Bioinformatics 2019; 35:1585–1587

18. Pires DEV, Ascher DB. CSM-lig: a web server for assessing and comparing protein-small molecule affinities. Nucleic Acids Res. 2016; 44:W557–W561

19. Li S, Shen Q, Su M, et al. Alloscore: A method for predicting allosteric ligand-protein interactions. Bioinformatics 2016; 32:1574–1576

20. Karlov DS, Sosnin S, Fedorov M V, et al. graphDelta: MPNN Scoring Function for the Affinity Prediction of Protein−Ligand Complexes. Cite This ACS Omega 2020; 5:5159

21. . DeltaDelta: CNN-based congeneric series delta delta free energy prediction [WEB APP].

22. Basse MJ, Betzi S, Morelli X, et al. 2P2Idb v2: Update of a structural database dedicated to orthosteric modulation of protein-protein interactions. Database 2016; 2016:

23. van Zundert GCP, Bonvin AMJJ. Modeling protein–protein complexes using the HADDOCK webserver "modeling protein complexes with HADDOCK". Methods Mol. Biol. 2014; 1137:163–179

24. Shen Q, Wang G, Li S, et al. ASD v3.0: Unraveling Allosteric regulation with structural mechanisms and biological networks. Nucleic Acids Res. 2016; 44:D527–D535

25. Jiménez-Luna J, Pérez-Benito L, Martínez-Rosell G, et al. DeltaDelta neural networks for lead optimization of small molecule potency. Chem. Sci. 2019; 10:10911–10918

26. Jones D, Kim H, Zhang X, et al. Improved Protein−Ligand Binding Affinity Prediction with Structure-Based Deep Fusion Inference. 2021;

27. Jiménez J, Škalič M, Martínez-Rosell G, et al. KDEEP: Protein-Ligand Absolute Binding Affinity Prediction via 3D-Convolutional Neural Networks. J. Chem. Inf. Model. 2018; 58:287–296

28. . Kdeep: a protein-ligand binding affinity predictor [WEB APP].

29. Öztürk H, Özgür A, Ozkirimli E. DeepDTA: Deep drug-target binding affinity prediction. Bioinformatics 2018; 34:i821–i829

30. Kwon Y, Shin WH, Ko J, et al. AK-score: Accurate protein-ligand binding affinity prediction using an ensemble of 3D-convolutional neural networks. Int. J. Mol. Sci. 2020; 21:1–16

31. Zhang H, Liao L, Saravanan KM, et al. DeepBindRG: A deep learning based method for estimating effective protein-ligand affinity. PeerJ 2019; 2019:e7362

32. Kim S, Oshima H, Zhang H, et al. CHARMM-GUI Free Energy Calculator for Absolute and Relative Ligand Solvation and Binding Free Energy Simulations. J. Chem. Theory Comput. 2020; 16:7207–7218

33. Oshima H, Re S, Sugita Y. Prediction of Protein−Ligand Binding Pose and Affinity Using the gREST+FEP Method. 2020;

34. Holderbach S, Adam L, Jayaram B, et al. RASPD+: Fast Protein-Ligand Binding Free Energy Prediction Using Simplified Physicochemical Features. Front. Mol. Biosci. 2020; 7:601065

35. Cournia Z, Allen BK, Beuming T, et al. Rigorous free energy simulations in virtual screening. J. Chem. Inf. Model. 2020; 60:4153–4169

36. . Glide | Schrödinger.

37. . FRED — Applications, v2020.2.2.

38. Li J, Fu A, Zhang L. An Overview of Scoring Functions Used for Protein–Ligand Interactions in Molecular Docking. Interdiscip. Sci. Comput. Life Sci. 2019; 11:320–328

39. Lohning AE, Levonis SM, Williams-Noonan B, et al. A Practical Guide to Molecular Docking and Homology Modelling for Medicinal Chemists. Curr. Top. Med. Chem. 2017; 17:

40. Trott O, Olson AJ. AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. J. Comput. Chem. 2009; 31:NA-NA

41. . Schrödinger | Schrödinger is the scientific leader in developing state-of-the-art chemical simulation software for use in pharmaceutical, biotechnology, and materials research.

42. . Docking and Scoring | Schrödinger.

43. McGann M. FRED pose prediction and virtual screening accuracy. J. Chem. Inf. Model. 2011; 51:578–596

44. Moustakas DT, Lang PT, Pegg S, et al. Development and validation of a modular, extensible docking program: DOCK 5. J. Comput. Aided. Mol. Des. 2006; 20:601–619

45. Morris GM, Ruth H, Lindstrom W, et al. Software news and updates AutoDock4 and AutoDockTools4:

Automated docking with selective receptor flexibility. J. Comput. Chem. 2009; 30:2785–2791

46. Jones G, Willett P, Glen RC, et al. Development and validation of a genetic algorithm for flexible docking. J. Mol. Biol. 1997; 267:727–748

47. Rarey M, Kramer B, Lengauer T, et al. A fast flexible docking method using an incremental construction algorithm. J. Mol. Biol. 1996; 261:470–489

48. Jain AN. Surflex-Dock 2.1: Robust performance from ligand energetic modeling, ring flexibility, and knowledge-based search. J. Comput. Aided. Mol. Des. 2007; 21:281–306

49. Grosdidier A, Zoete V, Michielin O. SwissDock, a protein-small molecule docking web service based on EADock DSS. Nucleic Acids Res. 2011; 39:

50. Wang R, Lai L, Wang S. Further development and validation of empirical scoring functions for structure-based binding affinity prediction. J. Comput. Aided. Mol. Des. 2002; 16:11–26

51. Friesner RA, Banks JL, Murphy RB, et al. Glide: A New Approach for Rapid, Accurate Docking and Scoring. 1. Method and Assessment of Docking Accuracy. J. Med. Chem. 2004; 47:1739–1749

52. Korb O, Stützle T, Exner TE. Empirical scoring functions for advanced Protein-Ligand docking with PLANTS. J. Chem. Inf. Model. 2009; 49:84–96

53. . Maestro | Schrödinger.

54. . OMEGA | Multi-Conformer Structure Databases | Database Preparation | OpenEye Scientific.

55. . LigPrep | Schrödinger.

56. . QUACPAC | Tautomer / Protomer Enumeration & Charge Assignment | Database Preparation.

57. . Protein Preparation Wizard | Schrödinger.

58. . OEDocking Software | Molecular Docking Tools | Fred Docking.

59. . RStudio | Open source & professional software for data science teams - RStudio.

SELF-EVALUATION

My Internship in the Cheminformatics and Nutrition research group has enabled myself to discover how is to be a member of a leader research group. I can feel satisfied to have this amazing experience as it has prepared me for my professional growth and future and has fulfilled my insatiable desire for learning new things.

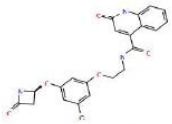
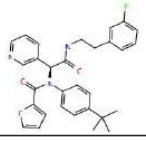
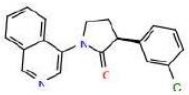
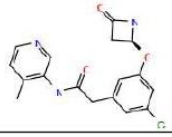
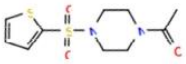
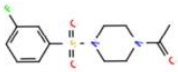
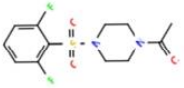
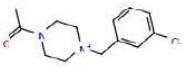
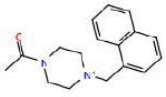
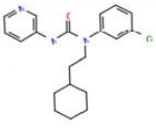
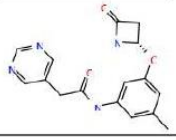
Literature research has allowed me to learn and understand all the knowledge related to this study, such as the properties of SARS-CoV-2 virus and its main protease (M-pro), the programs used to predict the binding affinity and the programs used to predict docked poses. Furthermore, I have learned experimentally how to use some of these programs/web servers and I have followed all the procedure that has been taken collaboratively with all the members of this research group. Moreover, I am delighted to write the results in a scientific article as a personal new challenge. Overall, I have discovered deeply the field of bioinformatics, one of the parts that I consider that will be the future of biotechnology and, as well as that, I am fully realized to have learned how to work independently and to organise myself as I did not go to a laboratory.

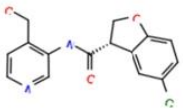
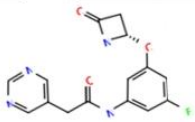
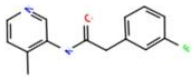
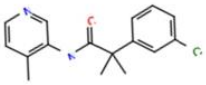
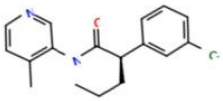
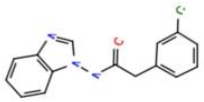
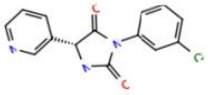
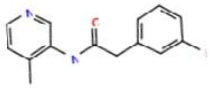
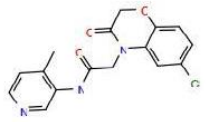
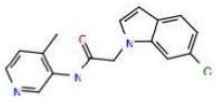
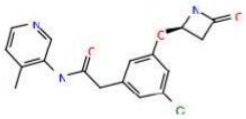
My weekly meetings have allowed me to share my work and to solve doubts and questions thanks to the learning received from the members of the research group. Being part of a team has awakened my interest in engaging in even more to the research world. I am confident that I have been a valuable asset to the team.

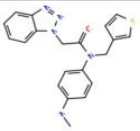
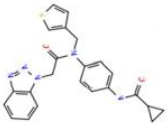
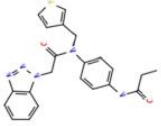
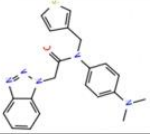
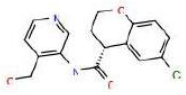
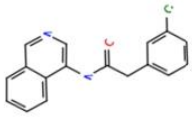
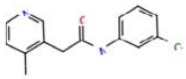
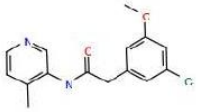
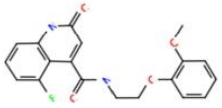
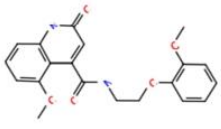
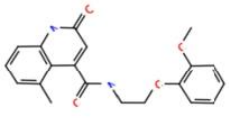
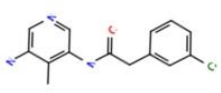
Finally, I feel gratified to take part in a study related to this global health emergency situation caused by COVID-19 disease. I have always tough that one of the most important parts of scientific research is to provide an appropriate solution to the current challenges. For this reason, I hope that this project can contribute to this matter.

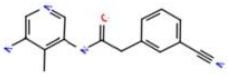
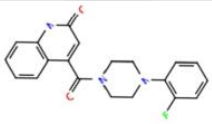
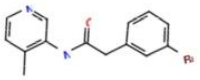
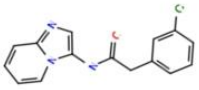
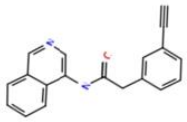
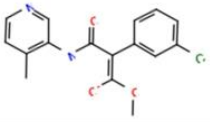
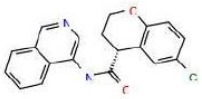
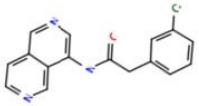
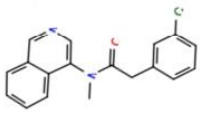
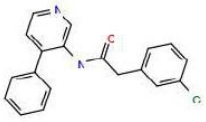
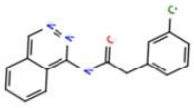
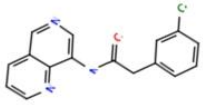
ANNEX I: Libraries

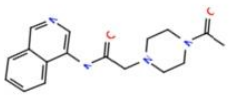
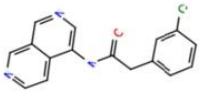
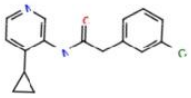
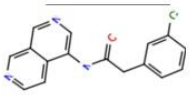
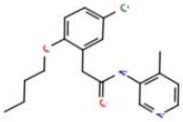
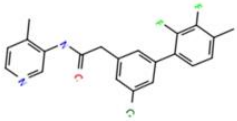
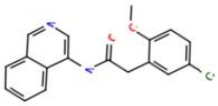
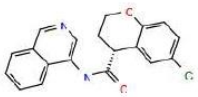
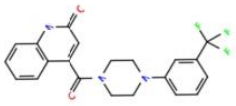
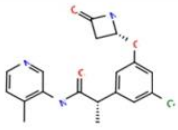
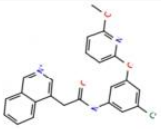
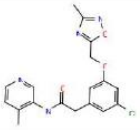
Table 2. Libraries of non-covalent compounds used in this study. The table shows the 111 compounds contained in the updated library (2021-02-26) and the 40 compounds contained in the old library (2020-12-02) are marked.

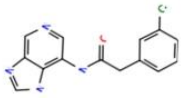
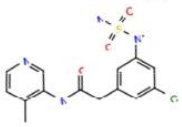
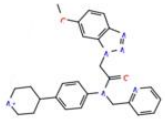
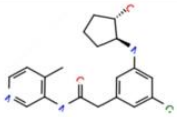
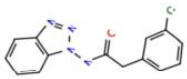
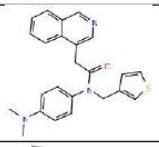
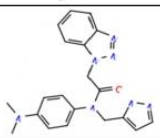
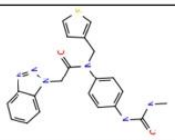
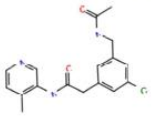
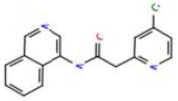
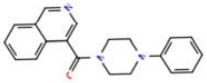
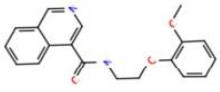
Compound ID	Dataset	2D structure	Rapid Fire inhibition at 50 uM	Rapid Fire avg IC50 (uM)	Rapid Fire pIC50	Fluores-cence inhibition at 50 uM	Fluores-cence avg IC50 (uM)	Fluores-cence pIC50
ERI-UCB-ce40166b-17	Mpro_Nterm-x0029			1.18014	5.92807		1.10393	5.95706
MAT-POS-f2460aef-1	Mpro_Nterm-x0050		100	0.41439	6.38259		0.10029	6.99876
PET-UNK-c9c1e0d8-4	Mpro_Nterm-x0066			0.86571	6.06263		1.26810	5.89685
TRY-UNI-2eddb1ff-7	Mpro_Nterm-x0077		87.19132	0.94075	6.02653	97.00562	3.60986	5.44251
LON-WEI-8f408cad-3	Mpro-x0689		97.79	0.55654	6.25450	99.55992	2.16183	5.66518
LON-WEI-8f408cad-4	Mpro-x0691		96.84	2.02356	5.69388	103.54665	4.55049	5.34194
LON-WEI-8f408cad-2	Mpro-x0755		99.205	2.62321	5.58117	103.96275	3.28451	5.48353
MED-COV-4280ac29-15	Mpro-x0770		100.4237	0.54545	6.26325	108.2425	3.13741	5.50343
MED-COV-4280ac29-31	Mpro-x0830		100.4237	0.90035	6.04559	98.5734	5.76179	5.23944
JOR-UNI-2fc98d0b-12*	Mpro-x10236		78.9447	3.30036	5.48144	88.791675	3.05667	5.51475
RAL-MED-2de63afb-2	Mpro-x10322		44.9512			72.72272	34.31991	4.46445

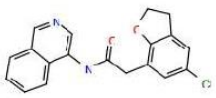
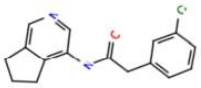
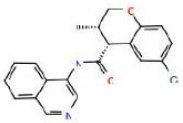
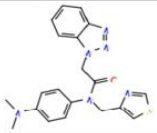
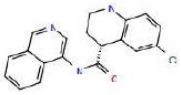
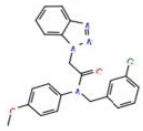
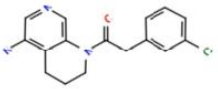
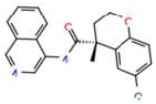
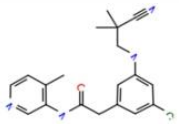
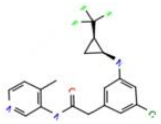
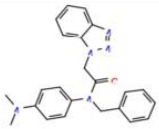
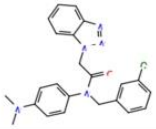
JAG-UCB-a3ef7265-3*	Mpro-x10338		58.62226	18.38796	4.73547	73.050635	42.35759	4.37307
RAL-MED-2de63afb-1	Mpro-x10371		26.44020			71.7735	88.34228	4.05383
RAL-MED-2de63afb-14	Mpro-x10387		45.74945			87.837815	24.94667	4.60299
ALP-POS-95b75b4d-1	Mpro-x10417		40.85454			65.25146	95.83308	4.01848
JAN-GHE-83b26c96-22*	Mpro-x10422		66.91404	24.52281	4.61043	67.95416	96.93832	4.01350
JAN-GHE-83b26c96-8*	Mpro-x10423		78.31820	5.47937	5.26127	90.64927	17.31035	4.76169
JAN-GHE-5a013bed-2*	Mpro-x10466		45.27875	4.26520	5.37006	87.70723	8.30347	5.08074
JAN-GHE-83b26c96-18	Mpro-x10535		69.55622	57.84842	4.23771	33.925115		
JAN-GHE-83b26c96-13	Mpro-x10565		64.32451	69.09189	4.16057	39.673875		
BAR-COM-0f94fc3d-48*	Mpro-x10638		67.195	38.92729	4.40975	49.274905	59.93313	4.22233
BAR-COM-0f94fc3d-41	Mpro-x10679		77.42	19.96818	4.69966	27.01315		
TRY-UNI-2eddb1ff-7*	Mpro-x10789		87.191316	0.94075	6.02653	97.00562	3.60986	5.44251

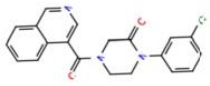
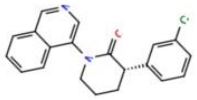
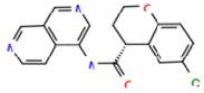
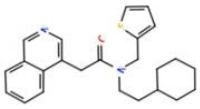
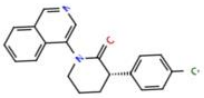
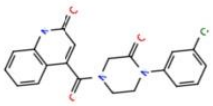
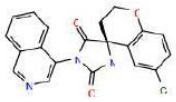
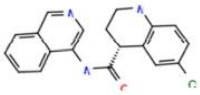
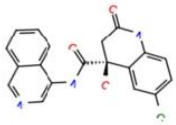
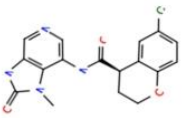
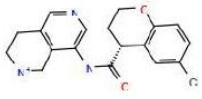
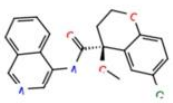
ALP-POS-c59291d4-4*	Mpro-x10820		70.49220	18.67688	4.72870	82.620135	4.32893	5.36362
ALP-POS-c59291d4-3*	Mpro-x10870		62.23556	28.35325	4.54740	70.195875	3.99496	5.39849
ALP-POS-c59291d4-2*	Mpro-x10871		78.12914	12.56226	4.90093	89.29024	1.62764	5.78844
ALP-POS-d2866bdf-1*	Mpro-x10876		90.28649	5.36883	5.27012	90.45516	1.79372	5.74624
LON-WEI-0a73fcb8-7*	Mpro-x10942		58.93036	66.35081	4.17815	84.22906	13.67199	4.86417
ADA-UCB-6c2cb422-1*	Mpro-x10959		84.91022	1.05672	5.97604	84.803335	0.72644	6.13880
EDJ-MED-49816e9b-1*	Mpro-x11011		88.60642	26.36213	4.57902	67.38706	29.85959	4.52492
MAT-POS-c9973a83-1*	Mpro-x11271			43.60280	4.36049	48.298705	36.78228	4.43436
EDJ-MED-6af13d92-2	Mpro-x11276			1.82765	5.73811	12.47232		
EDJ-MED-6af13d92-3*	Mpro-x11294			2.07679	5.68261	90.90139	2.02892	5.69274
EDJ-MED-6af13d92-1*	Mpro-x11313		89.65191	2.60150	5.58478	96.436315	10.62102	4.97383
TRY-UNI-714a760b-3*	Mpro-x11317			8.20170	5.08610		4.22234	5.37445

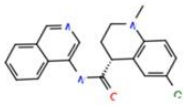
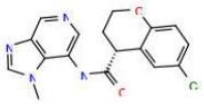
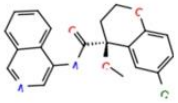
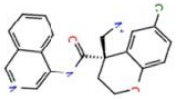
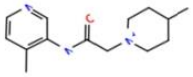
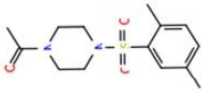
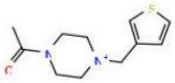
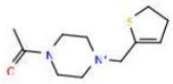
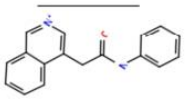
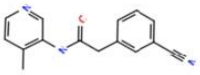
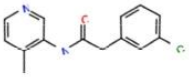
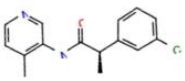
TRY-UNI-714a760b-19	Mpro-x11318			32.22106	4.49186		16.06556	4.79410
BEN-DND-7e92b6ca-16	Mpro-x11366		46.37896			75.578355	23.99960	4.61980
JAN-GHE-83b26c96-12*	Mpro-x11368		98.33862	15.94298	4.79743	77.679955	20.59394	4.68626
PET-UNK-8df914d1-2	Mpro-x11454						8.03395	5.09507
PET-UNK-c9c1e0d8-2	Mpro-x11458						9.38335	5.02764
MAT-POS-1e5f28a7-1*	Mpro-x11488			17.09140	4.76722		23.71230	4.62503
VLA-UCB-1dbca3b4-15*	Mpro-x11498			0.42396	6.37268		0.35948	6.44432
MAT-POS-f7918075-5*	Mpro-x11499			2.13050	5.67152		4.27326	5.36924
MAT-POS-bb423b95-7	Mpro-x11501			48.19355	4.31701		34.36825	4.46384
ALP-POS-0c2c77e1-1	Mpro-x11507			25.13378	4.59974		40.89174	4.38836
EDJ-MED-50fe53e8-1	Mpro-x11508						21.65137	4.66451
MAT-POS-bb423b95-2	Mpro-x11530			6.21925	5.20626		7.01875	5.15374

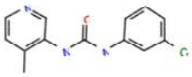
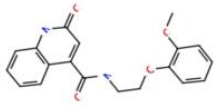
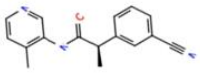
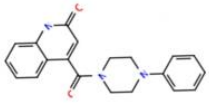
ALP-POS-6747fa38-1*	Mpro-x11541		42.445	72.54815	4.13937	71.68810	4.14455
MAT-POS-bb423b95-1	Mpro-x11542			3.70893	5.43075	7.94166	5.10009
ALP-POS-95b75b4d-5	Mpro-x11543			7.39797	5.13089	7.17364	5.14426
MIC-UNK-08cd9c58-1	Mpro-x11548			57.92701	4.23712	67.11236	4.17320
EDG-MED-0da5ad92-7*	Mpro-x11562			15.68583	4.80449	24.78240	4.60586
MAT-POS-f42f3716-6*	Mpro-x11564			56.18048	4.25041	31.91952	4.49594
MAT-POS-3b92565d-1	Mpro-x11609		79.31	1.34684	5.87068	1.20236	5.91997
MAT-POS-b3e365b9-1	Mpro-x11612		52.82	0.24552	6.60992	0.18959	6.72218
ERI-UCB-a0b0dbcb-4	Mpro-x11616		70.965	5.64647	5.24822	3.79972	5.42025
TRY-UNI-2eddb1ff-8	Mpro-x11641			88.09233	4.05506	77.13484	4.11275
MAT-POS-6344a35d-1*	Mpro-x11642			1.00390	5.99831	3.78369	5.42208
CHO-MSK-6e55470f-5*	Mpro-x11723			24.25844	4.61514	23.92453	4.62116

MAT-POS-f7918075-8	Mpro-x11742		79.45	7.96693	5.09871	13.26088	4.87743
WIL-MOD-03b86a88-6*	Mpro-x11743		87.94	14.47181	4.83948	11.25659	4.94859
ALP-POS-c0c213c9-1	Mpro-x11757		94.805	32.18979	4.49228	38.77391	4.41146
MAT-POS-044491d2-1*	Mpro-x11764		61.545	38.85593	4.41054	19.70250	4.70548
MAT-POS-bfb445d4-2	Mpro-x11789			31.04303	4.50804	5.80242	5.23639
PET-UNK-1901c25b-1*	Mpro-x11790			0.83421	6.07872	0.28463	6.54571
PET-UNK-7374c256-2*	Mpro-x11797			90.75757	4.04212	15.93610	4.79762
ROB-IMP-e811baff-1	Mpro-x11798			15.32726	4.81454	1.42043	5.84758
RAL-THA-6b94ceba-5*	Mpro-x11801			55.52098	4.25554	41.33004	4.38373
PET-UNK-3c72d439-1	Mpro-x11810			11.38069	4.94383	3.39727	5.46887
MAR-UCB-f313ec4d-6	Mpro-x11812			7.25370	5.13944	5.44560	5.26395
MAR-UCB-f313ec4d-2	Mpro-x11813			5.78068	5.23802	3.24130	5.48928

MAT-POS-bbbbc21a-3*	Mpro-x11831		78.32	0.49190	6.30812	1.44887	5.83897
MAT-POS-199e2e7c-1	Mpro-x12000			4.64891	5.33265	5.16341	5.28706
MAT-POS-8a69d52e-7*	Mpro-x12073			1.82971	5.73762	1.71349	5.76612
ALP-POS-87c86d55-1	Mpro-x12143			14.29698	4.84476	4.85168	5.31411
ALP-POS-477dc5b7-2	Mpro-x12171			0.72040	6.14242	0.25876	6.58711
CHO-MSK-00c5269a-2	Mpro-x12177			5.65406	5.24764	1.13776	5.94395
ALP-POS-cd485364-2	Mpro-x12202			54.22646	4.26579	32.50610	4.48804
EDJ-MED-e4b030d8-13	Mpro-x12207			0.33870	6.47019	0.28363	6.54725
MAT-POS-044491d2-7	Mpro-x12300			32.87800	4.48309	36.53174	4.43733
MAT-POS-044491d2-3	Mpro-x12321			19.72462	4.70499	25.31642	4.59660
ALP-POS-6d04362c-1	Mpro-x12419			1.56502	5.80548	5.05327	5.29643
ALP-POS-6d04362c-2	Mpro-x12423			0.39149	6.40728	0.49676	6.30385

ERI-UCB-d6de1f3c-2	Mpro-x12582		5.53266	5.25707	9.18999	5.03668
PET-UNK-c9c1e0d8-3	Mpro-x12587		0.73163	6.13571	1.14150	5.94252
ALP-POS-ce760d3f-2	Mpro-x12659		1.29106	5.88905		
ALP-POS-64a710fa-1	Mpro-x12661		3.02151	5.51978	9.86113	5.00607
ALP-POS-ce760d3f-8	Mpro-x12674		16.53652	4.78156	78.20944	4.10674
ERI-UCB-d6de1f3c-1	Mpro-x12679		9.61327	5.01713	15.82123	4.80076
VLA-UCB-29506327-1	Mpro-x12686		0.36447	6.43833	1.31600	5.88074
EDJ-MED-92e193ae-1	Mpro-x12692				0.23023	6.63785
MAT-POS-c7771779-1	Mpro-x12695				20.75590	4.68286
EDG-MED-0e5afe9d-1	Mpro-x12696				20.67477	4.68456
BEN-DND-c852c98b-10	Mpro-x12698				6.16251	5.21024
EDG-MED-0e5afe9d-3	Mpro-x12699				0.18561	6.73140

MIC-UNK-91acba05-6	Mpro-x12710						0.44298	6.35362
MAT-POS-2bb0cf2b-2	Mpro-x12715						13.04496	4.88456
EDG-MED-971238d3-5	Mpro-x12716						5.02057	5.29925
MAT-POS-2bb0cf2b-1	Mpro-x12731						82.24060	4.08491
EDJ-MED-3c65e9ce-4	Mpro-x12740						49.15014	4.30848
WAR-XCH-79d12f6e-6	Mpro-x1336		96.2867	1.62337	5.78958	101.1492	14.21649	4.84721
LON-WEI-8f408cad-7	Mpro-x1386			1.13730	5.94412	99.437475	3.49772	5.45621
LON-WEI-8f408cad-5	Mpro-x1418			3.31638	5.47934	98.998055	6.19211	5.20816
DAR-DIA-23aa0b97-20	Mpro-x2563		62.475	45.07747	4.34604	57.79387	57.46967	4.24056
TRY-UNI-714a760b-20	Mpro-x2572		52.68	43.84705	4.35806	58.318795	61.33027	4.21233
TRY-UNI-714a760b-6	Mpro-x2646			13.89516	4.85714	81.75696	24.56662	4.60965
TRY-UNI-714a760b-18	Mpro-x2649			12.96769	4.88714	78.712695	26.17964	4.58204

TRY-UNI-714a760b-12	Mpro-x2908		64.135	45.68926	4.34019	58.708675	64.41361	4.19102
MAT-POS-916a2c5a-2	Mpro-x2910		83.785	3.46160	5.46072	92.667755	7.23400	5.14062
TRY-UNI-714a760b-22	Mpro-x2912		58.415	45.31192	4.34379	61.411195	54.94210	4.26009
MAT-POS-916a2c5a-4	Mpro-x3303		78.785	10.40272	4.98285	54.07172	27.52490	4.56027

*40 non-covalent compounds contained on the library created on 2020-12-02