

Virtual Screening and binding mode analysis of selected FDA approved drugs against PL^{pro} target: An effort to identify therapeutics to combat COVID-19

Mitra Ashouri^{1*} and Rohoullah Firouzi^{2*}

¹ Department of Physical Chemistry, School of Chemistry, College of Science, University of Tehran, Tehran, Iran.

² Department of Physical Chemistry, Chemistry and Chemical Engineering Research Center of Iran, Tehran, Iran.

Corresponding Authors:

* m.ashouri@ut.ac.ir

* rfirouzi@ccerci.ac.ir and firouzi.chemist@yahoo.com

ABSTRACT

Papain-like protease (PL^{pro}) is one of the most promising targets for anti-SARS-CoV drugs. In this study, a bioactive library of FDA-approved drugs has virtually been screened using a new successful computational pipeline in the framework of the ensemble docking to identify potential binding molecules of PL^{pro}. Based on our protocol, 20 FDA-approved drugs were found to bind the target enzyme with significant affinities and good geometries, suggesting their potential to be utilized against the virus. Our computational studies identified IMATINIB, a well-characterized drug used in the treatment of lymphoblastic and chronic myeloid leukemia, as a potential inhibitor against PL^{pro}. Subsequently, SIMPREVIR, NALDEMEDINE, TUCATINIB, and some other FDA-approved drugs exhibit great affinities to the PL^{pro}. These drugs are used in the treatment of several diseases such as cancer, schizophrenia, hypertension, hepatitis C, HIV infection, and AIDS showed high affinities in the screening. The high affinity of ligands was rationalized by the newly identified allosteric site named “BL2 groove”. Occupying this groove may disrupt access to the catalytic site and affect the protein function. The best binding mode and the efficacy of hydrogen bonds and hydrophobic interactions on inhibitory activities of ligands were also disclosed. Furthermore, our studies provide significant molecular insight into PL^{pro} inhibition that could aid in the development of new drugs for COVID-19 treatment.

Keywords: COVID-19, Virtual Screening, Ensemble Docking Simulations, FDA-approved drugs, Drug Discovery, Papain-like protease (PL^{pro}).

1. Introduction

For months, an ongoing pandemic caused by a single-stranded RNA virus (SARS-CoV-2) has been one of the major challenges facing the world. However, efforts are underway to keep up with the flood of coronavirus and no definitive cure has been reported to date. In general, the multiple stages of the SARS-CoV-2 life cycle provide potential targets for drug therapy [1-10].

To fight the current pandemic, major attempts have been made to target the SARS-CoV-2 spike protein, RdRp (RNA-dependent RNA polymerase), and cysteine proteases (Chemotrypsin-like protease or main protease (3CL^{pro} or M^{pro}) and Papain-like protease (PL^{pro})). In contrast to others, there are very few potent inhibitors of PL^{pro}

with efficacy validated experimentally [11]. The PL^{pro} encoded within non-structural proteins (nsps) during the viral life cycle is one of the most promising targets for anti-SARS-CoV drugs [12]. Available crystal structures of PL^{pro} reveal that this functional protein is composed of three domains: (a) a classic catalytic triad cysteine cleavage domain (Cys111, His272, Asp286) (b) a Zn-binding domain (Cys189, Cys192, Cys224 and Cys226) and (c) ubiquitin domain (Asp164, Arg166, and Glu167). Besides these significant domains, distal from the cysteine active site, a newly identified binding pocket, “BL2 groove”, has been identified in recent articles (consisting of residues Leu162, Asp164, Arg166, Glu167, Pro247, Pro248, Tyr264, Gly266, Asn267, Tyr268, Gln269, Tyr273, and Pro299) [13-15]. Varieties of potent natural and synthetic compounds, from natural herbal ones or diverse antiviral drugs, are introduced as therapeutic compounds (e.g. Remdesivir designed for the Ebola virus, Lopinavir/Ritonavir designed for the HIV, chloroquine and hydroxychloroquine designed for anti-malarial action etc.). These compounds are found to be effective but their efficacy still remains disputed. In this way, exploring the repurposing of existing drugs and developing new drugs are both of considerable interest [16].

Computer-aided drug design has been a reliable technique in designing novel and potent therapeutics. The availability of multiple experimental protein structures has led to coping with the challenge of protein flexibility and induced fit at a lower computational cost via the ensemble docking technique. In this way, trial compounds are docked into binding sites of the protein ensemble and then ranked according to their strength of binding to identify initial hits for further experimental studies [17, 18].

The development of new antiviral drugs is time-consuming [19]. Therefore, to speed up the process, we used a new docking protocol in the framework of the ensemble docking strategy to identify SARS-CoV-2 PL^{pro} inhibitors from an unbiased existing bioactive library containing 1916 FDA-approved drugs [20]. Since the safety of these drugs is well established and their efficacy can be quickly tested, drug repurposing could be an efficient approach to find potential inhibitors against COVID-19. The performance of the docking protocol has been evaluated in predicting the correct binding modes for the available crystal structures of PL^{pro}-ligand complexes.

2. Methods

2.1. PL^{pro} Structure Preparation

To create a conformational ensemble of PL^{pro}, 45 PL^{pro} structures in the Protein Data Bank (PDB) were selected [21]. All protein chains were characterized and analysed based on information which recorded in their PDB files (e.g. mutated/modified residues, missing atoms/residues, ligands information, resolution, alternate locations etc.) (Table S1 in the Supporting Information). It is noteworthy that 12 protein-ligand complex were covalently bonded to the thiol group of Cys111 side chain in an irreversible manner. In addition, Cys/Ser mutation were observed in Cys111 catalytic residue in 16 protein-chain structures (6WRH_A, 6XG3_A, 6YVA_A, 7CJD_D, 7CJD_A, 7CJD_B, 7CJD_C, 7CJM_B, 7D47_A, 7D47_B, 7JIR_A, 7JIT_A, 7JIV_A, 7KOJ_A, 7KOK_A, 7KRX_A). The 87 monomeric chains were identified from the 45 PDB files of the PL^{pro}, which 36 of the chains were discarded due to having the missing functionally important segments in the catalytic pocket of the PL^{pro} (which includes a Cys111-His272-Asp286 catalytic triad).

The intersection of residues involved in the formation of β -sheets in the thumb and palm domains which contain 41 residues was used to drive the structural alignment process (Table S2). All PL^{pro} monomeric chains were aligned on the C α atoms of the selected residues of 6WRH chain A as a reference structure, which is a high-quality crystal structure of the PL^{pro} with a resolution of 1.6 Å. The pairwise RMSD values for all above C α atoms between every

pair of the aligned protein structures of the PL^{pro} did not exceed 0.48 Å. Even though we are all aware that the average nature of the RMSD measure does not capture local structural changes/dissimilarities between structures [22], this small amount of RMSD indicates the significant conformational stability of β -sheets (in the presence or absence of ligands) for use in alignment staging. In the following sections, to gauge conformational changes and structural differences between two protein structures, pairwise RMSD values between individual the same pairs of residues (including all backbone and side-chain atoms) belonging to various pairs of the aligned protein structures were evaluated. This subject will be clarified more in the next sections.

2.2. Defining PL^{pro} Active Site

The structural analysis of protein structures showed that 46 out of 51 protein chains were in complex with at least one ligand - from very small molecular fragments to very large peptidomimetic inhibitors. All the ligands were characterized by their approximated center and size that were defined by the centroid and the longest distance between two atoms of the ligands, respectively (Table S3). The distribution of ligands' position in the protein structure based on their scaled approximate bounding sphere is shown in Figure 1a. To create a lucid and tangible 2D-representation of ligand distribution in the Cartesian space of protein structure, a principle component analysis (PCA) was done on ligands' centroid [23, 24]. Figure 1b shows the projection of the ligands' centroid distribution onto the subspace spanned by the first two principal components (PC1 and PC2) obtained from PCA. In Figure 1b, closed and overlapped circles occupy the same region on protein surface. By considering Figures 1a and 1b, six separate regions are observed on the protein space, the most populated one (33 out of 46) is located at the same region that is known as the PL^{pro} binding site - a cleft between the thumb and palm domains [13-15]. Other sparsely distributed ligands should be discarded for the binding site analysis.

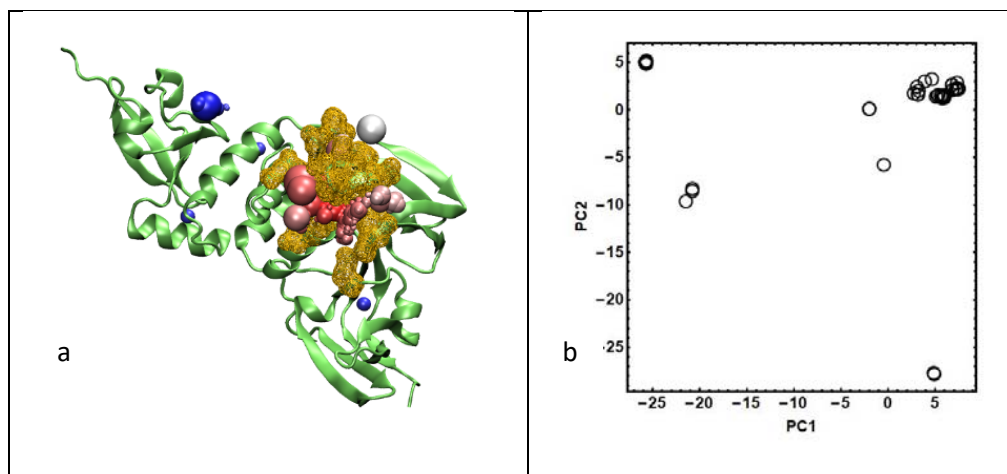


Figure 1: (a) Representation of ligands' position distribution in protein structure and PL^{pro} binding site. Defined active site is shown in orange and surface wired-frame representation. (b) Projection of ligands centroid distribution to the subspace spanned by PC1 and PC2 on Cartesian coordinates of ligands centers.

In order to define PL^{pro} general active site, for all holo forms, all interacting amino acids that had even one heavy atom at a cutoff 3.5 Å were included in the definition. Therefore, a pocket that was defined with 21 unique amino acids were considered as general binding site of PL^{pro} to use in further calculations (Table S4 and Figure 1a). This binding pocket definition is in good agreement with other studies based on different methods [15, 16]. It should be

noticed that after defining the general active site, 6W9C chain A was omitted from protein structure data due to existing some missing residues in the active site.

The “relative importance” of the general binding site residues was determined by the percentage of occurrence contacts of the residue in 33 protein-ligand complexes (see Table S4). For example, based on the “contact-importance” factor, Asp164 is the most important residue in PL^{pro} active site. This amino acid is located at ubiquitin domain of PL^{pro} and plays a crucial role in protein activity [14,15]. Considering Table S4 shows that the general binding site includes all previous definitions of different PL^{pro} domains (cysteine domain, ubiquitin domain, and BL2 groove) and is in good agreement with recent studies on PL^{pro} active site [15, 25]. Relative importance data could be employed to select an optimal sub-set of PL^{pro} residues for performing quantum mechanics (QM) calculations, especially for studying covalent inhibition mechanism [6, 26].

Another notable parameter for the active site residues is their flexibility. Accordingly, “residue flexibility” (RF) was defined as $RF = 100 * N_{RMSD} / (N_{Complex}^2 - N_{Complex})$, where N_{RMSD} is the number of pairwise RMSD matrix elements with values more than 0.7 Å and $N_{Complex}$ is the number of complex that RMSD calculation was done for their general binding site residues (33 complex). In this way, the more flexible residues would have more RMSD matrix elements with values higher than 0.7 Å. The “relative importance” and the “residue flexibility” of the residues belonging to the general binding site have been displayed in Figure 2. For example, based on values reported in Figure 2 and Table S4 Asp164, Tyr268, Tyr264, Gln269, Gly163 and Pro248 residues in more than 75% of the complexes have effective contacts with their ligands (within the distance cutoff 3.5 Å). consequently, such residues are important and should be considered for determination of the binding pocket that this finding is in excellent agreement with recent studies [13-15]. Moreover, Thr301, Gly163, Asp164, Tyr264, Glu167, Gly271, and Pro242 have RF values lower than 60% and the loop structure. it did not seem farfetched that these most flexible residues belong to the BL2 groove. On the other hand, in comparison with other residues, Cys111 in the catalytic triad has the lowest flexibility. It is important to note that quantifying flexibility results can be used to identify and select an optimal set of flexible binding site residues for induced-fit docking (IFD) and flexible docking methods [27, 28].

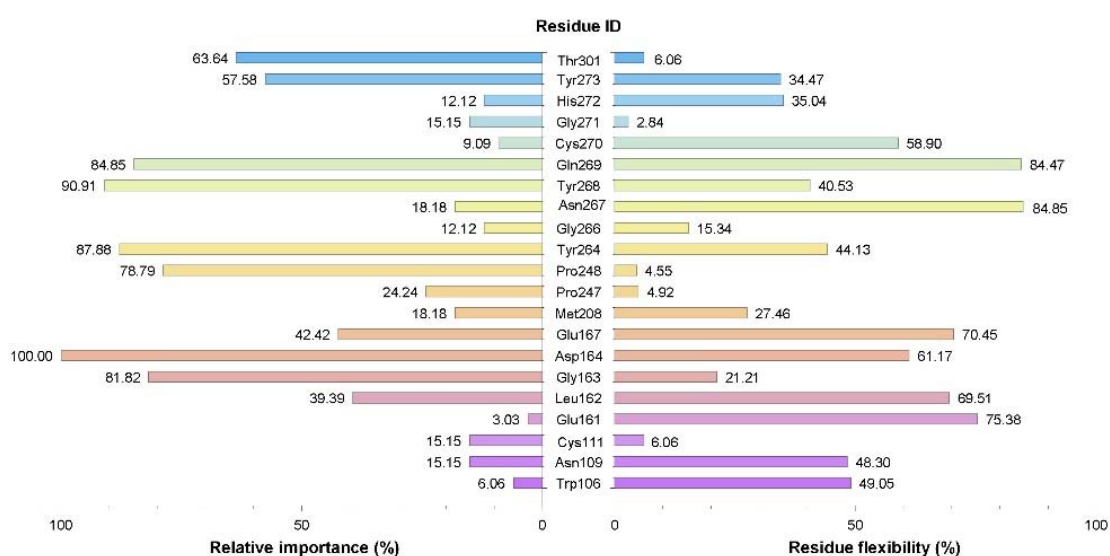


Figure 2. The “relative importance” and the “residue flexibility” of the residues belonging to the general binding site.

2.3. Conformational Ensemble Generation

In this step, the pairwise RMSD matrix of general binding site residues were calculated for all 51 PL^{pro} chains and the single-linkage hierarchical agglomerative clustering method based on the Ward Variance Minimization Algorithm was employed to detect representative protein structures [29-30]. A conformer belongs to a cluster if not only the pairwise RMSD average values with all members of the cluster are less than 1.2 Å but also the number of the pairwise RMSD values greater than 2.0 Å is less than three. The clustering result is depicted in Figure 3. As can be seen in this figure, the clustering yields 8 clusters that the population of each cluster is mentioned in the boxes. As it turns out, by loosening the clustering criteria clusters will merge and the number of them will decrease. Given the clustering constraints and the computational cost of calculations simultaneously, choosing eight conformers seems optimal for constructing a conformational ensemble of PL^{pro}. In each cluster, structures with the lowest average RMSD than the others were selected as representatives. In cases where there were more than one representative conformer ones that had better resolution was chosen. PDB IDs and their corresponding chains for the representative structures were depicted in Figure 3.

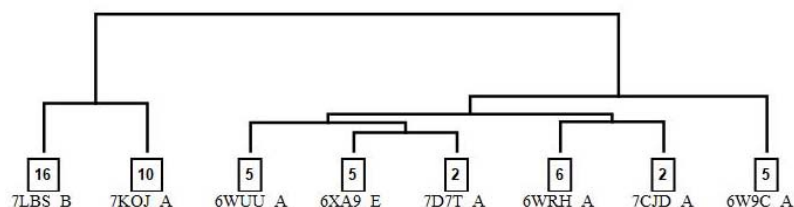


Figure 3: Bottom-up tree dendrogram of the clusters obtained using the Ward's hierarchical method. The population of each cluster is given in each box and the PDB IDs of the representative structures for each cluster are also displayed below the boxes.

2.4. Ensemble docking

For representative PL^{pro} structures, hydrogen atoms and missing atoms were retrieved using REDUCE software [31] and PSFGEN package within VMD software [32], respectively. During retrieving hydrogen atoms, side-chain flipping was considered based on the hydrogen-bonding patterns and the chemical environment, most of the histidine residues were protonated at $N\epsilon$ (i.e., HSE) and some of them were protonated at $N\delta$ (e.g. His275). The prediction of protonation states and side-chain flips for the histidine residues found near the substrate-binding region can be found in Table S5. In addition, the basic residues (Arginine/Lysine) and the acidic residues (Aspartic/Glutamic) of all the structures were assigned positive and negative charges, respectively.

Protonated structures were optimized using the NAMD.2.13 program [33] with the CHARMM36m force field and generalized Born implicit solvent (GBIS) to regulate the structure in the CHARMM36m force field, and in particular to reduce the steric conflicts that may occur in the system. Minimization was performed for 20000 steps of the conjugated gradient where during the first 10000 steps protein heavy atoms were restrained with a positional harmonic force of 200 kcal/mol·Å² followed by 10000 steps with the harmonic force of 100 kcal/mol·Å². In all minimizations, the gradient has been reduced below 0.05 (kcal/mol)/Å.

Three-dimensional structures of the FDA-approved drugs, containing 1993 compounds, were downloaded from the ChEMINFORMATIC Tools and Database for Pharmacology site in the structure-data file (SDF) format [34]. The

structures with more than 20 rotatable bonds were excluded from subsequent calculations, due to the well-known fact that the docking success rates decrease with increasing the number of active rotatable bonds [doi: 10.1038/nrd1549 and 10.1002/prot.20088]. In addition, six ligands that contain Boron element (B) were deleted from the ligands library due to the lack of force field parameters for this element. All 1910 remaining ligands were geometrically optimized in the gas phase using the PM7 [35] semi-empirical quantum mechanics (SQM) method with MOPAC2016 [36] while the gradient norm was set to 10 kcal mol⁻¹Å⁻¹. It is noteworthy that the PM7 is a fast and successful SQM method that reliably describes various types of noncovalent interactions and some important chemical observations, such as the planarity of conjugated rings or molecular fragments [37].

The initial docking search space was determined based on the active site definition of PL^{pro} – as described in section 2.2. This definition includes all the bound ligands from small-molecule fragments to large peptidomimetic inhibitors at the catalytic binding site. Then the box size was extended by 5 Å in each of the three dimensions to ensure that the docking search space is large enough for the ligands to rotate in [38, 39]. The final docking box was included in a box of 20 × 23 × 27 Å³, centered on the mean of the geometric centers of all of the considered ligands.

Docking simulations were performed for all 1910 ligands and 8 representative protein conformers with both AutoDock4.2.5.1 and AutoDock Vina1.1.2 software [38, 39]. Input files of docking with AutoDock were prepared automatically with Python scripts provided by MGLTools [40] and a Lamarckian genetic algorithm with an initial population of 500 was repeated 200 times for each ligand-receptor complex, therefore 200 models are built in each run. A grid spacing of 0.375 Å and a distance-dependent function of the dielectric constant was used for the calculation of the energetic maps with the program AutoGrid4[41]. The default settings were used for all other parameters. For Vina calculations, the exhaustiveness parameter was increased to 500 in order to enhance the probability of finding the proper ligand conformations [42]. Each Vina run generates 20 poses.

All optimized ligands (1910 ligands) were docked in the conformational ensemble of the 8 prepared representative PL^{pro} structures (described in Sections 2.3). In total, 1600 (= 8 × 200) poses by AutoDock and 160 (= 8 × 20) poses by Vina were calculated for each ligand. All the predicted docking poses for each ligand were collected separately for each docking program and re-clustered based on the symmetry corrected heavy-atom RMSD algorithm implemented in AutoDock4 with an RMSD cutoff of 2.0 Å [20, 47-51]. As a result, all predicted poses of a given ligand into multiple different conformations of the PL^{pro} binding site are simultaneously organized and clustered for identifying representative poses and subsequent analyses. It is noteworthy to mention that using an ensemble of rigid receptor conformations in docking simulations (i.e. ensemble docking) is the most common strategy to cope with the receptor flexibility problem in rigid docking procedures[18, 20, 43, 44]. Additionally, ligands flexibility are considered based on active rotatable bonds during the docking calculations, despite the fact that the docking success rates decrease with increasing the number of active rotatable bonds [45,46].

3. Results and discussion

Based on our recent researches, the top-ranked poses (with the lowest-energy poses) from the *first* and the *most populated* clusters are chosen as the representative poses for each docked ligand [20, 37, 47]. Therefore, we focus on the two representative docking poses for each ligand and use their docking scores, instead of considering the mean over docking score of the top-ranked poses of all representative PL^{pro} conformations or taking the best-scoring

binding poses from an aggregation of all predicted docking poses (the lowest scoring poses of ensemble docking), as routinely utilized in an ensemble docking protocol [18, 43-44, 49-50].

The above-mentioned computational pipeline (including the structural clustering strategy to construct the protein ensemble for performing docking calculations and the proposed manner to choose the representative docking poses from the ensemble docking results) has already been used for correctly predicting experimental binding poses and affinity ranking of M^{Pro}-ligand complexes [20] and thus can be utilized to properly produce a rank-ordered list of the FDA-approved drugs. Accordingly, two rank-ordered lists of the docked FDA-approved drugs have been produced for each of both AutoDock and Vina software, one for those called the *first clusters* and others for the *most populated clusters*. The top 20 FDA-approved drugs for both the first and the most populated clusters are shown in Table 1 (the top 100 compounds are given in Tables S6 and S7 in the Supporting Information).

Table 1. The top 20 FDA-approved drugs obtained from the top-ranked poses of the first and the most populated clusters (using AutoDock and Vina software).

Rank	First Clusters		Most Populated Clusters	
	AutoDock	Vina	AutoDock	Vina
1	VENETOCLAX	CONIVAPTAN	PLERIXAFOR	CONIVAPTAN
2	CAPMATINIB	OLAPARIB	GENTAMICIN	PAZOPANIB
3	IMATINIB	LUMACAFTOR	LURASIDONE	TROVAFLOXACIN
4	ZAFIRLUKAST	NILOTINIB	THIOTHIXENE	VEMURAFENIB
5	PLERIXAFOR	PAZOPANIB	PONATINIB	TELMISARTAN
6	CABOZANTINIB	TROVAFLOXACIN	M RIFAPENTINE	PHENYTOIN
7	IRINOTECAN	VEMURAFENIB	PIMECROLIMUS	DOLUTEGRAVIR
8	GENTAMICIN	TELMISARTAN	ERIBULIN	ADAPALENE
9	LURASIDONE	PHENYTOIN	BRIGATINIB	BEXAROTENE
10	SIMEPREVIR	DOLUTEGRAVIR	TUCATINIB	TIPRANAVIR
11	THIOTHIXENE	BICTEGRAVIR	NETARSUDIL	MAZINDOL
12	PONATINIB	ENTRECTINIB	ZANUBRUTINIB	PERAMPANEL
13	M RIFAPENTINE	ADAPALENE	CALCIFEDIOL	OLAPARIB
14	PANCURONIUM	TROGLITAZONE	CALCITRIOL	IMATINIB
15	NALDEMEDINE	BEXAROTENE	SIMEPREVIR	LUMACAFTOR
16	TUCATINIB	TIPRANAVIR	IVERMECTIN	NETARSUDIL
17	EVEROLIMUS	MAZINDOL	TIPRANAVIR	FEXOFENADINE
18	CILOSTAZOL	CARBENICILLIN INDANYL	IMATINIB	NILOTINIB
19	PIMECROLIMUS	TROGLITAZONE	PIPERACETAZINE	GLYBURIDE
20	ERDAFITINIB	PERAMPANEL	RALOXIFENE	BINIMETINIB

The analysis and comparison of the screening results show that there are several common drugs among the top 100 ranked-ordered lists of the aforementioned protocols. Eventually, 20 common FDA-approved drugs were found in all tables obtained from the representative poses of the *first clusters* and the *most populated clusters* (Tables S6-S7). These 20 common drugs and their medical uses in the treatment of different diseases and their 2D-chemical structures are shown in Table 2 and Figure S1. From the computational viewpoint, it means that these FDA-

approved drugs target the PL^{pro} protein and can be considered potential antiviral drug candidates in the treatment of COVID-19.

Table 2: Docking scores (in kcal.mol⁻¹) of the top 20 common FDA-approved drugs (obtained from the 100 top-ranked poses of Tables S6 and S7)

No.	Name	First Clusters		Most Populated Clusters		Used in the treatment of:
		AutoDock	Vina	AutoDock	Vina	
1	IMATINIB	-11.95	-9.30	-10.54	-9.30	Lymphoblastic and Chronic Cyeloid leukaemia
2	SIMEPREVIR	-11.20	-8.80	-10.59	-8.80	Chronic Viral Hepatitis C
3	NALDEMEDINE	-11.07	-9.00	-9.61	-8.50	Constipation
4	TUCATINIB	-11.07	-9.30	-10.71	-9.10	Breast Cancer
5	ERDAFITINIB	-10.94	-8.80	-10.41	-8.80	Metastatic Urothelial Carcinoma
6	BRIGATINIB	-10.81	-8.90	-10.81	-8.50	Lung Cancer
7	NILOTINIB	-10.80	-9.70	-10.17	-9.20	Chronic Myelogenous Leukemia
8	ELTROMBOPAG	-10.76	-9.10	-10.08	-9.10	Cancer
9	TELMISARTAN	-10.70	-9.60	-10.12	-9.60	Hypertension
10	NETARSUDIL	-10.66	-9.30	-10.66	-9.30	Glaucoma or Ocular Hypertension
11	ZANUBRUTINIB	-10.64	-9.40	-10.64	-8.70	Mantle Cell Lymphoma (MCL)
12	TIPRANAVIR	-10.56	-9.50	-10.56	-9.50	HIV infection and AIDS
13	RISPERIDONE	-10.53	-9.00	-9.72	-9.00	Mania and Schizophrenia
14	PALIPERIDONE	-10.48	-9.00	-9.85	-9.00	Schizophrenia
15	CERITINIB	-10.33	-8.90	-9.63	-8.80	Lung Cancer
16	LUMACAFOR	-10.31	-9.80	-9.93	-9.30	cystic fibrosis
17	AVAPRITINIB	-10.29	-8.90	-10.29	-8.90	Metastatic Gastrointestinal Stromal Tumours
18	MENTRECTINIB	-10.22	-9.20	-10.22	-9.20	Lung Cancer
19	BAZEDOXIFENE	-10.21	-9.00	-10.21	-9.00	Vasomotor Symptoms and Menopause
20	PERAMPANEL	-10.16	-9.40	-10.16	-9.40	Partial Onset Seizures

The screening results show that a number of cancer drugs (IMATINIB, TUCATINIB, ERDAFITINIB, BRIGATINIB, NILOTINIB, ELTROMBOPAG, ZANUBRUTINIB, CERITINIB, AVAPRITINIB, MENTRECTINIB), schizophrenia drugs (RISPERIDONE, PALIPERIDONE), hypertension drugs (TELMISARTAN, NETARSUDIL), hepatitis C drug (SIMEPREVIR) and HIV infection and AIDS drug (TIPRANAVIR) display high binding affinities to PL^{pro} protein. Our results are in agreement with others' work that introduce possible drug candidates as PL^{pro} inhibitors (e.g. Antivirus, Antihypertensive, Antibacterial, Antipsychotic, Anti-tumor, and anti-hepatic drugs) [52-54]. They demonstrate that some FDA-approved drugs that are used in the treatment of other diseases can be used as therapeutic agents in COVID-19. (e.g. Acetophenazine, Sulfasalazine, Dutasteride, Atovaquone). It is noteworthy that these drugs have also appeared in our 100 top-ranked lists.

A detailed survey of the chemical structures in Figure S1 demonstrates that most of the top-ranked compounds are *Naphthalene-based* that are of particular interest for their inhibitory activity against SARS-CoV-2 PL^{pro} [54,55]. In some inhibitors, the *Naphthyl ring* is replaced with a *biaryl* group scaffold. It has been shown that this replacement

is a promising development in this class of inhibitors as it likely represents a *metabolic liability* along the path to the clinic [56]. These findings can inspire the discovery of a new generation of naphthalene-based PL^{PRO} inhibitors. A detailed insight into the contacts and noncovalent interactions in docking-generated complexes has been obtained by LigPlot [57]. Careful inspection of the results shows that most of the 20 common ligands are located in the “BL2 Groove” binding pocket and Ubiquitin domain that are far from the classical catalytic triad. As mentioned before, the “BL2 Groove” is positioned at the N-terminal side of the BL2 loop and composed of hydrophobic amino acids such as Pro299, Pro248, Tyr264, Pro247, Gln269 and hydrogen bonding residues such as Gly266, Tyr268, Leu162, Asn267, Tyr273, Arg166, and Asp164. The strong hydrogen bonding and hydrophobic interaction between FDA-approved drugs with the enzyme imply that they can be potent PL^{PRO} inhibitors. (Figure4 and Figure S2). Additionally, the inhibitors' orientations show that the BL2 loop is in direction of the catalytic site and occupying the BL2 groove can disrupt access to this site and affect the protein function by preventing the substrate from entering the active site in a non-competitive inhibition manner [54]. These results are in agreement with others' works and confirm the possibility of identifying allosteric inhibitors against SARS-CoV-2 PL^{PRO} [58,14].

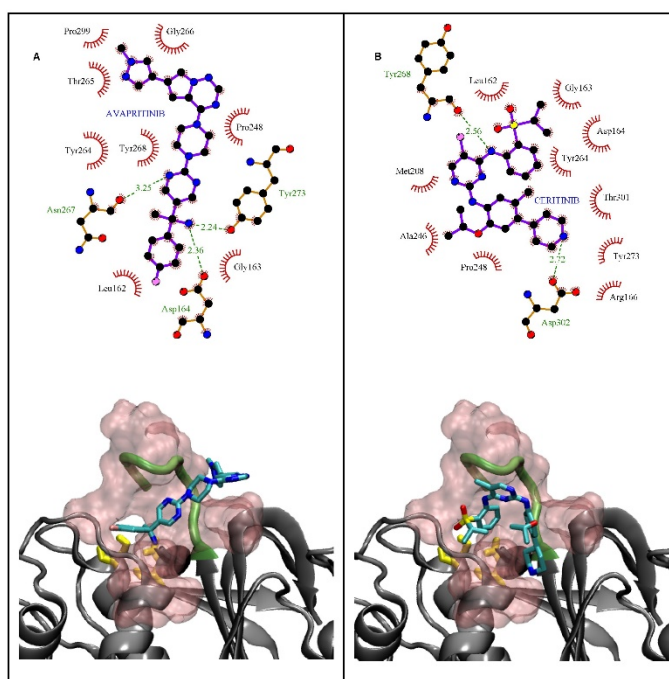


Figure 4: Docked complexes and binding details of the two drugs from 20 common FDA-approved drugs. Hydrogen bonds are shown as dotted green lines. Residues involved in hydrophobic interactions are represented as lines in red and ligand atoms in hydrophobic contacts are surrounded by red lines

4. CONCLUSION

The SARS-CoV-2 papain-like protease (PL^{PRO}), is a key target for the design of inhibitors to tackle virus activity in host cells. Lately, computer-aided drug design techniques are reliable, affordable, and less time-consuming in designing novel therapeutics.

The objective of this work is the identification of potential inhibitors of SARS-CoV-2 PL^{PRO} protease from FDA-approved drugs using molecular docking. In this work, we used a new docking protocol in the framework of the ensemble docking strategy to identify SARS-CoV-2 PL^{PRO} inhibitors from a library containing 1910 FDA-approved drugs against PL^{PRO} proteases. The computational pipeline and its reliability are discussed in our recent work for M^{PRO}-ligand complexes and has now been employed to properly rank 1910 FDA-approved drugs during their

screening against PL^{pro} target [20]. In this way, 100 top-ranked of drugs for four different used protocol were introduced and 20 common compounds were analyzed in details. It has been shown that the high flexible BL2 loop can regulate the binding cavity accessibility and besides the catalytic triad, the “BL2 Groove” should be considered as an allosteric pocket.

We demonstrate that some FDA-approved drugs which are used in the treatment of cancer, schizophrenia, hypertension, hepatitis C, HIV infection, and AIDS exhibit high binding affinity to PL^{pro} protein and can be used as therapeutic candidates in the treatment of COVID-19. Obviously, performing comprehensive molecular dynamics (MD) simulations, experimental assays, and clinical trials are necessary to confirm their actual activity against COVID-19.

From a chemical scaffold point of view, we show that naphthalene-based drugs along with biaryl group substituted compounds have inhibitory activity against SARS-CoV-2 PL^{pro}. Ensemble docking studies accented to the non-competitive inhibition mechanism and confirmed the existence of allosteric sites in the PL^{pro} structure. The binding mode analysis confirms the contribution of hydrogen bonds along with extensive hydrophobic interactions to inhibition of activity.

References:

- (1) Zhenming Jin, Xiaoyu Du, Yechun Xu, Yongqiang Deng, Meiqin Liu, Yao Zhao, Bing Zhang, Xiaofeng Li, Leike Zhang, Chao Peng, Yinkai Duan, Jing Yu, Lin Wang, Kailin Yang, Fengjiang Liu, Rendu Jiang, Xinglou Yang, Tian You, Xiaoce Liu, Xiuna Yang, Fang Bai, Hong Liu, Xiang Liu, Luke W. Guddat, Wenqing Xu, Gengfu Xiao, Chengfeng Qin, Zhengli Shi, Hualiang Jiang, Zihao Rao, and Haitao Yang, Structure of Mpro from SARS-CoV-2 and discovery of its inhibitors, *Nature* 2020, 582, 289–293.
- (2) Wenhao Dai, Bing Zhang, Xia-Ming Jiang, Haixia Su, Jian Li, Yao Zhao, Xiong Xie, Zhenming Jin, Jingjing Peng, Fengjiang Liu, Chunpu Li, You Li, Fang Bai, Haofeng Wang, Xi Cheng, Xiaobo Cen, Shulei Hu, Xiuna Yang, Jiang Wang, Xiang Liu, Gengfu Xiao, Hualiang Jiang, Zihao Rao, Lei-Ke Zhang, Yechun Xu, Haitao Yang, and Hong Liu, Structure-based design of antiviral drug candidates targeting the SARS-CoV-2 main protease, *Science* 2020, 368, 1331–1335.
- (3) Daniel W. Kneller, Stephanie Galanie, Gwyndalyn Phillips, Hugh M. O’Neill, Leighton Coates, and Andrey Kovalevsky, Malleability of the SARS-CoV-2 3CL Mpro Active-Site Cavity Facilitates Binding of Clinical Antivirals, *Structure* 2020, 28, 1313–1320.
- (4) Zhe Li, Xin Li, Yi-You Huang, Yaoxing Wu, Runduo Liu, Lingli Zhou, Yuxi Lin, Deyan Wu, Lei Zhang, Hao Liu, Ximing Xu, Kunqian Yu, Yuxia Zhang, Jun Cui, Chang-Guo Zhan, Xin Wang, and Hai-Bin Luo, Identify potent SARS-CoV-2 main protease inhibitors via accelerated free energy perturbation-based virtual screening of existing drugs, *PNAS* 2020, 117, 27381–27387.
- (5) Krishna Kumar and Tania J. Lupoli, Exploiting Existing Molecular Scaffolds for Long-Term COVID Treatment, *ACS Med. Chem. Lett.* 2020, 11, 1357–1360.
- (6) Katarzyna Świderek and Vicent Moliner, Revealing the molecular mechanisms of proteolysis of SARS-CoV-2 Mpro by QM/MM computational methods, *Chem. Sci.*, 2020, 11, 10626–10630.

- (7) Jared S. Morse, Tyler Lalonde, Shiqing Xu, and Wenshe Ray Liu, Learning from the Past: Possible Urgent Prevention and Treatment Options for Severe Acute Respiratory Infections Caused by 2019-nCoV, *ChemBioChem* 2020, 21,730–738.
- (8) Antonio Francés-Monerris, Cécilia Hognon, Tom Miclot, Cristina García-Iriepa, Isabel Iriepa, Alessio Terenzi, Stéphanie Grandemange, Giampaolo Barone, Marco Marazzi, and Antonio Monari, Molecular Basis of SARS-CoV-2 Infection and Rational Design of Potential Antiviral Agents: Modeling and Simulation Approaches, *J. Proteome Res.* 2020, 19, 4291–4315.
- (9) Hailei Su, Feng Zhou, Ziru Huang, Xiaohua Ma, Kathiresan Natarajan, Minchuan Zhang, Yong Huang, and Haibin Su, Molecular Insights into Small-Molecule Drug Discovery for SARS-CoV-2, *Angew. Chem. Int. Ed.* 2021, 60, 9789–9802.
- (10) Eugene N. Muratov, Rommie Amaro, Carolina H. Andrade, Nathan Brown, Sean Ekins, Denis Fourches, Olexandr Isayev, Dima Kozakov, José L. Medina-Franco, Kenneth M. Merz, Tudor I. Oprea, Vladimir Poroikov, Gisbert Schneider, Matthew H. Todd, Alexandre Varnek, David A. Winkler, Alexey V. Zakharov, Artem Cherkasov, and Alexander Tropsha, A critical overview of computational approaches employed for COVID-19 drug discovery, *Chem. Soc. Rev.*, 2021, 50, 9121–9151.
- (11) Sternberg, A., Naujokat, C. Hide, Structural features of coronavirus SARS-CoV-2 spike protein: Targets for vaccination, *Life Sciences*, Volume 257, 15 September 2020
- (12) Ramarao Poduri, Gaurav Joshi, Gowraganahalli Jagadeesh, Drugs targeting various stages of the SARS-CoV-2 life cycle: Exploring promising drugs for the treatment of Covid-19, *Cellular Signalling* 74 (2020) 109721
- (13) Biplab K Maiti, Can Papain-like Protease Inhibitors Halt SARS-CoV-2 Replication? *ACS Pharmacol. Transl. Sci.* 2020, 3, 1017–1019
- (14) Zhengnan Shen, Kiira Ratia¹, Laura Cooper, Deyu Kong, Hyun Lee, Youngjin Kwon, Yangfeng Li, Saad Alqarni, Fei Huang, Oleksii Dubrovskiy, Lijun Rong, Gregory Thatcher, Rui Xiong, Design of SARS-CoV-2 PLpro Inhibitors for COVID-19 Antiviral Therapy Leveraging Binding Cooperativity, *J. Med. Chem.* 2022, 65, 4, 2940–2955.
- (15) Sajjan Rajpoot, Manikandan Alagumuthu, Mirza S. Baig, Dual targeting of 3CLpro and PLpro of SARS-CoV-2: A novel structure-based design approach to treat COVID-19, *Current Research in Structural Biology* 3, 2021, 9–18.
- (16) *Bioorg. Med. Chem.* 2020, 10.1016/j.bmc.2020.115860, 4- M.D. Laskar, M.A. Choudhury, Search for therapeutics against COVID 19 targeting SARS-CoV-2 papain-like protease: an in silico study, 2020, 1–28.
- (17) P.K. Panda, M.N. Arul, P. Patel, S.K. Verma, W. Luo, H.G. Rubahn, Y.K. Mishra, M. Suar, R. Ahuja, Structure-based drug designing and immunoinformatics approach for SARS-CoV-2, *Sci. Adv.* 6, 2020.
- (18) Acharya A, Agarwal R, Baker MB, Baudry J, Bhowmik D, Boehm S, Byler KG, Chen SY, Coates L, Cooper CJ, Demerdash O, Daidone I, Eblen JD, Ellingson S, Forli S, Glaser J, Gumbart JC, Gunnels J, Hernandez O, Irle S, Kneller DW, Kovalevsky A, Larkin J, Lawrence TJ, LeGrand S, Liu SH, Mitchell JC, Park G, Parks JM, Pavlova A, Petridis L, Poole D, Pouchard L, Ramanathan A, Rogers DM, Santos-Martins D, Scheinberg A, Sedova A, Shen Y, Smith JC, Smith MD, Soto C, Tsaris A, Thavappiragasam M, Tillack AF, Vermaas JV, Vuong VQ, Yin J, Yoo S,

Zahran M, Zanetti-Polzi L. Supercomputer-Based Ensemble Docking Drug Discovery Pipeline with Application to Covid-19. *J Chem Inf Model.* 2020; 60:5832–5852.

(19) S. Pushpakom, F. Iorio, P.A. Eyers, K.J. Escott, S. Hopper, A. Wells, A. Doig, T. Williams, J. Latimer, C. McNamee, A. Norris, P. Sanseau, D. Cavalla, M. Pirmohamed, Drug repurposing: progress, challenges and recommendations, *Nat. Rev. Drug Discov.* 18 (1), 2019, 41–58.

(20) Rohoullah Firouzi, Mitra Ashouri, and Mohammad Hossein Karimi-Jafari, Structural insights into the substrate-binding site of main protease for the structure-based COVID-19 drug discovery, *Proteins.* 2022; 90:1090–1101.

(21) S. K. Burley, H. M. Berman, C. Bhikadiya, C. Bi, L. Chen, L. Di Costanzo, C. Christie, K. Dalenberg, J. M. Duarte, S. Dutta, Z. Feng, S. Ghosh, D. S. Goodsell, R. K. Green, V. Guranović, D. Guzenko, B. P. Hudson, T. Kalro, Y. Liang, R. Lowe, H. Namkoong, E. Peisach, I. Periskova, A. Prlić, C. Randle, A. Rose, P. Rose, R. Sala, M. Sekharan, C. Shao, L. Tan, Y.-P. Tao, Y. Valasatava, M. Voigt, J. Westbrook, J. Woo, H. Yang, J. Young, M. Zhuravleva and C. Zardecki, RCSB Protein Data Bank: biological macromolecular structures enabling research and education in fundamental biology, biomedicine, biotechnology and energy, *Nucleic Acids Res.*, 2019, 47, D464–D474.

(22) Behnaz Moghadam, Mitra Ashouri, Hossein Roohi, Mohammad Hosein Karimi-jafari, Computational evidence of new putative allosteric sites in the acetylcholinesterase receptor, *Journal of Molecular Graphics and Modelling* 107 (2021) 107981.

(23) Martin Krzywinski, and Naomi Altman, Principal component analysis, Jake Lever, *Nat. Methods* 2017, 14, 641–642.

(24) Alessandro Giuliani, The application of principal component analysis to drug discovery and biomedical data, *Drug Discov. Today* 2017, 22, 1069076.

(25) Brendan T. Freitas, Ian A. Durie, Jackelyn Murray, Jaron E. Longo, Holden C. Miller, David Crich, Robert Jeff Hogan, Ralph A. Tripp, and Scott D. Pegan; Characterization and Noncovalent Inhibition of the Deubiquitinase and deISGylase Activity of SARS-CoV-2 Papain-Like Protease; *ACS Infect. Dis.* 2020, 6, 8, 2099–2109.

(26) Kemel Arafet, Natalia Serrano-Aparicio, Alessio Lodola, Adrian J. Mulholland, Florenci V. González, Katarzyna Świderek, and Vicent Moliner, Mechanism of inhibition of SARS-CoV-2 Mpro by N3 peptidyl Michael acceptor explained by QM/MM simulations and design of new derivatives with tunable chemical reactivity, *Chem. Sci.*, 2021, 12,1433–1444.

(27) Elekofehinti, O.O., Iwaloye, O., Josiah, S.S. et al. Molecular docking studies, molecular dynamics and ADME/tox reveal therapeutic potentials of STOCK1N-69160 against papain-like protease of SARS-CoV-2. *Mol Divers* 25, 1761–1773 (2021).

(28) Yosef Masoudi-Sobhanzadeh, Aysan Salemi, Mohammad M ourseif, Behzad Jafari, Yadollah Omidi, and Ali Masoudi-Nejad, Structure-based drug repurposing against COVID-19 and emerging infectious diseases: methods, resources and discoveries, *Briefings in Bioinformatics*, 22, 6, 2021, bbab113.

(29) Murtagh F, Contreras P. Algorithms for hierarchical clustering: an overview, *WIREs Data Mining Knowl Discov.* 2017; 7:e1219.

- (30) Altman N, Krzywinski M. Clustering. *Nat Methods*. 14, 2017, 545–546.
- (31) J. M. Word, S. C. Lovell, J. S. Richardson and D. C. Richardson, Asparagine and glutamine: using hydrogen atom contacts in the choice of side-chain amide orientation 1 Edited by J. Thornton, *J. Mol. Biol.*, 1999, 285, 1735–1747.
- (32) W. Humphrey, A. Dalke and K. Schulten, VMD: Visual molecular dynamics, *J. Mol. Graph.*, 1996, 14, 33–38.
- (33) Phillips JC, Braun R, Wang W, Gumbart J, Tajkhorshid E, Villa E, Chipot C, Skeel RD, Kalé L, Schulten K. Scalable molecular dynamics with NAMD. *J Comput. Chem.* 26, 2005, 1781–1802.
- (34) Douguet D., Data sets representative of the Structures and Experimental Properties of FDA-approved Drugs, *ACS Med Chem Lett.*, 2018, 9(3): 204-209.
- (35) Stewart JJP. Optimization of parameters for semiempirical methods VI: more modifications to the NDDO approximations and re-optimization of parameters. *J Mol Model.* 2013; 19:1–32.
- (36) Stewart JJP, MOPAC2016, Stewart Computational Chemistry. Colorado Springs, CO, USA, [HTTP://OpenMOPAC.net](http://OpenMOPAC.net) (2016).
- (37) Bagheri S, Behnejad H, Firouzi R, Karimi-Jafari MH. Using the Semiempirical Quantum Mechanics in Improving the Molecular Docking: A Case Study with CDK2. *Mol Inf.* 2020; 39:2000036.
- (38) Morris GM, Huey R, Lindstrom W, Sanner MF, Belew RK, Goodsell DS, Olson AJ. AutoDock4 and AutoDockTools4: Automated Docking with Selective Receptor Flexibility. *J Comput Chem.* 2009; 30:2785–2791.
- (39) 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.* 2010;31:455–461.
- (40) M.F. Sanner, Python: a programming language for software integration and development, *J. Mol. Graph. Model.* 17 (1999) 57–61.
- (41) Goodford PJ. A computational procedure for determining energetically favorable binding sites on biologically important macromolecules. *J Med Chem.* 1985;28:849–857.
- (42) Gorgulla C, Boeszoermyeni A, Wang Z, Fischer PD, Coote PW, Das KMP, Malets YS, Radchenko DS, Moroz YS, Scott DA, Fackeldey K, Hoffmann M, Iavniuk I, Wagner G, Arthanari H. An open-source drug discovery platform enables ultra-large virtual screens. *Nature*, 2020;580:663–668.
- (43) Craig IR, Essex JW, Spiegel K. Ensemble Docking into Multiple Crystallographically Derived Protein Structures: An Evaluation Based on the Statistical Analysis of Enrichments. *J Chem Inf Model.* 2010;50:511–524.
- (44) Rao S, Sanschagrin PC, Greenwood JR, Repasky MP, Sherman W, Farid R. Improving database enrichment through ensemble docking. *J Comput Aided Mol Des.* 2008; 22:621–627.
- (45) Kitchen DB, Decornez H, Furr JR, Bajorath J. Docking and scoring in virtual screening for drug discovery: methods and applications. *Nat. Rev. Drug Discov.* 2004;3:935–949.
- (46) Perola E, Walters WP, Charifson PS. A Detailed Comparison of Current Docking and Scoring Methods on Systems of Pharmaceutical Relevance. *Proteins.* 2004;56:235–249.

- (47) Rohoullah Firouzi, Mitra Ashouri, Identification of potential anti-COVID-19 drug leads from Medicinal Plants through Virtual High-Throughput Screening. [10.26434/chemrxiv-2022-zsk04](https://doi.org/10.26434/chemrxiv-2022-zsk04)
- (48) Allen WJ, Rizzo RC. Implementation of the Hungarian Algorithm to Account for Ligand Symmetry and Similarity in Structure-Based Design. *J Chem Inf Model.* 2014; 54:518–529. <https://doi.org/10.1021/ci400534h>
- (49) Ballante F, Marshall GR, An Automated Strategy for Binding-Pose Selection and Docking Assessment in Structure-Based Drug Design. *J Chem Inf Model.* 2016; 56:54–72. <https://doi.org/10.1021/acs.jcim.5b00603>
- (50) Barril X, Morley SD. Unveiling the Full Potential of Flexible Receptor Docking Using Multiple Crystallographic Structures. *J Med Chem.* 2005;48:4432–4443.
- (51) Cavasotto CN, Abagyan RA. Protein Flexibility in Ligand Docking and Virtual Screening to Protein Kinases. *J Mol Biol.* 2004; 337:209–225.
- (52) Canrong Wu, Yang Liu, Yueying Yang, Peng Zhang, Wu Zhong, Yali Wang, Qiqi Wang, Yang Xu, Mingxue Li, Xingzhou Lic, Mengzhu Zheng, Lixia Chen, Hua Lia, Analysis of therapeutic targets for SARS-CoV-2 and discovery of potential drugs by computational methods, *Acta Pharmaceutica Sinica B* 2020;10 (5):766e788.
- (53) Simona De Vita, Maria Giovanna Chini, Gianluigi Lauro and Giuseppe Bifulco, Accelerating the repurposing of FDA-approved drugs against coronavirus disease-19 (COVID-19), *RSC Adv.*, 2020, 10, 40867
- (54) Haihai Jiang, Peiyao Yang and Jin Zhang, Potential Inhibitors Targeting Papain-Like Protease of SARS-CoV-2: Two Birds With One Stone *Frontiers in Chemistry* February 2022, Volume 10, Article 822785
- (55) Brendan T. Freitas, Ian A. Durie, Jackelyn Murray, Jaron E. Longo, Holden C. Miller, David Crich, Robert Jeff Hogan, Ralph A. Tripp, and Scott D. Pegan, Characterization and Noncovalent Inhibition of the Deubiquitinase and deISGylase Activity of SARS-CoV-2 Papain-Like Protease, *ACS Infect. Dis.* 2020, 6, 8, 2099–2109
- (56) Dale J. Calleja, Guillaume Lessene and David Komander, Inhibitors of SARS-CoV-2 PLpro, *Frontiers in Chemistry*, April 2022, 10, Article 876212
- (57) A.C. Wallace, R.A. Laskowski, J.M. Thornton, LIGPLOT: a program to generate schematic diagrams of protein-ligand interactions, *Protein Eng. Des. Sel.* 8 (1995) 127–134.
- (58) Haozhou Tan, Yanmei Hu, Prakash Jadhav, Bin Tan, and Jun Wang, Progress and Challenges in Targeting the SARS-CoV-2 Papain-like Protease, *J Med Chem.* 2022 Jun 9;65(11):7561-7580.