molecules

MDPI

Article

Prioritisation of Compounds for 3CL^{pro} Inhibitor Development on SARS-CoV-2 Variants

Marko Jukič [1,2], Blaž Škrlj [3], Gašper Tomšič [4], Sebastian Pleško [5], Črtomir Podlipnik [6,*] and Urban Bren [1,2,*]

1 Laboratory of Physical Chemistry and Chemical Thermodynamics, Faculty of Chemistry and Chemical Engineering, University of Maribor, Smetanova ulica 17, SI-2000 Maribor, Slovenia; marko.jukic@um.si
2 Faculty of Mathematics, Natural Sciences and Information Technologies, University of Primorska, Glagoljaška 8, SI-6000 Koper, Slovenia
3 Institute Jožef Stefan, Jamova cesta 39, SI-1000 Ljubljana, Slovenia; blaz.skrlj@ijs.si
4 Independent Researcher, Cesta Cirila Kosmača 66, SI-1000 Ljubljana, Slovenia; gasper.tomsic@icloud.com
5 Erudio, Litostrojska Cesta 40, SI-1000 Ljubljana, Slovenia; sebastian.plesko@erudio.si
6 Faculty of Chemistry and Chemical Technology, University of Ljubljana, Večna pot 113, SI-1000 Ljubljana, Slovenia
* Correspondence: crtomir.podlipnik@fkkt.uni-lj.si (Č.P.); urban.bren@um.si (U.B.); Tel.: +386-41-440-198 (Č.P.); +386-2-22-94-421 (U.B.)

Abstract: COVID-19 represents a new potentially life-threatening illness caused by severe acute respiratory syndrome coronavirus 2 or SARS-CoV-2 pathogen. In 2021, new variants of the virus with multiple key mutations have emerged, such as B.1.1.7, B.1.351, P.1 and B.1.617, and are threatening to render available vaccines or potential drugs ineffective. In this regard, we highlight 3CL^{pro}, the main viral protease, as a valuable therapeutic target that possesses no mutations in the described pandemically relevant variants. 3CL^{pro} could therefore provide trans-variant effectiveness that is supported by structural studies and possesses readily available biological evaluation experiments. With this in mind, we performed a high throughput virtual screening experiment using CmDock and the "*In-Stock*" chemical library to prepare prioritisation lists of compounds for further studies. We coupled the virtual screening experiment to a machine learning-supported classification and activity regression study to bring maximal enrichment and available structural data on known 3CL^{pro} inhibitors to the prepared focused libraries. All virtual screening hits are classified according to 3CL^{pro} inhibitor, viral cysteine protease or remaining chemical space based on the calculated set of 208 chemical descriptors. Last but not least, we analysed if the current set of 3CL^{pro} inhibitors could be used in activity prediction and observed that the field of 3CL^{pro} inhibitors is drastically underrepresented compared to the chemical space of viral cysteine protease inhibitors. We postulate that this methodology of 3CL^{pro} inhibitor library preparation and compound prioritisation far surpass the selection of compounds from available commercial "corona focused libraries".

Keywords: COVID-19; SARS-CoV-2; M^{pro}; 3CL^{pro}; 3C-like protease; high-throughput; virtual screening; inhibitors; in silico drug design; chemical library design; machine learning; compound prioritisation

Citation: Jukič, M.; Škrlj, B.; Tomšič, G.; Pleško, S.; Podlipnik, Č.; Bren, U. Prioritisation of Compounds for 3CL^{pro} Inhibitor Development on SARS-CoV-2 Variants. *Molecules* **2021**, *26*, 3003. https://doi.org/10.3390/molecules26103003

Academic Editor: Elena Cichero

Received: 23 March 2021
Accepted: 14 May 2021
Published: 18 May 2021

1. Introduction

Coronavirus disease (COVID-19) is an infectious disease caused by a novel severe acute respiratory syndrome coronavirus 2 or SARS-CoV-2. COVID-19 was initially reported in Wuhan province in China and was declared as a global pandemic [1]. COVID-19 is a severe illness similar to flu, with major symptoms being cough, fever and breathing difficulty. Furthermore, the illness can cause systemic inflammation [2,3]. The pathogen SARS-CoV-2 belongs to the *Coronaviridae* family, an enveloped positive-sense single-stranded (+ssRNA) RNA virus, and is closely related to the previously described SARS-CoV and MERS-CoV coronaviruses [4]. The SARS-CoV-2 genome shares 82% sequence identity with SARS-CoV and 90% identity with MERS-CoV and shares common pathogenesis mechanisms [5].

Currently, there are registered vaccines available to fight this global crisis, and multiple vaccine development programs are underway [6–9]. However, there are only a handful of therapeutic options for COVID-19 treatment and no registered antiviral drugs against SARS-CoV-2 at the moment [10–14]. Furthermore, research suggests a minimal variation in the genome sequence of SARS-CoV-2 pathogen may translate to changes in the structures of viral proteins rendering available vaccines or even medicines ineffective [15]. In late 2020, early 2021, the emergence of the new SARS-CoV-2 variants was reported; namely the B.1.1.7 variant, dubbed the UK variant, the B.1.351 variant or South African variant and B.1.617, known as the Indian variant [16–18]. Both variants are reported to possess N501Y mutation in the RBD (receptor binding domain) of the Sprot (spike protein) that is associated with increased viral transmission [19]. The South African variant also possesses K417N and E484K mutations in the Sprot that are potentially responsible for the diminished binding of viral Sprot to host antibodies [20]. In Brazil, the P.1 variant with known N501Y, E484K and novel K417T mutation at the Sprot was identified [21]. A SARS-CoV-2 variant summary is presented in Table 1.

Table 1. Summary of dominant SARS-CoV-2 variants and relevant mutations.

Variant [1]	Alternative Name	Sprot/ All Mutations	Key Mutations	Comment	$3CL^{pro}/PL^{pro}$ Mutations [2]
B.1.1.7	UK Variant	8/23	E69/70 del 144Y del N501Y (RBD interface) A570D P681H	higher transmissibility	none/A1708D
B.1.351	South African Variant	9/21	K417N (RBD) E484K (RBD) N501Y (RBD) orf1b del	escape host immune response	none/K1655N
P.1	Brasil Variant	10/17	K417N/T (RBD) E484K (RBD) N501Y (RBD) orf1b del	under research	none/K1795Q
B.1.617	Indian Variant	7/23	G142D E154K L452R (RBD) E484Q (RBD) D614G P681R Q1071H	under research	none/under research

[1] Other known variants are COH.20G, S Q677H (Midwest variant) and L452R, B1429; [2] The mutations on PL^{pro} are located far outside the enzyme's active site.

It is of key interest that no mutations have been observed on SARS-CoV-2 $3CL^{pro}$ main protease (location 3292 → 3582 on ORF1ab polyprotein; YP_009724389.1) and only one mutation on SARS-CoV-2 PL^{pro} protease (location 1564 → 1882 on ORF1ab polyprotein; YP_009724389.1) for each variant [22]. It should be stressed, however, that this does not mean that these proteins cannot mutate in the future. In this context, $3CL^{pro}$ is an attractive target for novel antiviral discovery with a potential for trans-variant activity [23–26]

$3CL^{pro}$ (Picornain 3C-like protease, also referred to as M^{pro} for main protease) is a homodimeric cysteine protease (EC 3.4.22.69) and is 96% sequence identical with the SARS-CoV M^{pro} [27]. The enzyme belongs to the family C30 peptidases in the PA peptidases clan. It consists of two 306 residues long polypeptide chains, which fold into three domains (I, II and III). Domains I and II have an antiparallel β-barrel structure, while domain III is composed of 5 α-helices, which connect to domain II by a long loop region. Judging by its dimer interface, it seems the dimer (comprised of protomers A and B) is an active form which is considerably less efficient when isolated in its monomer form. The active

site is readily accessible to the solvent and is located distal to the dimer interface [28]. The substrate binding site is comprised of pockets P1, P1', P2 and P3. The P1 subsite is formed with Phe140, Asn142, Glu166, His163 and His172 residues (Figure 1) and two conserved water molecules. P2 is a deep pocket comprised of His41, Met49, Tyr54, Met165 and Asp187 residues while P3 is defined by Leu168 and flanked by Glu166, Pro168 and Gly170 [29]. Proteolysis occurs via a catalytic dyad defined by Cys145 and His41 [30]. The enzyme is responsible for cleavage on no less than 11 sites on the large viral polyprotein 1ab. Cleavage generally follows the pattern Leu/Phe/Met-Gln ↓ Gly/Ser/Ala (↓ denotes the cleavage site). Glutamine at the P1 position is crucial for proteolysis to occur. As there are no known native human enzymes with such cleavage sites, M^{pro} looks to be an ideal drug target, since there is a low risk for toxic effects on host cells [31–33]. Structural data on the enzyme is available and reporter assays developed making this target suitable for novel antiviral design [34,35].

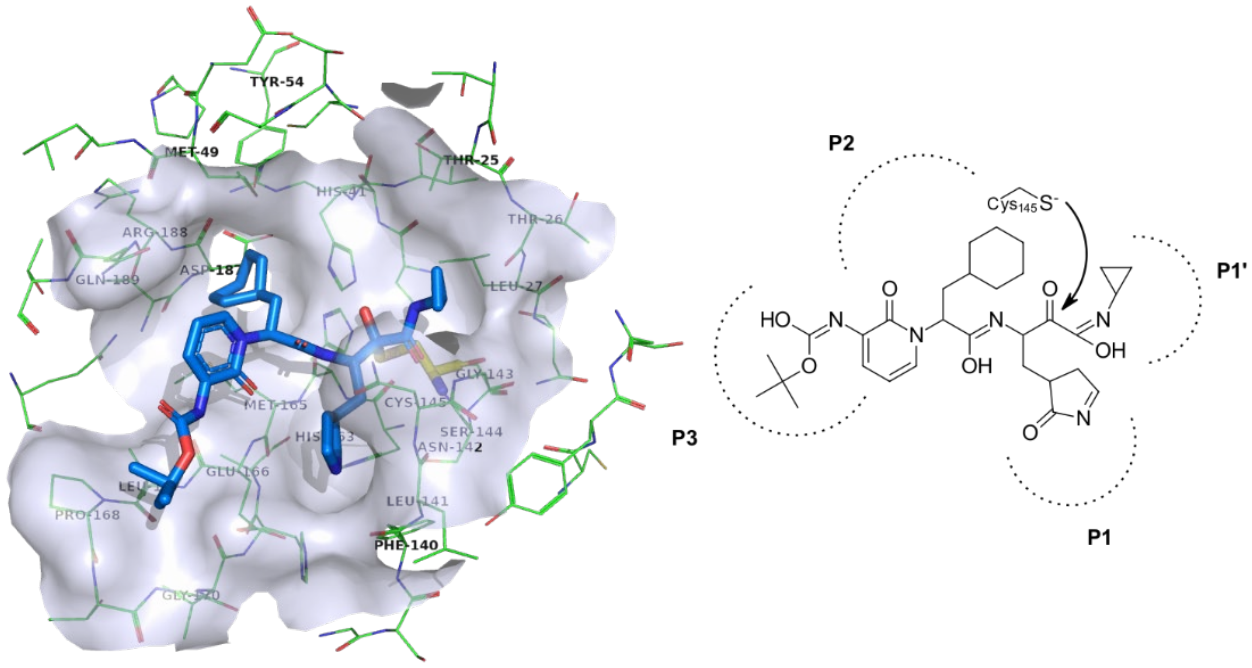


Figure 1. Active site of SARS-CoV-2 $3Cl^{pro}$ or M^{pro} enzyme (PDB ID: 6Y7M) with emphasised small molecule and pocket designation on the right. Active site residues are depicted in green coloured line model with emphasized transparent blue-white surface 6 Å around the small molecule inhibitor depicted in blue coloured stick model.

In this study, we perform a high-throughput virtual screening (HTVS) experiment coupled with machine learning classification to offer a prioritisation approach for compounds with potential activity on SARS-CoV-2 $3CL^{pro}$. We employ fast methodologies to cover a comprehensive chemical space based on molecular docking scores in order to offer prioritisation lists of compounds for further free energy calculations and suitable for biological evaluation [36]. We focus on identifying novel potential non-covalent protease inhibitors, as we firmly believe they offer the flexibility of optimisation and synthetic or commercial availability [37,38].

2. Results and Discussion

2.1. Database Preparation

For our HTVS (high-throughput virtual screening) experiment, we employed the ZINC 15 library [39]. To produce robust results that would enable downstream experimental support and biological evaluation, we specifically selected the "*In-Stock*" library subset that includes commercially readily-available compounds. The subset was trenched to exclude small fragments below the molecular weight of 200 g/mol and included 9,232,022

compounds in total (Figure 2). The next step was compound ionisation at the pH of 7.4, the calculation of initial 3D conformations for the whole database, the enumeration of undefined chiral centres, and the removal of structural faults. Smiles strings were syntactically validated (all rings/branches closed, no illegal atom types) with RDKit software (MolFromSmiles). Ionisation was performed using OpenEye QUACPAC 2.1.1.0 software (OpenEye Scientific Software, Inc., Santa Fe, NM, USA; www.eyesopen.com; accessed on 8 May 2021) with FixpKa module and ionise 7.4 parameters. The initial 3D conformation was calculated with OpenEye OMEGA 2.5.1.4 software (OpenEye Scientific Software, Inc., Santa Fe, NM, USA; www.eyesopen.com). The following parameters were used: maxconfs 1 and flipper: true [40].

Molecular Weight (up to, Daltons)

LogP (up to)	200	250	300	325	350	375	400	425	450	500	>500	Totals, by LogP
-1	6,983	4,872	6,160	3,072	3,195	2,316	1,605	1,227	1,047	1,818	5,512	37,807
0	25,064	14,280	20,924	12,273	12,788	9,576	5,675	3,104	2,636	2,174	3,659	112,153
1	68,295	57,603	89,093	55,185	58,890	58,207	25,604	13,457	7,984	11,812	7,080	453,210
2	95,368	135,194	219,877	149,361	177,006	141,881	90,155	51,537	34,593	28,715	16,417	1,140,104
2.5	35,264	80,162	154,491	115,080	147,089	125,312	87,880	57,871	40,950	55,901	16,682	916,682
3	20,563	68,510	155,591	123,995	172,255	155,063	120,576	86,103	64,255	62,238	27,773	1,056,922
3.5	8,919	44,795	127,299	112,524	159,869	162,782	141,042	111,698	89,637	92,146	47,962	1,098,673
4	2,757	22,199	85,077	83,773	125,581	143,906	140,363	126,445	149,236	122,298	75,666	1,077,301
4.5	575	8,077	45,618	54,011	85,000	107,846	120,572	121,883	153,755	141,350	108,908	947,595
5	94	2,293	18,611	28,113	48,883	69,054	85,673	95,141	134,181	144,525	136,747	763,315
>5	28	793	8,656	16,434	35,453	64,780	100,483	127,892	199,811	340,649	733,281	1,628,260
Totals, by Weight	263,910	438,778	931,397	753,821	1,026,009	1,040,723	919,628	796,358	878,085	1,003,626	1,179,687	**9,232,022** Substances / **960** Tranches

Figure 2. Detailed ZINC 15 database "*In-Stock*" subset tranche description with a total of 9,322,002 compounds used for further calculations (As obtained from the https://zinc15.docking.org; accessed on 8 May 2021; at the time of smiles compound download; the database on the master server is continuously updating).

2.2. Binding Site Identification

Upon examination of available SARS-CoV-2 3CL^{pro} crystal structures at the time, we selected the complex (PDB ID: 6Y7M) with an excellent resolution of 1.9 Å deposited by Zhang L. et al. [41]. This complex contained a relatively large peptidomimetic inhibitor OEW (MW = 585.69 g/mol) occupying major pockets at the enzyme active site, essentially keeping the S1 pocket in the correct shape and the enzyme in the active conformation. After superimposition to a reference structure (PDB ID:6LU7) by Yang et al., the catalytic binding pocket was defined around Cys145 as published previously by Jukic et al. [29,42]. Finally, the target was prepared as a sphere of 7 Å around the ligand for docking volume calculation using the CmDock docking package (https://gitlab.com/Jukic/cmdock/; accessed on 8 May 2021; Figure 3; details are in Supplementary Materials).

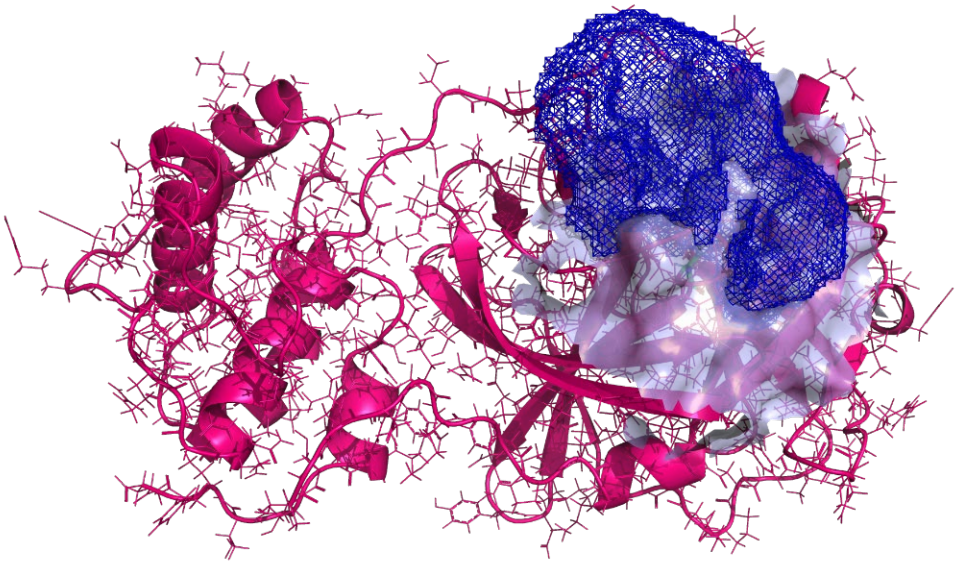


Figure 3. Generated receptor volume with the $3CL^{pro}$ PDB ID: 6Y7M, the active site near the Cys145 residue and the docking volume defined as a sphere of 7 Å around the reference ligand OEW. Protein is depicted as a pink-magenta-coloured cartoon model with residues emphasised in the line model and coloured atoms (carbon in green, oxygen in red, nitrogen in blue and hydrogen in white colour) with a blue-white transparent active site surface. Docking volume boundary mesh is depicted in blue colour. Isomesh (0.99) was constructed using PyMol 2.1.0 using a grid calculated with cmgrid software (v 0.1.1).

2.3. HTVS

For the HTVS campaign, the complete pre-prepared database was docked using CmDock into the prepared receptor binding site to afford 1‰ of top-scoring compounds with the Z-score cutoff of –7.7 (Figure 4).

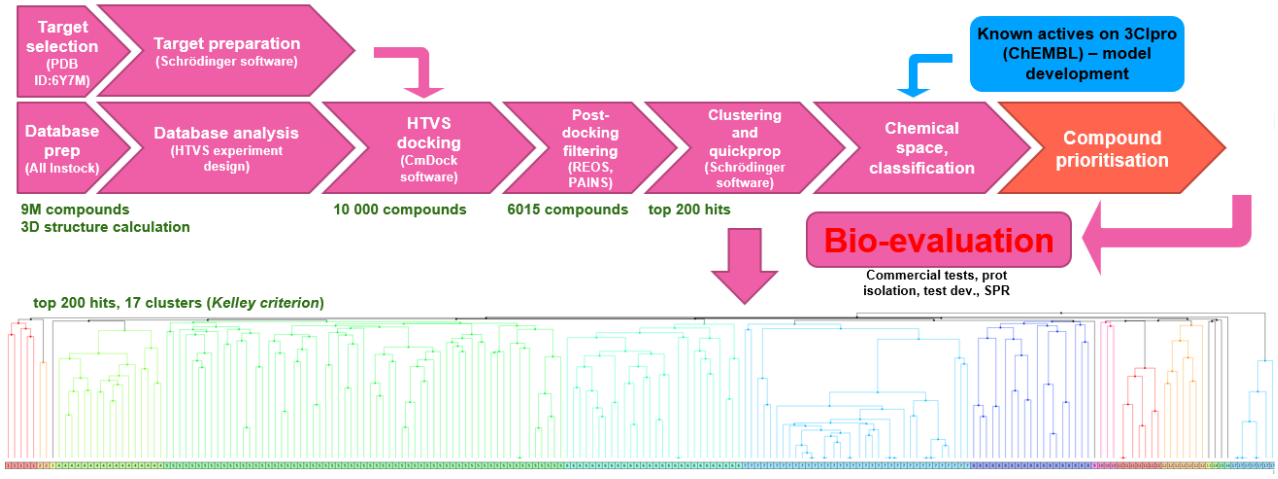


Figure 4. HTVS workflow with post-docking filtering, cluster analysis, and compound classification according to the chemical space of $3CL^{pro}$ inhibitors collected in the ChEMBL database.

In order to inspect all chemical space, filtering was done post-docking, where subsequently PAINS [43–45] and REOS structures were filtered out [46,47]. KNIME software with RDKit nodes was applied to compare all structures in the library to the selection of SMARTS-formatted PAINS and to remove flagged hits from the database followed by REOS filtering with Schrödinger SMD software (Schrödinger LLC., New York, NY, USA). In this way, approximately 40% of the pan-assay interfering and reactive compounds (with

labile functional groups) were filtered out, and the top 200 scoring compounds were examined by cluster analysis and classified in the current 3CL^{pro}-actives space. Hierarchical clustering was performed using Schrödinger SMD (Schrödinger LLC., New York, NY, USA) using Molprint2D hashed fingerprints, Tanimoto similarity metric and average cluster linkage method; the 17 clusters have been estimated by the Kelley criterion [48]. The final compound selection was performed with top-scoring hits with consideration of QuickProp QPlogS (Schrödinger LLC., New York, NY, USA) descriptor (>−6.5) in order to focus on compounds that have a greater potential to be soluble. Complete QuickProp descriptor set was calculated to flag all un-/desired properties and enable further custom prioritisation of compounds, and the whole set of top 200 hits is supplied in the Supplementary Materials for the convenience of future research. Upon examination of all top-scoring compounds, we identified that they all conform to the classic P1-P2-P3 binding pockets, as described previously [42]. Predicted bound conformations for the top-scoring hit as well as for the first ten hit compounds are analogous (and in accordance to the crystal ligand OEW), and compounds interact with key Thr25, Leu27, Gly143, Ser144, Cys145, His163, His164, Met165, Glu166, Asp187, Thr190, Gln189 and Gln192 residues at the 3CL^{pro} active site (Table 2, Figure 5).

Table 2. Identified top-scoring compounds in the HTVS on the SARS-CoV-2 main protease for further compound prioritisation in biological evaluation experiments.

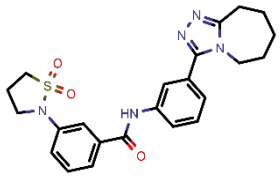
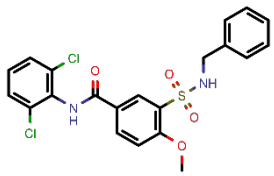
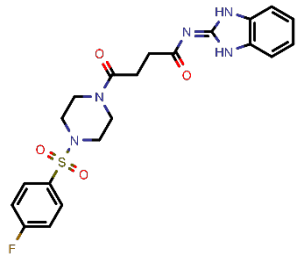
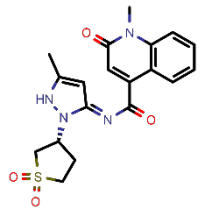
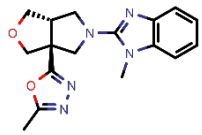
no	Structure	Mr (g/mol)	Cluster/QPlogS [1]	CmDock Docking Score [2]	Classification [3]
1		451.54	5/−6.42	−32.51	*general*
2		465.35	5/−6.22	−29.02	*general*
3		459.49	5/−3.89	−26.80	*viral_cys_prot*
4		400.45	4/−4.54	−25.58	*viral_cys_prot*
5		325.37	5/−3.59	−25.53	*viral_cys_prot*

Table 2. *Cont.*

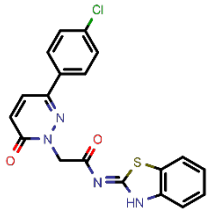
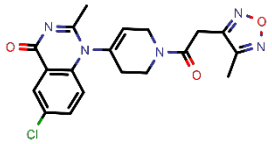
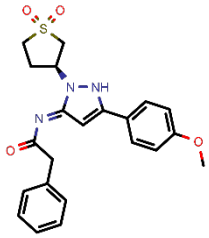
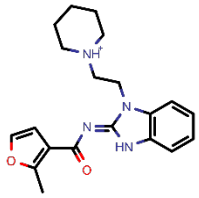
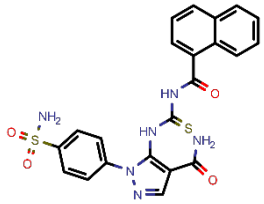
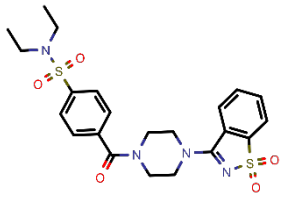
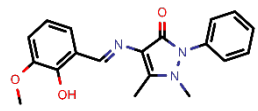
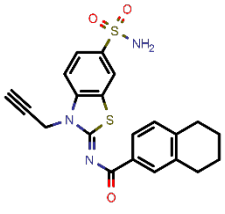
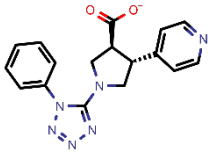
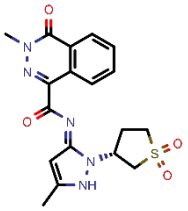
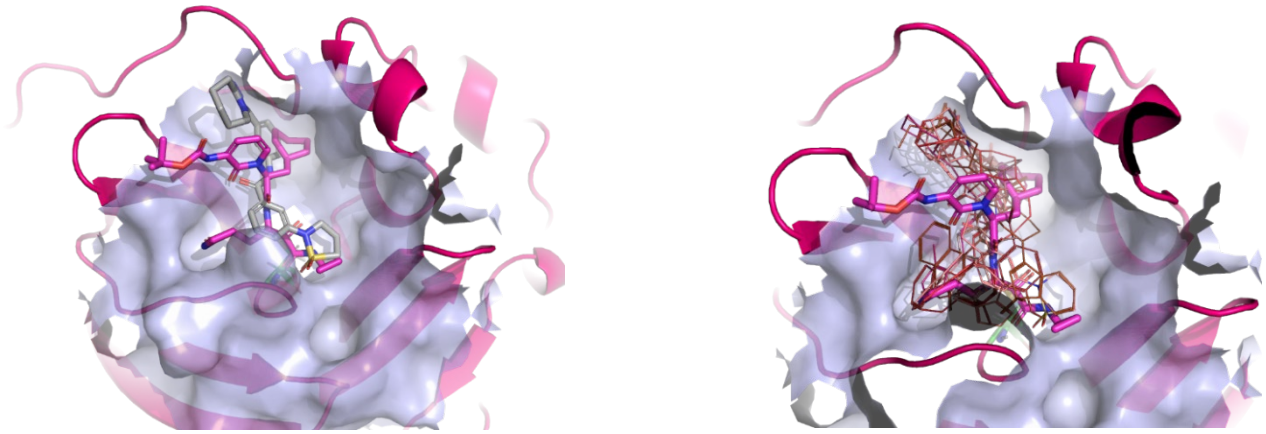
no	Structure	Mr (g/mol)	Cluster/QPlogS [1]	CmDock Docking Score [2]	Classification [3]
6		396.85	5/−6.47	−25.05	*general*
7		399.83	2/−3.44	−24.76	*general*
8		425.50	4/−5.38	−24.51	*general*
9		353.44	5/−4.40	−24.17	*general*
10		494.55	5/−5.60	−24.01	*general*
11		490.60	5/−2.87	−23.98	*general*
12		337.37	5/−4.75	−23.61	*viral_cys_prot*
13		425.52	5/−5.66	−23.53	*general*

Table 2. *Cont.*

no	Structure	Mr (g/mol)	Cluster/QPlogS [1]	CmDock Docking Score [2]	Classification [3]
14		335.34	5/−3.90	−23.26	*general*
15		401.44	4/−5.00	−23.18	*general*

[1] QPlogS Predicted aqueous solubility, log S. S in mol dm^{-3} is the solute concentration in a saturated solution in equilibrium with the crystalline solid (recommended value range by QuickProp is between –6.5 and –0.5); [3] CmDock INTER-molecular docking score.; [3] As per NeuralNet model.



(**A**) (**B**)

Figure 5. Calculated bound conformations in the $3CL^{pro}$ active site of the top-scoring hit compounds. Protein is presented in a cartoon model coloured pink with an emphasised molecular surface in light-blue colour. (**A**); the reference OEW ligand is presented in stick model cored magenta, while the top-scoring hit is coloured grey. (**B**) the reference OEW ligand is presented in stick model cored magenta while the top 10 scoring compounds are depicted in red-coloured line representations to emphasise their analogous binding mode in the P1-P2-P3 pockets of the active site.

2.4. Chemical Space Prediction and Classification Supported by Machine Learning

We also investigated whether state-of-the-art machine learning algorithms could be of use for compound prioritisation. In this work, we implemented four different machine learning algorithms capable of detecting various complexities and evaluated their performance on the training data. The training data for the regression experiment consisted of all compounds with activity on the $3CL^{pro}$ enzyme (ChEMBL standard value IC50 in nM, 133 compounds with 1k decoy set, available in Supplementary Materials) present in the ChEMBL database, while the training data for classification consisted of chemical libraries active on $3CL^{pro}$ as well as actives on all viral cysteine proteases (3145 compounds with 1k decoy set, available in Supplementary Materials) [49]. For all compounds, we used RDKit.Chem package. RDKit.Chem.Descriptors module calculated a set of 208 chemical descriptors in order to capture the respective chemical space. In this manner, we could effectively classify compounds in the context of their representative chemical space: similar to known $3CL^{pro}$ inhibitors, similar to known viral cysteine protease inhibitors, or belonging to a different chemical space to help in future design directions for individual scaffolds.

The learning algorithms were (from simplest to more complex ones): a majority classifier (prediction of the most common class), a logistic regression classifier (linear classifier), and two non-linear classifiers, namely a deep feedforward neural network, designed specifically for this task, as well as the extreme gradient boosting machines (XGB), which are considered a strong learner in contemporary chemoinformatics [50–53].

3. Materials and Methods

3.1. Target Preparation

Complex (PDB ID: 6Y7M) contains {tert}-butyl ~{N}-[1-[(2~{S})-3-cyclohexyl-1-[[(2~{S},3~{R})-4-(cyclopropylamino)-3-oxidanyl-4-oxidanylidene-1-[(3~{R})-2-oxidanylidene-3,4-dihydropyrrol-3-yl]butan-2-yl]amino]-1-oxidanylidene-propan-2-yl]-2-oxidanylidene-pyridin-3-yl]carbamate (OEW), a peptide-like covalent inhibitor where thiohemiketal is formed by the nucleophilic attack of the catalytic cysteine (Cys145) onto the carbonyl group of the inhibitor. Residue conformation at the active site was checked by superposition with PDB ID: 6M2N and 6XMK with an all atom RMSD of 0.330 and 0.662 Å, respectively [54,55]. For the docking receptor preparation, the covalent bond was cleaved, a small molecule removed, and the Cys145 amino-acid residue regenerated (Open Source PyMOL, release 2.1). The target was prepared using Schrödinger Small-Molecule Discovery Suite (Schrödinger LLC., New York), protein preparation module. Missing hydrogens were added, h-bonding network was optimised using PROPKA tool at pH 7.4, waters were removed, and restrained minimisation was performed with a convergence of heavy atoms towards 0.3 Å. Finally, using the cmcavity program from CmDock docking package (https://gitlab.com/Jukic/cmdock/; accessed on 8 May 2021), we generated a docking receptor [56–58]. The reference ligand method was employed for cavity calculation (*receptor* definition), where we used the OEW-cleaved regenerated ligand as a reference and a sphere of 7 Å around the ligand for the docking volume (cavity volume) calculation. We calculated a total cavity volume of 3106.25 A^3 and included the calculated cavity (Cavity #1) in the *docking receptor* definition. Cavity #1 parameters were the size of 24850 points; min = (−33.5,−53.5,−8.5); max = (−13,−26.5,12); centre = (−24.4138,−38.9632,−0.179235); coordinates and extent = (20.5,27,20.5) Å (Figure 3, details are in Supplementary Materials).

3.2. Virtual Screening Experiment Design

For the virtual screening experiment (HTVS), a docking approach with a robust CmDock software, the prepared compound database, and the docking receptor (Cavity #1), as described in the previous section, were employed [56,57]. Firstly, we conducted a redocking experiment. The reference-regenerated OEW peptidomimetic ligand (PDB ID: 6Y7M) was prepared as a SMILES string and energy-minimised in Ligprep tool from Schrödinger SMD using the OPLS 3e forcefield. The minimised structure was subsequently used as an input for the redocking experiment into the prepared receptor in a non-covalent manner. Applied parameters for the CmDock software (v 0.1.1) were standard docking protocol (dock.prm) with 100 runs, no constraints, and no score filters. We successfully retrieved the crystal-complex binding conformation of the OEW ligand with an RMSD of 1.34 Å. Furthermore, we calculated the receiver operating characteristic (ROC) curve to validate the performance of the classifier docking method. We selected a set of known SARS-CoV $3CL^{pro}$ inhibitors from the ChEMBL database with experimental IC_{50} < 100 µM values and created a testing database by the addition of negative control compounds that were calculated decoys based on employed actives using DUD-E: A Database of Useful (Docking) Decoys [59]. Upon using 1% and 10% of actives in the test database, we obtained a ROC AUC of 0.80 and 0.79, respectively. We also employed the activity data from the PostEra Covid Moonshot project (https://covid.postera.ai/covid/activity_data; accessed on 8 May 2021). We selected compounds with pIC_{50} above 7 as true actives and compounds with pIC_{50} up to 4.00436 as inactives or experimental decoys (compounds with no data were left out). When using 2% and 10% of actives in the test set, we obtained ROC AUCs

of 0.61, indicating that our docking protocol can indeed identify active compounds and produce enriched libraries.

In order to effectively utilise CPU-time in downstream calculations, we analysed the chemical library performance in HTVS. We sampled a random 10% population of the designed library and performed an exhaustive docking experiment on 977,600 molecules (dock.prm protocol, 100 runs per molecule, no constraints, and no score filter). We then analysed the docking results using the sdreport script (part of CmDock package) and KNIME software, where the mean docking score was −12.317, and the standard deviation was 4.273. Upon passthrough analysis using rbhtfinder script (part of CmDock package, parameters: −15 and −20), the HTVS analysis afforded an optimal HTVS experiment workflow where up to 5 runs were performed for molecules that possessed docking score of −18 and up to 15 runs for molecules that were found with docking scores of −25. In the case of docking scores lower than −25, 50 runs were performed. The first filter was passed by 25.7% of molecules, the second filter was passed by 1.71% of molecules, and the run was on average 7.8% CPU-time compared to the exhaustive docking run-time, thereby effectively improving the HTVS efficiency.

3.3. Machine Learning

The deep neural network consisted of six hidden layers, where each layer was activated with the ReLU activation, followed by dropout-based regularisation, set to 0.3 [60]. The final layer of the neural network was activated with a softmax function to obtain a distribution across the space of possible classes. During prediction, the most probable class is selected (argmax across class probabilities). The extreme gradient boosting machines are a type of tree-based ensemble classifier capable of fast learning and robust and accurate predictions. The evaluation regime was designed as follows: we randomly sampled ten different stratified splits, where the training data was used to learn the models, followed by the evaluation of their predictions on the test data. The splits were in ratio 8:2. We report the average performance with a standard deviation for the macro F1 score (the task considered is a multiclass classification). Apart from the classification task, the point was to differentiate between chemical spaces occupied by known 3Clpro inhibitors reported in ChEMBL, known viral cysteine protease inhibitors in the ChEMBL database, and other chemical space. We also conducted a similar series of experiments to assess to what extent machine learning can predict the inhibition of $3CL^{pro}$ directly (on the basis of the ChEMBL standard value IC50 in nM), which is a regression task. The regression variants of the algorithms mentioned above were considered, namely the XGB with the "req:squarederror" loss, the neural network's classification head was replaced with a regression one, and instead of logistic regression, we used a simple linear regressor. The neural network that performed best consisted of seven (regularized) hidden layers; other hyperparameters were the same as in the classification scenario. As the neural network's expected output is a positive real number, the final activation used was a ReLU as well, as we observed faster convergence compared to using no activation at all. Mean squared errors with standard deviation across ten random splits (see the previous section) are reported (Table 3).

Table 3. Machine learning model accuracy comparison table.

Macro F1/mse [1]	NeuralNet [2]	XGB [2]	Linear [2]	Majority/Average [3]
Classification	0.895 ± 0.05	0.889 ± 0.014	0.667 ± 0.015	0.283 ± 0.001
Regression	0.002 ± 0.001	0.003 ±0.001	0.012 ± 0.002	0.005 ± 0.001

[1] macro F1 for classification and mse for regression; [2] Learner; [3] majority class classifier in classification and average of the training target space in a regression model.

The neural network and XGB performed, as expected, the best, followed by a simpler linear learner. Compared to the majority baseline, e.g., the neural network's performance is substantially higher, indicating that it learned to differentiate between the classes. On the contrary, the average-training baseline (averaged train predictions) performed better than

the linear classifier, indicating that the regression problem at hand is relatively hard and potentially not linearly separable. The XGB and neural networks performed better than this baseline, albeit not by a large margin, indicating the actives $3CL^{pro}$ dataset was insufficient for effective training. However, the regression task similarly indicates that it is possible to use non-linear models for direct activity prediction necessitating good data collection on $3CL^{pro}$ actives in the future. Furthermore, the authors are aware of the generalisation of the idea on other well-represented datasets and postulate its great value in future in silico drug design.

4. Conclusions

Herein, we presented a novel in silico approach towards compound prioritisation in the design of novel $3CL^{pro}$ inhibitors. Rather than relying on commercial "corona-focused libraries", we performed a robust and efficient HTVS experiment to obtain virtual hit compounds. We then coupled the method to a machine learning classification experiment where each compound is classified into the chemical space of $3CL^{pro}$ inhibitors, into general viral cysteine protease inhibitors, or into a completely novel unrelated chemical space. In this way, medicinal chemists can turn their attention towards compounds that would otherwise escape and derive insight into the possible mechanism of action. Therefore, we employed a complete library of $3CL^{pro}$ inhibitors and viral cysteine protease inhibitors obtained from the ChEMBL library using the calculated ensemble of 208 descriptors for each compound in the library. We reported ten top-scoring compounds as viable binders supported by CmDock docking calculations considering the QuickProp QPlogS descriptor that indicates possible soluble compounds. Namely, solubility forms one of the serious problems when transitioning from in silico enrichment lists towards physical samples for in vitro evaluation. We supply full lists of hit compounds in the Supplementary Materials for the reader's benefit, especially if alternative compound selection criteria are desired, all with the aim of providing the medicinal chemistry community with a viable prioritisation library of potential $3CL^{pro}$ inhibitors tailored for further molecular dynamics (MD) studies [61] as well as experimental design and development. Namely, reported enriched lists of compounds can serve as starting points, whereas compounds can be purchased commercially or be a subject of a synthetic campaign. Compounds can thus serve as experimental decoys or, if successful, progress as leads or probes.

Supplementary Materials: Details on target preparation and virtual screening. References [62] and [63] are cited in the supplementary materials.

Author Contributions: Conceptualization, M.J., Č.P. and U.B.; data curation, M.J. and Č.P.; formal analysis, M.J., B.Š. and Č.P.; funding acquisition, Č.P. and U.B.; investigation, M.J., B.Š., G.T., S.P., Č.P. and U.B.; methodology, M.J. and B.Š.; project administration, M.J., U.B. and Č.P.; supervision, Č.P. and U.B.; validation, M.J., Č.P. and U.B.; writing—original draft, M.J.; writing—review and editing, all authors. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the Slovenian Ministry of Science and Education infrastructure project grant HPC-RIVR, by the Slovenian Research Agency (ARRS) programme and project grants P2-0046 and J1-2471, the Physical Chemistry programme grant P1-0201, the Knowledge Technologies programme group (P2-0103 and P6-0411) as well as Slovenian Ministry of Education, Science and Sports programme grant OP20.04342. The work of the second author was funded via the ARRS junior research grant.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data is contained within the article or supplementary materials.

Acknowledgments: We gratefully acknowledge NVIDIA Corporation's support with the donation of GPU hardware used in this research. We thank OpenEye for the academic licencing of their software and their support. We thank Rita Podzuna and support from Schrödinger LLC. Thanks to Boštjan Laba for support in the community project. Last but not least: a heartfelt thanks from all authors to all participants in the COVID.SI community (www.covid.si and www.sidock.si) for supporting our work. This represents a wonderful example of citizen science, and we are grateful for all joint scientific and community work. Thank You All!

Conflicts of Interest: The authors declare no conflict of interest.

Sample Availability: Not available (compounds are commercially available).

Abbreviations

$3CL^{pro}$	3C-like protease
MD	Molecular Dynamics
VS	Virtual Screening
HTVS	High-Throughput Virtual Screening
LIE	Linear Interaction Energy
Xgb	Gradient boosting framework library

References

1. Huang, C.; Wang, Y.; Li, X.; Ren, L.; Zhao, J.; Hu, Y.; Cao, B. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *Lancet* **2020**, *395*, 497–506. [CrossRef]
2. Wang, D.; Hu, B.; Hu, C.; Zhu, F.; Liu, X.; Zhang, J.; Peng, Z. Clinical characteristics of 138 hospitalized patients with 2019 novel coronavirus–infected pneumonia in Wuhan, China. *JAMA* **2020**, *323*, 1061–1069. [CrossRef]
3. Wang, C.; Horby, P.W.; Hayden, F.G.; Gao, G.F. A novel coronavirus outbreak of global health concern. *Lancet* **2020**, *395*, 470–473. [CrossRef]
4. Chen, Y.; Liu, Q.; Guo, D. Emerging coronaviruses: Genome structure, replication, and pathogenesis. *J. Med. Virol.* **2020**, *92*, 418–423. [CrossRef]
5. Lu, R.; Zhao, X.; Li, J.; Niu, P.; Yang, B.; Wu, H.; Tan, W. Genomic characterisation and epidemiology of 2019 novel coronavirus: Implications for virus origins and receptor binding. *Lancet* **2020**, *395*, 565–574. [CrossRef]
6. Widge, A.T.; Rouphael, N.G.; Jackson, L.A.; Anderson, E.J.; Roberts, P.C.; Makhene, M.; Beigel, J.H. Durability of responses after SARS-CoV-2 mRNA-1273 vaccination. *N. Engl. J. Med.* **2021**, *384*, 80–82. [CrossRef] [PubMed]
7. Amanat, F.; Krammer, F. SARS-CoV-2 vaccines: Status report. *Immunity* **2020**, *52*, 583–589. [CrossRef] [PubMed]
8. Krammer, F. SARS-CoV-2 vaccines in development. *Nature* **2020**, *586*, 516–527. [CrossRef] [PubMed]
9. Chen, W.H.; Strych, U.; Hotez, P.J.; Bottazzi, M.E. The SARS-CoV-2 vaccine pipeline: An overview. *Curr. Trop. Med. Rep.* **2020**, *7*, 61–64. [CrossRef] [PubMed]
10. Wu, C.; Liu, Y.; Yang, Y.; Zhang, P.; Zhong, W.; Wang, Y.; Li, H. Analysis of therapeutic targets for SARS-CoV-2 and discovery of potential drugs by computational methods. *Acta Pharm. Sin. B* **2020**, *10*, 766–788. [CrossRef] [PubMed]
11. McKee, D.L.; Sternberg, A.; Stange, U.; Laufer, S.; Naujokat, C. Candidate drugs against SARS-CoV-2 and COVID-19. *Pharmacol. Res.* **2020**, *157*, 104859. [CrossRef] [PubMed]
12. Li, H.; Zhou, Y.; Zhang, M.; Wang, H.; Zhao, Q.; Liu, J. Updated approaches against SARS-CoV-2. *Antimicrob. Agents Chemother.* **2020**, *64*, e00483-20. [CrossRef] [PubMed]
13. Gordon, D.E.; Jang, G.M.; Bouhaddou, M.; Xu, J.; Obernier, K.; White, K.M.; Krogan, N.J. A SARS-CoV-2 protein interaction map reveals targets for drug repurposing. *Nature* **2020**, *583*, 459–468. [CrossRef] [PubMed]
14. Riva, L.; Yuan, S.; Yin, X.; Martin-Sancho, L.; Matsunaga, N.; Pache, L.; Chanda, S.K. Discovery of SARS-CoV-2 antiviral drugs through large-scale compound repurposing. *Nature* **2020**, *586*, 113–119. [CrossRef]
15. Naqvi, A.A.T.; Fatima, K.; Mohammad, T.; Fatima, U.; Singh, I.K.; Singh, A.; Hassan, M.I. Insights into SARS-CoV-2 genome, structure, evolution, pathogenesis and therapies: Structural genomics approach. *Biochim. Biophys. Acta (BBA) Mol. Basis Dis.* **2020**, *10*, 165878. [CrossRef]
16. Volz, E.; Swapnil, M.; Meera, C.; Barrett, J.C.; Johnson, R.; Geidelberg, L.; Hinsley, W.R.; Laydon, D.J.; Dabrera, G.; O'Toole, Á.; et al. Transmission of SARS-CoV-2 Lineage, B. 1.1. 7 in England: Insights from linking epidemiological and genetic data. *MedRxiv* **2021**, 2020-12. [CrossRef]
17. Tegally, H.; Wilkinson, E.; Giovanetti, M.; Iranzadeh, A.; Fonseca, V.; Giandhari, J.; Doolabh, D.; Pillay, S.; San, E.J.; Msomi, N.; et al. Emergence and rapid spread of a new severe acute respiratory syndrome-related coronavirus 2 (SARS-CoV-2) lineage with multiple spike mutations in South Africa. *MedRxiv* **2020**. [CrossRef]
18. Gupta, R.K. Will SARS-CoV-2 variants of concern affect the promise of vaccines? *Nat. Rev. Immunol.* **2021**, 1–2. [CrossRef]
19. Hongjing, G.; Chen, Q.; Yang, G.; He, L.; Fan, H.; Deng, Y.-Q.; Wang, Y.; Teng, Y.; Zhao, Z.; Cui, Y.; et al. Adaptation of SARS-CoV-2 in BALB/c mice for testing vaccine efficacy. *Science* **2020**, *369*, 1603–1607.

20. Wibmer, C.K.; Ayres, F.; Hermanus, T.; Madzivhandila, M.; Kgagudi, P.; Lambson, B.E.; Vermeulen, M.; van den Berg, K.; Rossouw, T.; Boswell, M.; et al. SARS-CoV-2 501Y. V2 escapes neutralization by South African COVID-19 donor plasma. *BioRxiv* **2021**, *27*, 622–625.
21. Nuno, R.F.; Morales Claro, I.; Candido, D.; Franco, L.A.M.; Andrade, P.S.; Coletti, T.M.; Silva, C.A.M.; Sales, F.C.; Manuli, E.R.; Aguiar, R.S. Genomic characterisation of an emergent SARS-CoV-2 lineage in Manaus: Preliminary findings. *Virological* **2021**.
22. Klemm, T.; Ebert, G.; Calleja, D.J.; Allison, C.C.; Richardson, L.W.; Bernardini, J.P.; Lu, B.G.; Kuchel, N.W.; Grohmann, C.; Shibata, Y.; et al. Mechanism and inhibition of SARS-CoV-2 PLpro. *Biorxiv* **2020**. [CrossRef]
23. Sumit, K.; Sharma, P.P.; Shankar, U.; Kumar, D.; Joshi, S.K.; Pena, L.; Durvasula, R.; Kumar, A.; Kempaiah, P.; Rathi, P.; et al. Discovery of new hydroxyethylamine analogs against 3CLpro protein target of SARS-CoV-2: Molecular docking, molecular dynamics simulation, and structure–activity relationship studies. *J. Chem. Inf. Model.* **2020**, *60*, 5754–5770.
24. ul Qamar, M.T.; Alqahtani, S.M.; Alamri, M.A.; Chen, L.L. Structural basis of SARS-CoV-2 3CLpro and anti-COVID-19 drug discovery from medicinal plants. *J. Pharm. Anal.* **2020**, *10*, 313–319. [CrossRef] [PubMed]
25. Koulgi, S.; Jani, V.; Uppuladinne, M.; Sonavane, U.; Nath, A.K.; Darbari, H.; Joshi, R. Drug repurposing studies targeting SARS-CoV-2: An ensemble docking approach on drug target 3C-like protease (3CLpro). *J. Biomol. Struct. Dyn.* **2020**, 1–21. [CrossRef]
26. Macchiagodena, M.; Pagliai, M.; Procacci, P. Inhibition of the main protease 3cl-pro of the coronavirus disease 19 via structure-based ligand design and molecular modeling. *arXiv* **2020**, arXiv:2002.09937.
27. Malcolm, B.A. The picornaviral 3C proteinases: Cysteine nucleophiles in serine proteinase folds. *Prot. Sci.* **1995**, *4*, 1439–1445. [CrossRef]
28. Shi, J.; Sivaraman, J.; Song, J. Mechanism for Controlling the Dimer-Monomer Switch and Coupling Dimerization to Catalysis of the Severe Acute Respiratory Syndrome Coronavirus 3C-Like Protease. *J. Virol.* **2008**, *82*, 4620–4629. [CrossRef]
29. Jin, Z.; Du, X.; Xu, Y.; Deng, Y.; Liu, M.; Zhao, Y.; Zhang, B.; Li, X.; Zhang, L.; Peng, C.; et al. Structure of M^{pro} from COVID-19 virus and discovery of its inhibitors. *Nature* **2020**, *582*, 289–293. [CrossRef] [PubMed]
30. Anand, K.; Ziebuhr, J.; Wadhwani, P.; Mesters, J.R.; Hilgenfeld, R. Coronavirus Main Proteinase (3CLpro) Structure: Basis for Design of Anti-SARS Drugs. *Science* **2003**, *300*, 1763–1767. [CrossRef]
31. Hilgenfeld, R. From SARS to MERS: Crystallographic studies on coronaviral proteases enable antiviral drug design. *FEBS* **2014**, *281*, 4085–4096. [CrossRef]
32. Zhang, L.; Lin, D.; Kusov, Y.; Nian, Y.; Ma, Q.; Wang, J.; von Brunn, A.; Leyssen, P.; Lanko, K.; Neyts, J.; et al. α-Ketoamides as Broad-Spectrum Inhibitors of Coronavirus and Enterovirus Replication: Structure-Based Design, Synthesis, and Activity Assessment. *J. Med. Chem.* **2020**, *63*, 4562–4578. [CrossRef]
33. Zhang, C.H.; Stone, E.A.; Deshmukh, M.; Ippolito, J.A.; Ghahremanpour, M.M.; Tirado-Rives, J.; Spasov, K.A.; Zhang, S.; Takeo, Y.; Kudalkar, S.N.; et al. Potent Noncovalent Inhibitors of the Main Protease of SARS-CoV-2 from Molecular Sculpting of the Drug Perampanel Guided by Free Energy Perturbation Calculations. *ACS central science* **2021**, *7*, 467–475. [CrossRef] [PubMed]
34. Froggatt, H.M.; Heaton, B.E.; Heaton, N.S. Development of a fluorescence-based, high-throughput SARS-CoV-2 3CLpro reporter assay. *J. Virol.* **2020**, *94*, e01265-20. [CrossRef] [PubMed]
35. Mariusz, J.; Dauter, Z.; Shabalin, I.G.; Gilski, M.; Brzezinski, D.; Kowiel, M.; Rupp, B.; Wlodawer, A. Crystallographic models of SARS-CoV-2 3CLpro: In-depth assessment of structure quality and validation. *IUCrJ* **2021**, *8*, 238–256.
36. Macchiagodena, M.; Pagliai, M.; Karrenbrock, M.; Guarnieri, G.; Iannone, F.; Procacci, P. Virtual Double-System Single-Box: A Nonequilibrium Alchemical Technique for Absolute Binding Free Energy Calculations: Application to Ligands of the SARS-CoV-2 Main Protease. *J. Chem. Theory Comput.* **2020**, *16*, 7160–7172. [CrossRef] [PubMed]
37. Bauer, R.A. Covalent inhibitors in drug discovery: From accidental discoveries to avoided liabilities and designed therapies. *Drug Disc. Today* **2015**, *20*, 1061–1073. [CrossRef]
38. Baillie, T.A. Targeted covalent inhibitors for drug design. *Angew. Chem. Int. Ed.* **2016**, *55*, 13408–13421. [CrossRef]
39. Sterling, T.; Irwin, J.J. ZINC 15–ligand discovery for everyone. *J. Chem. Inf. Model.* **2015**, *55*, 2324–2337. [CrossRef]
40. Hawkins, P.C.D.; Skillman, A.G.; Warren, G.L.; Ellingson, B.A.; Stahl, M.T. Conformer Generation with OMEGA: Algorithm and Validation Using High Quality Structures from the Protein Databank and the Cambridge Structural Database. *J. Chem. Inf. Model.* **2010**, *50*, 572–584. [CrossRef]
41. Zhang, L.; Lin, D.; Sun, X.; Curth, U.; Drosten, C.; Sauerhering, L.; Becker, S.; Rox, K.; Hilgenfeld, R. Crystal structure of SARS-CoV-2 main protease provides a basis for design of improved α-ketoamide inhibitors. *Science* **2020**, *368*, 409–412. [CrossRef] [PubMed]
42. Jukič, M.; Janežič, D.; Bren, U. Ensemble Docking Coupled to Linear Interaction Energy Calculations for Identification of Coronavirus Main Protease ($3CL^{pro}$) Non-Covalent Small-Molecule Inhibitors. *Molecules* **2020**, *25*, 5808. [CrossRef] [PubMed]
43. Shoichet, B.K. Interpreting steep dose-response curves in early inhibitor discovery. *J. Med. Chem.* **2006**, *49*, 7274–7277. [CrossRef] [PubMed]
44. Baell, J.B.; Holloway, G.A. New substructure filters for removal of pan assay interference compounds (PAINS) from screening libraries and for their exclusion in bioassays. *J. Med. Chem.* **2010**, *53*, 2719–2740. [CrossRef] [PubMed]
45. Saubern, S.; Guha, R.; Baell, J.B. KNIME workflow to assess PAINS filters in SMARTS format. Comparison of RDKit and Indigo cheminformatics libraries. *Mol. Inform.* **2011**, *30*, 847–850. [CrossRef] [PubMed]
46. Walters, W.P.; Stahl, M.T.; Murcko, M.A. Virtual screening—An overview. *Drug Disc. Today* **1998**, *3*, 160–178. [CrossRef]

47. Zhu, T.; Cao, S.; Su, P.C.; Patel, R.; Shah, D.; Chokshi, H.B.; Hevener, K.E. Hit identification and optimization in virtual screening: Practical recommendations based on a critical literature analysis: Miniperspective. *J. Med. Chem.* **2013**, *56*, 6560–6572. [CrossRef]
48. Kelley, L.A.; Gardner, S.P.; Sutcliffe, M.J. An automated approach for clustering an ensemble of NMR-derived protein structures into conformationally-related subfamilies. *Protein Eng.* **1996**, *9*, 1063–1065. [CrossRef] [PubMed]
49. Gaulton, A.; Bellis, L.J.; Bento, A.P.; Chambers, J.; Davies, M.; Hersey, A.; Overington, J.P. ChEMBL: A large-scale bioactivity database for drug discovery. *Nucleic Acids Res.* **2012**, *40*, D1100–D1107. [CrossRef]
50. Goodfellow, I.; Bengio, Y.; Courville, A.; Bengio, Y. *Deep Learning*; MIT Press: Cambridge, MA, USA, 2016; Volume 1, No. 2.
51. Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Duchesnay, E. Scikit-learn: Machine learning in Python. *J. Mach. Learn. Res.* **2011**, *12*, 2825–2830.
52. Chen, T.; Guestrin, C. Xgboost: A Scalable Tree Boosting System. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, 13–17 August 2016; pp. 785–794.
53. Chollet, F. Keras. 2015. Available online: https://keras.io (accessed on 8 May 2021).
54. Su, H.X.; Yao, S.; Zhao, W.F.; Li, M.J.; Liu, J.; Shang, W.J.; Xie, H.; Ke, C.; Hu, H.; Gao, M.; et al. Anti-SARS-CoV-2 activities in vitro of Shuanghuanglian preparations and bioactive ingredients. *Acta Pharmacol. Sin.* **2020**, *41*, 1167–1177. [CrossRef] [PubMed]
55. Rathnayake, A.D.; Zheng, J.; Kim, Y.; Perera, K.D.; Mackin, S.; Meyerholz, D.K.; Kashipathy, M.M.; Battaile, K.P.; Lovell, S.; Perlman, S.; et al. 3C-like protease inhibitors block coronavirus replication in vitro and improve survival in MERS-CoV–infected mice. *Sci. Transl. Med.* **2020**, *12*, 1–11. [CrossRef] [PubMed]
56. Morley, S.D.; Afshar, M. Validation of an empirical RNA-ligand scoring function for fast flexible docking using RiboDock. *J. Comput. Aided Mol. Des.* **2004**, *18*, 189–208. [CrossRef] [PubMed]
57. Ruiz-Carmona, S.; Alvarez-Garcia, D.; Foloppe, N.; Garmendia-Doval, A.B.; Juhos, S.; Schmidtke, B.; Barril, X.; Hubbard, R.E.; Morley, S.D. rDock: A Fast, Versatile and Open Source Program for Docking Ligands to Proteins and Nucleic Acids. *PLoS Comput. Biol.* **2014**, *10*, e1003571. [CrossRef]
58. DeLano, W.L. Pymol: An open-source molecular graphics tool. *CCP4 Newsl. Protein Crystallogr.* **2002**, *40*, 82–92.
59. Mysinger, M.M.; Carchia, M.; Irwin, J.J.; Shoichet, B.K. Directory of useful decoys, enhanced (DUD-E): Better ligands and decoys for better benchmarking. *J. Med. Chem.* **2012**, *55*, 6582–6594. [CrossRef]
60. Srivastava, N.; Hinton, G.; Krizhevsky, A.; Sutskever, I.; Salakhutdinov, R. Dropout: A simple way to prevent neural networks from overfitting. *J. Mach. Learn. Res.* **2014**, *15*, 1929–1958.
61. Graf, M.M.; Bren, U.; Haltrich, D.; Oostenbrink, C. Molecular dynamics simulations give insight into D-glucose dioxidation at C2 and C3 by Agaricus meleagris pyranose dehydrogenase. *Comput. Aided Mol. Des.* **2013**, *27*, 295–304. [CrossRef]
62. Jukic, M.; Ilc, N.; Sluga, D.; Tomšič, G.; Podlipnik, Č. CmDock. 2021. Available online: https://gitlab.com/Jukic/cmdock/ (accessed on 8 May 2021).
63. Tosco, P.; Stiefl, N.; Landrum, G. Bringing the MMFF force field to the RDKit: Implementation and validation. *J. Cheminformatics* **2014**, *6*, 1–4. [CrossRef]