



Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.



Research paper

# Discovery of novel SARS-CoV-2 3CL protease covalent inhibitors using deep learning-based screen

Liyang Wang<sup>a,1</sup>, Zhongtian Yu<sup>a,1</sup>, Shiwei Wang<sup>a</sup>, Zheng Guo<sup>a</sup>, Qi Sun<sup>a,c,\*</sup>, Luhua Lai<sup>a,b,c,\*\*</sup>

<sup>a</sup> BNLMs, Peking-Tsinghua Center for Life Sciences at the College of Chemistry and Molecular Engineering, Peking University, Beijing, 100871, PR China

<sup>b</sup> Center for Quantitative Biology, Academy for Advanced Interdisciplinary Studies, Peking University, Beijing, 100871, PR China

<sup>c</sup> Research Unit of Drug Design Method, Chinese Academy of Medical Sciences (2021RU014), Beijing, 100871, PR China



## ARTICLE INFO

## Keywords:

SARS-CoV-2  
3C-like protease inhibitors  
Deep learning  
Covalent warheads

## ABSTRACT

SARS-CoV-2 3CL protease is one of the key targets for drug development against COVID-19. Most known SARS-CoV-2 3CL protease inhibitors act by covalently binding to the active site cysteine. Yet, computational screens against this enzyme were mainly focused on non-covalent inhibitor discovery. Here, we developed a deep learning-based stepwise strategy for selective covalent inhibitor screen. We used a deep learning framework that integrated a directed message passing neural network with a feed-forward neural network to construct two different classifiers for either covalent or non-covalent inhibition activity prediction. These two classifiers were trained on the covalent and non-covalent 3CL protease inhibitors dataset, respectively, which achieved high prediction accuracy. We then successively applied the covalent inhibitor model and the non-covalent inhibitor model to screen a chemical library containing compounds with covalent warheads of cysteine. We experimentally tested the inhibition activity of 32 top-ranking compounds and 12 of them were active, among which 6 showed IC<sub>50</sub> values less than 12 μM and the strongest one inhibited SARS-CoV-2 3CL protease with an IC<sub>50</sub> of 1.4 μM. Further investigation demonstrated that 5 of the 6 active compounds showed typical covalent inhibition behavior with time-dependent activity. These new covalent inhibitors provide novel scaffolds for developing highly active SARS-CoV-2 3CL covalent inhibitors.

## 1. Introduction

Caused by the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) [1], the ongoing global coronavirus disease 2019 (COVID-19) pandemic has led to more than 565 million confirmed cases and over 6 million deaths as of July 2022 according to World Health Organization. Despite the relentless effort of researchers, COVID-19 may still pose a great threat to human life in the future. On the one hand, the continuous emergence of virus variants compromises the efficacy of available vaccines [2], while the durability and long-term side effects of vaccines are unknown currently [3]. On the other hand, although several drugs such as remdesivir [4], molnupiravir [5], and Paxlovid [6] have been approved for emergent use by FDA for the treatment of patients with mild-to-moderate COVID-19, only Paxlovid performs

relatively well. Therefore, it is still imperative to develop new effective therapies targeting this pandemic.

SARS-CoV-2 and other coronaviruses encode a chymotrypsin-like protease (3C-like protease (3CL<sup>Pro</sup>) or main protease (M<sup>Pro</sup>)), which cleaves the viral polyproteins at 11 sites and plays a pivotal role in the replication as well as transcription of viruses [7,8]. Its substrate cleaving sites involve a conserved glutamine at the P1 position that is essential to hydrolysis [9–11]. Such strict substrate specificity has never been found in human-host protease [12,13], enabling 3CL<sup>Pro</sup> as an ideal target for developing drugs against COVID-19. A number of 3CL<sup>Pro</sup> inhibitors for other coronaviruses had been studied before [14–19], which have been repurposed or further developed to treat COVID-19. Despite the therapeutic benefits, fear of toxicity caused by inherent chemical reactivity has impeded the development of covalent drugs for a long time and few

\* Corresponding author. BNLMs, Peking-Tsinghua Center for Life Sciences at the College of Chemistry and Molecular Engineering, Peking University, Beijing, 100871, PR China.

\*\* Corresponding author. BNLMs, Peking-Tsinghua Center for Life Sciences at the College of Chemistry and Molecular Engineering, Peking University, Beijing, 100871, PR China.

E-mail addresses: [qsun2015@pku.edu.cn](mailto:qsun2015@pku.edu.cn) (Q. Sun), [lhlai@pku.edu.cn](mailto:lhlai@pku.edu.cn) (L. Lai).

<sup>1</sup> These authors contributed equally to this work.

studies were dedicated to covalent inhibition 30 years ago [20]. Nevertheless, due to the development of characterization techniques and the discovery of some covalent inhibitors with moderate chemical reactivity, the design of covalent drugs becomes more prevalent [21,22]. At present, the known SARS-CoV-2 3CL<sup>PRO</sup> inhibitors are also mainly covalent, such as PF-00835231 [22], PF-07321332 [6], myricetin [21], GC376 [23], MI-23 [24], 11a [25], calpain inhibitors I, II, and XII [26], carmofur [10], boceprevir [26], N3 [10], most of which contain the warheads such as aldehyde, alpha-ketoamide, nitrile, acrylamide, Michael acceptors, and chloroacetamide to react with the active site Cys145. There are also a few non-covalent inhibitors, such as S-217622 [27], baicalin [28], and masitinib [29]. In view of the enhanced therapeutic potency and long-lasting effects of these covalent warheads [30, 31], developing covalent inhibitors of SARS-CoV-2 3CL<sup>PRO</sup> provides an effective way to combat the pandemic.

In addition to rational design [32], molecular docking-based virtual screening [33], high-throughput experimental screening [34], and artificial intelligence (AI) have been used to screen for anti-SARS-CoV-2 compounds. For example, Wang et al. [35] developed a transferable deep learning method to screen a large compound library and suggested a list of potential anti-SARS-CoV-2 compounds. Duc et al. [36] used algebraic topology and deep learning to understand the molecular mechanism of SARS-CoV-2 3CL<sup>PRO</sup> inhibition from 137 complex crystal structures and predicted 71 possible covalent binding inhibitors, for which no experimental validation has been carried out. Hu et al. [37] developed a novel framework, AIMEE, that integrated an AI model with enzymological experiments to screen a bioactive chemical library and found four SARS-CoV-2 3CL<sup>PRO</sup> inhibitors with half-maximal inhibitory concentration (IC<sub>50</sub>) at micromole level. Despite that the most successful SARS-CoV-2 3CL<sup>PRO</sup> inhibitors until now all covalently bind to the active site Cys145, currently reported AI models for SARS-CoV-2 3CL<sup>PRO</sup> inhibitor screen all used datasets of experimentally identified inhibitors that contain both covalent and non-covalent inhibitors. Most covalent inhibitors screened by these models can covalently bind to various targets without selectivity, which may easily result in off-target effects. Consequently, AI models that can be used to screen selective covalent inhibitors need to be developed.

Given the contribution of non-covalent interactions to the selectivity of ligands and the increased efficacy when two binding mechanisms exist simultaneously [20], the existence of covalent warheads and the binding of non-covalent sub-structures should be taken into account together when screening new covalent inhibitors. In the present study, we developed a strategy that incorporates deep learning models and *in vitro* experiments to quickly identify potential selective covalent SARS-CoV-2 3CL<sup>PRO</sup> inhibitors. Firstly, we applied a deep learning framework that integrated a directed message passing neural network with a feed-forward neural network to construct two different classifiers for the activity prediction of molecules. These two classifiers were trained on the covalent and non-covalent 3CL<sup>PRO</sup> inhibitors dataset separately. Both models achieved satisfactory performance on the test sets. We then successively applied the covalent inhibitor model and the non-covalent inhibitor model to screen a library with more than 39,000 compounds that contain covalent warheads of cysteine. Among the 32 top-ranking compounds that were experimentally tested, 6 showed IC<sub>50</sub> values less than 12 μM and the strongest one inhibited SARS-CoV-2 3CL<sup>PRO</sup> with an IC<sub>50</sub> of 1.4 μM. Further investigation demonstrated that 5 of the 6 active compounds showed typical covalent inhibition behavior with time-dependent activity. These new inhibitors may provide useful guidance for the design of new drugs towards COVID-19.

## 2. Material and methods

### 2.1. Dataset

Since SARS-CoV-2 3CL<sup>PRO</sup> has up to 96% sequence identity with SARS-CoV 3CL<sup>PRO</sup>, inhibitors targeting SARS-CoV 3CL<sup>PRO</sup> are likely to

have the same effect on SARS-CoV-2 3CL<sup>PRO</sup> [38]. Therefore, we collected not only SARS-CoV-2 3CL<sup>PRO</sup> inhibitors from recent literature (from 2020 to 2021) but also SARS-CoV 3CL<sup>PRO</sup> inhibitors from previous studies (from 2004 to 2021) [11,14–19,23,25,39–42]. Experimental data of inhibition assays are available for all collected inhibitors in literature and IC<sub>50</sub> of 50 μM was chosen as the cutoff to distinguish positive and negative samples. In addition, we divided this curated dataset into covalent and non-covalent datasets based on the binding mechanism of inhibitors. After dividing the dataset into training and test sets in a ratio of approximately 10:1, we finally constructed a covalent training dataset (Training Set 1) with 463 molecules (209 positives and 254 negatives), a covalent test dataset (Test Set 1) with 53 molecules (25 positives and 28 negatives), a non-covalent training dataset (Training Set 2) with 1086 molecules (224 positives and 862 negatives) and a non-covalent test dataset (Test Set 2) with 108 molecules (18 positives and 90 negatives) (Data S1).

Cysteine Targeted Covalent Library from ChemDiv (<https://www.chemdiv.com>) contains 39,301 compounds that have specific warheads, which were designed to react with cysteine. After removing overlaps with the training set and test set, the remaining 39,014 compounds of this library were used to screen potential covalent SARS-CoV-2 3CL<sup>PRO</sup> inhibitors.

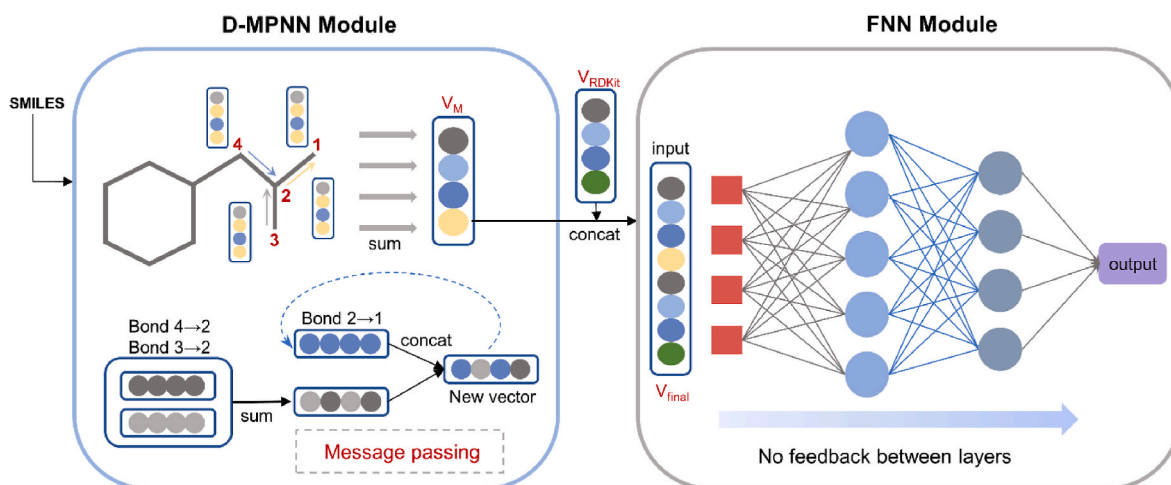
### 2.2. Deep learning models

We developed one classification model for covalent inhibitor classification (COVCL) and one classification model for non-covalent inhibitor classification (NOVCL) for SARS-CoV-2 3CL<sup>PRO</sup>. The COVCL model, trained on Training Set 1, can be used to predict whether a compound has covalent inhibitory activity against SARS-CoV-2 3CL<sup>PRO</sup>. The NOVCL model, trained on Training Set 2, can be used to predict whether a compound has non-covalent inhibitory activity against SARS-CoV-2 3CL<sup>PRO</sup>. Successive application of these two models can screen out compounds that may specifically bind SARS-CoV-2 3CL<sup>PRO</sup> and form covalent bond with Cys145.

We used the Chemprop architecture to train the models and encoded the molecules using SMILES representation. Chemprop model mainly includes a directed message passing neural network (D-MPNN) module for molecular feature extraction and a feed-forward neural network (FNN) for property prediction. Initially, Chemprop takes molecular SMILES as input and converts it into a molecular graph with atoms regarded as nodes and bonds as edges. After the message passing stage and readout stage, the D-MPNN module finally extracts all the atomic features and bond features in the molecular graph to generate a single feature vector representing the whole molecule [43]. In this work, we concatenated an additional vector containing 200 descriptors computed by RDKit [44] to introduce more molecular-level features. These descriptors contain partial charge, number of heavy atoms, number of hydrogen-bond donors, etc (Table S1). The FNN structure means that information is transferred unidirectionally from the input layer to the output layer step by step, with no feedback between layers [45]. The final feature vector of D-MPNN module was then fed to the FNN module to predict the activity of candidate molecules (Fig. 1).

### 2.3. Inhibition assay of SARS-CoV-2 3CL<sup>PRO</sup>

The expression and purification of SARS-CoV-2 3CL<sup>PRO</sup> were carried out using the reported protocol [46]. A fluorescent substrate Dabcyl-KTSAVLQSGFRKM-E(Edans)-NH<sub>2</sub> (GL Biochemistry Ltd) and assay buffer (40 mM PBS, 100 mM NaCl, 1 mM EDTA, 0.1% Triton 100, pH 7.3) were used for the inhibition assay. Stock solutions of the inhibitors were prepared with DMSO. 0.5 μM SARS-CoV-2 3CL<sup>PRO</sup> was pre-incubated for 180 min or other lengths of time with 5 μL DMSO or inhibitor at various concentrations. Then, 20 μM fluorescent substrate was added into the system to initiate the reaction. The reaction system was excited at 360 nm and an increase in absorbance at 460 nm was



**Fig. 1.** Schematic of the framework used for two deep classifiers.  $V_M$  represents a vector of the whole molecular features.  $V_{RDKit}$  represents a vector containing 200 descriptors calculated by RDKit.  $V_M$  and  $V_{RDKit}$  are concatenated to generate a final vector  $V_{final}$ , which is then used as the input of FNN module.

recorded for 20 min at an interval of 37 s with a kinetics mode program using a 96-well plate reader (Synergy, Biotek). The inhibition rate was calculated by  $V_i/V_0$ , where  $V_0$  or  $V_i$  represents the mean reaction rate of the protease incubated with DMSO or compounds.  $IC_{50}$  was fitted with Prism GraphPad 8.0.

For the inhibition test in presence of DTT, the final concentration of DTT in assay buffer was 5 mM.

#### 2.4. Inhibition kinetics analysis and reversibility assay

We studied the enzyme kinetics properties by adding different concentrations of the fluorescent substrate to initiate the reaction. Mean velocities were collected and plotted then fitted with the Michaelis-Menten equation to obtain values of  $K_m$  and  $V_{max}$ .

To investigate the reversibility of inhibition, 1  $\mu$ M SARS-CoV-2 3CL<sup>pro</sup> was incubated with 10  $\mu$ M compound stock solution on ice for 180 min and then divided into different Millipore tubes for various times of ultrafiltration. For each time of ultrafiltration, equal volume of assay buffer used in the inhibition assay was added to elute the protease then ultrafiltered at 4 °C, 12,000 rpm for 5 min and collected for the inhibition assay.

#### 2.5. Half-life of compounds reacting with GSH

Evaluation of the intrinsic reactivity of warheads was conducted by measuring the half-life of the compounds reacting with GSH. 500  $\mu$ M compound was incubated with 10 mM GSH for 0–60 min and the remaining compounds were quantified by HPLC (Agilent 1200). Ln (the percentage of the remaining compound) was plotted against incubation time to generate the half-life time of the compound reacting with GSH.

#### 2.6. Cathepsin L inhibition assay

The inhibition assay of Cathepsin L was performed as previously reported [47]. Briefly, **compound 9** and **compound 13** were tested using the commercial Cathepsin L Inhibitor Assay Kit (Abcam, Cat# ab197012). FF-FMK, a known inhibitor for Cathepsin L was used as a positive control.

#### 2.7. Mass spectrometry analysis

For mass spectrometry analysis of the protease, 1  $\mu$ M SARS-CoV-2 3CL<sup>pro</sup> was incubated with DMSO or 10  $\mu$ M compound on ice for 180 min. The solution was ultrafiltered three times at 4 °C, 12,000 rpm for 5

min and analyzed by Quadrupole-TOF LC-MS/MS System using the ESI (+) mode. Signals of observed mass were collected and deconvoluted.

#### 2.8. Covalent docking

The crystal structure of SARS-CoV-2 3CL<sup>pro</sup> in complex with compound 4 (PDB ID: 7JT7) was used as the target structure [48]. Protein and ligands were first prepared by the Protein Prep Wizard module [49] and LigPrep module in Schrodinger [50], respectively. Covalent Docking module [51] in Schrodinger was then used to carry out covalent docking of compounds with reaction type set to Michael addition and docking mode set to pose prediction.

### 3. Results and discussions

#### 3.1. Models for prediction of covalent and non-covalent SARS-CoV-2 3CL<sup>pro</sup> inhibitors

We first trained and tested the COVCL and NOVCL models. Chemprop is a deep learning model for molecular property prediction that outperforms existing strong baselines on 16 proprietary datasets and 19 publicly available datasets [52]. Given the outstanding performance of Chemprop, we considered using it as the framework of our models. Since the performance of deep learning model is usually limited by the size of dataset, we additionally tested the performance of several classical machine learning models including Random Forest (RF) [53], Support Vector Classification (SVC) [54], eXtreme Gradient Boosting (XGB) [55], and Artificial Neural Network (ANN) [56]. Firstly, 10-fold cross-validation was carried out on training sets to evaluate the performance. Since ensemble methods can improve the stability of model by aggregating the predictions of multiple classifiers [57], we further constructed an ensemble of 10 models. The ensemble was trained by different splits of training data and validation data (9:1) from training sets. Then, performance of this ensemble was tested on the independent test sets and the final prediction was the average result of these 10 models. The receiver operating characteristic curve (ROC-AUC) and accuracy (ACC) were used as metrics to evaluate the performance of different models. As indicated by Table S2 and Table S3, the performance of classical models was close to Chemprop on both covalent and non-covalent datasets, but overall Chemprop performed slightly better. Consequently, considering that our small dataset does not diminish the effect of deep learning and that Chemprop incorporates more molecular structure information as described in methods, we finally chose Chemprop as the basic architecture of COVCL and NOVCL models.

The default parameters of Chemprop model were used except for four hyperparameters: the number of message-passing steps (depth), the dropout values of all layers, the size of bond message vectors (hidden size), and the number of FNN layers. We used the automated hyperparameter optimization tool in Chemprop, which is achieved by the Bayesian optimization method, to identify the optimal hyperparameters (Table 1).

As shown in Fig. 2A and Fig. 2C, AUC and ACC of ten-fold cross-validation on Training Set 1 were all above 0.850, while the AUC on Test set 1 achieved 0.987, indicating the good performance of COVCL. As for NOVCL, the AUC and ACC of ten-fold cross-validation this time were all above 0.880, while the AUC of Test set 2 could reach 0.853 (Fig. 2B and D). In summary, both COVCL and NOVCL showed satisfactory performance, and thus can be used for the following screening of SARS-CoV-2 3CL<sup>pro</sup> inhibitors.

### 3.2. Virtual screening of Cysteine Targeted Covalent Library to identify potential 3CL<sup>pro</sup> inhibitors

We constructed a generally applicable 5-step screening architecture that integrates deep learning and further properties analysis to identify potential selective covalent inhibitors (Fig. 3A).

We first screened the Cysteine Targeted Covalent Library with the COVCL model for compounds that might covalently bind SARS-CoV-2 3CL<sup>pro</sup>. The top 727 compounds with scores higher than 0.8 were selected (Fig. 3B) and subjected to the NOVCL model filtering for compounds that specifically bind to the enzyme and form covalent bond with Cys145. Using a cutoff score of 0.7, 298 molecules were selected (Fig. 3C) for further analysis.

Since some of these 298 compounds are structurally similar, a cluster analysis of the compounds was performed to select representative compounds from each cluster. We used the Butina [58] module of RDKit to cluster the above 298 molecules with a distance cutoff of 0.4. This allowed the compounds to be divided into 102 clusters, among which two clusters contain more than 20 compounds. We further clustered these two groups of compounds with a cutoff of 0.2 and the compound with the highest NOVCL score was taken as a representative for each class. These analyses gave 113 compounds for further study.

As the physicochemical properties of a molecule can have a significant impact on its absorption, distribution, metabolism, excretion (ADME), toxicity, and pharmacological activity, they were considered in the fourth step of the screening process. In the present study, we simply used the octanol-water partition coefficient (logP) to ensure the absorption and bioavailability property of a compound [59]. There are 79 compounds among the 117 with logP values not higher than 5. We also limited the molecular weight to the range of 300–600 and 32 candidate compounds were selected for experimental study.

In order to visualize the model prediction results and further analyze the structural information of the screened molecules, we applied t-distributed stochastic neighbor embedding (t-SNE) on the dataset to visualize the chemical space distribution of compounds. t-SNE is a dimensionality reduction method that can utilize Tanimoto similarity to quantify the chemical distance. The spatial proximity of two points on the t-SNE plot graph implies structural similarity. As shown in Fig. 4A, the molecules in the training set were widely distributed in chemical space and contained various structures. The distribution of 32 screened molecules was dispersed and 29 of these molecules had the maximum similarity of less than 0.5 to the positive compounds in the training set, indicating that COVCL and NOVCL can identify novel structures to some

extent (Fig. 4B). This indicates that the models learned certain structure-activity relationships from the training set, and the learned features can be recombined into some new compounds. Certainly, the performance of deep learning models is closely related to the training data, so increasing the structural diversity of training data should facilitate the model to learn more information and hence enhance its prediction ability.

### 3.3. Enzyme inhibition activity of the virtually screened compounds

We then tested the inhibitory activity of the 32 compounds against SARS-CoV-2 3CL<sup>pro</sup> according to our previously published methods [46], using tideglusib as a positive control (Fig. S1) [10]. Among the 32 compounds, 12 showed more than 50% inhibition at the concentration of 100  $\mu$ M (Table S4) and 6 exhibited considerable inhibitory activity with IC<sub>50</sub> < 12  $\mu$ M with 180 min incubation time (Fig. 5). These 6 active compounds contain covalent warheads of  $\alpha$ ,  $\beta$ -unsaturated ketone (ester), disulfide bond, and heteroaromatic ester. To the best of our knowledge, no anti-SARS-CoV-2 activities nor other biological activities of these molecules have been reported before. We also used the D3Similarity web server to calculate the molecular similarity between our compounds and the reported bioactive compounds against coronaviruses [60]. The maximum two-dimensional similarity of the 6 compounds were all below 0.45, indicating their structural novelty compared to known inhibitors, which provides good starting points for designing novel SARS-CoV-2 3CL<sup>pro</sup> covalent inhibitors.

### 3.4. Mode of action analysis of the active compounds

To elucidate the mode of action of hit compounds against SARS-CoV-2 3CL<sup>pro</sup>, we carried out a series of enzymatic studies and mass spectrometry analysis. As time-dependent inhibition is a good indicator of covalent binding, we incubated the 6 compounds with SARS-CoV-2 3CL<sup>pro</sup> for different lengths of time and measured their inhibition activity. Five of them exhibited obvious time-dependent increase of inhibition activity (Fig. 5 and Table S5). Among the 5 compounds, **compound 9** and **compound 13** with the most potent inhibitory activities were selected for further analysis.

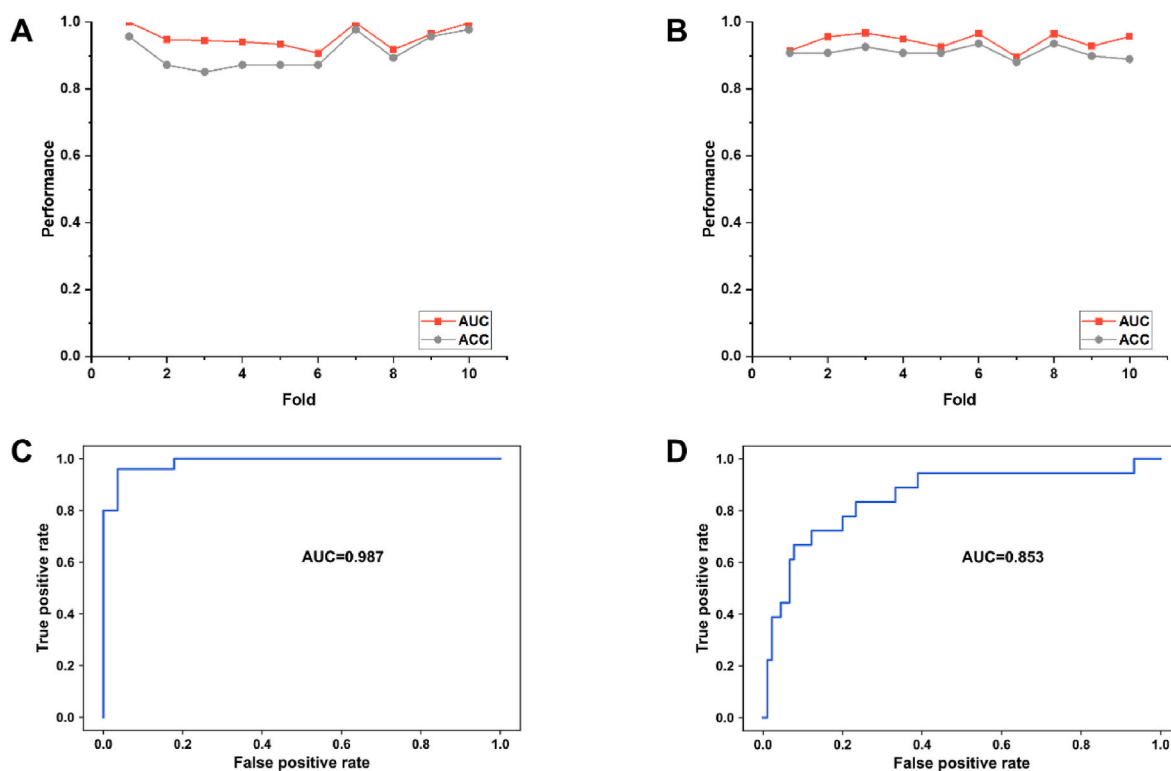
As shown in Fig. 6A, the addition of dithiothreitol (DTT) could reverse the inhibitory effect of **compound 9** and **compound 13**, which is typical of covalent inhibition through Cys binding. Furthermore, reversibility assay of both compounds to SARS-CoV-2 3CL<sup>pro</sup> revealed that the ultrafiltration of inhibitors could recover enzymatic activity to a certain extent (Fig. 6B), which indicated that they were reversible inhibitors. As exhibited in Fig. 6C and Table S6, the Michaelis-Menten kinetics analysis of **compound 9** and **compound 13** yielded both Km and Vmax values altered, indicating that they are mix-type inhibitors of SARS-CoV-2 3CL<sup>pro</sup>. Overall, the enzymatic studies demonstrated that both **compound 9** and **compound 13** were reversible covalent inhibitors of SARS-CoV-2 3CL<sup>pro</sup>.

We have also explored the selectivity of **compound 9** and **compound 13** by investigating their inhibitory activity against Cathepsin L, a key host cysteine protease utilized by coronaviruses for cell entry [3, 61,62], using the reported method (the known Cathepsin L inhibitor FF-FMK was used as the positive compound, Fig. S2) [63]. At the concentration of 50  $\mu$ M, the inhibition rates of Cathepsin L by **compound 9** and **compound 13** are  $20.3 \pm 5.3\%$  and  $26.6 \pm 5.8\%$ , respectively. The weak inhibitory activity of both compounds against Cathepsin L suggested their high selectivity for coronavirus protease.

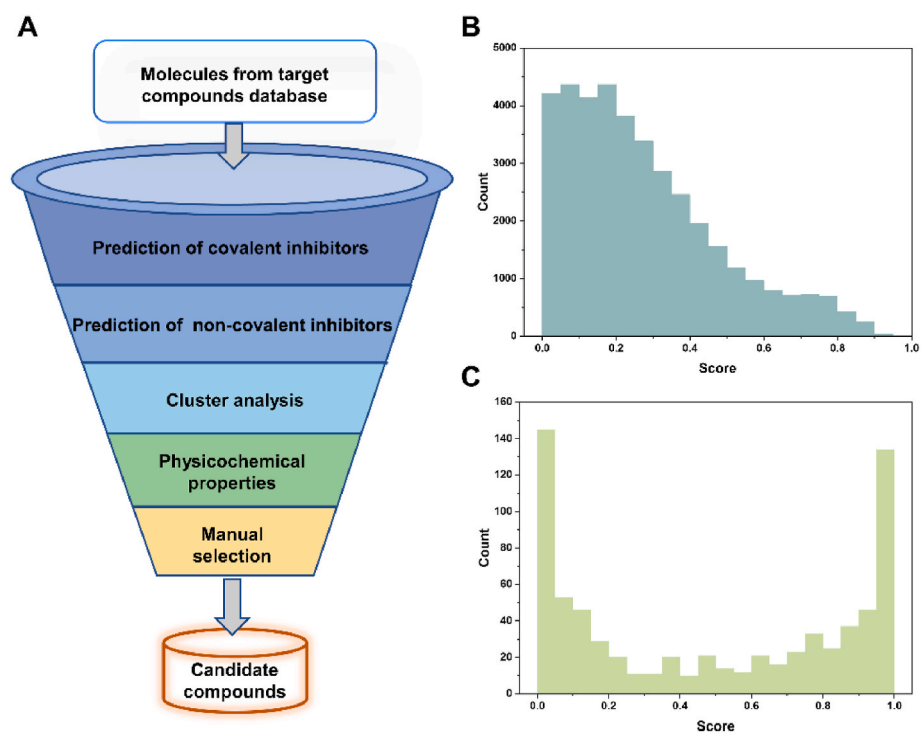
To reduce possible off-target covalent binding that may lead to toxicity risks, the intrinsic reactivity of the warheads should be low. The intrinsic reactivity of **compound 9** and **compound 13** was measured using a GSH assay which is routinely used to assess the reactivity of cysteine-targeted warheads [64]. The reaction half-life ( $t_{1/2}$ ) of **compound 9** and **compound 13** were 1.4 min and 49.6 min, respectively (Fig. S3). Regarding the fact that  $t_{1/2}$ s of clinical covalent kinase inhibitors determined in GSH assay were within the range of 30–512 min

**Table 1**  
Optimal hyperparameters for the two models.

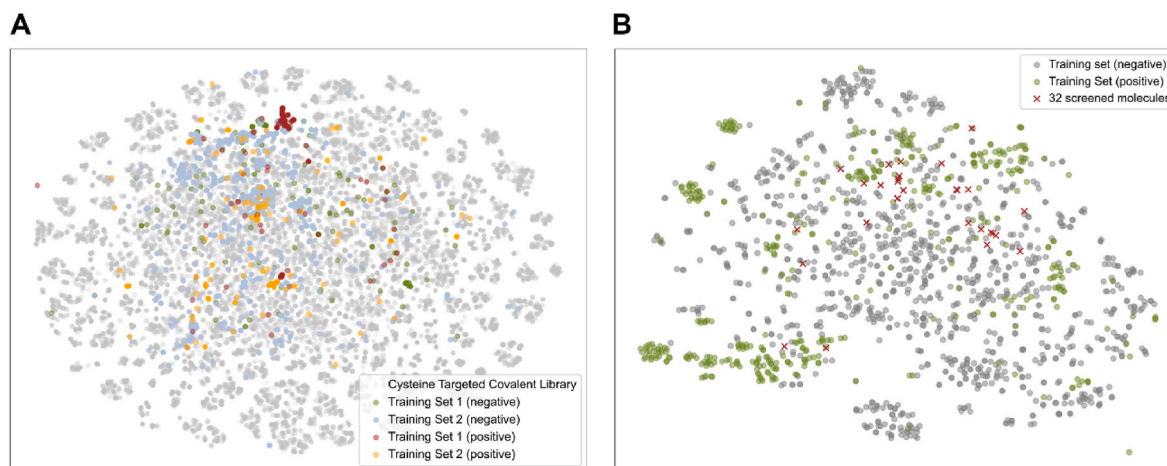
model	depth	dropout	FNN layers	hidden size
COVCL	5	0.1	2	1100
NOVCL	3	0.1	2	1300



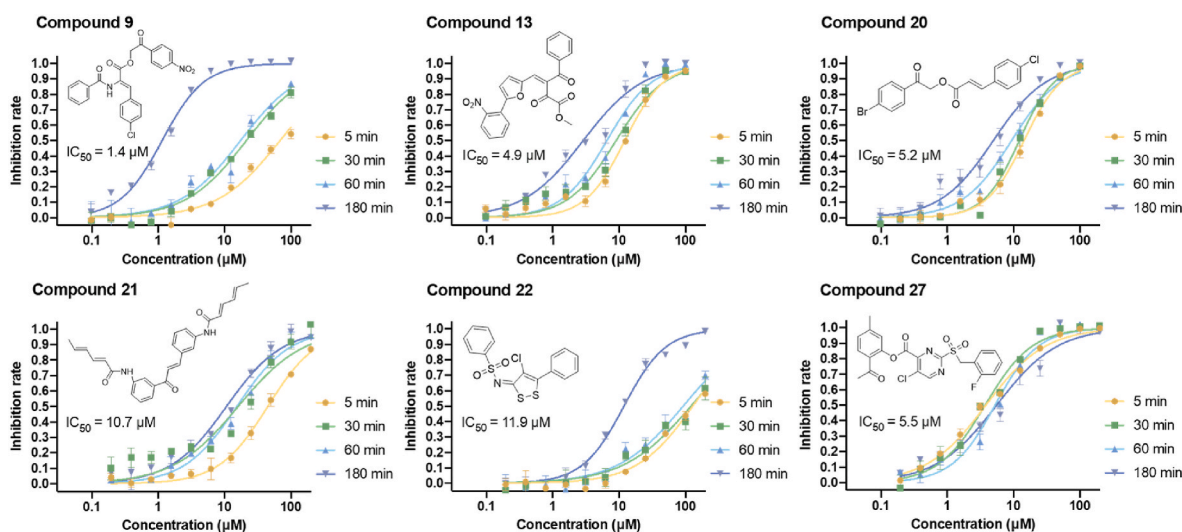
**Fig. 2.** Performance of two classification models (COVCL and NOVCL). (A) Result of 10-fold cross-validation on Training Set 1 for COVCL. (B) Result of 10-fold cross-validation on Training Set 2 for NOVCL. (C) Receiver operating characteristic curve of COVCL on Test Set 1. AUC is the area under the curve, which is usually used to evaluate the performance of classification model. (D) Receiver operating characteristic curve of NOVCL on Test Set 2.



**Fig. 3.** Library screening for potential inhibitors. (A) Overall schematic of the screening architecture. (B) Distribution of predicted scores for Cysteine Targeted Covalent Library using the COVCL model. Count indicates the number of molecules. (C) Distribution of the predicted scores for the 727 selected molecules using the NOVCL model.



**Fig. 4.** Chemical space distribution of molecules. **(A)** Visualization of the molecular structures in two dimensions space for Cysteine Targeted Covalent Library (grey dots) and training set (green, blue, red and orange dots) by t-SNE plot. The spatial proximity of two points on the graph implies structural similarity. **(B)** Visualization of the molecular structures in two dimensions space for training set (green and grey dots) and 32 screened molecules (red crosses) by t-SNE plot.



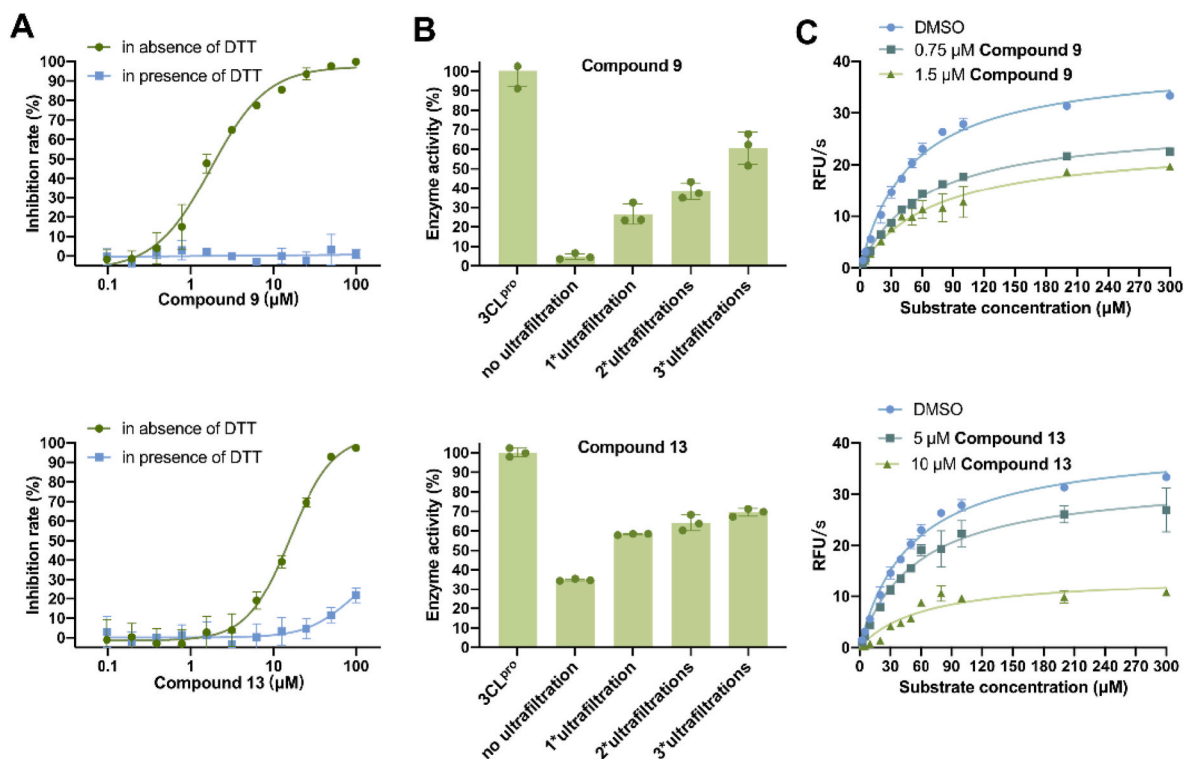
**Fig. 5.** Dose-response curves of the inhibition of SARS-CoV-2 3CL<sup>PRO</sup> by the 6 hits under different incubation times. IC<sub>50</sub> values under 180 min incubation time are shown.

[65], **compound 13** demonstrated moderate reactivity, implying its potential for further development of more potent SARS-CoV-2 3CL<sup>PRO</sup> covalent inhibitors.

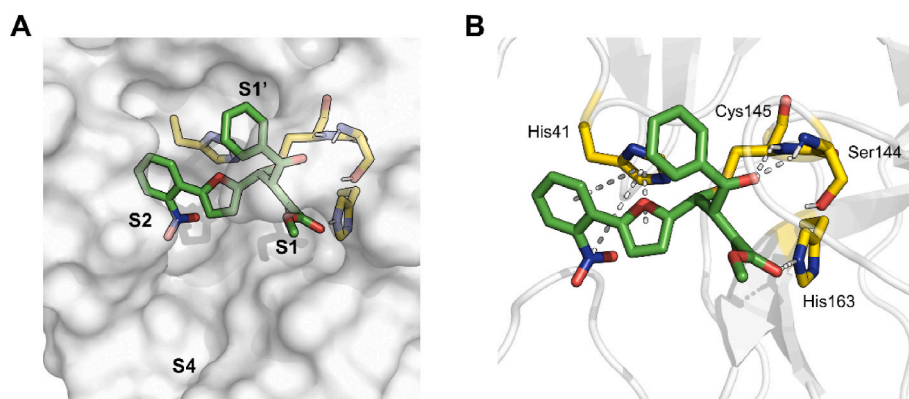
We further carried out liquid chromatography-tandem mass spectrometry (LC-MS/MS) analysis of covalent binding between SARS-CoV-2 3CL<sup>PRO</sup> and **compound 13**. We found that **compound 13** could only partially modify SARS-CoV-2 3CL<sup>PRO</sup> (Fig. S4), perhaps due to the instability of reversible covalent binding manner. Intact protein mass spectrometry study confirmed that a covalent bond is formed between the protease and the  $\alpha$ ,  $\beta$ -unsaturated ketone unit of **compound 13**, detected as a peak with MW shift of +405 Da (C<sub>22</sub>H<sub>15</sub>NO<sub>7</sub>). Besides, a peak with MW shift of +319 Da (C<sub>19</sub>H<sub>13</sub>NO<sub>4</sub>), which is equal to the mass of the SARS-CoV-2 3CL<sup>PRO</sup>/arylfurylpropenone complex along with the removal of methyl glyoxylate was detected.

Eventually, we performed covalent docking to understand the interactions of **compound 13** with SARS-CoV-2 3CL<sup>PRO</sup>. As shown in Fig. 7A, **compound 13** mainly occupied the S1', S1, and S2 pockets. A covalent bond is formed between Cys145 and the  $\alpha$ ,  $\beta$ -unsaturated ketone unit. In the docking model, two  $\pi$ - $\pi$  stackings are observed between the furan unit and the benzene ring of **compound 13** and His41 of 3CL<sup>PRO</sup>. His41 also interacts with the nitro group of **compound 13** by

cation- $\pi$  interaction. In addition, **compound 13** forms three hydrogen bonds with Ser144, Cys145, and His163, respectively (Fig. 7B). Apart from the covalent bond formed between Michael acceptors and Cys145, those mentioned non-covalent interactions of **compound 13** also contribute to its inhibitory activity and selectivity against SARS-CoV-2 3CL<sup>PRO</sup>. In the docking model, **compound 13** occupies a similar position as myricetin, a natural product that was reported to target Cys145 with its pyrogallol warhead [21]. Myricetin bound to 3CL<sup>PRO</sup> distinctively (PDB ID: 7DPP), mainly occupying the S1' and S2 pockets. The benzene ring in the case of compound 13 binding to 3CL<sup>PRO</sup> is similar to the binding position of the pyrogallol moiety. On the other hand, different from the cyclic moieties adopted in the reported peptidomimetic covalent inhibitors of 3CL<sup>PRO</sup> [6,23,25,66,67], the methyl glyoxylate moiety in **compound 13** occupies the S1 pocket (Fig. S8). In addition, the occupied S1' pocket and currently unoccupied S4 pocket are relatively novel situations among cases of binding to 3CL<sup>PRO</sup> with covalent inhibitors, which would hopefully provide inspiration for developing novel 3CL<sup>PRO</sup> covalent inhibitors.



**Fig. 6.** Investigation of inhibitory properties of **compound 9** and **compound 13** against SARS-CoV-2 3CL<sup>Pro</sup>. (A) Inhibition test of **compound 9** and **compound 13** in the presence or absence of DTT. (B) Enzyme activity of SARS-CoV-2 3CL<sup>Pro</sup> co-incubated with **compound 9** or **compound 13** after different times of ultrafiltration (C). Michaelis-Menten kinetics study of **compound 9** and **compound 13**.



**Fig. 7.** Interaction diagram of **compound 13** with SARS-CoV-2 3CL<sup>Pro</sup>. (A) The overall structure of **compound 13** binding to the active pocket of SARS-CoV-2 3CL<sup>Pro</sup>. (B) Detailed interactions between **compound 13** and SARS-CoV-2 3CL<sup>Pro</sup>. Ligands are shown as green sticks and SARS-CoV-2 3CL<sup>Pro</sup> is displayed as grey cartoon in 3D structure. Amino acid residues interacting with the compounds are shown as yellow sticks.

#### 4. Conclusion

We have developed a generally applicable deep learning screening strategy that uses both covalent and non-covalent classifiers to identify selective covalent enzyme inhibitors. In the present study, we trained the two classifiers on covalent and non-covalent SARS-CoV-2 3CL<sup>Pro</sup> inhibitor datasets, respectively, and successfully applied them to identify 5 novel covalent SARS-CoV-2 3CL<sup>Pro</sup> inhibitors. Such a deep learning approach only requires molecular SMILES as input files, making it convenient to be applied. Compared to experimental and docking-based screening, it is highly efficient. The inhibitors that we found provide new scaffolds for developing the next generation of specific covalent SARS-CoV-2 3CL<sup>Pro</sup> inhibitors. Especially, our framework can be easily extended to other targets or diseases in principle although we only

presented the study of SARS-CoV-2 3CL<sup>Pro</sup> here. For instance, papain-like protease (PL<sup>Pro</sup>) of SARS-CoV-2, which is another promising target for the treatment of coronavirus, can also form a covalent bond with ligands through Cys111 in the catalytic site [68,69]. The covalent inhibitors of PL<sup>Pro</sup> can be screened out by our framework as long as researchers replace the original datasets with PL<sup>Pro</sup>-related data. Overall, the deep learning-based framework for rapid identification of covalent inhibitors established in this work is feasible and efficient, providing a significant complementary approach to guide drug development.

#### Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence



the work reported in this paper.

## Data availability

Data will be made available on request.

## Acknowledgments

This study was supported in part by the Chinese Academy of Medical Sciences (2021-I2M5–014) and the Ministry of Science and Technology of China (2016YFA0502303).

## Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.ejmech.2022.114803>.

## References

- [1] F. Wu, S. Zhao, B. Yu, Y.M. Chen, W. Wang, Z.G. Song, Y. Hu, Z.W. Tao, J.H. Tian, Y.Y. Pei, M.L. Yuan, Y.L. Zhang, F.H. Dai, Y. Liu, Q.M. Wang, J.J. Zheng, L. Xu, E. C. Holmes, Y.Z. Zhang, A new coronavirus associated with human respiratory disease in China, *Nature* 579 (7798) (2020) 265, <https://doi.org/10.1038/s41586-020-2008-3>.
- [2] W.T. Harvey, A.M. Carabelli, B. Jackson, R.K. Gupta, E.C. Thomson, E.M. Harrison, C. Ludden, R. Reeve, A. Rambaut, S.J. Peacock, D.L. Robertson, C.-G.U.C.-U. Conso, SARS-CoV-2 variants, spike mutations and immune escape, *Nat. Rev. Microbiol.* 19 (7) (2021) 409–424, <https://doi.org/10.1038/s41579-021-00573-0>.
- [3] C. Ma, Z. Xia, M.D. Sacco, Y. Hu, J.A. Townsend, X. Meng, J. Choza, H. Tan, J. Jang, M.V. Gongora, X. Zhang, F. Zhang, Y. Xiang, M.T. Marty, Y. Chen, J. Wang, Discovery of Di- and trihaloacetamides as covalent SARS-CoV-2 main protease inhibitors with high target specificity, *J. Am. Chem. Soc.* 143 (49) (2021) 20697–20709, <https://doi.org/10.1021/jacs.1c08060>.
- [4] J.H. Beigel, K.M. Tomashek, L.E. Dodd, A.K. Mehta, B.S. Zingman, A.C. Kaili, E. Hohmann, H.Y. Chu, A. Luetkemeyer, S. Kline, D.L. de Castilla, R.W. Finberg, K. Dieterberg, V. Tapson, L. Hsieh, T.F. Patterson, R. Paredes, D.A. Sweeney, W. R. Short, G. Touloumi, D.C. Lye, N. Ohmagari, M.D. Oh, G.M. Ruiz-Palacios, T. Benfield, G. Fatkenheuer, M.G. Kortepeter, R.L. Atmar, C.B. Creech, J. Lundgren, A.G. Babiker, S. Pett, J.D. Neaton, T.H. Burgess, T. Bonnett, M. Green, M. Makowski, A. Osinusi, S. Nayak, H.C. Lane, A.-S. Grp, Remdesivir for the treatment of covid-19-final report, *N. Engl. J. Med.* 383 (19) (2020) 1813–1826, <https://doi.org/10.1056/NEJMoa2007764>.
- [5] A.J. Bernal, M.M.G. da Silva, D.B. Musungaie, E. Kovalchuk, A. Gonzalez, V. Delos Reyes, A. Martin-Quiros, Y. Caraco, A. Williams-Diaz, M.L. Brown, J. Du, A. Pedley, C. Assaid, J. Strizki, J.A. Grobler, H.H. Shamsuddin, R. Tipping, H. Wan, A. Paschke, J.R. Butterson, M.G. Johnson, C. De Anda, M.-O.S. Grp, Molnupiravir for oral treatment of covid-19 in nonhospitalized patients, *N. Engl. J. Med.* (2021), <https://doi.org/10.1056/NEJMoa2116044>.
- [6] D.R. Owen, C.M.N. Allerton, A.S. Anderson, L. Aschenbrenner, M. Avery, S. Berritt, B. Boras, R.D. Cardin, A. Carlo, K.J. Coffman, A. Dantonio, L. Di, H. Eng, R. Ferre, K.S. Gajiwala, S.A. Gibson, S.E. Greasley, B.L. Hurst, E.P. Kadar, A.S. Kalgutkar, J. C. Lee, J. Lee, W. Liu, S.W. Mason, S. Noell, J.J. Novak, R.S. Obach, K. Ogilvie, N. C. Patel, M. Pettersson, D.K. Rai, M.R. Reese, M.F. Sammons, J.G. Sathish, R.S. P. Singh, C.M. Steppan, A.E. Stewart, J.B. Tuttle, L. Updyke, P.R. Verhoest, L. Wei, Q. Yang, Y. Zhu, An oral SARS-CoV-2 M(pro) inhibitor clinical candidate for the treatment of COVID-19, *Science* 374 (6575) (2021) 1586–1593, <https://doi.org/10.1126/science.abb4784>.
- [7] D.A. Brian, R.S. Baric, Coronavirus genome structure and replication, *Curr. Top. Microbiol. Immunol.* 287 (2005) 1–30, <https://doi.org/10.1007/3-540-26765-4-1>.
- [8] C. Li, Y. Qi, X. Teng, Z. Yang, P. Wei, C. Zhang, L. Tan, L. Zhou, Y. Liu, L. Lai, Maturation mechanism of severe acute respiratory syndrome (SARS) coronavirus 3C-like proteinase, *J. Biol. Chem.* 285 (36) (2010) 28134–28140, <https://doi.org/10.1074/jbc.M109.095851>.
- [9] K. Fan, L. Ma, X. Han, H. Liang, P. Wei, Y. Liu, L. Lai, The substrate specificity of SARS coronavirus 3C-like proteinase, *Biochem. Biophys. Res. Commun.* 329 (3) (2005) 934–940, <https://doi.org/10.1016/j.bbrc.2005.02.061>.
- [10] Z. Jin, X. Du, Y. Xu, Y. Deng, M. Liu, Y. Zhao, B. Zhang, X. Li, L. Zhang, C. Peng, Y. Duan, J. Yu, L. Wang, K. Yang, F. Liu, R. Jiang, X. Yang, T. You, X. Liu, X. Yang, F. Bai, H. Liu, X. Liu, L.W. Guddat, W. Xu, G. Xiao, C. Qin, Z. Shi, H. Jiang, Z. Rao, H. Yang, Structure of M(pro) from SARS-CoV-2 and discovery of its inhibitors, *Nature* 582 (7811) (2020) 289–293, <https://doi.org/10.1038/s41586-020-2223-y>.
- [11] K. Akaji, H. Konno, H. Mitsui, K. Teruya, Y. Shimamoto, Y. Hattori, T. Ozaki, M. Kusunoki, A. Sanjoh, Structure-based design, synthesis, and evaluation of peptide-mimetic SARS 3CL protease inhibitors, *J. Med. Chem.* 54 (23) (2011) 7962–7973, <https://doi.org/10.1021/jm200870n>.
- [12] C. Wu, Y. Liu, Y. Yang, P. Zhang, W. Zhong, Y. Wang, Q. Wang, Y. Xu, M. Li, X. Li, M. Zheng, L. Chen, H. Li, Analysis of therapeutic targets for SARS-CoV-2 and discovery of potential drugs by computational methods, *Acta Pharm. Sin.* 41 (10) (2020) 766–788, <https://doi.org/10.1016/j.apsb.2020.02.008>.
- [13] R. Hilgenfeld, From SARS to MERS: crystallographic studies on coronaviral proteases enable antiviral drug design, *FEBS J.* 281 (18) (2014) 4085–4096, <https://doi.org/10.1111/febs.12936>.
- [14] V. Kumar, J.S. Shin, J.J. Shie, K.B. Ku, C. Kim, Y.Y. Go, K.F. Huang, M. Kim, P. H. Liang, Identification and evaluation of potent Middle East respiratory syndrome coronavirus (MERS-CoV) 3CL(Pro) inhibitors, *Antivir. Res.* 141 (2017) 101–106, <https://doi.org/10.1016/j.antiviral.2017.02.007>.
- [15] C.J. Kuo, H.G. Liu, Y.K. Lo, C.M. Seong, K.I. Lee, Y.S. Jung, P.H. Liang, Individual and common inhibitors of coronavirus and picornavirus main proteases, *FEBS Lett.* 583 (3) (2009) 549–555, <https://doi.org/10.1016/j.febslet.2008.12.059>.
- [16] Z. Lu, J.I.N. Feng, L.I.U. Ying, S. Er-Chang, W.E.I. Ping, L.I. Chun-Mei, L.A.I. Lu-Hua, Isatin dual functional inhibitors: modulating the aggregation state and enzyme activity of SARS-3CL proteinase, *Acta Phys. Chim. Sin.* 28 (10) (2012) 2418–2422, <https://doi.org/10.3866/pku.Whxb201209143>.
- [17] J.Y. Park, J.A. Ko, D.W. Kim, Y.M. Kim, H.J. Kwon, H.J. Jeong, C.Y. Kim, K.H. Park, W.S. Lee, Y.B. Ryu, Chalcones isolated from Angelica keiskei inhibit cysteine proteases of SARS-CoV, *J. Enzym. Inhib. Med. Chem.* 31 (1) (2016) 23–30, <https://doi.org/10.3109/14756366.2014.1003215>.
- [18] C.Y. Wu, K.Y. King, C.J. Kuo, J.M. Fang, Y.T. Wu, M.Y. Ho, C.L. Liao, J.J. Shie, P. H. Liang, C.H. Wong, Stable benzotriazole esters as mechanism-based inactivators of the severe acute respiratory syndrome 3CL protease, *Chem Biol* 13 (3) (2006) 261–268, <https://doi.org/10.1016/j.chembiol.2005.12.008>.
- [19] J. Zhang, C. Huitema, C. Niu, J. Yin, M.N. James, L.D. Eltis, J.C. Vederas, Aryl methylene ketones and fluorinated methylene ketones as reversible inhibitors for severe acute respiratory syndrome (SARS) 3C-like proteinase, *Bioorg. Chem.* 36 (5) (2008) 229–240, <https://doi.org/10.1016/j.bioorg.2008.01.001>.
- [20] S. De Cesco, J. Kurian, C. Dufresne, A.K. Mittermaier, N. Moïstessier, Covalent inhibitors design and discovery, *Eur. J. Med. Chem.* 138 (2017) 96–114, <https://doi.org/10.1016/j.ejmech.2017.06.019>.
- [21] H. Su, S. Yao, W. Zhao, Y. Zhang, J. Liu, Q. Shao, Q. Wang, M. Li, H. Xie, W. Shang, C. Ke, L. Feng, X. Jiang, J. Shen, G. Xiao, H. Jiang, L. Zhang, Y. Ye, Y. Xu, Identification of pyrogallol as a warhead in design of covalent inhibitors for the SARS-CoV-2 3CL protease, *Nat. Commun.* 12 (1) (2021) 3623, <https://doi.org/10.1038/s41467-021-23751-3>.
- [22] R.L. Hoffman, R.S. Kania, M.A. Brothers, J.F. Davies, R.A. Ferre, K.S. Gajiwala, M. He, R.J. Hogan, K. Kozminski, L.Y. Li, J.W. Lockner, J. Lou, M.T. Marra, L. J. Mitchell Jr., B.W. Murray, J.A. Nieman, S. Noell, S.P. Planken, T. Rowe, K. Ryan, G.J. Smith 3rd, J.E. Solowiej, C.M. Steppan, B. Taggart, Discovery of ketone-based covalent inhibitors of coronavirus 3CL proteases for the potential therapeutic treatment of COVID-19, *J. Med. Chem.* 63 (21) (2020) 12725–12747, <https://doi.org/10.1021/acs.jmedchem.0c01063>.
- [23] W. Vuong, M.B. Khan, C. Fischer, E. Arutyunova, T. Lamer, J. Shields, H.A. Saffran, R.T. McKay, M.J. van Belkum, M.A. Joyce, H.S. Young, D.L. Tyrrell, J.C. Vederas, M.J. Lemieux, Feline coronavirus drug inhibits the main protease of SARS-CoV-2 and blocks virus replication, *Nat. Commun.* 11 (1) (2020) 4282, <https://doi.org/10.1038/s41467-020-18096-2>.
- [24] J. Qiao, Y.S. Li, R. Zeng, F.L. Liu, R.H. Luo, C. Huang, Y.F. Wang, J. Zhang, B. Quan, C. Shen, X. Mao, X. Liu, W. Sun, W. Yang, X. Ni, K. Wang, L. Xu, Z. Li, Duan, Q. C. Zou, H.L. Zhang, W. Qu, Y.H. Long, M.H. Li, R.C. Yang, X. Liu, J. You, Y. Zhou, R. Yao, W.P. Li, J.M. Liu, P. Chen, Y. Liu, G.F. Lin, X. Yang, J. Zou, L. Li, Y. Hu, G. W. Lu, W.M. Li, Y.Q. Wei, Y.T. Zheng, J. Lei, S. Yang, SARS-CoV-2 M(pro) inhibitors with antiviral activity in a transgenic mouse model, *Science* 371 (6536) (2021) 1374–1378, <https://doi.org/10.1126/science.abb1611>.
- [25] W. Dai, B. Zhang, X.M. Jiang, H. Su, J. Li, Y. Zhao, X. Xie, Z. Jin, J. Peng, F. Liu, C. Li, Y. Li, F. Bai, H. Wang, X. Cheng, X. Cen, S. Hu, X. Yang, J. Wang, X. Liu, G. Xiao, H. Jiang, Z. Rao, L.K. Zhang, Y. Xu, H. Yang, H. Liu, Structure-based design of antiviral drug candidates targeting the SARS-CoV-2 main protease, *Science* 368 (6497) (2020) 1331–1335, <https://doi.org/10.1126/science.abb4489>.
- [26] C. Ma, M.D. Sacco, B. Hurst, J.A. Townsend, Y. Hu, T. Sze, X. Zhang, B. Tarbet, M.T. Marty, Y. Chen, J. Wang, Boceprevir, GC-376, and calpain inhibitors II, XII inhibit SARS-CoV-2 viral replication by targeting the viral main protease, *Cell Res.* 30 (8) (2020) 678–692, <https://doi.org/10.1038/s41422-020-0356-z>.
- [27] Y. Unoh, S. Uehara, K. Nakahara, H. Nobori, Y. Yamatsu, S. Yamamoto, Y. Maruyama, Y. Taoda, K. Kasamatsu, T. Suto, K. Kouki, A. Nakahashi, S. Kawashima, T. Sanaki, S. Toba, K. Uemura, T. Mizutare, S. Ando, M. Sasaki, Y. Orba, H. Sawa, A. Sato, T. Sato, T. Kato, Y. Tachibana, Discovery of S-217622, a noncovalent oral SARS-CoV-2 3CL protease inhibitor clinical candidate for treating COVID-19, *J. Med. Chem.* (2022), <https://doi.org/10.1021/acs.jmedchem.2c00117>.
- [28] H.X. Su, S. Yao, W.F. Zhao, M.J. Li, J. Liu, W.J. Shang, H. Xie, C.Q. Ke, H.C. Hu, M. N. Gao, K.Q. Yu, H. Liu, J.S. Shen, W. Tang, L.K. Zhang, G.F. Xiao, L. Ni, D. W. Wang, J.P. Zuo, H.L. Jiang, F. Bai, Y. Wu, Y. Ye, Y.C. Xu, Anti-SARS-CoV-2 activities in vitro of Shuanghuanglian preparations and bioactive ingredients, *Acta Pharmacol. Sin.* 41 (9) (2020) 1167–1177, <https://doi.org/10.1038/s41401-020-0483-6>.
- [29] N. Drayman, J.K. DeMarco, K.A. Jones, S.A. Azizi, H.M. Froggatt, K. Tan, N. I. Maltseva, S. Chen, V. Nicolaescu, S. Dvorkin, K. Furlong, R.S. Kathayat, M. R. Firpo, V. Mastrodomenico, E.A. Bruce, M.M. Schmidt, R. Jedrzyszczak, M. A. Munoz-Alia, B. Schuster, V. Nair, K.Y. Han, A. O'Brien, A. Tomatsidou, B. Meyer, M. Vignuzzi, D. Missiakas, J.W. Botten, C.B. Brooker, H. Lee, S.C. Baker, B. C. Mounce, N.S. Heatson, W.E. Severson, K.E. Palmer, B.C. Dickinson, A. Joachimiak, G. Randall, S. Tay, Masitinib is a broad coronavirus 3CL inhibitor that blocks replication of SARS-CoV-2, *Science* 373 (6557) (2021) 931–936, <https://doi.org/10.1126/science.abb5827>.

- [30] A. Chaikwad, P. Koch, S.A. Laufer, S. Knapp, The cysteinome of protein kinases as a target in drug development, *Angew Chem. Int. Ed. Engl.* 57 (16) (2018) 4372–4385, <https://doi.org/10.1002/anie.201707875>.
- [31] P. Abranyi-Balogh, L. Petri, T. Imre, P. Sziñj, A. Scarpino, M. Hrst, A. Mitrovic, U. P. Fonovic, K. Nemeth, H. Barreteau, D.I. Roper, K. Horvati, G.G. Ferenczy, J. Kos, J. Ilas, S. Gobec, G.M. Keseru, A road map for prioritizing warheads for cysteine targeting covalent inhibitors, *Eur. J. Med. Chem.* 160 (2018) 94–107, <https://doi.org/10.1016/j.ejmech.2018.10.010>.
- [32] J. Chen, F. Ali, I. Khan, Y.Z. Zhu, Recent progress in the development of potential drugs against SARS-CoV-2, *Curr Res Pharmacol Drug Discov* 2 (2021), 100057, <https://doi.org/10.1016/j.crphar.2021.100057>.
- [33] G. Macip, P. Garcia-Segura, J. Mestres-Truyol, B. Saldivar-Espinoza, M.J. Ojeda-Montes, A. Gimeno, A. Cereto-Massague, S. Garcia-Vallve, G. Pujadas, Haste makes waste: a critical review of docking-based virtual screening in drug repurposing for SARS-CoV-2 main protease (M-pro) inhibition, *Med. Res. Rev.* 42 (2) (2022) 744–769, <https://doi.org/10.1002/med.21862>.
- [34] B.L. Vittoria, C. Imbesi, G. Irene, G. Cali, A. Bitto, New approaches and repurposed antiviral drugs for the treatment of the SARS-CoV-2 infection, *Pharmaceuticals* 14 (6) (2021), <https://doi.org/10.3390/phi14060503>.
- [35] S. Wang, Q. Sun, Y. Xu, J. Pei, L. Lai, A transferable deep learning approach to fast screen potential antiviral drugs against SARS-CoV-2, *Briefings Bioinf.* 22 (6) (2021), <https://doi.org/10.1093/bib/bbab301>.
- [36] D.D. Nguyen, K. Gao, J. Chen, R. Wang, G.W. Wei, Unveiling the molecular mechanism of SARS-CoV-2 main protease inhibition from 137 crystal structures using algebraic topology and deep learning, *Chem. Sci.* 11 (44) (2020) 12036–12046, <https://doi.org/10.1039/d0sc04641h>.
- [37] F. Hu, L. Wang, Y. Hu, D. Wang, W. Wang, J. Jiang, N. Li, P. Yin, A novel framework integrating AI model and enzymological experiments promotes identification of SARS-CoV-2 3CL protease inhibitors and activity-based probe, *Briefings Bioinf.* 22 (6) (2021), <https://doi.org/10.1093/bib/bbab301>.
- [38] S. Ullrich, C. Nitsche, The SARS-CoV-2 main protease as drug target, *Bioorg. Med. Chem. Lett* 30 (17) (2020), 127377, <https://doi.org/10.1016/j.bmcl.2020.127377>.
- [39] Y. Liu, C. Liang, L. Xin, X. Ren, L. Tian, X. Ju, H. Li, Y. Wang, Q. Zhao, H. Liu, W. Cao, X. Xie, D. Zhang, Y. Wang, Y. Jian, The development of coronavirus 3C-Like protease (3CL(pro)) inhibitors from 2010 to 2020, *Eur. J. Med. Chem.* 206 (2020), 112711, <https://doi.org/10.1016/j.ejmech.2020.112711>.
- [40] Q. Sun, F. Ye, H. Liang, H. Liu, C. Li, R. Lu, B. Huang, L. Zhao, W. Tan, L. Lai, Bardoxolone and bardoxolone methyl, two Nrf2 activators in clinical trials, inhibit SARS-CoV-2 replication and its 3C-like protease, *Signal Transduct. Targeted Ther.* 6 (1) (2021) 212, <https://doi.org/10.1038/s41392-021-00628-x>.
- [41] A.S. Brown, D.F. Ackerley, M.J. Calcott, High-throughput screening for inhibitors of the SARS-CoV-2 protease using a FRET-biosensor, *Molecules* 25 (20) (2020), <https://doi.org/10.3390/molecules25204666>.
- [42] PubChem, SARS-CoV-2 3CL-Pro protease inhibition IC50 determined by FRET kind of response from peptide substrate, <https://pubchem.ncbi.nlm.nih.gov/bioassay/1640021>.
- [43] J. Gilmer, S.S. Schoenholz, P.F. Riley, O. Vinyals, G.E. Dahl, Neural message passing for quantum chemistry, in: P. Doina, T. Yee Whye (Eds.), *Proceedings of the 34th International Conference on Machine Learning*, PMLR: Proceedings of Machine Learning Research, vol. 70, 2017, pp. 1263–1272.
- [44] RDKit, open-source cheminformatics. <https://www.rdkit.org>.
- [45] D. Svovil, V. Kvasnicka, J.I. Pospichal, Introduction to multi-layer feed-forward neural networks, *Chemometr. Intell. Lab. Syst.* 39 (1) (1997) 43–62, [https://doi.org/10.1016/s0169-7439\(97\)00061-0](https://doi.org/10.1016/s0169-7439(97)00061-0).
- [46] H. Liu, F. Ye, Q. Sun, H. Liang, C. Li, S. Li, R. Lu, B. Huang, W. Tan, L. Lai, Scutellaria baicalensis extract and baicalein inhibit replication of SARS-CoV-2 and its 3C-like protease in vitro, *J. Enzym. Inhib. Med. Chem.* 36 (1) (2021) 497–503, <https://doi.org/10.1080/14756366.2021.1873977>.
- [47] K.H. Baek, R. Karki, E.S. Lee, Y. Na, Y. Kwon, Synthesis and investigation of dihydroxychalcones as calpain and cathepsin inhibitors, *Bioorg. Chem.* 51 (2013) 24–30, <https://doi.org/10.1016/j.bioorg.2013.09.002>.
- [48] S. Iketani, F. Forouhar, H. Liu, S.J. Hong, F.Y. Lin, M.S. Nair, A. Zask, Y. Huang, L. Xing, B.R. Stockwell, A. Chavez, D.D. Ho, Lead compounds for the development of SARS-CoV-2 3CL protease inhibitors, *Nat. Commun.* 12 (1) (2021), <https://doi.org/10.1038/s41467-021-22362-2>, 2016.
- [49] G. Madhavi Sastry, M. Adzhigirey, T. Day, R. Annabhimoju, W. Sherman, Protein and ligand preparation: parameters, protocols, and influence on virtual screening enrichments, *J. Comput. Aided Mol. Des.* 27 (3) (2013) 221–234, <https://doi.org/10.1007/s10822-013-9644-8>.
- [50] I.J. Chen, N. Foloppe, Drug-like bioactive structures and conformational coverage with the LigPrep/ConfGen suite: comparison to programs MOE and catalyst, *J. Chem. Inf. Model.* 50 (5) (2010) 822–839, <https://doi.org/10.1021/ci100026x>.
- [51] K. Zhu, K.W. Borrelli, J.R. Greenwood, T. Day, R. Abel, R.S. Farid, E. Harder, Docking covalent inhibitors: a parameter free approach to pose prediction and scoring, *J. Chem. Inf. Model.* 54 (7) (2014) 1932–1940, <https://doi.org/10.1021/ci500118s>.
- [52] K. Yang, K. Swanson, W. Jin, C. Coley, P. Eiden, H. Gao, A. Guzman-Perez, T. Hopper, B. Kelley, M. Mathea, A. Palmer, V. Settels, T. Jaakkola, K. Jensen, R. Barzilay, Analyzing learned molecular representations for property prediction, *J. Chem. Inf. Model.* 59 (8) (2019) 3370–3388, <https://doi.org/10.1021/acs.jcim.9b00237>.
- [53] L. Breiman, Random forests, *Mach. Learn.* 45 (1) (2001) 5–32, <https://doi.org/10.1023/a:1010933404324>.
- [54] C. Cortes, V. Vapnik, Support-vector networks, *Mach. Learn.* 20 (3) (1995) 273–297, <https://doi.org/10.1023/a:1022627411411>.
- [55] J.H. Friedman, Stochastic gradient boosting, *Comput. Stat. Data Anal.* 38 (4) (2002) 367–378, [https://doi.org/10.1016/s0167-9473\(01\)00065-2](https://doi.org/10.1016/s0167-9473(01)00065-2).
- [56] S.-C. Wang, Artificial neural network, in: S.-C. Wang (Ed.), *Interdisciplinary Computing in Java Programming*, Springer US, Boston, MA, 2003, pp. 81–100.
- [57] T.G. Dietterich, In Ensemble Methods in Machine Learning, Springer Berlin Heidelberg, Berlin, Heidelberg, 2000, pp. 1–15. Berlin, Heidelberg.
- [58] D. Butina, Unsupervised data base clustering based on Daylight’s fingerprint and Tanimoto similarity: a fast and automated way to cluster small and large data sets, *J. Chem. Inf. Comput. Sci.* 39 (4) (1999) 747–750, <https://doi.org/10.1021/ci9803381>.
- [59] R.C. Young, R.C. Mitchell, T.H. Brown, C.R. Ganellin, R. Griffiths, M. Jones, K. K. Rana, D. Saunders, I.R. Smith, N.E. Sore, et al., Development of a new physicochemical model for brain penetration and its application to the design of centrally acting H2 receptor histamine antagonists, *J. Med. Chem.* 31 (3) (1988) 656–671, <https://doi.org/10.1021/jm00398a028>.
- [60] Y. Yang, Z. Zhu, X. Wang, X. Zhang, K. Mu, Y. Shi, C. Peng, Z. Xu, W. Zhu, Ligand-based approach for predicting drug target sets and for virtual screening against COVID-19, *Briefings Bioinf.* 22 (2) (2021) 1053–1064, <https://doi.org/10.1093/bib/bbaa422>.
- [61] M.D. Sacco, C. Ma, P. Lagarias, A. Gao, J.A. Townsend, X. Meng, P. Dube, X. Zhang, Y. Hu, N. Kitamura, B. Hurst, B. Tarbet, M.T. Marty, A. Kolocouris, Y. Xiang, Y. Chen, J. Wang, Structure and inhibition of the SARS-CoV-2 main protease reveal strategy for developing dual inhibitors against M(pro) and cathepsin L, *Sci. Adv.* 6 (50) (2020), <https://doi.org/10.1126/sciadv.abe0751>.
- [62] A.S. Ashhurst, A.H. Tang, P. Fajtová, M.C. Yoon, A. Aggarwal, M.J. Bedding, A. Stoye, L. Beretta, D. Pwee, A. Drellich, D. Skinner, L. Li, T.D. Meek, J. H. McKerrow, V. Hook, C.T. Tseng, M. Larance, S. Turville, W.H. Gerwick, A. J. O’Donoghue, R.J. Payne, Potent anti-SARS-CoV-2 activity by the natural product gallinamide A and analogues via inhibition of cathepsin L, *J. Med. Chem.* 65 (4) (2022) 2956–2970, <https://doi.org/10.1021/acs.jmedchem.1c01494>.
- [63] R. Vicik, M. Busemann, C. Gelhaus, N. Stiefel, J. Scheiber, W. Schmitz, F. Schulz, M. Mladenovic, B. Engels, M. Leippe, K. Baumann, T. Schirmeister, Aziridine-based inhibitors of cathepsin L: synthesis, inhibition activity, and docking studies, *ChemMedChem* 1 (10) (2006) 1126–1141, <https://doi.org/10.1002/cmde.200600106>.
- [64] M. Xiong, T. Nie, Q. Shao, M. Li, H. Su, Y. Xu, In silico screening-based discovery of novel covalent inhibitors of the SARS-CoV-2 3CL protease, *Eur. J. Med. Chem.* 231 (2022), 114130, <https://doi.org/10.1016/j.ejmech.2022.114130>.
- [65] Y. Shin, J.W. Jeong, R.P. Wurz, P. Achanta, T. Arvedson, M.D. Bartberger, I.D. G. Campuzano, R. Fucini, S.K. Hansen, J. Ingersoll, J.S. Iwig, J.R. Lipford, V. Ma, D. J. Kopecky, J. McCarter, T. San Miguel, C. Mohr, S. Sabet, A.Y. Saiki, A. Sawayama, S. Sethofer, C.M. Tegley, L.P. Volak, K. Yang, B.A. Lanman, D.A. Erlanson, V.J. Cee, Discovery of N-(1-Acryloylazetidino-3-yl)-2-(1H-indol-1-yl)acetamides as covalent inhibitors of KRAS(G12C), *ACS Med. Chem. Lett.* 10 (9) (2019) 1302–1308, <https://doi.org/10.1021/acsmedchemlett.9b00258>.
- [66] S. Konno, K. Kobayashi, M. Senda, Y. Funai, Y. Seki, I. Tamai, L. Schäkel, K. Sakata, T. Pillaiyar, A. Taguchi, A. Taniguchi, M. Gütschow, C.E. Müller, K. Takeuchi, M. Hirohama, A. Kawaguchi, M. Kojima, T. Senda, Y. Shirasaka, W. Kamitani, Y. Hayashi, 3CL protease inhibitors with an electrophilic arylketone moiety as anti-SARS-CoV-2 agents, *J. Med. Chem.* 65 (4) (2022) 2926–2939, <https://doi.org/10.1021/acs.jmedchem.1c00665>.
- [67] B.X. Qian, H. Shuai, A.J. Xia, Y. Hou, R. Zeng, X.L. Liu, G.F. Lin, J.X. Qiao, W.P. Li, F.L. Wang, K. Wang, R.J. Zhou, T.T. Yuen, M.X. Chen, C. Yoon, M. Wu, S.Y. Zhang, C. Huang, Y.F. Wang, W. Yang, C. Tian, W.M. Li, Y.Q. Wei, K.Y. Yuen, J.F. Chan, J. Lei, H. Chu, S. Yang, An orally available M(pro) inhibitor is effective against wild-type SARS-CoV-2 and variants including Omicron, *Nat Microbiol* 7 (5) (2022) 716–725, <https://doi.org/10.1038/s41564-022-01119-7>.
- [68] W. Rut, Z. Lv, M. Zmudzinski, S. Patchett, D. Nayak, S.J. Snipas, F. El Oualid, T. T. Huang, M. Bekes, M. Drag, S.K. Olsen, Activity profiling and crystal structures of inhibitor-bound SARS-CoV-2 papain-like protease: a framework for anti-COVID-19 drug design, *Sci. Adv.* 6 (42) (2020), <https://doi.org/10.1126/sciadv.abd4596>.
- [69] C. Hognon, M. Marazzi, C. Garcia-Iriepa, Atomistic-level description of the covalent inhibition of SARS-CoV-2 papain-like protease, *Int. J. Mol. Sci.* 23 (10) (2022), <https://doi.org/10.3390/ijms23105855>.