



RESEARCH PAPER



## Ligand-based discovery of coronavirus main protease inhibitors using MACAW molecular embeddings

Jie Dong<sup>a</sup> , Mihayl Varbanov<sup>b,c</sup>, Stéphanie Philippot<sup>b</sup>, Fanny Vreken<sup>b</sup>, Wen-bin Zeng<sup>a</sup> and Vincent Blay<sup>d</sup> 

<sup>a</sup>Xiangya School of Pharmaceutical Sciences, Central South University, Changsha, P. R. China; <sup>b</sup>Université de Lorraine, CNRS, Nancy, France; <sup>c</sup>Laboratoire de Virologie, CHRU de Nancy Brabois, Vandoeuvre-lès-Nancy, France; <sup>d</sup>Department of Microbiology and Environmental Toxicology, University of California at Santa Cruz, Santa Cruz, CA, USA

### ABSTRACT

Ligand-based drug design methods are thought to require large experimental datasets to become useful for virtual screening. In this work, we propose a computational strategy to design novel inhibitors of coronavirus main protease, M<sup>Pro</sup>. The pipeline integrates publicly available screening and binding affinity data in a two-stage machine-learning model using the recent MACAW embeddings. Once trained, the model can be deployed to rapidly screen large libraries of molecules *in silico*. Several hundred thousand compounds were virtually screened and 10 of them were selected for experimental testing. From these 10 compounds, 8 showed a clear inhibitory effect on recombinant M<sup>Pro</sup>, with half-maximal inhibitory concentration values (IC<sub>50</sub>) in the range 0.18–18.82 μM. Cellular assays were also conducted to evaluate cytotoxic, haemolytic, and antiviral properties. A promising lead compound against coronavirus M<sup>Pro</sup> was identified with dose-dependent inhibition of virus infectivity and minimal toxicity on human MRC-5 cells.

### ARTICLE HISTORY

Received 5 August 2022  
Revised 24 September 2022  
Accepted 29 September 2022

### KEYWORDS

Coronavirus; ligand-based drug design; machine learning; cheminformatics; drug discovery

### Introduction



With over 610 million reported human infections and 6.5 million deaths<sup>1</sup>, SARS-CoV-2, also denoted 2019-nCoV, is a positive-sense single-stranded RNA enveloped virus responsible for the ongoing COVID-19 pandemic. The virus mainly infects the respiratory system, where it can cause acute respiratory distress syndrome (ARDS) and fatal respiratory failure in some patients<sup>2</sup>, and it can also have long-lasting effects on other organs and systems, including long-term comorbidities (neurological disorders, memory loss, gastrointestinal distress, fatigue, insomnia, dyspnea) and post-acute sequelae of COVID-19 (PASC)<sup>3–5</sup>. Given the spread and severity of the disease, it is crucial to develop efficient treatments and rapidly available solutions that can supplement active immunisation efforts, which are still challenged by high viral transmissibility, re-infection, and immune escape variants<sup>6,7</sup>.

A variety of medicinal targets to fight infection by SARS-CoV-2 are being investigated<sup>8,9</sup>, and the main viral protease M<sup>Pro</sup> is particularly promising. After the virus infects and enters a human host cell, the two main ORF1a/b of its RNA genome first translate and express two polyprotein precursors (pp1a and pp1ab) with the help of the host cell machinery<sup>10–14</sup>. The polyprotein precursor undergoes intramolecular cleavage under the action of the main protease of SARS-CoV-2, M<sup>Pro</sup> (also known as 3C-like protease or 3CL<sup>Pro</sup>), and the papain-like protease, PL<sup>Pro</sup>, to produce multiple non-structural proteins (Nsps), Nsp1 to Nsp16. Some of the non-structural proteins produced participate in the production of viral subgenomic RNA encoding the four major structural proteins (Envelope/E protein, Membrane/M protein, Spike/S protein, and

Nucleocapsid/N protein), which are needed to complete the reproduction and release of progeny viruses<sup>10–14</sup>.

At present, few antiviral drugs against SARS-CoV-2 are in or close to clinical use<sup>15–17</sup>. These include molnupiravir (Merck)<sup>18,19</sup>, Paxlovid (Pfizer)<sup>20</sup>, and PF-07304814 (Novartis)<sup>21</sup>. Paxlovid is a combination of the M<sup>Pro</sup> inhibitor nirmatrelvir (PF-07321332) and ritonavir (a CYP3A4 inhibitor that slows down clearance of nirmatrelvir)<sup>20</sup>. The potential of drug–drug interactions from ritonavir, however, may limit its use by many patients<sup>22</sup>. PF-07304814 is a prodrug under clinical trials that improves the pharmacokinetics of PF-00835231, another M<sup>Pro</sup> inhibitor<sup>21</sup>. Notwithstanding the advances, the possibility of resistance, the potential of combination therapies for treatment and prophylaxis, and the complexities of global logistics of the pandemic demand the development of additional antiviral therapies<sup>23–25</sup>. M<sup>Pro</sup> plays a vital role in mediating virus replication and transcription, and there is no homologous protein in humans. Besides, M<sup>Pro</sup> is highly conserved across different coronaviruses (alpha-, beta-, and gamma-coronaviruses)<sup>26,27</sup> that might cause epidemics in the future. Thus, M<sup>Pro</sup> is an excellent target for the development of novel antiviral drugs against SARS-CoV-2, its variants, and other coronaviruses<sup>16,28–31</sup>.

Computational tools have played a prominent role in proposing potential M<sup>Pro</sup> inhibitors recently<sup>32–40</sup>. Most of these have leveraged structure-based drug design (SBDD) approaches, such as molecular docking, which rely on a structural model of M<sup>Pro</sup> to propose new ligands. However, few of these proposals have been experimentally validated. Besides, the protease is a complex target and computational and experimental efforts so far have faced low success rates<sup>40–43</sup>. On the other hand, the potential of

**CONTACT** Vincent Blay  [vroger@ucsc.edu](mailto:vroger@ucsc.edu)  Department of Microbiology and Environmental Toxicology, University of California at Santa Cruz, Santa Cruz, CA 95064-1077, USA.

 Supplemental data for this article is available online at <https://doi.org/10.1080/14756366.2022.2132486>.

© 2022 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

ligand-based drug design (LBDD) has been little explored due to the scarcity of experimental data available early on during the pandemic. In the past months, some initiatives, such as NCATS' COVID-19 OpenData Portal (<https://opendata.ncats.nih.gov/covid19/>) or PostEra's Moonshot ([https://covid.postera.ai/covid/activity\\_data](https://covid.postera.ai/covid/activity_data)), have been facilitating larger datasets that can enable this type of ligand-based approaches.

In this study, we explore a two-stage ligand-based drug design strategy for the discovery of novel M<sup>Pro</sup> inhibitors. We use MACAW embeddings, a recently proposed method to describe molecules computationally<sup>44</sup>, and build two predictive models on a curated compilation of experimental data. The models are applied in series to computationally assess hundreds of thousands of drug-like compounds. The most promising compounds are then sourced and evaluated experimentally for their ability to inhibit the viral protease and to arrest infection of human cells. The experimental results confirm the effectiveness of the virtual screening strategy, which can lead to rapid discovery of hit compounds with clinical potential against coronaviruses.

## Materials and methods

### Datasets

To model binding to SARS-CoV-2 M<sup>Pro</sup>, we compiled two datasets from a variety of sources: a "regression dataset" and a "classification dataset". For the regression dataset, we retrieved IC<sub>50</sub> values from different studies<sup>45–49</sup>. IC<sub>50</sub> values were converted to pK<sub>i</sub> values using the Cheng-Prusoff equation, taking into account the substrate concentration and corresponding K<sub>M</sub> values for each study (Equation (2)). The dataset was complemented with molecules from BindingDB (<https://www.bindingdb.org/>) with K<sub>i</sub> values against M<sup>Pro</sup>. As a result, we compiled 1716 molecules with their corresponding pK<sub>i</sub> values (regression dataset).

$$pK_i = -\log_{10} \left( \frac{IC_{50}}{1 + \frac{[S]}{K_M}} \right) \quad (1)$$

To generate the classification dataset, we combined fluorescence primary HTS data from the Kuzikov and Zhu studies<sup>46,47</sup>. For the Zhu study, we considered as hits those molecules with pK<sub>i</sub> >5.0 (labelled as 1). Molecules with lower pK<sub>i</sub> values or with class 4 curves were considered non-binders (labelled as 0). For the Kuzikov study, we considered hits those molecules from the primary screening that were selected for subsequent hit confirmation, and non-hits otherwise. This corresponds roughly to the top 3% of the molecules tested. In addition, to complete the binary screening data, we looked at the pK<sub>i</sub> data from the regression dataset compiled above (excluding the molecules from the Kuzikov and Zhu studies). We labelled the molecules as hits if they showed pK<sub>i</sub> >5.0 and non-binders otherwise. Finally, we added molecules from BindingDB. Those molecules with pK<sub>i</sub> >5.0 or pIC<sub>50</sub> ≥5.0 were labelled as hits, whereas those with pK<sub>i</sub> ≤5.0 or pIC<sub>50</sub> <4.0 were labelled as non-binders. Note that, from Equation (2), pIC<sub>50</sub> ≤ pK<sub>i</sub>. This way, the classification dataset comprised 22 376 molecules labelled as hits (1) or non-binders (0).

Lastly, to discover new potential inhibitors of M<sup>Pro</sup>, we compiled a virtual library composed of the Enamine Premium Collection (45 664 compounds), the Asinex Gold and Platinum Collections (261 120 compounds), and the DrugBank database of compounds (11 834 compounds)<sup>50</sup>. These are all lead- and drug-like molecules with favourable physicochemical properties (high Fsp3, low LogP, and MW, most of them satisfying Lipinski's rule of 5).

### Hit classification

4261 molecules were downsampled from the screening dataset to attain a better balance between classes (1261 hits and 3000 non-hits). 20-D MACAW predictors were computed for each molecule. MACAW is a small-molecule embedding method that allows projecting molecules into a continuous, low-dimensional numerical space while extracting relevant molecular characteristics from the training dataset<sup>44</sup>. The distance between molecules in the embedding was defined by the combination of MACCS fingerprints and the Sokal similarity metric. 10% of the molecules were randomly held out as a test set. The MACAW predictors were then used to train a distance-weighted *k*-Nearest Neighbour classifier from scikit-learn 0.24.1<sup>51</sup>, with *k* = 10 and using a Euclidean distance metric (*p* = 2). Details are provided in the accompanying Jupyter Notebook 1.

### Affinity prediction

All molecules in the affinity dataset were projected into a 20-D embedding space using MACAW 1.0<sup>44</sup>. The distance between molecules in this case was specified by the combination of featurised Morgan fingerprints of radius 2 (featMorgan2) and the Tanimoto similarity metric. 10% of the molecules were randomly held out as the test set. The MACAW embeddings were used as inputs to a support vector regressor (SVR) from scikit-learn 0.24.1, with hyperparameters *C* = 3 and *ε* = 0.2. Details are provided in the accompanying Jupyter Notebook 1.

### ADMET property prediction

Poor pharmacokinetics is a key cause of drug candidate attrition. We used ADMETlab 2.0<sup>52</sup> (<https://admetmesh.scbdd.com>) and SwissADME<sup>53</sup> (<http://www.swissadme.ch>) to predict ADMET properties and help prioritise the compounds after the two-stage target binding predictions. We devised a custom scoring rule (ADMET\_score) using KNIME (<https://www.knime.com>), which assigns specific weights to different ADMET properties based on the desirable range for each property. Higher ADMET\_score values indicate a better predicted pharmacokinetic profile (see file SI01.xlsx).

### Molecular docking

Molecular docking was used to predict the binding mode of selected molecules against SARS-CoV-2 M<sup>Pro</sup> using the software MOE (Molecular Operating Environment) version 2019 (<https://www.chemcomp.com>). The crystal structure of SARS-CoV-2 M<sup>Pro</sup> in complex with N3 (PDB: 7BQY)<sup>54</sup> was used as the template. The protein structure was first prepared by correcting amino acid residues, assigning ionisation states, and positioning hydrogens. Then the binding site of N3 was used as the pocket and each selected compound was docked against the pocket. Docking settings were chosen to reproduce the crystal pose of N3. In particular, the "Triangle Matcher" method with "London dG" score was used for placement, and the "Rigid Receptor" with "GBVI/WSA dG" score was used for replacement. 100 and 30 were set for the retained pose for the two stages. The best-scoring poses was considered the most likely binding mode. Further, to explore the potential of selected compounds binding to the II and III domains of the protein, we used the "site finder" to find a proper binding pocket there, and then the same pipeline was applied as above.

### Compound prioritisation

Compounds from the virtual screening libraries were prioritised for experimental testing based on their predicted classification probability of being a hit, their predicted binding affinity to M<sup>Pro</sup>, and their predicted ADMET profile. In particular, compounds were considered of interest if they had a predicted hit probability  $\geq 65\%$  and a predicted  $pK_i \geq 6.1$ . This led to 105 compounds (Supporting Information file SI01.xlsx). The compounds were then grouped in 15 clusters based on the pairwise similarity of their FragFp fingerprints using the software OSIRIS Datawarrior 5.2.1. At this stage, compounds were deemed most desirable if they exhibited predicted  $pK_i \geq 6.5$ , hit probability  $\geq 70\%$ , ADMET\_Score  $\geq 25$  (or coming from the DrugBank collection), and belonged to different clusters. The results were inspected, and 10 compounds were manually selected for experimental testing. The compounds were sourced from Topscience and Cayman and dissolved in DMSO as 10 mM stocks. All compounds were at least 90% as analysed by the supplier.

### Enzymatic assays

A FRET-based enzymatic cleavage test (Beyotime P0315M) was used to assess the inhibitory capacity of 10 compounds that had been prioritised computationally (Table 1). Each well of a 96-well fluorescence plate was dispensed with 92  $\mu\text{L}$  of buffer, 1  $\mu\text{L}$  M<sup>Pro</sup>, and 5  $\mu\text{L}$  of test compound in DMSO at a suitable concentration, such that the DMSO concentration was low and constant. Each test compound was tested at 10 different concentrations (0.01, 0.02, 0.05, 0.1, 0.5, 1, 5, 10, 15, and 20  $\mu\text{M}$  final concentrations in the reaction mix). Lastly, 2  $\mu\text{L}$  of the fluorogenic peptide substrate (MCA-AVLQSGFR-Lys(Dnp)-Lys-NH<sub>2</sub>) was introduced to each well to initiate the experiment. After adding the substrate as the final component, the reaction mix was incubated in the dark at 37 °C for 5 min. Fluorescence of each well was then measured at an excitation wavelength of 325 nm and an emission wavelength of 393 nm using a Thermo Varioskan LUX spectrophotometer. Ebselen was used as a positive control, which is known to bind covalently to the active site of M<sup>Pro</sup><sup>55</sup>. Cleavage of the fluorogenic substrate by M<sup>Pro</sup> leads to a separation of the FRET donor (Dnp) and acceptor (MCA) labels, leading to an increase in fluorescence from the acceptor label. Inhibitory compounds reduce the rate at which the fluorogenic substrate is cleaved and the fluorescence signal produced. The average fluorescence value of different wells can be recorded as RFU<sub>blank</sub>, RFU<sub>100% enzyme</sub>, RFU<sub>positive</sub>, and RFU<sub>sample</sub>, respectively, where RFU stands for Relative Fluorescence Units. The percentage of inhibition for each compound *i* was calculated as follows:

$$\text{Inhibition}_i(\%) = \frac{\text{RFU}_{100\% \text{ enzyme}} - \text{RFU}_i}{\text{RFU}_{100\% \text{ enzyme}} - \text{RFU}_{\text{blank}}} \times 100 \quad (2)$$

To exclude inhibitors possibly acting as aggregators, a detergent-based control was performed on selected compounds by adding 0.1% of freshly prepared Triton X-100 to the reaction mixture<sup>54</sup>. All experiments were performed in triplicate, and the experimental data was analysed using GraphPad Prism (<https://www.graphpad.com>).

### Cell model and culture

For virus propagation, MRC-5 pulmonary fibroblasts (ECACC, ref. 05090501) were grown in antibiotic-free Minimum Essential Media

(MEM) with nonessential amino acids (Thermo Fisher Scientific), complemented with 10% foetal bovine serum (FBS, Eurobio), 2 mM glutamine (Sigma Aldrich), 1% nonessential amino acids (Gibco), and 1% sodium pyruvate (GE Healthcare). For the antiviral assays, the same medium was used, containing only 2% FBS.

### Virus

Human coronavirus hCoV-229 (PHE/NCPV 0310051v) was propagated and quantified in MRC-5 cells. Initial virus was titrated at 10<sup>4</sup> IP/mL according to the Reed and Muench method<sup>56</sup>. All virus stocks were stored at -70 °C until used.

### Cytotoxicity

First, cellular toxicity was evaluated on MRC-5 in culture for 72 h in the presence of the selected compounds. Cells seeded the day before at 10 000 cells/well in a 96-well plate were treated with decreasing compound concentrations, from 100 to 3.12  $\mu\text{g}/\text{mL}$ , in a culture medium containing only 2% FCS. Control wells of untreated or solvent-treated cells (DMSO), as well as blank cell-free wells, were added. After 72 h of incubation, cell viability was evaluated by the MTT test<sup>57</sup>.

### Cytopathogenic effect

The cells were seeded in 5 plates with 96 wells at 10 000 cells/well. After 24 h incubation, three dilutions of virus were carried out in cascade at 1/10, in culture medium (2% serum) with or without the test compound (at the highest concentration not inducing toxicity). The cell plates were emptied then the mats were treated with 100  $\mu\text{L}$  of these dilutions (a column per condition, i.e.  $n = 8$ ). An untreated control and a DMSO solvent control were included in each plate. The plates were incubated at 33 °C for 72 h. The cytopathic effect of the virus was observed under the microscope and the viral titres were determined according to the Reed and Muench method<sup>56</sup>. The plates were then stained with crystal violet and the cytopathic effect in each well was quantified thanks to the optical density at 540 nm. The percentage of cytopathic effect (% CPE) was calculated for treated and non-treated infected wells according to the formula:

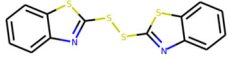
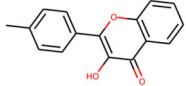
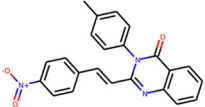
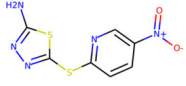
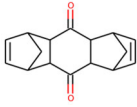
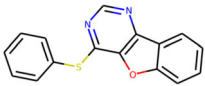
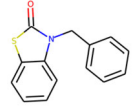
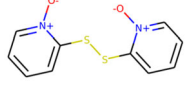
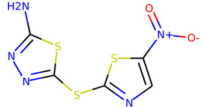
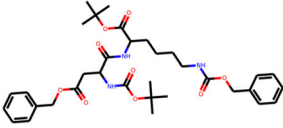
$$\text{CPE}_i(\%) = \frac{\text{OD}_{\text{sample},i} - \text{mean}(\text{OD}_{\text{blank}})}{\text{mean}(\text{OD}_{\text{control}})} \times 100 \quad (3)$$

where the control was the non-treated and non-infected cells.

### Viral titres

The infectivity of HCoV 229E was determined by titration in triplicate on 96-well microtiter plates containing 100  $\mu\text{L}$  of confluent MRC-5 cells. MRC-5 cells were added 100  $\mu\text{L}$  from serial 10-fold dilutions of the virus from 10<sup>1</sup> to 10<sup>-8</sup> in MEM medium with 2% FCS. The infected cells were incubated at 37 °C in 5% CO<sub>2</sub> for 72 h. The appearance of cytopathic effect (CPE) was recorded daily. The tissue culture infectious dose (50%) (TCID<sub>50</sub>), defined as the dilution of the virus required to infect 50% of the cell culture, was determined using the Reed and Muench method<sup>56</sup> and expressed as TCID<sub>50</sub>/mL.

**Table 1.** Information about the 10 compounds selected from the virtual screening of 408 935 lead-like molecules.

Id.	Structure	Catalog id. / CAS	Virtual screening	ADMET
1		316789	p(hit) = 1	MW = 331.96 HBD = 0
2		120-78-5	pKi = 6.453	logP = 5.298 TPSA = 25.78 HBA = 2 QED = 0.449 HBD = 1
		134210	p(hit) = 0.673	MW = 252.08
3		19275-68-4	pKi = 6.186	logP = 3.967 TPSA = 50.44 HBA = 3 QED = 0.721 HBD = 0
		103542	p(hit) = 0.791	MW = 383.13
4		433262-70-5	pKi = 6.134	logP = 4.105 TPSA = 78.03 HBA = 6 QED = 0.375 HBD = 2
		244527	p(hit) = 0.868	MW = 254.99
5		79134-17-1	pKi = 7.079	logP = 1.616 TPSA = 108.56 HBA = 7 QED = 0.636 HBD = 2
		126615	p(hit) = 0.798	MW = 240.12
6		5439-22-5	pKi = 6.143	logP = 3.873 TPSA = 40.46 HBA = 2 QED = 0.681 HBD = 0
		155928	p(hit) = 0.706	MW = 278.05
7		65023-97-4	pKi = 6.313	logP = 4.511 TPSA = 38.92 HBA = 3 QED = 0.505 HBD = 0
		230318	p(hit) = 0.705	MW = 241.06
8		22291-74-3	pKi = 6.661	logP = 3.514 TPSA = 22 HBA = 2 QED = 0.675 HBD = 0
		314478	p(hit) = 0.961	MW = 252
9		3696-28-4	pKi = 6.508	logP = 0.382 TPSA = 53.88 HBA = 4 QED = 0.474 HBD = 2
		317502 40045-50-9	p(hit) = 1	MW = 260.94
10		316703	pKi = 7.266	logP = 1.886 TPSA = 108.56 HBA = 7 QED = 0.642 HBD = 3
		174630-04-7	p(hit) = 0.904	MW = 641.33
			pKi = 6.143	logP = 4.649 TPSA = 158.36 HBA = 12 QED = 0.139

## Haemolysis

A haemolysis assay of the compounds was performed using human red blood cells (RBCs) to assess potential effects on the integrity of red blood cells. Selected compounds were diluted in PBS to relevant concentrations. RBCs were for 60 min at 37 °C. Two controls were used: a negative control containing RBCs and PBS, and a positive control in which RBCs were mixed with a product inducing haemolysis (Triton X-100 at 10%). After RBCs treatment, cells were centrifuged at 800 g for 5 min. The supernatant was recovered and measured with a spectrophotometer at 540 nm (Multiskan GO, Thermo Scientific, Saint Herblain, France).

## Results and discussion

### Virtual screens

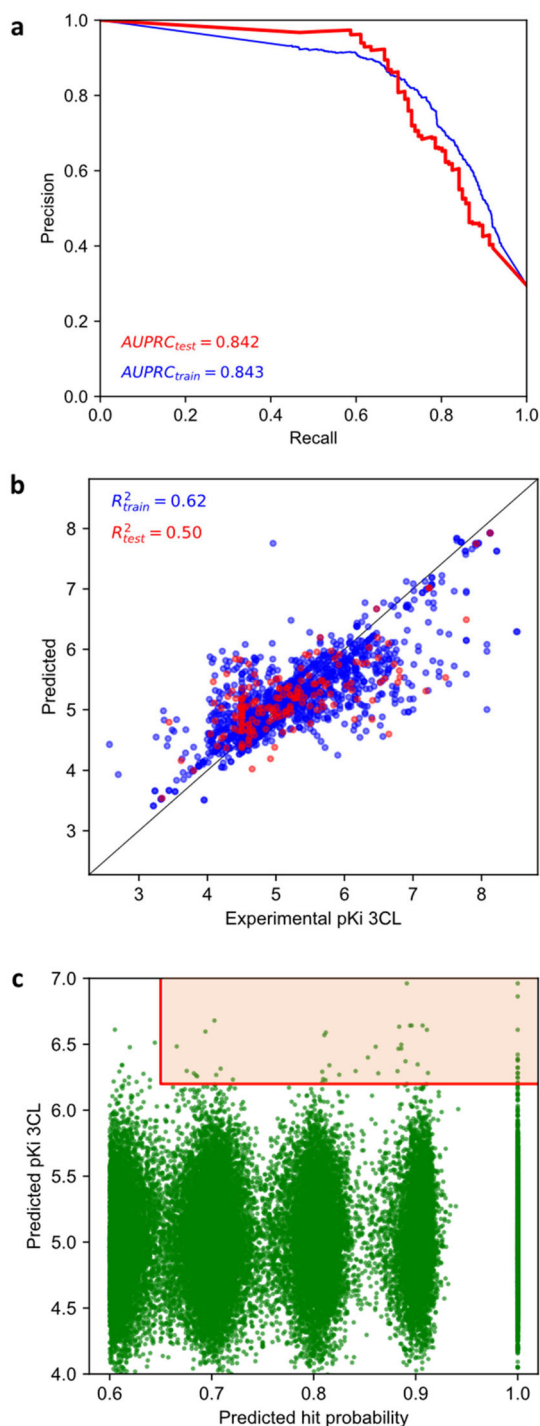
MACAW is a recent computational tool that allows the featurisation of molecules for use in predictive machine-learning models, as well as the generation of molecules based on a predefined property specification<sup>44</sup>. In this work, we used the featurisation capabilities of MACAW. Low-dimensional MACAW embeddings can be directly used as inputs to machine-learning models, without the need of variable cleaning or feature selection steps that are often needed with conventional molecular descriptors.

A two-stage virtual screening of compounds was used. Firstly, a classifier trained on HTS data mainly rejected molecules unlikely to make promising hits. Given the large class imbalance in the classification dataset, we used a random downsampling strategy. As a classifier, we used a simple *k*-nearest neighbour classifier, which attained a notable AUPRC of 84% on the test set (Figure 1(a)). This case illustrates the usefulness of MACAW embeddings to rapidly deploy powerful classifiers.

Secondly, a regressor trained on affinity binding constants sought to prioritise potent molecules amongst those that may show some binding to the target. For this, a SVR model was trained on the respective MACAW embeddings for the regression dataset. In this case, the challenge is that the number of molecules for which quantitative affinity binding data is currently available is limited. Moreover, data is generally obtained in the form of  $IC_{50}$ , which complicates comparison across different studies, as they often used different experimental conditions. The Cheng-Prusoff correction applied to the data can partially account for the difference (see “Methods” section), although it does not correct for other possible artefacts, like the potential dimerisation of  $M^{pro}$  *in vitro*<sup>58</sup>. Still, the regressor trained was able to achieve an  $R_{train}$  of almost 0.8 and an  $R_{test}$  above 0.7 (Figure 1(b)), which may contribute to increasing the enrichment provided by the virtual screening pipeline, which is the goal in this case.

With the two models (classifier and regressor) in place, we then interrogated a custom virtual library containing over 400 000 lead-like molecules. MACAW embeddings can be computed on a modest laptop in a few minutes and interesting molecules can be identified for further study (Figure 1(c)). The pipeline is elaborated in detail in the accompanying Jupyter Notebook 1.

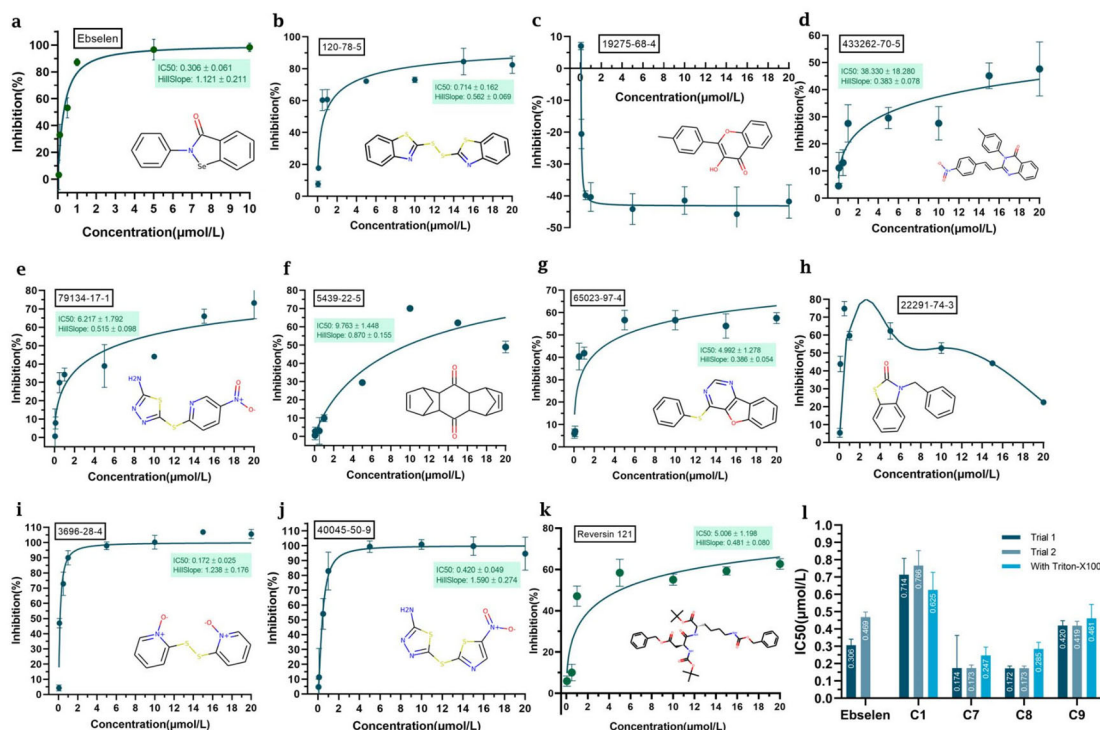
From this library, 105 compounds were first extracted based on their predicted scores by the first and second models (see Supporting Information file S101.xlsx). Some of the drugs highlighted in this list were identified as promising by early repurposing efforts. Notably, one of the highest ranked drugs in this list is ebselen, which has been reported as a potent inhibitor of  $M^{pro}$ <sup>55</sup>. Another highly ranked compound is tideglusib, which has also been reported to inhibit the viral protease<sup>59</sup>. Thiram has also been shown to have antiviral activity against the virus, although



**Figure 1.** MACAW embeddings can help identify molecules able to bind to SARS-CoV-2  $M^{pro}$ . (a) Precision-recall curve of a kNN hit classifier trained on MACAW embeddings applied to a test set of molecules. (b) A SVR regressor was also trained to predict  $pK_i$  values for promising molecules. (c) We computationally screened a custom library of 408 935 lead-like molecules and prioritised 105 that both the classifier and the regressor considered promising (orange region). See Jupyter Notebook 1 for details.

possibly through another mechanism<sup>60</sup>. Taken together, these observations suggest that the two-stage screening pipeline may have potential in identifying novel small-molecule inhibitors of  $M^{pro}$ .

Compounds were further shortlisted based on their predicted ADMET properties and chemical diversity. In particular, we kept



**Figure 2.** Selected compounds inhibit the activity of SARS-CoV-2  $\text{M}^{\text{Pro}}$ . The hydrolytic activity of SARS-CoV-2  $\text{M}^{\text{Pro}}$  was measured in the presence of increasing concentrations of different test compounds. (a) Ebselen. (b) Compound 1. (c) Compound 2. (d) Compound 4. (e) Compound 5. (f) Compound 6. (g) Compound 7. (h) Compound 8. (i) Compound 9. (j) Compound 10. (k) Compound 10. (l)  $\text{IC}_{50}$  values for Ebselen, Compound 1, Compound 7, Compound 8, and Compound 9. The dose-response curves and  $\text{IC}_{50}$  values were determined by nonlinear regression. All data are shown as mean  $\pm$  SEM,  $n = 3$  biological replicates.

compounds with  $\text{pK}_i \geq 6.5$ , hit probability  $\geq 70\%$ , and ADMET Score  $\geq 25$ , which were further selected to increase diversity (see “Methods” section). The 10 compounds finally selected, along with some of their predicted properties, like the number of hydrogen bond donors (HBD) and acceptors (HBA) or the quantitative estimate of druglikeness (QED)<sup>61</sup>, are shown in Table 1.

### Enzymatic assays

We evaluated the ability of the compounds shortlisted to inhibit the viral protease in an *in vitro* FRET-based assay using recombinant  $\text{M}^{\text{Pro}}$ . Active  $\text{M}^{\text{Pro}}$  can cleave the labelled peptide substrate, leading to a fluorescent signal. If a compound that inhibits the proteolytic activity is present, the fluorescence signal will be diminished. Out of the 10 selected compounds, 8 have a clear inhibitory effect on  $\text{M}^{\text{Pro}}$ , with half-maximal inhibitory concentration values ( $\text{IC}_{50}$ ) in the range 0.18–18.82  $\mu\text{M}$  (Figure 2). Among them, compounds 2 and 7 showed atypical inhibition results. Compound 8 (dipyrrithione) displayed the strongest inhibitory effect in this assay, with an  $\text{IC}_{50}$  value of 0.18  $\mu\text{M}$ . Notably, this value is even lower than that for the positive control with ebselen. Compounds 1 and 9 also displayed relatively low  $\text{IC}_{50}$  values of 0.71 and 0.42  $\mu\text{M}$ , respectively. Interestingly, compound 7 seemed to be highly potent at low concentrations, although the fluorescent signal increased at high compound concentrations. Thus, we decided to not discard it and keep it in our pipeline instead.

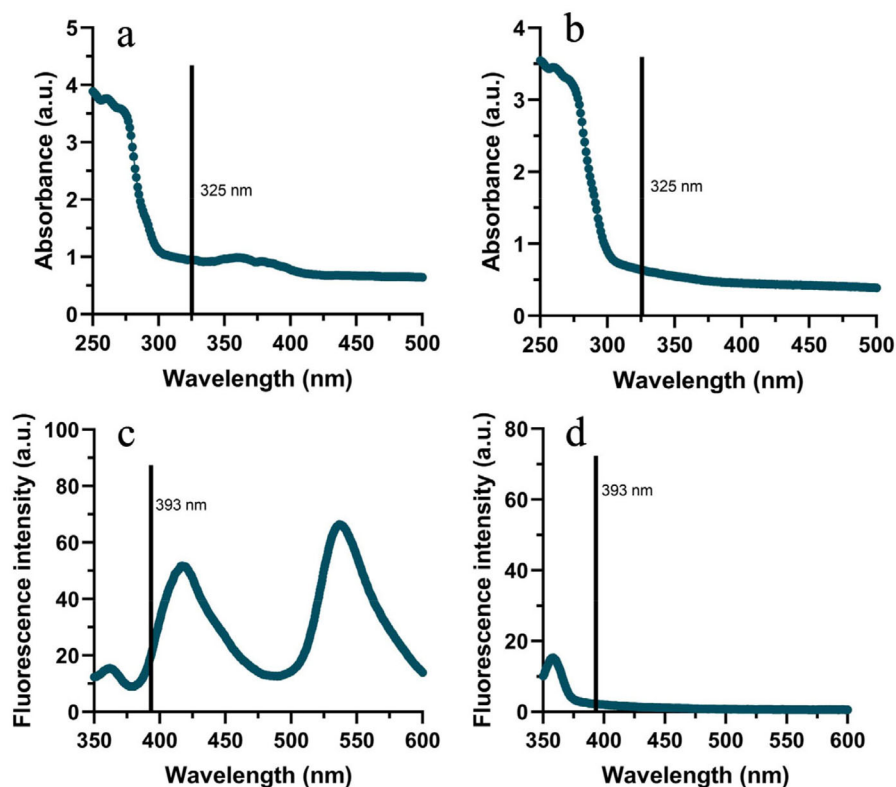
To analyse the potential effect of autofluorescence, UV-vis absorption spectra and fluorescent spectra of compounds 2 and 7 were measured (Figure 3). From Figure 3(a,c), we can see that compound 2 had some absorption at 325 nm (the excitation

wavelength used in the inhibition assay) and a significant fluorescence emission at 393 nm (the emission wavelength corresponding to the labelled substrate). Thus, autofluorescence may contribute to the atypical inhibition curve observed for compound 2 (Figure 2(c)). From Figure 4(b), we see that compound 7 has very weak absorption at 325 nm and it showed no fluorescence. This indicates that the inhibition curve for compound 7 is not affected by autofluorescence and other phenomena must explain its unusual trend.

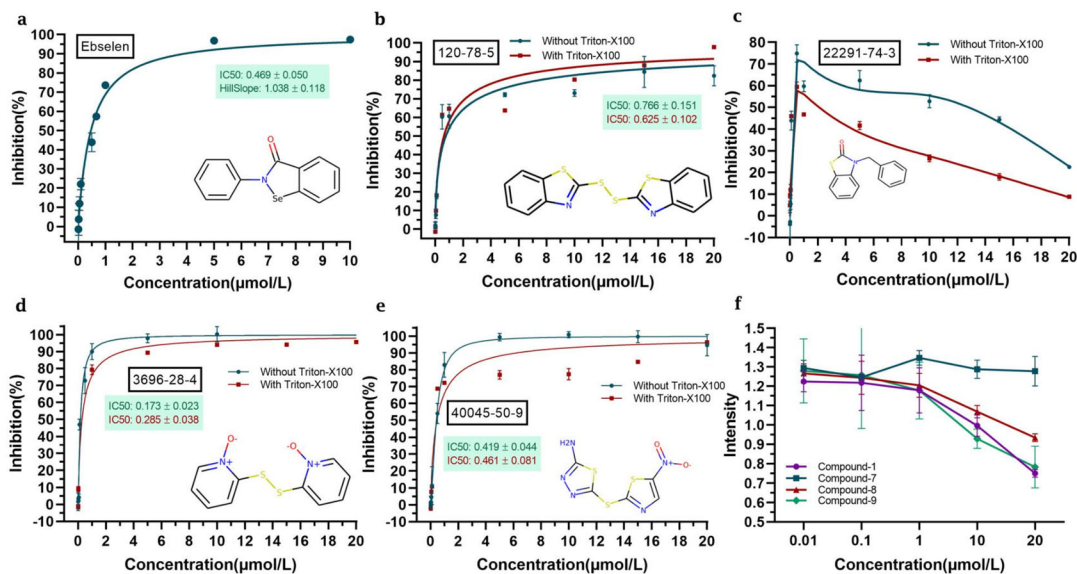
On the other hand, upon addition of 0.1% Triton X-100 detergent, the results for compounds 1, 7, 8, and 9 did not change significantly, confirming that they do not inhibit the viral protease by inducing aggregation (Figure 4). Compounds 1, 7, 8, and 9 were thus selected for further study.

### Cellular assays

The cellular toxicity of compounds 1, 7, 8, and 9 was evaluated on MRC-5 cells in culture for 72 h at varying concentrations, using cellular viability as a proxy (see “Methods” section). The results, shown in Figure 5, revealed that compound 9 is toxic. Compound 8 had limited toxicity, whereas compounds 1 and 7 showed minimal toxicity. The ADMET predictions in this work failed to flag compound 9 as toxic, indicating that it might be beneficial to conduct additional toxicity predictions *in silico* before prioritising compounds. In particular, nitroaromatic groups often display toxicity<sup>62</sup>. Interestingly, compound 9 was recently given the name “halicin” and found to have antibiotic properties through an unusual membrane associated-mechanism of action<sup>63</sup>. It is possible that this mechanism is toxic to some human cells as well.



**Figure 3.** Spectra of compounds 2 and 7. UV-vis absorption spectra of compound 2 (a) and compound 7 (b). Fluorescence emission spectra of compound 2 (c) and compound 7 (d) at an excitation wavelength of 325 nm.

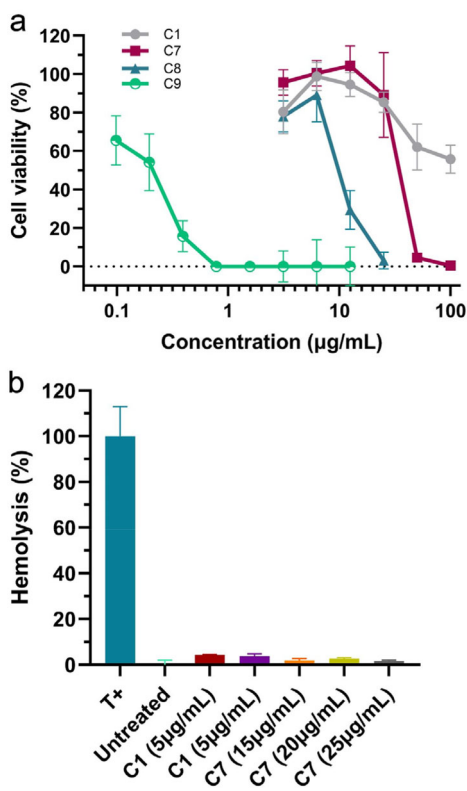


**Figure 4.** Addition of detergent does not affect the inhibition of SARS-CoV-2 M<sup>Pro</sup> by selected compounds. (a) The hydrolytic activity of SARS-CoV-2 M<sup>Pro</sup> was measured in the presence of increasing concentrations of ebselen. (b–e) IC<sub>50</sub> values were determined in the presence and absence of 0.1% Triton X-100. (b) Compound 1. (c) Compound 7. (d) Compound 8. (e) Compound 9. (f) Autofluorescence of selected compounds at different concentrations. All data are shown as mean ± SEM, *n* = 3 biological replicates.

Our results thus clearly show that halicin can be highly toxic to human cells. Furthermore, pyriothione zinc complexes, related to compound **8**, have been very recently reported to be inhibitors of another viral protease, PL<sup>pro64</sup>.

Compounds **1** and **7** were also evaluated for their ability to inhibit the cytopathogenic effect of the virus (see Methods). The concentrations for this assay were selected based on the cytotoxicity results in Figure 5. The results for compound **1** (Figure 6(a))

show that it has no inhibitory effect on the cytopathogenic effect (CPE) of the virus at high multiplicity of infection (i.e. high viral load). Nonetheless, it may elicit some reduction in CPE at a low multiplicity of infection (MOI) of 0.003. Notably, compound **7** shows a very significant reduction in CPE across all the viral doses tested (Figure 6(b)). We evaluated the effect of compound **7** on viral titres (Figure 6(c)), confirming that the reduction in the cytopathic effect is associated with a substantial reduction in the viral



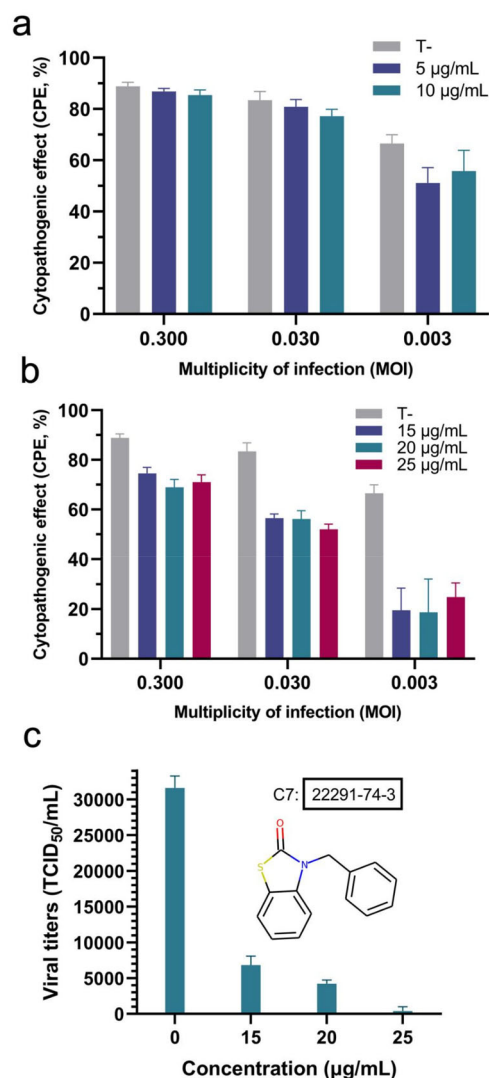
**Figure 5.** (a) Evaluation of cytotoxicity of compounds 1, 7, 8, and 9 on MRC-5 cells at 72 h post-treatment by MTT test,  $n=3$ . (b) Effects of compounds 1 and 7 on the haemolysis of red blood cells compared to a positive control (T+, 10% Triton X-100) and a negative control (Untreated),  $n=3$ .

load. This would be consistent with the main mechanism of action of compound **7** being the inhibition of  $M^{pro}$ . Compound **7** thus appears as a promising antiviral hit compound against SARS-CoV-2, warranting further studies and optimisation efforts. Of note, 2(3*H*)-benzoxazolone and bioisosteres like 2(3*H*)-benzothiazolone in compound **7** are considered a “privileged scaffold”, which has been used in commercial drugs such as Benzolone, Paraflex, Vinizene, and Tiaramide<sup>65</sup>.

### Molecular docking results

We attempted to predict the potential binding pose of compound **7** on  $M^{pro}$  using molecular docking. Previous studies highlighted that the catalytic site of SARS-CoV-2  $M^{pro}$  includes the regions S1', S1, S2, S4, and the surface depression S3<sup>54,66</sup>, depicted in Figure 7. Since the complex structure of inhibitor N3 and  $M^{pro}$  has been resolved<sup>54</sup>, we first optimised the docking parameters to recapitulate the native pose of the inhibitor N3 (RMSD = 2.15 Å, Figure 7(a, d)). After optimising the docking parameters, we docked the positive control drug ebselen and compound **7**. The resulting docking poses and ligand interaction maps are shown in Figure 7(b–f).

From the docking results, we can see that ebselen and compound **7** may have similar binding poses. They both would occupy the S2 and S4 sites. Ebselen would form a pi-H interaction with residue Gln189, while compound **7** would form a pi-H interaction with Glu166. Actually, the second ranked pose of compound **7** also showed a pi-H interaction with Gln189, albeit with a



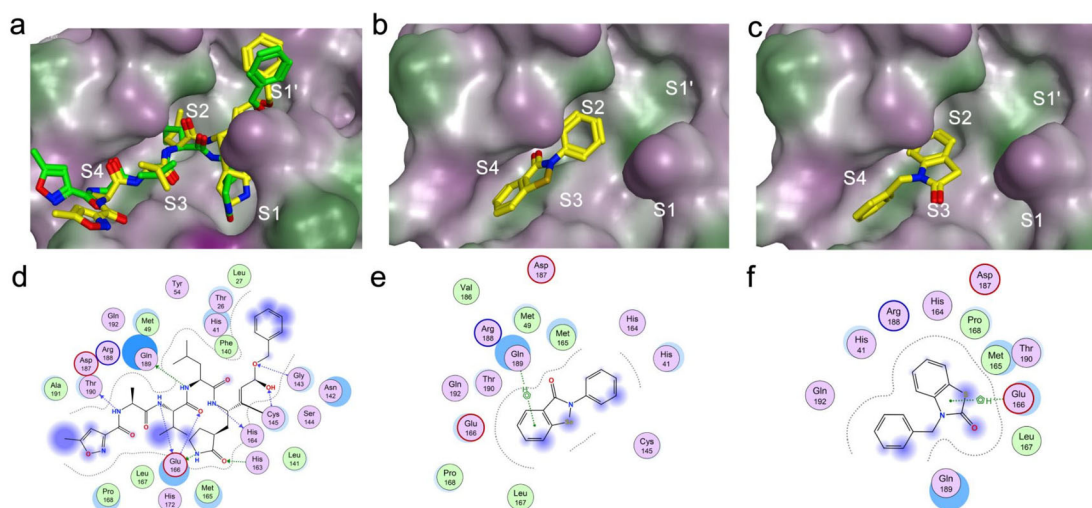
**Figure 6.** Evaluation of antiviral activity of compounds 1 and 7. Effect of (a) compound 1 and (b) compound 7 on the cytopathogenic effect by coronavirus hCoV-229 at different viral loads,  $n=3$ . (c) Tissue culture infectious dose (50%) (TCID<sub>50</sub>), defined as the dilution of the virus required to infect 50% of the cell culture, in the presence of different concentrations of compound 7,  $n=3$ .

poor docking score. This suggests that compound **7** may adopt a similar binding conformation to ebselen.

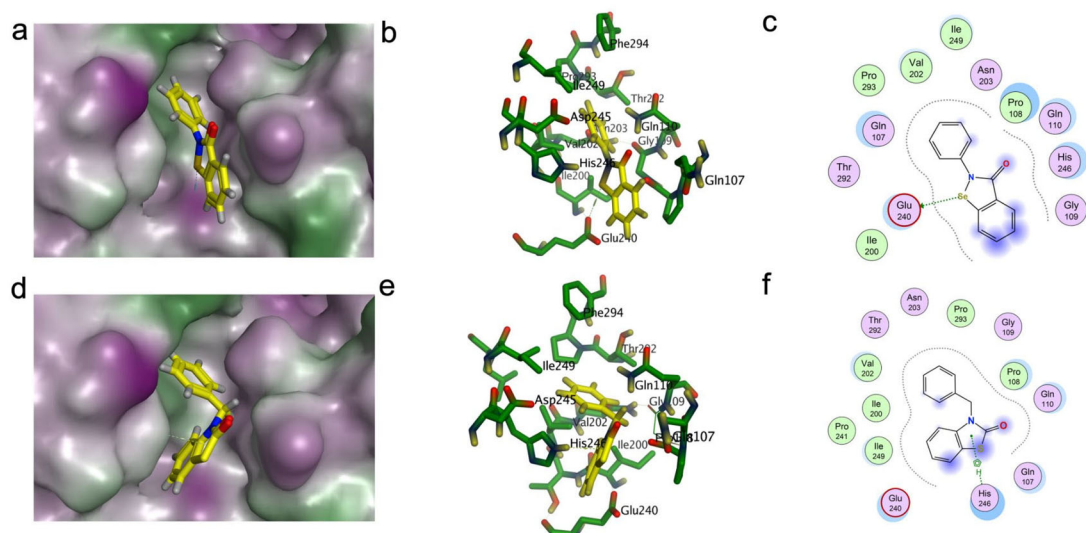
Given its small size, ebselen cannot form extensive interactions like N3, which occupies the entire S1 to S4 sites and obtains a high docking score, and yet ebselen still has high inhibitory activity. This is possible because ebselen can donate a selenium atom after a ring-opening reaction, forming a covalent bond and blocking the histidine-Cys catalytic dyad, thus acting as a covalent inhibitor. Although the docking of compound **7** suggests that it may have similar interactions as ebselen and its five-membered ring might also open, further studies are necessary to evaluate this possibility.

Recently, another study has shown that ebselen can bind not just to the catalytic site of SARS-CoV-2  $M^{pro}$ , but also to between the domains II and III of the protein, where it can act as an allosteric regulator<sup>67</sup>. In fact, allosteric inhibition of  $M^{pro}$  is increasingly recognised as a promising treatment modality<sup>68</sup>. This led us to reconsider the *in vitro* binding results above (Figure 2(h)). Compound **7** showed high inhibitory activity at low concentration, but its inhibition activity changed significantly at higher concentration. A possible explanation could be that compound **7** also





**Figure 7.** Docking results of N3, ebselen, and compound **7** against the catalytic site of  $M^{Pro}$ . (a) Highest scoring pose (yellow) and the crystallographic N3 pose (green). (b) Highest scoring pose of ebselen. (c) Highest scoring pose of compound **7**. (d–f) 2D interaction maps for the highest scoring poses of N3, ebselen and compound **7** with  $M^{Pro}$ , respectively.



**Figure 8.** Docking results of ebselen and compound **7** in the region between domains II and III of  $M^{Pro}$ . The best pose of ebselen (a) and compound **7** (d) in that site. 3D visualisation of the key residues involved in the interaction between  $M^{Pro}$  and ebselen (b) or compound **7** (e). 2D interaction maps for the best poses of ebselen (c) and compound **7**(f) on  $M^{Pro}$ .

acted as an allosteric modulator of the enzyme. Thus, we conducted a second docking study of ebselen and compound **7** against the site between domains II and III. The results are shown in Figure 8. Ebselen would form a hydrogen bond between its selenium atom and Glu240. It would also establish hydrophobic contacts with Gln107, Pro108, Gln110, Ile200, Val202, His246, Ile249, Thr292, and Phe294. These interactions are in good agreement with the findings of Menéndez et al.<sup>67</sup>. The molecular dynamics simulations by the researchers point to highly dynamic hydrogen bonds between ebselen and the side chains of Gln107, Gln110, and His246 side. From Figure 8(b) we can see the oxygen atom approaching Gln107 and Gln110, likely forming hydrogen bonds under dynamic conditions. Compared to ebselen, compound **7** obtained a better docking score (0.4 kcal mol<sup>-1</sup> lower) and a very similar interaction profile (Figure 8), supporting a possible effect in this allosteric site. Inspection of the electrostatic surface potential of the two pockets using the APBS formalism<sup>69</sup> reveals some

resemblances between the sites (Supporting Information Figures S1 and S2), suggesting that both might participate in binding.

In summary, compound **7** showed a better binding potential than ebselen in both two sites of the viral protein given its docking scores and interaction profiles. Our results thus warrant further studies on the inhibitory mechanism of compound **7**.

## Conclusion

Viruses leverage biochemical machinery of host cells, which makes the design of viral inhibitors a challenging problem. Viral inhibitors would be greatly beneficial to complement vaccination approaches, particularly in the context of the current SARS-CoV-2 pandemic. The resolution of SARS-CoV-2 protein structures enabled the structure-based design of potential inhibitors of the main protease,  $M^{Pro}$ , as well as other viral targets. However, these

approaches had limited success. The application of machine-learning ligand-based approaches is potentially more accurate, but it is thought to require large training datasets, which may not be available in a new epidemic. In this work, we show that a machine-learning ligand-based strategy to design small-molecule inhibitors of the coronavirus main protease can be useful even with limited data. We demonstrate a two-stage computational strategy based on the recent MACAW embeddings, which leverages both screening and affinity binding measurements publicly available to increase the probability of success. The pipeline can be used to make predictions for large libraries of compounds *in silico*. Out of 10 compounds selected with this pipeline, as many as 8 showed a clear inhibitory effect on M<sup>pro</sup> when tested *in vitro*, with half-maximal inhibitory concentration values (IC<sub>50</sub>) in the range 0.18–18.82  $\mu$ M. Additional assays were conducted to evaluate the cytotoxicity, haemolysis potential, and antiviral activity of selected compounds on MRC-5 cells. The results highlight the importance of considering ADMET properties early on in drug discovery. Most importantly, we identify compound **7**, 3-benzyl-1,3-benzothiazol-2-one, as a promising novel candidate against coronaviruses, demonstrating the utility of MACAW embeddings for molecular design.

## Acknowledgements

The authors thank Dr. Alejandro Cabrera-García (Universidad de La Laguna, Spain), Prof. Miguel Ángel González-Cardenete (Instituto de Tecnología Química, UPV-CSIC, Spain), Dr. Ariadna Montero-Blay and Prof. Luis Serrano (Centre for Genomic Regulation, Spain), and Dr. Nuria Izquierdo-Useros and Prof. Bonaventura Clotet (IrsiCaixa – Institute for AIDS Research, Spain) for helpful discussions.

## Disclosure statement

No potential conflict of interest was reported by the author(s).

## Funding

This work was financially supported by grants from the Project of Innovation-driven Plan in Central South University. The authors gratefully acknowledge the University of Lorraine, CNRS and the FEDER-FSE “Fire Light”.

## ORCID

Jie Dong  <http://orcid.org/0000-0002-3324-9000>

Vincent Blay  <http://orcid.org/0000-0001-9602-2375>

## References

- Dong E, Du H, Gardner L. An interactive web-based dashboard to track COVID-19 in real time. *Lancet Infect Dis*. 2020;20(5):533–534.
- Selickman J, Vrettou CS, Mentzelopoulos SD, Marini JJ. COVID-19-related ARDS: key mechanistic features and treatments. *J Clin Med*. 2022;11(16):4896.
- Zhang Y, Geng X, Tan Y, Li Q, Xu C, Xu J, Hao L, Zeng Z, Luo X, Liu F, et al. New understanding of the damage of SARS-CoV-2 infection outside the respiratory system. *Biomed Pharmacother*. 2020;127:110195.
- Fazzini B, Battaglini D, Careno L, Pelosi P, Cecconi M, Puthuchery Z. Physical and psychological impairment in survivors of acute respiratory distress syndrome: a systematic review and meta-analysis. *Br J Anaesthesia*. 2022;S0007-0912(22)00455-X.
- Munipalli B, Seim L, Dawson NL, Knight D, Dabrh AMA. Post-acute sequelae of COVID-19 (PASC): a meta-narrative review of pathophysiology, prevalence, and management. *SN Compr Clin Med*. 2022;4(1):90.
- Delshad M, Sanaei M-J, Pourbagheri-Sigaroodi A, Bashash D. Host genetic diversity and genetic variations of SARS-CoV-2 in COVID-19 pathogenesis and the effectiveness of vaccination. *Int Immunopharmacol*. 2022;111:109128.
- Chen Q, Zhang J, Wang P, Zhang Z. The mechanisms of immune response and evasion by the main SARS-CoV-2 variants. *iScience*. 2022;25(10):105044.
- Gil C, Ginex T, Maestro I, Nozal V, Barrado-Gil L, Cuesta-Geijo MÁ, Urquiza J, Ramírez D, Alonso C, Campillo NE, et al. COVID-19: drug targets and potential treatments. *J Med Chem*. 2020;63(21):12359–12386.
- Li S, Li S, Disoma C, Zheng R, Zhou M, Razzaq A, Liu P, Zhou Y, Dong Z, Du A, et al. SARS-CoV-2: mechanism of infection and emerging technologies for future prospects. *Rev Med Virol*. 2021;31(2):e2168.
- Jamison DA, Anand Narayanan S, Trovão NS, Guarnieri JW, Topper MJ, Moraes-Vieira PM, Zaksas V, Singh KK, Wurtele ES, Beheshti A. A comprehensive SARS-CoV-2 and COVID-19 review, part 1: intracellular overdrive for SARS-CoV-2 infection. *Eur J Hum Genet*. 2022;30(8):889–898.
- Jackson CB, Farzan M, Chen B, Choe H. Mechanisms of SARS-CoV-2 entry into cells. *Nat Rev Mol Cell Biol*. 2022;23(1):3–20.
- V'kovski P, Kratzel A, Steiner S, Stalder H, Thiel V. Coronavirus biology and replication: implications for SARS-CoV-2. *Nat Rev Microbiol*. 2021;19(3):155–170.
- Hu B, Guo H, Zhou P, Shi Z-L. Characteristics of SARS-CoV-2 and COVID-19. *Nat Rev Microbiol*. 2021;19(3):141–154.
- Chan JF-W, Kok K-H, Zhu Z, Chu H, To KK-W, Yuan S, Yuen K-Y. Genomic characterization of the 2019 novel human-pathogenic coronavirus isolated from a patient with atypical pneumonia after visiting Wuhan. *Emerg Microbes Infect*. 2020;9(1):221–236.
- Chia CSB, Xu W, Shuyi Ng P. A patent review on SARS coronavirus main protease (3CLpro) inhibitors. *ChemMedChem*. 2022;17(1):e202100576.
- Vangeel L, Chiu W, De Jonghe S, Maes P, Slechten B, Raymenants J, André E, Leyssen P, Neyts J, Jochmans D. Remdesivir, molnupiravir and nirmatrelvir remain active against SARS-CoV-2 Omicron and other variants of concern. *Antiviral Res*. 2022;198:105252.
- Siemieniuk RA, Bartoszko JJ, Zeraatkar D, Kum E, Qasim A, Martinez JPD, Izcovich A, Lamontagne F, Han MA, Agarwal A, et al. Drug treatments for covid-19: living systematic review and network meta-analysis. *BMJ*. 2020;370:m2980.
- Zhou S, Hill CS, Sarkar S, Tse LV, Woodburn BMD, Schinazi RF, Sheahan TP, Baric RS, Heise MT, Swanstrom R.  $\beta$ -d-N<sup>4</sup>-hydroxycytidine inhibits SARS-CoV-2 through lethal mutagenesis but is also mutagenic to mammalian cells. *J Infect Dis*. 2021;224(3):415–419.
- Fischer WA, Eron JJ, Holman W, Cohen MS, Fang L, Szwedczyk LJ, Sheahan TP, Baric R, Mollan KR, Wolfe CR, et al. A phase 2a clinical trial of molnupiravir in patients with COVID-19 shows accelerated SARS-CoV-2 RNA clearance and

- elimination of infectious virus. *Sci Transl Med*. 2022;14(628): eabl7430.
20. Owen DR, Allerton CMN, Anderson AS, Aschenbrenner L, Avery M, Berritt S, Boras B, Cardin RD, Carlo A, Coffman KJ, et al. An oral SARS-CoV-2 Mpro inhibitor clinical candidate for the treatment of COVID-19. *Science*. 2021;374 (6575): eabl4784–1593.
  21. Boras B, Jones RM, Anson BJ, Arenson D, Aschenbrenner L, Bakowski MA, Beutler N, Binder J, Chen E, Eng H, et al. Preclinical characterization of an intravenous coronavirus 3CL protease inhibitor for the potential treatment of COVID-19. *Nat Commun*. 2021;12(1):6055.
  22. Wanounou M, Caraco Y, Levy RH, Bialer M, Perucca E. Clinically relevant interactions between ritonavir-boosted nirmatrelvir and concomitant antiepileptic medications: implications for the management of COVID-19 in patients with epilepsy. *Clin Pharmacokinet*. 2022;61(9):1219–1236.
  23. Yang KS, Leeuwon SZ, Xu S, Liu WR. Evolutionary and structural insights about potential SARS-CoV-2 evasion of nirmatrelvir. *J Med Chem*. 2022;65(13):8686–8698.
  24. Narayanan D, Parimon T. Current therapeutics for COVID-19, what we know about the molecular mechanism and efficacy of treatments for this novel virus. *IJMS*. 2022;23(14):7702.
  25. Md Khairi LNH, Fahrni ML, Lazzarino AI. The race for global equitable access to COVID-19 vaccines. *Vaccines*. 2022;10(8): 1306.
  26. Hegyi A, Ziebuhr J. Conservation of substrate specificities among coronavirus main proteases. *J Gen Virol*. 2002;83(Pt 3):595–599.
  27. Drayman N, DeMarco JK, Jones KA, Azizi S-A, Froggatt HM, Tan K, Maltseva NI, Chen S, Nicolaescu V, Dvorkin S, et al. Masitinib is a broad coronavirus 3CL inhibitor that blocks replication of SARS-CoV-2. *Science*. 2021;373(6557):931–936.
  28. Liu H, Iketani S, Zask A, Khanizeman N, Bednarova E, Forouhar F, Fowler B, Hong SJ, Mohri H, Nair MS, et al. Development of optimized drug-like small molecule inhibitors of the SARS-CoV-2 3CL protease for treatment of COVID-19. *Nat Commun*. 2022;13(1):1891.
  29. Xiu S, Dick A, Ju H, Mirzaie S, Abdi F, Cocklin S, Zhan P, Liu X. Inhibitors of SARS-CoV-2 entry: current and future opportunities. *J Med Chem*. 2020;63(21):12256–12274.
  30. Yan F, Gao F. An overview of potential inhibitors targeting non-structural proteins 3 (PLpro and Mac1) and 5 (3CLpro/Mpro) of SARS-CoV-2. *Comput Struct Biotechnol J*. 2021;19: 4868–4883.
  31. Pavan M, Bassani D, Sturlese M, Moro S. Bat coronaviruses related to SARS-CoV-2: What about their 3CL proteases (MPro)? *J Enzyme Inhib Med Chem*. 2022;37(1):1077–1082.
  32. Clyde A, Galanie S, Kneller DW, Ma H, Babuji Y, Blaiszik B, Brace A, Brettin T, Chard K, Chard R, et al. High-throughput virtual screening and validation of a SARS-CoV-2 main protease noncovalent inhibitor. *J Chem Inf Model*. 2022;62(1): 116–128.
  33. Gajjar ND, Dhameliya TM, Shah GB. In search of RdRp and Mpro inhibitors against SARS CoV-2: molecular docking, molecular dynamic simulations and ADMET analysis. *J Mol Struct*. 2021;1239:130488.
  34. Galvez-Llompant M, Zanni R, Galvez J, Basak SC, Goyal SM. COVID-19 and the importance of being prepared: a multidisciplinary strategy for the discovery of antivirals to combat pandemics. *Biomedicines*. 2022;10(6):1342.
  35. Hosseini M, Chen W, Xiao D, Wang C. Computational molecular docking and virtual screening revealed promising SARS-CoV-2 drugs. *Precis Clin Med*. 2021;4(1):1–16.
  36. Mercorelli B, Desantis J, Celegato M, Bazzacco A, Siragusa L, Benedetti P, Eleuteri M, Croci F, Cruciani G, Goracci L, et al. Discovery of novel SARS-CoV-2 inhibitors targeting the main protease Mpro by virtual screenings and hit optimization. *Antiviral Res*. 2022;204:105350.
  37. Mishra B, Ballaney P, Saha G, Shinde A, Banerjee S, Thimmakonda VS, Aduri R. An in silico discovery of potential 3CL protease inhibitors of SARS-CoV-2 based upon inactivation of the cysteine 145-Histidine 41 catalytic dyad. *J Biomol Struct Dyn*. 2022;0(0):1–20.
  38. Rao P, Shukla A, Parmar P, Rawal RM, Patel BV, Saraf M, Goswami D. Proposing a fungal metabolite-flaviolin as a potential inhibitor of 3CLpro of novel coronavirus SARS-CoV-2 identified using docking and molecular dynamics. *J Biomol Struct Dyn*. 2022;40(1):348–360.
  39. Sobhia ME, Ghosh K, Sivangula S, Kumar S, Singh H. Identification of potential SARS-CoV-2 Mpro inhibitors integrating molecular docking and water thermodynamics. *J Biomol Struct Dyn*. 2022;40(11):5079–5089.
  40. Zev S, Raz K, Schwartz R, Tarabeh R, Gupta PK, Major DT. Benchmarking the ability of common docking programs to correctly reproduce and score binding modes in SARS-CoV-2 protease Mpro. *J Chem Inf Model*. 2021;61(6):2957–2966.
  41. Cui W, Yang K, Yang H. Recent progress in the drug development targeting SARS-CoV-2 main protease as treatment for COVID-19. *Front Mol Biosci*. 2020;7:616341.
  42. Mengist HM, Dilnessa T, Jin T. Structural basis of potential inhibitors targeting SARS-CoV-2 main protease. *Front Chem*. 2021;9:622898.
  43. Yang H, Yang J. A review of the latest research on Mpro targeting SARS-COV inhibitors. *RSC Med Chem*. 2021;12: 1026–1036.
  44. Blay V, Radivojevic T, Allen JE, Hudson CM, Garcia Martin H. MACAW: an accessible tool for molecular embedding and inverse molecular design. *J Chem Inf Model*. 2022;62(15): 3551–3564.
  45. Gao K, Nguyen DD, Chen J, Wang R, Wei G-W. Repositioning of 8565 existing drugs for COVID-19. *J Phys Chem Lett*. 2020;11(13):5373–5382.
  46. Zhu W, Xu M, Chen CZ, Guo H, Shen M, Hu X, Shinn P, Klumpp-Thomas C, Michael SG, Zheng W. Identification of SARS-CoV-2 3CL protease inhibitors by a quantitative high-throughput screening. *ACS Pharmacol Transl Sci*. 2020;3(5): 1008–1016.
  47. Kuzikov M, Costanzi E, Reinshagen J, Esposito F, Vangeel L, Wolf M, Ellinger B, Claussen C, Geisslinger G, Corona A, et al. Identification of inhibitors of SARS-CoV-2 3CL-Pro enzymatic activity using a small molecule in vitro repurposing screen. *ACS Pharmacol Transl Sci*. 2021;4(3):1096–1110.
  48. Vatansever EC, Yang KS, Drelich AK, KratchKC, Cho C-C, KempaiahKR, Hsu JC, Mellott DM, Xu S, Tseng C-TK, et al. Bepridil is potent against SARS-CoV-2 in vitro. *Proc Natl Acad Sci USA*. 2021;118(10):e2012201118.
  49. PostEra. 2021. MPro Activity Data [Internet]. [https://covid.postera.ai/covid/activity\\_data](https://covid.postera.ai/covid/activity_data).
  50. Wishart DS, Feunang YD, Guo AC, Lo EJ, Marcu A, Grant JR, Sajed T, Johnson D, Li C, Sayeeda Z, et al. DrugBank 5.0: a major update to the DrugBank database for 2018. *Nucleic Acids Res*. 2018;46(D1):D1074–D1082.

51. Pedregosa F, et al. Scikit-learn: machine learning in Python. *JMLR*. 2011;12:2825–2830.
52. Xiong G, WuZ, Yi J, Fu L, Yang Z, Hsieh C, Yin M, Zeng X, WuC, Lu A, et al. ADMETlab 2.0: an integrated online platform for accurate and comprehensive predictions of ADMET properties. *Nucleic Acids Res*. 2021;49(W1):W5–W14.
53. Daina A, Michielin O, Zoete V. SwissADME: a free web tool to evaluate pharmacokinetics, drug-likeness and medicinal chemistry friendliness of small molecules. *Sci Rep*. 2017;7(1):42717.
54. Jin Z, Du X, Xu Y, Deng Y, Liu M, Zhao Y, Zhang B, Li X, Zhang L, Peng C, et al. Structure of Mpro from SARS-CoV-2 and discovery of its inhibitors. *Nature*. 2020;582(7811):289–293.
55. Amporndanai K, Meng X, Shang W, Jin Z, Rogers M, Zhao Y, Rao Z, Liu Z-J, Yang H, Zhang L, et al. Inhibition mechanism of SARS-CoV-2 main protease by ebselen and its derivatives. *Nat Commun*. 2021;12(1):3061.
56. Reed LJ, Muench H. A simple method of estimating fifty per cent ENDPOINTS. *Am J Epidemiol*. 1938;27(3):493–497.
57. Mosmann T. Rapid colorimetric assay for cellular growth and survival: application to proliferation and cytotoxicity assays. *J Immunol Methods*. 1983;65(1–2):55–63.
58. Silvestrini L, Belhaj N, Comez L, Gerelli Y, Lauria A, Libera V, Mariani P, Marzullo P, Ortore MG, Palumbo Piccionello A, et al. The dimer-monomer equilibrium of SARS-CoV-2 main protease is affected by small molecule inhibitors. *Sci Rep*. 2021;11(1):9283.
59. Ma C, Hu Y, Townsend JA, Lagarias PI, Marty MT, Kolocouris A, Wang J. Ebselen, Disulfiram, Carmofur, PX-12, Tideglusib, and Shikonin are nonspecific promiscuous SARS-CoV-2 main protease inhibitors. *ACS Pharmacol Transl Sci*. 2020;3(6):1265–1277.
60. Baddock H, Brolih S, Yosaatmadja Y, Ratnaweera M, Bielinski M, Swift L, Cruz-Migoni A, Morris G, Schofield C, Gileadi O, et al. 2020. Characterisation of the SARS-CoV-2 ExoN (nsp14ExoN-nsp10) complex: implications for its role in viral genome stability and inhibitor identification. *bioRxiv*. <https://ora.ox.ac.uk/objects/uuid:3899d504-d33f-4e5c-8cec-588fc841b784>.
61. Bickerton GR, Paolini GV, Besnard J, Muresan S, Hopkins AL. Quantifying the chemical beauty of drugs. *Nat Chem*. 2012;4(2):90–98.
62. Nepali K, Lee H-Y, Liou J-P. Nitro-group-containing drugs. *J Med Chem*. 2019;62(6):2851–2893.
63. Stokes JM, Yang K, Swanson K, Jin W, Cubillos-Ruiz A, Donghia NM, MacNair CR, French S, Carfrae LA, Bloom-Ackermann Z, et al. A deep learning approach to antibiotic discovery. *Cell*. 2020;180(4):688–702.e13.
64. Kladnik J, Dolinar A, Kljun J, Perea D, Grau-Expósito J, Genescà M, Novinec M, Buzon MJ, Turel I. Zinc pyrithione is a potent inhibitor of PLPro and cathepsin L enzymes with ex vivo inhibition of SARS-CoV-2 entry and replication. *J Enzyme Inhib Med Chem*. 2022;37(1):2158–2168.
65. Poupaert J, Carato P, Colacino E, Yous S. 2(3H)-benzoxazolone and bioisosters as “privileged scaffold” in the design of pharmacological probes. *Curr Med Chem*. 2005;12(7):877–885.
66. Qiao J, Li Y-S, Zeng R, Liu F-L, Luo R-H, Huang C, Wang Y-F, Zhang J, Quan B, Shen C, et al. SARS-CoV-2 Mpro inhibitors with antiviral activity in a transgenic mouse model. *Science*. 2021;371(6536):1374–1378.
67. Menéndez CA, Byléhn F, Perez-Lemus GR, Alvarado W, de Pablo JJ. Molecular characterization of ebselen binding activity to SARS-CoV-2 main protease. *Sci Adv*. 2020;6(37):eabd0345.
68. Samrat SK, Xu J, Xie X, Gianti E, Chen H, Zou J, Pattis JG, Elokely K, Lee H, Li Z, et al. Allosteric inhibitors of the main protease of SARS-CoV-2. *Antiviral Res*. 2022;205:105381.
69. Baker NA, Sept D, Joseph S, Holst MJ, McCammon JA. Electrostatics of nanosystems: application to microtubules and the ribosome. *Proc Natl Acad Sci USA*. 2001;98(18):10037–10041.