# Computational models identify several FDA approved or experimental drugs as putative agents against SARS-CoV-2

Tesia Bobrowski[a], Vinicius M. Alves[b], Cleber C. Melo-Filho[a], Daniel Korn[a,c], Scott Auerbach[d], Charles Schmitt[b], Eugene N. Muratov[a,*], Alexander Tropsha[a,*]

[a] Laboratory for Molecular Modeling, Division of Chemical Biology and Medicinal Chemistry, UNC Eshelman School of Pharmacy, University of North Carolina, Chapel Hill, NC, 27599, USA.

[b] Office of Data Science, National Toxicology Program, NIEHS, Morrisville, NC, 27560, USA.

[c] Department of Computer Science, University of North Carolina, Chapel Hill, NC, 27599, USA.

[d] Toxinformatics Group, National Toxicology Program, NIEHS, Morrisville, NC, 27560, USA.

**Corresponding Author**

* Address for correspondence: 100K Beard Hall, UNC Eshelman School of Pharmacy, University of North Carolina, Chapel Hill, NC, 27599, USA; Telephone: (919) 966-2955; FAX: (919) 966-0204; E-mail: murik@email.unc.edu and alex_tropsha@unc.edu.

## Abstract

The outbreak of a novel human coronavirus (SARS-CoV-2) has evolved into global health emergency, infecting hundreds of thousands of people worldwide. In an effort to find antiviral medications, many computational groups have pursued the 3C-like protease of the virus, also known as main protease ($M^{pro}$), as a drug target. We have identified experimental data on the inhibitory activity of compounds tested against closely related (96% sequence identity, 100% active site conservation) protease of SARS-CoV and employed this data to build Quantitative Structure-Activity Relationships (QSAR) models for this dataset. We employed these models for virtual screening of all marketed, withdrawn, experimental, and investigational drugs from DrugBank, including compounds in clinical trials. Molecular docking and similarity search approaches were explored in parallel with QSAR modeling, but molecular docking failed to correctly discriminate between experimentally active and inactive compounds, so we did not rely on this approach in prospective virtual screening. As a result of our studies, we recommended 41 approved, experimental, or investigational drugs as potential agents against SARS-CoV-2 acting as putative inhibitors of $M^{pro}$>. Ten compounds with feasible prices were purchased and are awaiting the experimental validation. This manuscript will be updated once results are available and submitted for peer-review publication if compounds are found to be active in SARS-CoV-2 phenotypic screen.

## Introduction

On December 8th, 2019, Chinese health authorities in Hubei detected the first case of an infection caused by a novel coronavirus since named SARS-CoV-2.[1,2] On January 31, less than two months later, the World Health Organization declared the SARS-CoV-2 outbreak a global health emergency.[3] The new coronavirus is most similar to a bat betacoronavirus that does not infect humans, but it is also in the same family as the notorious human coronaviruses SARS-CoV (sudden acute respiratory syndrome coronavirus) and MERS-CoV (Middle Eastern Respiratory Syndrome coronavirus), which have reported fatality rates of 10% and 35%, respectively.[4,5] Current (as of April 16th, 2020) estimates of the fatality rate of COVID-19 vary per age cohort and the virus to date is estimated to have infected over two million people, though these statistics are approximate due to established asymptomatic transmission of the disease or likely underreporting or lack of testing by health authorities.[6,7] While the fatality rate of the current virus is estimated to be less than that of SARS and MERS-CoV, it has been shown to be highly transmissible, infecting the first 1,000 patients in only 48 days, whereas SARS took 130 days and MERS took 2.5 years.[8] The initial velocity of the spread of SARS-CoV-2 was enough to indicate pandemic potential at the start of the outbreak, and now and hundreds of thousands of cases have been reported worldwide despite strict quarantine and travel protocols set in place in many countries.

No antivirals or vaccines exist against SARS-CoV-2 or past epidemic betacoronaviruses, which represents a larger-scale paucity of data on this genus of viruses.[9] Genomic sequences of the SARS-CoV-2 continue to be uploaded to GenBank, hosted by the National Center for Biotechnology Information (NCBI), and there are 1084 distinct sequences listed there to date.[10] The first protein crystal structure for SARS-CoV-2 deposited in the Protein Data Bank in February 2020 was the 2019-nCoV main protease (also known as 3C-like protease or $M^{pro}$) in complex with an inhibitor N3 (PDB ID: 6LU7).[11] One of the only papers to date investigating compounds with anti-SARS-CoV-2 activities tested seven compounds total and reported four hits, most notably remdesivir and chloroquine.[12] Other studies have reported other compounds with anti-SARS-CoV-2 activities such as ivermectin[13] and β-D-N4-hydroxycytidine (NHC, EIDD-1931).[14] Another study identified six compounds to have activity against SARS-CoV-2 $M^{pro}$, but only only ebselen showed activity in phenotypic screen.[15] Already COVID-19 clinical trials are being performed that utilize repurposing of existing experimental nucleoside analogs such as remdesivir, ribavirin, and favipiravir that have demonstrated past antiviral activities.[16]

Past research has identified several targets for coronavirus drug development, namely nonstructural protein 14 (nsp14-ExoN) and the proteins involved in the coronaviral RNA replication process (replicase polyprotein 1ab and $M^{pro}$)[17]. The replicase polyprotein 1ab is responsible for the synthesis of the large, functional polyproteins pp1a and pp1ab, which are precursors of 16 non-structural proteins that are important in the replication of coronavirus RNA.[18–20] The replicase polyprotein 1ab (CHEMBL5118) is a precursor of 16 non-structural proteins,[21] such as RNA polymerase, helicase, 3’-5’ exonuclease, and 2’-O-ribose methyltransferase. The polyprotein 1ab along with polyprotein 1a are precursors of all proteins that form the viral replication complex (e.g., 1ab has 7,095 aminoacids). These are not functional unless proteases ($M^{pro}$ and papain-like proteinase) cleave them into those 16 smaller proteins.[22] The virus-encoded $M^{pro}$ is integral to the proteolytic processing of these polyproteins and is highly conserved in coronaviruses, as are the cleavage sites and lengths of the polyproteins themselves.[19,23,24] Furthermore, $M^{pro}$ has been considered before in the design of broad-spectrum antiviral compounds as demonstrated in a 2012 study by Kim et al.[25] that reported in vitro inhibition of SARS-CoV replication by inhibitors of this protease.[19]

Given the lack of publicly available data on the new coronavirus, we emphasize the message of the recent editorial titled "Calling all coronavirus researchers: keep sharing, stay open," that calls for researchers to collaborate and share all data on the new coronavirus to better prevent its spread and morbidity.[26] Many studies reporting compounds identified by computational approaches have been published in both peer-reviewed[27,28] and arXiv journals[29,30] since the outbreak of SARS-CoV-2 was reported. In line with this call, we curated all available open-source data on SARS-CoV-2 and SARS-CoV and employed both structure- and ligand-based computational approaches to select a set of compounds that may have the potential to inhibit SARS-CoV-2 replication. In this initial investigation, we have exclusively focused on FDA approved medications or experimental/investigational compounds because these could be quickly repurposed as COVID-19 treatments if their experimental validation is successful.

## Materials and Methods

The workflow employed in this study can be seen in **Figure 1**.

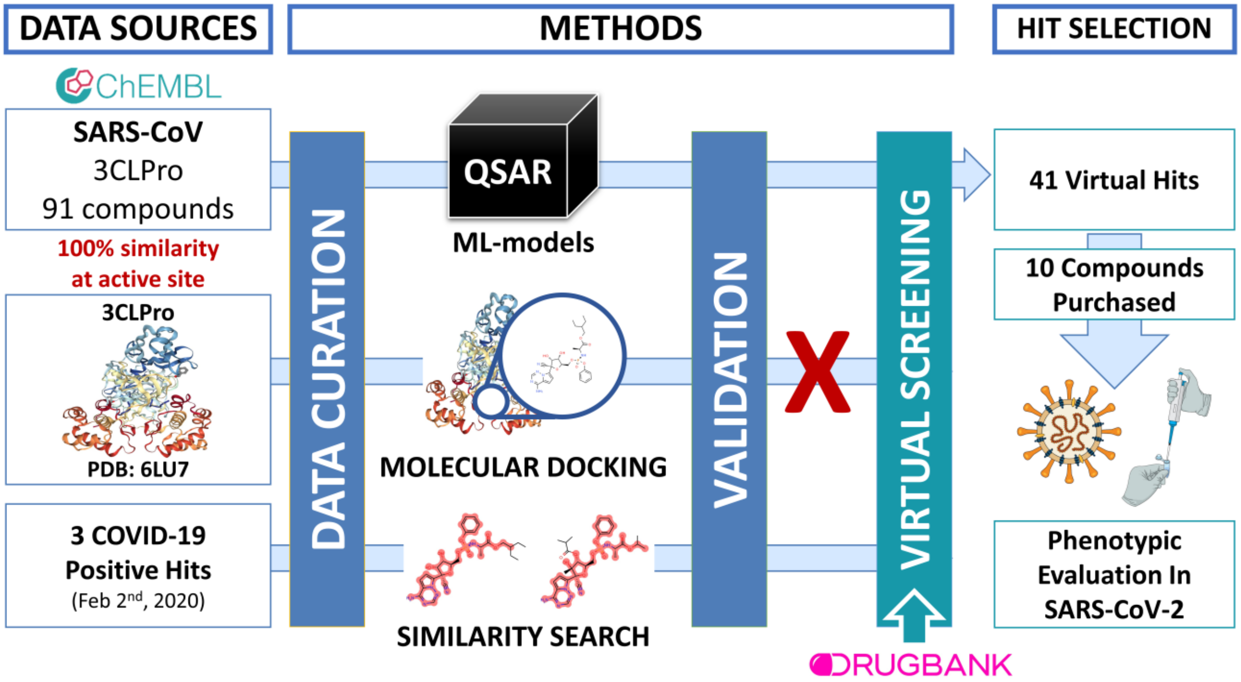

**Figure 1.** Study design.

## Quantitative Structure-Activity Relationship (QSAR) modeling

### *Data collection and curation*

We collected 201 datapoints for the SARS-CoV M$^{pro}$ assay (ChEMBL ID: X) and, after curation, 91 compounds (27 actives and 64 inactives, considering a threshold of 10 µM) were kept. We found 22 additional compoudns in PDB (13 actives and 9 inactives) that were not available in ChEMBL. At the end, 113 compounds (40 actives and 73 inactives) were kept for modeling. All chemical structures and correspondent biological information were carefully standardized using Standardizer v.20.8.0 (ChemAxon, Budapest, Hungary, http://www.chemaxon.com) according to the protocols proposed by Fourches and colleagues.[31,32] Briefly, inorganics, counterions, metals, organometallic compounds, and mixtures were removed. In addition, specific chemotypes such as aromatic rings and nitro groups were normalized. Furthermore, we performed the analysis and exclusion of duplicates: (i) if duplicates presented discordance in biological activity, both entries would be excluded; and (ii) if the reported outcomes of the duplicates were the same, one entry would be retained in the dataset and the other excluded.

### *Molecular descriptors*

The QSAR models were developed using three types of descriptors: Morgan fingerprints,[33] 2D Simplex Representation of Molecular Structure (SiRMS) descriptors[34] and Dragon (v.7 Kode Chemoinformatics srl – Pisa, Italy). The open-source Morgan fingerprints with 2048 bits and an atom radius of 3 calculated in RDKit (http://www.rdkit.org) using Python 3.6. SiRMS were calculated using HiTQSAR[35] at the 2D level. SiRMS descriptors account not only for the atom type, but also for other atomic characteristics that may impact biological activity of molecules, e.g., partial charge, lipophilicity, refraction, and atom ability for being a donor/acceptor in hydrogen-bond formation (H-bond). Detailed description of HiTQSAR and SiRMS can be found elsewhere.[35] Dragon descriptors were calculated at 2D level as well. For both SiRMS and Dragon, descriptors with less than 0.01 variance were removed. Correlated descriptors were also removed.

### *Model generation*

QSAR models were built and rigorously validated following best practices.[36] The models were built using the Random Forest (RF) algorithm[37] implemented in scikit-learn (http://scikit-learn.org). Random Forest hyperparameters were tuned using the GridSearchCV module implemented in scikit-learn. Trees were decorrelated by randomly bootstrapping compound instances used in modeling with replacement and selecting a random sample of root(N)-many features for each tree, where N is the total number of features available. Trees were configured to evaluate features on classification accuracy at the median value and to use gini as the split criterion.

A 5-fold external cross-validation procedure was performed using the following protocol. The full set of compounds with known experimental activity is randomly divided into five subsets of equal size. One of these subsets (20% of all compounds) is set aside as the external validation set, while the remaining four sets form the modeling set (80% of all compounds). This procedure is repeated five times, allowing each of the five subsets to be used as an external validation set. Models are built using the training set only, and it is important to emphasize that compounds are never simultaneously part of both the training and external validation set.

Two types of consensus were performed: consensus is a majority average of predictions from the independent models developed with Morgan, SiRMS, and Dragon. Consensus AD is a majority average prediction from independent models when predictions are inside the applicability domain of that model. The local (tree) applicability domain approach[38] setting a threshold of 70% was used for all RF models developed in this study.

### Molecular Docking

Molecular docking experiments were performed using the structure of $M^{pro}$ from SARS-CoV-2 (PDB ID: 6LU7). To enable these calculations, the structure was prepared in Maestro[39] under pH 7.0±2.0 and optimized with OPLS3e force field. All ligands were prepared under the same conditions and submitted to molecular docking using Glide[12] with the standard precision (SP) option.

### Similarity Search

Similarity search was performed in the KNIME platform (https://www.knime.com/) using Morgan fingerprints using the three compounds described by Wang et al.[12] as active in the phenotypic screen (remdesivir, chloroquine, and nitazoxanide). A threshold of 75% similarity in Tanimoto coefficient was employed to select compounds from DrugBank as putative actives.

## Results and Discussion

As seen in Figure 1, we employed three different computational strategies to screen a wide array of compounds from DrugBank in order to suggest preexisting compounds with possible inhibitory activities against SARS-CoV-2. We started by collecting all publicly available data on the SARS-CoV-2 and other coronaviruses. We excluded all phenotypic assays from modeling on the basis of a recent study by Wang et al.[40] which demonstrated that some compounds active against SARS-CoV were not active against SARS-CoV-2 in a phenotypic screen. The replicase polyprotein 1ab was discarded because its whole structure is not available in PDB, but just its derivatives. Using Basic Local Alignment Search Tool (BLAST) available in UniProt (https://www.uniprot.org/blast/)[41], we observed that the primary sequences of $M^{pro}$ in both SARS-CoV and SARS-CoV-2 had 96% identity (Figure 2a). The crystal structure of SARS-CoV-2 $M^{pro}$ was recently elucidated and superposition of the respective 3D protein structures (PDB IDs: 5N19, 6LU7) revealed a conserved binding site around the co-crystallized inhibitors including the catalytic dyad represented by His41 and Cys145 (Figures 2b and 2c).[42]

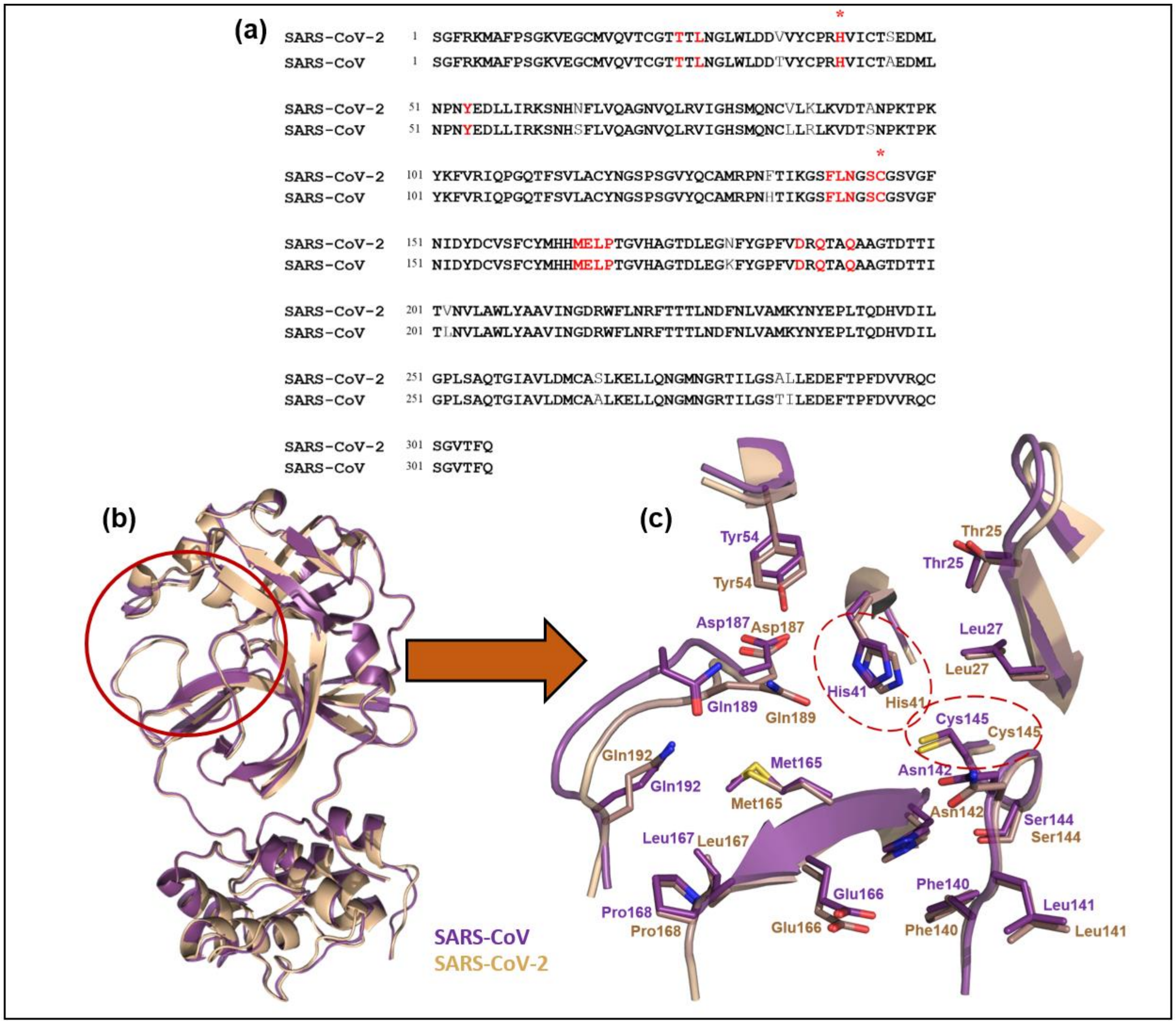

**Figure 2.** Alignment of SARS-CoV and SARS-CoV-2 $M^{pro}$ monomers. **(a)** Primary sequence alignment highlighting the conserved residues in bold font. The binding site residues are shown in red and the catalytic dyad, represented by His41 and Cys145, is marked with asterisks. **(b)** Alignment of $M^{pro}$ monomers available in PDB (IDs: 5N19, 6LU7). **(c)** Visualization of the overlap between residues at the $M^{pro}$ active site for SARS and SARS-CoV-2. The red dashed

circles show the conserved catalytic dyad and the remarkable conservation of the binding site of $M^{pro}$ between the coronaviruses.

The 113 compounds (40 actives and 73 inactives) kept after curation were used for binary QSAR modeling. The statistical characteristics of our QSAR models are available in Table 1. Due to the limited size of the dataset, models were only validated by 5-fold external cross validation and achieved external correct classification rate of 71-83% (sensitivity = 55-72%, positive predicted value = 72-100%, specificity = 88-100%, negative predicted value = 78-85%). Models were generatated with the entire (unbalanced) dataset. Although sensitivity was only acceptable[36] (> 60% for majority of the models) and below this threshold for Dragon models, we decided to proceed with this model because the PPV was higher. This guarantees that a lower number of hits would be found, but a higher confidence is expected.

**Table 1.** Statistical characteristics of QSAR models for SARS-CoV $M^{pro}$ assessed by 5-fold external validation.

| Model | CCR | Sensitivity | PPV | Specificity | NPV | Coverage |
|---|---|---|---|---|---|---|
| **Morgan** | 0.78 | 0.65 | 0.81 | 0.92 | 0.83 | 1.00 |
| **Morgan AD** | 0.80 | 0.62 | 0.94 | 0.98 | 0.85 | 0.69 |
| **SiRMS** | 0.76 | 0.65 | 0.72 | 0.86 | 0.82 | 1.00 |
| **SiRMS AD** | 0.83 | 0.72 | 0.86 | 0.93 | 0.85 | 0.61 |
| **Dragon** | 0.71 | 0.55 | 0.71 | 0.88 | 0.78 | 1.00 |
| **Dragon AD** | 0.78 | 0.56 | 1.00 | 1.00 | 0.87 | 0.54 |
| **Consensus** | 0.74 | 0.60 | 0.73 | 0.88 | 0.80 | 1.00 |
| **Consensus (AD)** | 0.78 | 0.62 | 0.86 | 0.95 | 0.83 | 0.77 |

Recently, Wang et al.[39] demonstrated that remdesivir and chloroquine were highly active; nitazoxanide was moderately active; and ribavirin, penciclovir, nafamostat, faviparir were inactive against SARS-CoV-2 in phenotypic assays. The SiRMS models predicted remdesivir and ribavirin as active, while Dragon predicted ribarin only. Currently, there are no evidence none of these targets act on M$^{pro}$; remdesivir is a known RNA polymerase inhibitor.[43]

In addition, Jin et al.[44] submitted a library of ~ 10,000 compounds to a high-throughput screening (HTS) and identified six inhibitors of SARS-CoV-2 M$^{pro}$, namely, ebselen, disulfiram, tideglusib, carmofur, shikonin, and PX-12. After additional phenotypic assays, only ebselen inhibited *in vitro* viral replication. Despite the large amount of compounds tested in HTS, only the activity of those six inhibitors was reported, so there is no publicly available data on SARS-CoV-2 M$^{pro}$ yet that could enable the development of QSAR models.

Due to the small amount of publicly available SARS-CoV-2 M$^{pro}$ assay data and the high similarity 96% identity sequence of M$^{pro}$ in SARS-CoV and SARS-CoV-2, including conserved active site (see above), we hypothesized that compounds predicted to be active in the SARS-CoV M$^{pro}$ assay[45] (used for compounds in our modeling set) could be active against SARS-CoV-2.

In addition, we have also predicted M$^{pro}$ activity for twenty three compounds reported to undergo clinical trials (as of March 23, 2020)[46] (See Table S1 in Supplementary Materials). Of these compounds, lopinavir, ritonavir, tetrandrine, cobicistat, losartan, ribavirin, remdesivir, aviptadil, and danoprevir were predicted as active by SiRMS models. Lopinavir was also predicted as active by Dragon. None of the molecules were predicted as active by Morgan models. Lopinavir is an established protease inhibitor that approved for use in HIV patients and is usually used in conjunction with ritonavir, another protease inhibitor.[47] Lopinavir and lopinavir/ritonavir have been tested previously on SARS[48] and MERS-CoV[49], but recent clinical trials suggest that the drug combination is not as successful as expected against SARS-CoV-2.[50]

Since no data is available to build models for SARS-CoV-2 M$^{pro}$ and considering the high similarity between these targets, we we decided to employ these models to virtually screen the curated DrugBank dataset and submit these molecules for experimental evaluation.. Applying our models to screen this dataset of 9,615 compounds yielded 41 compounds predicted as actives using a Consensus and Consensus AD models.

In parallel, we have also conducted molecular docking exeriments using the structure of $M^{pro}$ from SARS-CoV-2 (PDB ID: 6LU7).[11] Before using docking as a virtual screening tool, it is crucial to validate the approach with known experimental data. Therefore, known inhibitors and non-inhibitors of $M^{pro}$ were used to evaluate if the docking score was capable of ranking active compounds better than inactives. For this purpose, the curated dataset (CHEMBL3927) used for QSAR modeling was applied in a docking validation run. Then, compound ranking by the docking score was compared with ranking by activity in the ChEMBL assay. The results suggested that docking scores were poorly correlated with the binding affinity as indicated by the area under the receiver operating characteristic (ROC) score of 0.49 (Figure 3), implying that docking scores randomly assigned compounds as actives and inactives. Additionally, the early enrichment was poor with sensitivity of only 0.11 for the top 10% ranked compounds, i.e., actives were ranked poorly while inactives were occupying the top of the list of virtual hits. The top 15% also presented poor sensitivity (0.14) . Only after the top 69% of the list was considered, the sensitivity reached reasonable values (0.70). Based on these results, docking was discarded as a virtual screening approach.

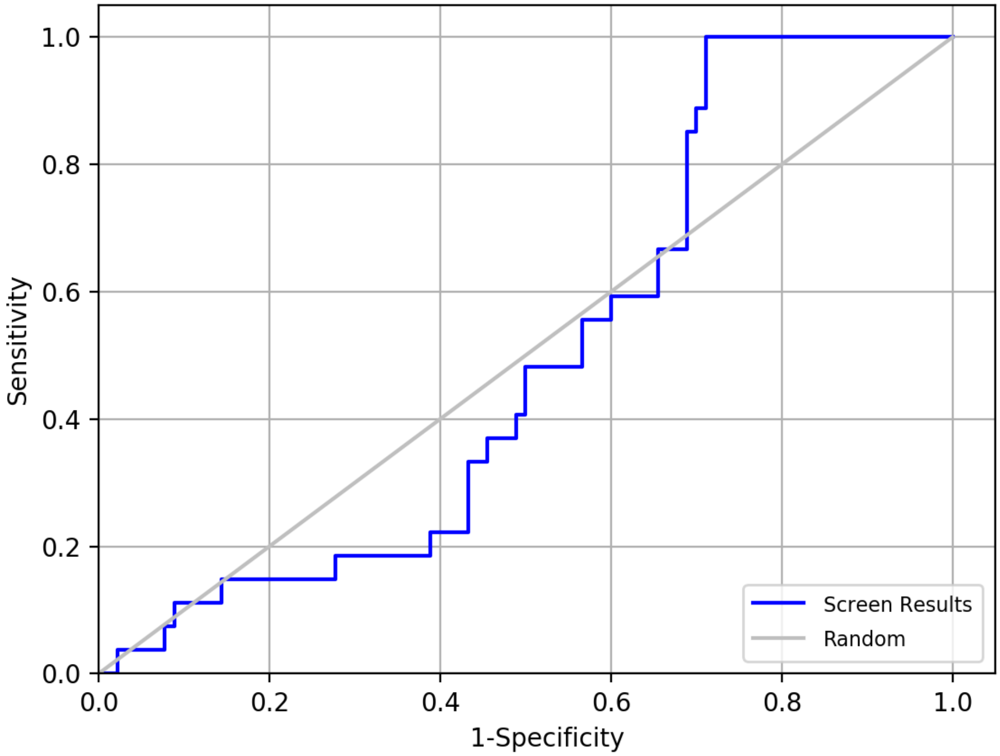

**Figure 3.** Receiver operating characteristic (ROC) after running the docking validation screening with known inhibitors and non-inhibitors of $M^{pro}$.

We also employed a similarity search using three compounds described by Wang et al.[12] as[39] active in the phenotypic screen (remdesivir, chloroquine, and nitazoxanide). We found that only the following 13 compounds from the curated DrugBank dataset had Tanimoto similarity coefficient higher than 75% to any of those three drugs: anhydrovinblastine, GS-6620, hydroxychloroquine, lurbinectedin, quinacrine, quinacrine mustard, rifalazil, vinblastine, vincristine, vindesine, vinflunine, vinorelbine, and 3''-(beta-chloroethyl)-2'',4''-dioxo-3, 5''-spiro-oxazolidino-4-deacetoxy-vinblastine.

Five out of 13 compounds were predicted as active by SiRMS models, including anhydrovinblastine, vincristine, vindesine, vinflunine, vinorelbine. SiRMS and Dragon together also predicted lurbinectedin, rifalazil, vinblastine and 3''-(beta-chloroethyl)-2'',4''-dioxo-3, 5''-

spiro-oxazolidino-4-deacetoxy-vinblastine as active. Most of these compounds are vinca alkaloids. Most literature on this class of alkaloids concerns cancer biology, since many are chemotherapy drugs, but other classes of alkaloids have been noted to have antiviral activities.[51–54] Interestingly, ritonavir, a protease inhibitor used in the treatment of HIV and that is being tested currently in clinical trials for COVID-19 boosts the levels of chemotherapy drugs, including vinca alkaloids.[55] Vinca alkaloids are used as chemotherapy drugs, but can have problematic side effects.[56] Lurbinectedin and rifalazil are both potent RNA polymerase inhibitors; lurbinectedin is used as an anticancer agent[57] while rifalazil has shown success in treating Chlamydia trachomatis infections.[58]

Thus, we selected 41 hits from DrugBank based on QSAR predictions, including four compounds identified by similarity search and predicted by both SiRMS and Dragon. These hits have been found among commercially available compounds listed in ZINC database[59] and the vendors selling these compouds were identified using our in-house ZINC-Express software (https://zincexpress.mml.unc.edu/) (Table S1 in Supplemental Materials). We purchased 10 compounds (Table 2) that were financially feasible for testing and submitted them for experimental evaluation by our collaborators at the University of Kentucky. The experimental data for testing these compounds in $M^{pro}$ assay will be reported in the updated version of this manuscript once the results become available. The complete list of hits is available in the supplementary materials.

**Table 2.** Selected hits for experimental evaluation.

| Generic name | Primary use | DrugBank ID |
|---|---|---|
| Ipamorelin | postoperative ileus | investigational |
| Tilmicosin | antibiotic | investigational; vet_approved |
| Budipine | antiparkinson | experimental |
| Atazanavir | HIV | approved; investigational |
| Pentagastrin | stimulates gastric acid secretion | approved |

| | | |
|---|---|---|
| Indinavir | HIV | approved |
| Vinblastine | Anti-cancer | approved |
| Afimoxifene | estrogen receptor modulator/anti-cancer | investigational |
| Navitoclax | Bcl-2 inhibitor/anti-cancer | investigational |
| Venetoclax | chronic lymphocytic leukemia | approved; investigational |

Of the model's top hits, two of the most promising are camostat and nitazoxanide, which are currently being tested in clinical trials[60,61] and have demonstrated anti-coronaviral activities in past studies.[62] Camostat is a serine protease inhibitor[63] and nitazoxanide is a broad-spectrum antiviral drug.[62,64,65] Analysis of the literature suggests that selumetinib, PD-0325901, and leflunomide (see Table 3) are also promising candidates, as they are known kinase inhibitors that also have suggested antiviral activity.[43,65] Leflunomide is an anti-rheumatic drug that has shown past antiviral activity against cytomegalovirus as well as immunosupressivity. Its metabolite, A77 1726, can inhibit protein kinase activity and the activity of dihydroorotate dehydrogenase (DHODH), the latter which has been suggested as a possible host antiviral target for SARS-CoV-2.[64] Selumetinib and PD-0325901 are MEK inhibitors; of the two, selumetinib is the only to have demonstrated anticoronaviral activity (against SARS- and MERS-CoV) in past studies.[66] In combination with another hit from Table 3, oseltamivir, PD-0325901 has shown antiviral activity against the influenza virus,[62] though it has been suggested that it could serve as a possible antiviral drug by itself.[43]

## Conclusions

In this study, we utilized previous experimental data on SARS-CoV $M^{pro}$ to develop a QSAR model that was used to virtually screen DrugBank in the search for novel potential hits against SARS-CoV-2 $M^{pro}$. As shown in Figure 2, the binding site of $M^{pro}$ is conserved across SARS-CoV and SARS-CoV-2. Collectively, the high conservation of $M^{pro}$ among coronaviruses has been noted in the past and previous studies have explored the potential of developing broad-spectrum antivirals by targeting this enzyme.[16] Molecular docking was not sufficient to discriminate between experimental actives and inactives and was ultimately not used to select hits. The generation of QSAR models according to best practices resulted in 41 virtual hits. Of the other top hits, several compounds currently being tested in clinical trials such as lopinavir and ritonavir were predicted to be active by our models.

The 41 virtual hits were analyzed for availability and price feasibility using our in-house ZINC Express software (*https://zincexpress.mml.unc.edu/*). At the end, 10 compounds (Table 2) were selected for experimental testing by our collaborators at the University of Kentucky. Our group has also selected compound combinations through other methods that will be tested at the National Center for Advancing Translational Sciences. All collected and curated data, models, and virtual screening results are publicly available in the Supplementary Materials of this paper and at GitHub (*https://github.com/alvesvm/sars-cov-mpro*). The curated data are also available in the Chembench web portal (*https://chembench.mml.unc.edu/*).

## Associated Content

Supporting information includes curated datasets and virtual screening results.

## Acknowledgments

This study was inspired by “Calling all coronavirus researchers” Nature editorial[26] and represents goodwill toward the contribution of the authors.

## Conflicts of Interest

The authors declare no actual or potential conflicts of interest.

## References

1. Chen, N. *et al.* Articles Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: a descriptive study. (2020) doi:10.1016/S0140-6736(20)30211-7.
2. Coronavirus latest: Scientists clash over virus name. *Nature 2020* (2020).
3. Statement on the second meeting of the International Health Regulations (2005) Emergency Committee regarding the outbreak of novel coronavirus (2019-nCoV). https://www.who.int/news-room/detail/30-01-2020-statement-on-the-second-meeting-of-the-international-health-regulations-(2005)-emergency-committee-regarding-the-outbreak-of-novel-coronavirus-(2019-ncov).
4. CDC. SARS | Basics Factsheet | CDC. *Centers for Disease Control and Prevention* https://www.cdc.gov/sars/about/fs-sars.html (2017).
5. Donnelly, C. A., Malik, M. R., Elkholy, A., Cauchemez, S. & van Kerkhove, M. D. Worldwide reduction in MERS cases and deaths since 2016. *Emerging Infectious Diseases* **25**, 1758–1760 (2019).
6. Rothe, C. *et al.* Transmission of 2019-NCOV infection from an asymptomatic contact in Germany. *New England Journal of Medicine* vol. 382 970–971 (2020).
7. *Novel Coronavirus(2019-nCoV) Situation Report – 22*. https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200211-sitrep-22-ncov.pdf?sfvrsn=fb6d49b1_2 (2020).
8. Comparing the Wuhan coronavirus outbreak with SARS and MERS. https://graphics.reuters.com/CHINA-HEALTH-VIRUS-COMPARISON/0100B5BY3CY/index.html.
9. Coronavirus latest: global infections surge past 30,000. *Nature 2020* (2020).
10. 2019-nCoV Sequences (Wuhan coronavirus). https://www.ncbi.nlm.nih.gov/genbank/2019-ncov-seqs/.
11. Liu, X., Zhang, B., Jin, Z., Yang, H. & Rao, Z. The crytal structure of 2019-nCoV main protease in complex with an inhibitor N3. *TO BE PUBLISHED* doi:10.2210/PDB6LU7/PDB.
12. Wang, M. *et al.* Remdesivir and chloroquine effectively inhibit the recently emerged novel coronavirus (2019-nCoV) in vitro. doi:10.1038/s41422-020-0282-0.
13. Caly, L., Druce, J. D., Catton, M. G., Jans, D. A. & Wagstaff, K. M. The FDA-approved Drug Ivermectin inhibits the replication of SARS-CoV-2 in vitro. *Antiviral Research* 104787 (2020) doi:10.1016/j.antiviral.2020.104787.
14. Sheahan, T. P. *et al.* An orally bioavailable broad-spectrum antiviral inhibits SARS-CoV-2 in human airway epithelial cell cultures and multiple coronaviruses in mice. *Science translational medicine* (2020) doi:10.1126/scitranslmed.abb5883.
15. Jin, Z. *et al.* Structure of Mpro from COVID-19 virus and discovery of its inhibitors. *bioRxiv* (2020) doi:10.1101/2020.02.26.964882.
16. Li, G. & de Clercq, E. Therapeutic options for the 2019 novel coronavirus (2019-nCoV). *Nature Reviews Drug Discovery* (2020) doi:10.1038/d41573-020-00016-0.
17. Smith, E. C., Blanc, H., Vignuzzi, M. & Denison, M. R. Coronaviruses Lacking Exoribonuclease Activity Are Susceptible to Lethal Mutagenesis: Evidence for

Proofreading and Potential Therapeutics. *PLoS Pathogens* (2013) doi:10.1371/journal.ppat.1003565.

18. Heusipp, G., Gro, C., Herold, J., Siddell, S. G. & Ziebuhr, J. *Printed in Great Britain Identification and subcellular localization of a 41 kDa, polyprotein 1ab processing product in human coronavirus 229E-infected cells*. *Journal of General Virology* vol. 78 (1997).
19. Ziebuhr, J. & Siddell, S. G. Processing of the human coronavirus 229E replicase polyproteins by the virus-encoded 3C-like proteinase: identification of proteolytic products and cleavage sites common to pp1a and pp1ab. *Journal of virology* **73**, 177–85 (1999).
20. Fang, S. G., Shen, H., Wang, J., Tay, F. P. L. & Liu, D. X. Proteolytic processing of polyproteins 1a and 1ab between non-structural proteins 10 and 11/12 of Coronavirus infectious bronchitis virus is dispensable for viral replication in cultured cells. *Virology* **379**, 175–180 (2008).
21. Chen, Y., Liu, Q. & Guo, D. Emerging coronaviruses: Genome structure, replication, and pathogenesis. *Journal of Medical Virology* (2020) doi:10.1002/jmv.25681.
22. Fang, S. G., Shen, H., Wang, J., Tay, F. P. L. & Liu, D. X. Proteolytic processing of polyproteins 1a and 1ab between non-structural proteins 10 and 11/12 of Coronavirus infectious bronchitis virus is dispensable for viral replication in cultured cells. *Virology* **379**, 175–180 (2008).
23. Gao, F., Ou, H.-Y., Chen, L.-L., Zheng, W.-X. & Zhang, C.-T. Prediction of proteinase cleavage sites in polyproteins of coronaviruses and its applications in analyzing SARS-CoV genomes. *FEBS Letters* **553**, 451–456 (2003).
24. Deng, X. *et al.* Coronaviruses Resistant to a 3C-Like Protease Inhibitor Are Attenuated for Replication and Pathogenesis, Revealing a Low Genetic Barrier but High Fitness Cost of Resistance. *Journal of Virology* **88**, 11886–11898 (2014).
25. Kim, Y. *et al.* Broad-Spectrum Antivirals against 3C or 3C-Like Proteases of Picornaviruses, Noroviruses, and Coronaviruses. *Journal of Virology* **86**, 11754–11762 (2012).
26. Calling all coronavirus researchers: keep sharing, stay open. *Nature* **578**, 7 (2020).
27. Ton, A.-T., Gentile, F., Hsing, M., Ban, F. & Cherkasov, A. Rapid Identification of Potential Inhibitors of SARS- CoV-2 Main Protease by Deep Docking of 1.3 Billion Compounds. *Molecular Informatics* 1–18 (2020) doi:10.1002/minf.202000028.
28. Zhang, D. hai, Wu, K. lun, Zhang, X., Deng, S. qiong & Peng, B. In silico screening of Chinese herbal medicines with the potential to directly inhibit 2019 novel coronavirus. *Journal of Integrative Medicine* **18**, 152–158 (2020).
29. Ekins, S. *et al.* Deja vu: Stimulating Open Drug Discovery For SARS-CoV-2. *OSF Preprints* 1–50.
30. Ge, Y. *et al.* A data-driven drug repositioning framework discovered a potential therapeutic agent targeting COVID-19. *bioRxiv* 2020.03.11.986836 (2020) doi:10.1101/2020.03.11.986836.
31. Fourches, D., Muratov, E. & Tropsha, A. Trust, but verify: On the importance of chemical structure curation in cheminformatics and QSAR modeling research. *Journal of Chemical Information and Modeling* (2010) doi:10.1021/ci100176x.
32. Fourches, D., Muratov, E. & Tropsha, A. Trust, but Verify II: A Practical Guide to Chemogenomics Data Curation. *Journal of Chemical Information and Modeling* (2016) doi:10.1021/acs.jcim.6b00129.

33. Figueras, J. Morgan Revisited. *Journal of Chemical Information and Computer Sciences* (1993) doi:10.1021/ci00015a009.
34. Muratov, E. N. *et al.* Per aspera ad astra: Application of Simplex QSAR approach in antiviral research. *Future Medicinal Chemistry* (2010) doi:10.4155/fmc.10.194.
35. Kuz'min, V. E. *et al.* The effects of characteristics of substituents on toxicity of the nitroaromatics: HiT QSAR study. *Journal of Computer-Aided Molecular Design* (2008) doi:10.1007/s10822-008-9211-x.
36. Tropsha, A. Best practices for QSAR model development, validation, and exploitation. *Molecular Informatics* (2010) doi:10.1002/minf.201000061.
37. Breiman, L. Random Forests. *Mach Learn* 5–32 (2001).
38. Artemenkoa, A. G. *et al.* Qsar analysis of the toxicity of nitroaromatics in tetrahymena pyriformis: Structural factors and possible modes of action. *SAR and QSAR in Environmental Research* (2011) doi:10.1080/1062936X.2011.569950.
39. Friesner, R. A. *et al.* Glide: A New Approach for Rapid, Accurate Docking and Scoring. 1. Method and Assessment of Docking Accuracy. *Journal of Medicinal Chemistry* (2004) doi:10.1021/jm0306430.
40. Hilgenfeld, R. From SARS to MERS: crystallographic studies on coronaviral proteases enable antiviral drug design. *The FEBS journal* (2014) doi:10.1111/febs.12936.
41. Bateman, A. UniProt: A worldwide hub of protein knowledge. *Nucleic Acids Research* (2019) doi:10.1093/nar/gky1049.
42. Schrödinger. Maestro | Schrödinger. *Schrödinger Release 2018-1* (2018).
43. Li, G. & de Clercq, E. Therapeutic options for the 2019 novel coronavirus (2019-nCoV). *Nature Reviews Drug Discovery* **19**, 149–150 (2020).
44. Jin, Z. *et al.* Structure of Mpro from COVID-19 virus and discovery of its inhibitors. *bioRxiv* (2020) doi:10.1101/2020.02.26.964882.
45. Target Report Card. https://www.ebi.ac.uk/chembl/target_report_card/CHEMBL612575/.
46. Home - ClinicalTrials.gov. https://clinicaltrials.gov/ct2/home.
47. Kempf, D. J. *et al.* Identification of Genotypic Changes in Human Immunodeficiency Virus Protease That Correlate with Reduced Susceptibility to the Protease Inhibitor Lopinavir among Viral Isolates from Protease Inhibitor-Experienced Patients. *Journal of Virology* (2001) doi:10.1128/jvi.75.16.7462-7469.2001.
48. Chu, C. M. *et al.* Role of lopinavir/ritonavir in the treatment of SARS: Initial virological and clinical findings. *Thorax* (2004) doi:10.1136/thorax.2003.012658.
49. Chan, J. F. W. *et al.* Treatment with lopinavir/ritonavir or interferon-β1b improves outcome of MERSCoV infection in a nonhuman primate model of common marmoset. *Journal of Infectious Diseases* (2015) doi:10.1093/infdis/jiv392.
50. Cao, B. *et al.* A Trial of Lopinavir-Ritonavir in Adults Hospitalized with Severe Covid-19. *The New England journal of medicine* (2020) doi:10.1056/NEJMoa2001282.
51. Moradi, M. T., Karimi, A., Rafieian-Kopaei, M. & Fotouhi, F. In vitro antiviral effects of Peganum harmala seed extract and its total alkaloids against Influenza virus. *Microbial Pathogenesis* (2017) doi:10.1016/j.micpath.2017.06.014.
52. Özçelik, B., Kartal, M. & Orhan, I. Cytotoxicity, antiviral and antimicrobial activities of alkaloids, flavonoids, and phenolic acids. *Pharmaceutical Biology* (2011) doi:10.3109/13880209.2010.519390.

53. Farnsworth, N. R., Svoboda, G. H. & Blomster, R. N. Antiviral activity of selected Catharanthus alkaloids. *Journal of Pharmaceutical Sciences* (1968) doi:10.1002/jps.2600571235.

54. Houghton, P. J., Woldemariam, T. Z., Khan, A. I., Burke, A. & Mahmood, N. Antiviral activity of natural and semi-synthetic chromone alkaloids. *Antiviral Research* (1994) doi:10.1016/0166-3542(94)90006-X.

55. Srirangam, A. *et al.* The human immunodeficiency virus protease inhibitor ritonavir inhibits lung cancer cells, in part, by inhibition of survivin. *Journal of Thoracic Oncology* (2011) doi:10.1097/JTO.0b013e31820c9e3c.

56. Moudi, M., Go, R., Yien, C. Y. S. & Nazre, M. Vinca alkaloids. *International Journal of Preventive Medicine* (2013) doi:10.2165/00128415-200711380-00080.

57. Jiménez, C. Marine Natural Products in Medicinal Chemistry. *ACS Medicinal Chemistry Letters* (2018) doi:10.1021/acsmedchemlett.8b00368.

58. Rothstein, D. M., Shalish, C., Murphy, C. K., Sternlicht, A. & Campbell, L. A. Development potential of rifalazil and other benzoxazinorifamycins. *Expert Opinion on Investigational Drugs* (2006) doi:10.1517/13543784.15.6.603.

59. Irwin, J. J. & Shoichet, B. K. ZINC - A free database of commercially available compounds for virtual screening. *Journal of Chemical Information and Modeling* (2005) doi:10.1021/ci049714+.

60. Rossignol, J. F. Nitazoxanide, a new drug candidate for the treatment of Middle East respiratory syndrome coronavirus. *Journal of Infection and Public Health* **9**, 227–230 (2016).

61. Zhou, Y. *et al.* Protease inhibitors targeting coronavirus and filovirus entry. *Antiviral Research* **116**, 76–84 (2015).

62. Hsieh, H.-P. & Hsu, J. Strategies of Development of Antiviral Agents Directed Against Influenza Virus Replication. *Current Pharmaceutical Design* (2007) doi:10.2174/138161207782794248.

63. Rossignol, J. F. Nitazoxanide: A first-in-class broad-spectrum antiviral agent. *Antiviral Research* vol. 110 94–103 (2014).

64. Schor, S. & Einav, S. Repurposing of Kinase Inhibitors as Broad-Spectrum Antiviral Drugs. *DNA and Cell Biology* **37**, 63–69 (2018).

65. Chong, A. S. *et al.* Concurrent Antiviral and Immunosuppressive Activities of Leflunomide in Vivo. *American Journal of Transplantation* (2006) doi:10.1111/j.1600-6143.2005.01152.x.

66. Haasbach, E., Hartmayer, C. & Planz, O. Combination of MEK inhibitors and oseltamivir leads to synergistic antiviral effects after influenza A virus infection in vitro. *Antiviral Research* **98**, 319–324 (2013).

67. Kandil, S. *et al.* ProTides of BVdU as potential anticancer agents upon efficient intracellular delivery of their activated metabolites. *Bioorganic and Medicinal Chemistry Letters* (2016) doi:10.1016/j.bmcl.2016.10.077.