


Quantitative Structure–Activity Relationship Machine Learning Models and their Applications for Identifying Viral 3CLpro- and RdRp-Targeting Compounds as Potential Therapeutics for COVID-19 and Related Viral Infections

Julian Ivanov,^{*,†} Dmitrii Polshakov,[†] Junko Kato-Weinstein, Qiongqiong Zhou, Yingzhu Li, Roger Granet, Linda Garner, Yi Deng, Cynthia Liu, Dana Albaiu, Jeffrey Wilson, and Christopher Aultman

 Cite This: *ACS Omega* 2020, 5, 27344–27358

 Read Online

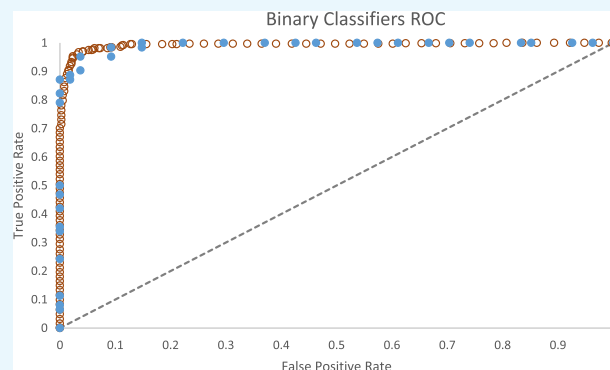
ACCESS |

 Metrics & More

 Article Recommendations

 Supporting Information

ABSTRACT: In response to the ongoing COVID-19 pandemic, there is a worldwide effort being made to identify potential anti-SARS-CoV-2 therapeutics. Here, we contribute to these efforts by building machine-learning predictive models to identify novel drug candidates for the viral targets 3 chymotrypsin-like protease (3CLpro) and RNA-dependent RNA polymerase (RdRp). Chemist-curated training sets of substances were assembled from CAS data collections and integrated with curated bioassay data. The best-performing classification models were applied to screen a set of FDA-approved drugs and CAS REGISTRY substances that are similar to, or associated with, antiviral agents. Numerous substances with potential activity against 3CLpro or RdRp were found, and some were validated by published bioassay studies and/or by their inclusion in upcoming or ongoing COVID-19 clinical trials. This study further supports that machine learning-based predictive models may be used to assist the drug discovery process for COVID-19 and other diseases.



1. INTRODUCTION

The ongoing COVID-19 pandemic has challenged the health system of many countries with its high rate of morbidity and mortality. Developing effective treatments against the disease is critical in the effort to save human lives and help society return to normal. In order to respond quickly, scientists have been exploring various ways to accelerate the drug development process, including drug repurposing and computational approaches in drug discovery. One promising way to achieve this goal is to build machine learning models by applying the quantitative structure–activity relationship (QSAR) methodology to suitable protein targets of the SARS-CoV-2 virus, which could predict possible drug candidates for treating COVID-19.

Among all the proteins in SARS-CoV-2, the 3 chymotrypsin-like protease (3CLpro) and RNA-dependent RNA polymerase (RdRp) are two ideal protein targets for QSAR modeling. 3CLpro is a protease that is required in order for the coronavirus to cleave the polyprotein peptides into individual functional nonstructural proteins.¹ In addition, by comparing the amino acid sequences and protein structures, the 3CLpro was found to be highly conserved among SARS-CoV-2 and other human coronaviruses, with sequence identities of 96% with SARS-CoV-1, 87% with MERS-CoV, and 90% with Human-CoV.^{2–5}

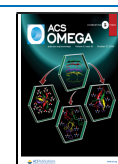
Therefore, the 3CLpro inhibitors identified in previous coronavirus-related research are promising inhibitors for SARS-CoV-2 3CLpro and their associated SAR study data are valuable information as training material for machine learning models in searching for new inhibitors of SARS-CoV-2 3CLpro. Consequently, it is reasonable to propose that broad-spectrum inhibitors against human coronaviruses are promising drug candidates that can be developed into suitable drugs for targeting SARS-CoV-2 and many other human coronaviruses.

RdRp is the major enzyme that is responsible for replicating viral genomic RNA in host cells. The amino acid residues of active sites in RdRp are highly conserved among single-stranded, positive-sense RNA [(+)ssRNA] viruses, including SARS-CoV-1 and hepatitis C virus (HCV).⁶ In addition, RdRp of SARS-CoV-2 shares an almost identical protein sequence with that of

Received: July 31, 2020

Accepted: September 29, 2020

Published: October 14, 2020



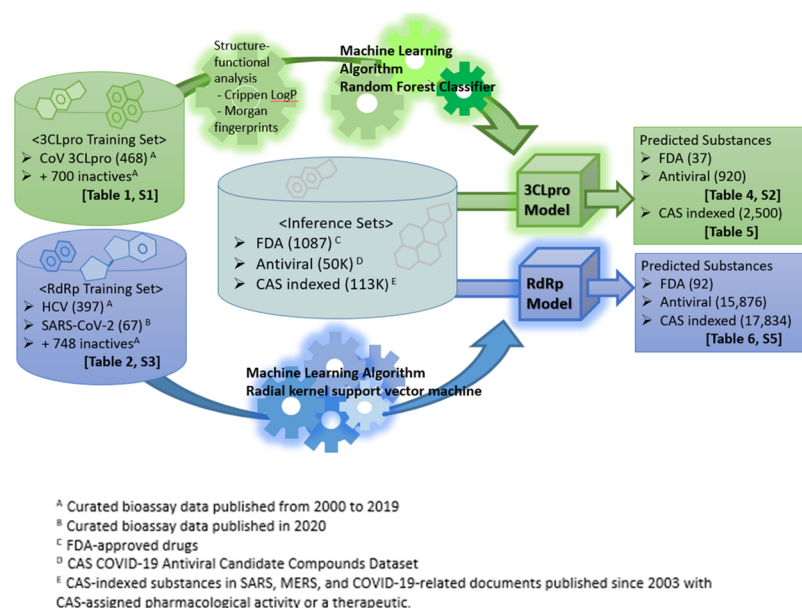


Figure 1. Overall process flow.

SARS-CoV-1, indicating that it also shares the highly conserved amino acid residues in the active sites among the (+)ssRNA viruses.¹ Indeed, the EM-cryo structure of SARS-CoV-2 RdRp reveals that this enzyme contains the classic divalent-cation-binding residue D618, which is conserved in most viral polymerases including HCV (residue D220), and the catalytic residues [759 to 761 (SDD)], which are also conserved in most viral RdRps, such as HCV.⁷ Fortunately, various viral RdRps have been widely studied for inhibitors in the (+)ssRNA viruses, especially in HCV-related research.⁸ Therefore, these existing RdRp inhibitors for the (+)ssRNA viruses, such as HCV, may provide valuable insights for drug development for SARS-CoV-2 RdRp inhibition.⁸

QSAR and machine learning^{9a} modeling have increasingly been used to facilitate drug discovery in recent years. QSAR is usually one of the first steps in the drug discovery process, in which large databases of chemical structures are screened through a variety of predictive mathematical models in order to narrow down the number of potential drug candidates. Simply put, QSARs are mathematical models approximating rather complicated biological or physicochemical properties of chemicals based on quantitative measures of the corresponding molecular structures. The underlying assumption is that the activity is directly related to the structure of the chemicals. Thus, molecules with similar structural features will exhibit similar physical properties or similar biological effects. One major challenge in QSAR is the selection of appropriate structural features to be used as molecular descriptors. To a large degree, the development of an accurate machine learning model depends on the results of the analysis of the factors that may be meaningful in the mechanism of action (MOA). Relevant molecular descriptors can be defined and calculated only after a solid understanding of the MOA has been achieved.

Many techniques have been developed to describe molecular structures in QSAR. Hansch et al.^{13–15} used 1-octanol/water partition coefficients and Hammett substituent constants to model hydrophobic and electronic effects. Methods that utilize topological indexes,^{16–20} 3D geometry descriptors,^{21–24} electronic structural descriptors,^{25–27} molecular shape descriptors,²⁸ fragment-based approaches,^{29–31} and 3D QSAR^{32,33}

have been extensively used to model a wide range of biological end points.

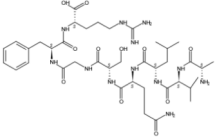
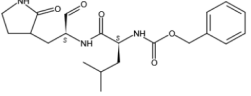
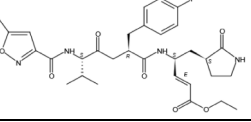
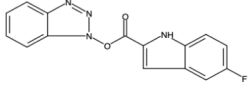
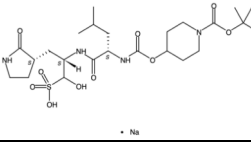
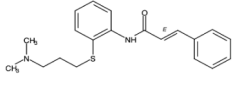
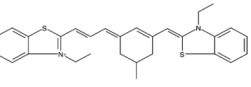
Since the outbreak of COVID-19, different QSAR studies have been employed to predict the possibilities of FDA-approved drugs or investigational drugs to be repurposed for COVID-19, as well as the likelihood of repurposing inhibitors for other viruses related to SARS-CoV-2. QSAR modeling has been proven to be useful in virtual screening and lead optimization for drug discovery. Monte Carlo optimization-based and classification QSAR models of SARS-CoV main protease (Mpro) were developed and used for screening of some natural products.^{9b} Multiple linear regression analysis combined with 2D-QSAR modeling was successfully employed to identify potential SARS-CoV-3CLpro enzyme inhibitors.^{10–12a} Classification QSAR data mining of diverse SARS-CoV papain-like protease (PLpro) inhibitors were utilized to generate predictive QSAR models and later used for virtual screening to identify molecules that could be effective against SARS-CoV PLpro.^{12b}

Development of a QSAR model also requires thorough collaboration between computational and experimental scientists who have deep knowledge of their respective fields and have made effort to understand and develop working knowledge of the complementary field. In this study, computational scientists and chemists with related background closely collaborated to curate training data and build highly predictive QSAR models for 3CLpro and RdRp. The two QSAR models presented here were validated with high sensitivity, specificity, and accuracy. After modeling protease inhibitor activity as a function of the substance structure, we identified some of the most promising candidates among substances predicted to be active inhibitors against coronavirus 3CLpro.

2. METHODS

In this study, as illustrated in Figure 1, scientists curated over 1000 inhibitors with structure–bioactivity data as training molecules for 3CLpro and RdRp protein targets. We collected these data from the most current SARS-CoV-2 bioassay studies as well as existing studies with SARS-CoV-1, MERS-CoV, and other related viruses in the CAS data collection. Using this data,

Table 1. List of Example Training Substances with *In Vitro* Inhibitory Effect on SARS-CoV-1 3CLpro

Substance	CAS Registry Number®	Structure	Inhibition Value (μM)	Source
Octapeptide AVLQSGFR	608531-54-0		$\text{IC}_{50} = 0.031$	Current Medicinal Chemistry (2006), 13(17), 2003-2020
GC 373	1333231-44-9		$\text{IC}_{50} = 3.48$	US2014024334
Rupintrivir	223537-30-2		$\text{IC}_{50} = 3.48$	WO2013049382
1H-Indole-2-carboxylic acid, 5-fluoro-, 1H-benzotriazol-1-yl ester	890402-79-6		$\text{IC}_{50} = 0.0138$	Journal of Medicinal Chemistry (2016), 59(14), 6595-6628
GC868	2170616-55-2		$\text{IC}_{50} = 0.4$	European Journal of Medicinal Chemistry (2018), 150, 334-346
(2E)-N-[2-[[3-(Dimethylamino)propyl]thio]phenyl]-3-phenyl-2-propenamide	33464-86-7		$\text{IC}_{50} = 4.92$	Journal of Virology (2005), 79(11), 7095-7103
3-Ethyl-2-[3-[3-(3-ethyl-2(3H)-benzothiazolylidene)methyl]-5-methyl-2-cyclohexen-1-ylidene]-1-propen-1-yl]benzothiazolium	393540-75-5		$\text{IC}_{50} = 2$	Antimicrobial Agents and Chemotherapy (2007), 51(11), 3924-3931

we applied a variety of machine learning algorithms to build several dozen QSAR models—selecting from among these, the strongest performing models—one targeting 3CLpro and one targeting RdRp.

We used the resulting models to screen 1087 FDA-approved drugs,³⁴ nearly 50,000 substances from the CAS COVID-19 Antiviral Candidate Compounds Dataset,^{35a} and a list of 113,000 substances with CAS-assigned pharmacological activity or a therapeutic role indexed in SARS, MERS, and COVID-19-related documents published since 2003. Some predicted molecules of these models were validated by published bioassay studies and clinical trials as a positive indication of the predictive models.

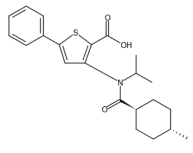
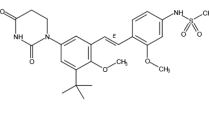
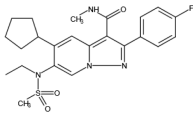
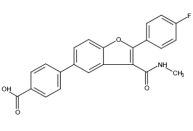
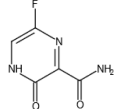
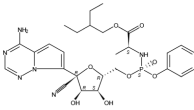
2.1. 3CLpro & RdRp Training Data set Preparation. To develop training sets for predicting active substances to target SARS-CoV-2 3CLpro and RdRp, we examined bioassay data published from 2000 to 2020 in the CAS data collections. This included substance information, targets, activity measures [half maximal inhibitory concentration (IC_{50}), half maximal effective concentration (EC_{50}), inhibition constant (K_i), and dissociation constant (K_d)], source organisms, and assay details. We selected

substances with assay data of $\text{IC}_{50} \leq 10 \mu\text{M}$, $\text{EC}_{50} \leq 10 \mu\text{M}$, $K_i \leq 10 \mu\text{M}$, and/or $K_d \leq 10 \mu\text{M}$ toward the targets as active substances from these bioassay data—a threshold suggested by Deng et al.^{35b} and by Zhou et al.^{35c} Substances with IC_{50} , EC_{50} , K_i , and K_d values above $100 \mu\text{M}$ from these bioassay data were considered to be inactive.

2.1.1. 3CLpro Inhibitors of SARS-CoV-1 and other Coronaviruses. For the training set for the 3CLpro model, we collected 468 diverse active substances from 76 different documents published prior to 2020—and 700 inactive substances. Over 70% of active substances are from SARS-CoV-1 3CLpro studies. The rest of the active substances are from MERS-CoV, common human coronavirus, feline infectious peritonitis, transmissible gastroenteritis coronavirus, and infectious bronchitis virus 3CLpro studies. Seven unique small molecules representing the structural diversity of SARS-CoV-1 3CLpro inhibitors are shown in Table 1. See Supporting Information Table S1 for the complete training set.

2.1.2. RdRp Inhibitors of SARS-CoV-1 and Other Viruses. For the training set for the RdRp model, we collected 1212 diverse active substances from 51 documents published prior to

Table 2. Examples of Training Substances with Inhibitory Effect on RdRp of HCV (a Positive-Stranded ssRNA Virus) in Various Biological Studies

Substance	CAS Registry Number	Structure	Inhibition Values (μM)	Assay Detail	Source
3-[Isopropyl(trans-4-methylcyclohexylcarbonyl)amino]-5-phenylthiophene-2-carboxylic acid	478025-29-5		$\text{IC}_{50} = 0.028$	<i>In vitro</i> inhibition of HCV 1b RdRp	Bioorganic & Medicinal Chemistry Letters (2010), 20(15), 4614-4619
			$\text{IC}_{50} = 0.009$	<i>In vitro</i> inhibition of HCV 1a RdRp	Journal of Medicinal Chemistry (2009), 52(16), 5217-5227
N-[4-[(1E)-2-[3-(1,1-Dimethylethyl)-2-methoxy-5-(tetrahydro-2,4-dioxo-1(2H)-pyrimidinyl)phenyl]ethenyl]-3-methoxyphenyl]methanesulfonamide	1132937-23-5		$\text{IC}_{50} \leq 0.01$	<i>In vitro</i> inhibition of HCV 1a RdRp	US20110070193
			$\text{IC}_{50} \leq 0.01$	<i>In vitro</i> inhibition of HCV 1a RdRp	US20090176784
5-Cyclopentyl-6-[ethyl(methylsulfonyl)amino]-2-(4-fluorophenyl)-N-methylpyrazolo[1,5-a]pyridine-3-carboxamide	1215105-09-1		$\text{IC}_{50} = 0.25\text{--}1.0$	<i>In vitro</i> inhibition of HCV 1a RdRp	US8198449
			$\text{IC}_{50} > \text{R}$	<i>In vitro</i> inhibition of HCV 1b RdRp	WO2011112186
			$\text{IC}_{50} > 0.25\text{--}1.0$	<i>In vitro</i> inhibition of HCV RdRp	US20100184800
4-[2-(4-Fluorophenyl)-3-[(methylamino)carbonyl]-5-benzofuranyl]benzoic acid	1217334-55-8		$\text{IC}_{50} = 0.25\text{--}1.0$	<i>In vitro</i> inhibition of HCV 1b RdRp	US7994171 B2
			$\text{IC}_{50} > 0.25\text{--}1.0$	<i>In vitro</i> inhibition of HCV 1b RdRp	WO2011112191
Favipiravir	259793-96-9		$\text{EC}_{50} = 61.88$	<i>In vitro</i> inhibition of SARS-CoV-2 virus yield in	Cell Research (2020), 30(3), 269-271
Remdesivir	1809249-37-3		$\text{EC}_{50} = 0.77$	Vero E6 cells	

2020. We also collected 67 SARS-CoV-2 RdRp inhibitors from journal articles published in 2020. Four unique small molecules representing the structural diversity of HCV RdRp inhibitors and two unique small molecules representing SARS-CoV-2 RdRp inhibitors are shown in Table 2. A full list of the training substances can be seen in Table S4.

2.2. Machine Learning Methods. An automated machine learning platform, DataRobot v6.0.3 <https://www.datarobot.com/>, was used to train and evaluate performance of more than 40 different machine learning algorithms. DataRobot is a commercial tool that we used to build informative features selected from molecular descriptors.³⁸ Model training and testing with fivefold cross-validation was utilized in the model selection process. Grid search was used as the default method for hyperparameter optimization.

Support Vector Classifier (Radial Kernel) is a robust algorithm that has been actively used to model biological activity of small molecules for various targets.^{39,40} DataRobot implementation is based on scikit-learn.⁴¹ This algorithm searches the optimal hypersurface in multidimensional feature space³⁸ that separates the active and inactive classes of compounds. DataRobot tool identified SVM with Radial Kernel as the best model for our RdRp data set based on highest cross-validation area under curve (AUC) values among the other

explored models. Hyperparameters Gamma and C were optimized using default DataRobot grid search.

XGBoost is a decision-tree-based ensemble ML algorithm that uses a gradient boosting framework.^{42,43} For building RdRp models, default DataRobot optimized parameters were used. Previously XGBoost demonstrated both high accuracy and robustness on bioactivity prediction tasks with highly imbalanced class distributions.⁴⁴

Many ensemble models that are called “blenders” in DataRobot also demonstrated excellent performance that was comparable to top selected models. For example, the ENET Blender model combines the predictions generated by SVM and XGBoost models followed by Elastic-Net Classifier⁴⁵—which is a linear model trained with L1 and L2 prior as regularizers. Very similar to that, average blender also combines the output of SVM and XGBoost predictions and returns mean that can improve performance of individual models.⁴⁶ Ensemble-based models also were found to be an improvement for building QSAR models.⁴⁷

2.3. Binary Classifiers. In our models, bioactivity prediction is a binary classification task that generates the probability of compounds, represented as a structural feature vector, being a member of the active or inactive class. An automated machine

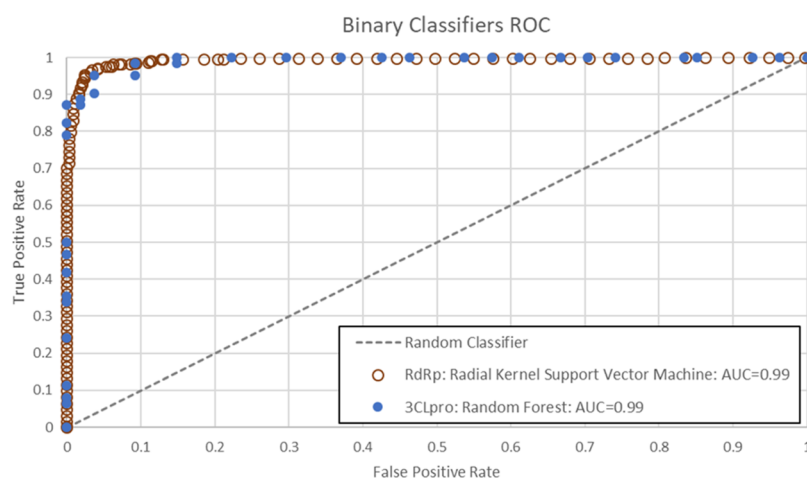


Figure 2. ROC curves for 3CLpro and RdRp trained binary classifiers.

learning tool was used to train and evaluate performance of more than 40 different machine learning algorithms.

A number of common metrics were used to evaluate statistical performance of our models, including:

$$\text{sensitivity} = \text{TP}/(\text{TP} + \text{FN})$$

$$\text{specificity} = \text{TN}/(\text{TN} + \text{FP})$$

$$\text{accuracy} = (\text{TP} + \text{TN})/(\text{TP} + \text{TN} + \text{FP} + \text{FN})$$

Receiver Operating Characteristic³⁵ (ROC)

where: TP—true positives. TN—true negatives. FP—false positives. FN—false negatives.

To estimate the statistical performance of the models, a common fivefold cross-validation procedure has been used for all models in this study. In each of five iterations of this approach, 80% of the data is used to build a model, and 20% is held as a test set. Across the whole process, then, every record is held out as validation in one part of the process, yet all records are made available to the model.

To measure the statistical quality of our models, we used the receiver–operator-characteristic (ROC)-AUC³⁶ that is a measure of model fit ranging from 0 to 1 and created by plotting the true positive rate against the false positive rate at various threshold settings, with a perfect model having an ROC-AUC of 1.

Many binary classifiers from our training data showed high AUC, accuracy, sensitivity, and specificity in the cross-validation tests. AUCs are shown in Supporting Information Table S5, with a description of validation metrics. ROC plots for the final models are shown in Figure 2. Many models—including Support Vector Machine with Radial Kernel (SVM), Gradient Boosted Trees, ENET, and variations of AVG Blenders—showed AUC values of 0.99.

The high performance of these binary classifier models provides us with the ability to identify the most relevant active molecules in large data sets. This can be attributed to (i) the more separated distributions of actives and inactives (observations with a midrange of $10 \mu\text{M} < \text{IC}_{50} < 100 \mu\text{M}$ were excluded from the training data sets) and (ii) the high diversity of active and inactive examples in the training data that were prepared by CAS scientists.

2.4. Molecular Descriptors. In order to train the binary classification machine learning models that predict the

probability of whether a new chemical will inhibit the human coronavirus 3CLpro or RdRp, the abovementioned training set molecules were used, and their molecular features were studied for the models. Molecular descriptors available in the software library RDKit,³⁷ including Morgan⁴⁸ (radius = 3; length = 2048), MACCS keys,⁴⁹ Atom Pairs,⁵⁰ Topological Torsion⁵¹ fingerprints, as well as Crippen LogP⁵² and MR,⁵² Molecular Quantum Numbers (MQN),³⁸ PEOE_VSA and SMR_VSA,⁵³ and FractionCSP3 (fraction of C atoms that are SP3 hybridized) were utilized to build the models.

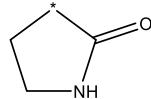
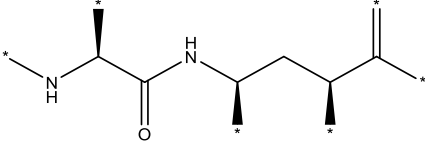
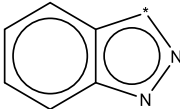
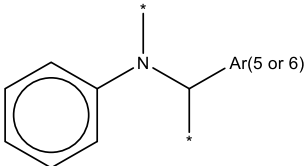
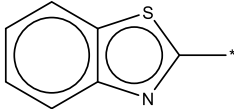
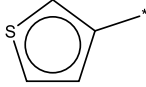
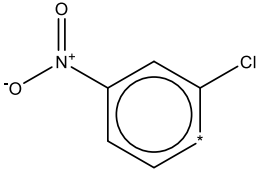
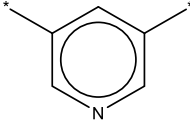
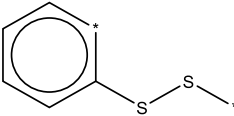
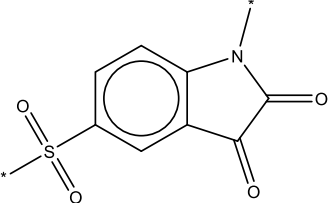
2.5. Structure–Functional Analysis. In order to calculate and evaluate structural alerts, the following structure–functional analysis approach has been developed. The molecular graph of a given chemical structure is split into subfragments similar to Klopman.²⁹ However, beyond just linear fragments, we further generated all possible subfragments with extended connectivity length from 1 to 9 atoms. The activity of the molecular structure is assigned to any of the generated subfragments. The above procedure is applied to every single molecule structure in the training set; thus, we end up with an extended list of fragments that may belong to several different molecules. In this study, we calculate the activity of a substructure as the mean value of the activities of the molecules that contain the fragment. Once the activity of the fragments is calculated, we order the list of fragments so that at the top of the list, are the most active structural alerts and at the bottom, the most inactive ones.

3. RESULTS AND DISCUSSION

3.1. 3CLpro Model. After the calculation of the molecular descriptors, several machine learning algorithms, including Random Forest, Gradient Boosting, Neural Networks, and Support Vector Machine (SVM), to name a few (see Methods Section 2.2 above), were exploited in order to obtain robust machine learning models. The best model was obtained by using a Random Forest Classifier algorithm and utilizing Crippen LogP and Morgan fingerprints as molecular representations. This model achieved a ROC-AUC of 0.99, as previously shown in Figure 2.

To assess the predictive ability of the 3CLpro model, we then performed a thorough structure–functional analysis of the substances in the training set. The following salient structural features, or “alerts”, were found to be prevalent among active substances and might be partly responsible for the 3CLpro inhibition activity (Table 3).

Table 3. Structural Alerts for the 3CLpro Model

Structural alert	Active substances in training set (N=468)
	138
	85
	60
	52
	40
	35
	26
	24
	18
	10

In order to predict the possibilities of FDA-approved drugs to be repurposed for COVID-19, this 3CLpro model was applied to predict inhibition activity in a data set of 1087 FDA-approved drugs. As approved drugs, these substances would be expected to have an acceptable ADME profile and side effects and might be

likely to secure faster FDA approval as coronavirus or COVID-19 therapeutics. The model predicted that 37 of these drugs are likely to be active against 3CLpro of the coronavirus and, by extension, could be used as potential inhibitors of 3CLpro in SARS-CoV-2.

Table 4. Examples of Substances Predicted to be Active Using 3CLpro Model

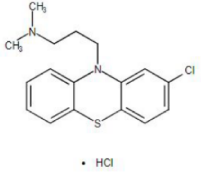
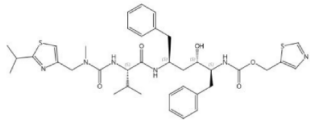
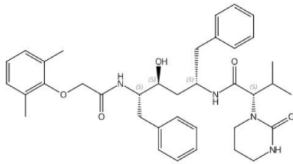
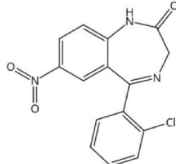
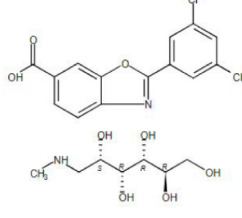
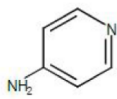
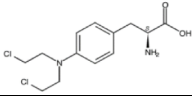
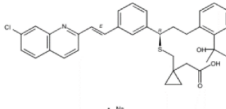
Substance ^{34, 63}	CAS Registry Number	Structure	Mode of action	Current use	In COVID-19 clinical trials ⁶⁵⁻⁶⁷	Data Set ^{34,35}
Chlorpromazine hydrochloride (Thorazine hydrochloride)	69-09-0		dopamine receptor blocker ^{34,64}	treat schizophrenia ^{34,64}	Yes	FDA
Ritonavir (Norvir)	155213-67-5		HIV-1 protease ^{34,64}	treat HIV-1 infection/AIDS ^{34,64}	Yes	FDA
Lopinavir	192725-17-0		HIV-1 protease ^{34,64}	treat HIV-1 infection/AIDS ^{34,64}	Yes	FDA
Clonazepam (Klonopin)	1622-61-3		neurotransmitter GABA enhancer ^{34,64}	treat seizure, panic disorder & movement disorder ^{34,64}	No	FDA
Tafamidis meglumine (Vyndaqel)	951395-08-7		transthyretin stabilizer ^{34,64}	treat familial amyloid polyneuropathy and cardiomyopathy ^{34,64}	No	FDA
Dalfampridine (Ampyra)	504-24-5		voltage-gated potassium channel blocker ^{34,64}	treat multiple sclerosis and Lambert-Eaton myasthenic syndrome ^{34,64}	Yes	FDA
Melphalan (Alkeran)	148-82-3		DNA and RNA synthesis inhibitor by altering guanine through alkylation ^{34,64}	treat multiple myeloma, ovarian cancer, melanoma, and amyloidosis ^{34,64}	Yes	FDA
Montelukast sodium (Singulair)	151767-02-1		leukotriene receptor antagonist; blocking action of leukotriene D ₄ ^{34,64}	treat asthma ^{34,64}	Yes	FDA

Table 4. continued

Substance ^{34,63}	CAS Registry Number	Structure	Mode of action	Current use	In COVID-19 clinical trials ⁶⁵⁻⁶⁷	Data Set ^{34,35}
Metyrapon (Metopirone)	54-36-4		reversible inhibitor of 11 β -hydroxylase ^{34,64}	treat adrenal insufficiency & Cushing's syndrome ^{34,64}	No	FDA
Celecoxib (Celebrex)	169590-42-5		cyclooxygenase Cox-2 inhibitor ^{34,64}	treat pain and inflammation in osteoarthritis ^{34,64}	No	FDA
Fluphenazine dihydrochloride (Prolixin)	146-56-5		post-synaptic D2 receptor blocker; α 1-adrenergic receptor, muscarinic-1 receptor and histamine-1 receptor blocker ^{34,64}	treat psychoses such as schizophrenia ^{34,64}	No	FDA
Amifampridine phosphate (Firdapse)	446254-47-3		potassium channel blocker in membranes ^{34,64}	treat rare muscle diseases such as Lambert-Eaton myasthenic syndrome ^{34,64}	No	FDA
Lansoprazole (Prevacid)	103577-45-3		H ⁺ /K ⁺ -ATPase blocker ^{34,64}	reduce stomach acid, treat peptic ulcer disease, gastroesophageal reflux disease, and Zollinger-Ellison syndrome ^{34,64}	Yes	FDA
Diltiazem hydrochloride (Cardizem)	33286-22-5		calcium channel blocker ^{34,64}	treat high blood pressure, angina, and certain heart arrhythmias ^{34,64}	Yes	FDA
Axitinib (Inlyta)	319460-85-0		tyrosine kinase inhibitor ^{34,64}	treat advanced renal cell carcinoma and other tumor types ^{34,64}	No	FDA
Ethacrynic acid (Edecrin)	58-54-8		sodium-potassium-chloride cotransporter NKCC2 inhibitor ^{34,64}	treat edema associated with congestive heart failure ^{34,64}	No	FDA
Anagrelide hydrochloride (Agrylin)	58579-51-4		cyclic AMP phosphodiesterase inhibitor ^{34,64}	treat essential thrombocytosis, overproduction of blood platelets, chronic myeloid leukemia ^{34,64}	No	FDA

Table 4. continued

Substance ^{34,63}	CAS Registry Number	Structure	Mode of action	Current use	In COVID-19 clinical trials ⁶⁵⁻⁶⁷	Data Set ^{34,35}
Tadalafil (Cialis)	171596-29-5		phosphodiesterase 5 inhibitor ^{34,64}	treat erectile dysfunction, benign prostatic hyperplasia, and pulmonary arterial hypertension ^{34,64}	No	FDA
Pexidartinib	1029044-16-3		CSF1R and c-kit receptor kinase inhibitor ^{34,64}	treat symptomatic tenosynovial giant cell tumor ^{34,64}	No	FDA
Leflunomide (Arava)	75706-12-6		dihydroorotate dehydrogenase inhibitor ^{34,64}	treat rheumatoid and psoriatic arthritis ^{34,64}	Yes	FDA
Benzenepropanamide, 4-fluoro-N-[(1S)-1-formyl-2-[(3S)-2-oxo-3-pyrrolidinyl]ethyl]-α-[[2E)-1-oxo-3-phenyl-2-propen-1-yl]amino]-, (αS)-	2243743-58-8		3CLpro inhibitor ⁵⁵	anti-Enterovirus A71 agent ⁵⁵	No	Antiviral Dataset
2-Pentenoic acid, 4-[[[(2R,5S)-2-[(4-fluorophenyl)methyl]-6-methyl-5-[[[(5-methyl-3-isoxazolyl)carboxyl]amino]-1,4-dioxoheptyl]amino]-5-[(3S)-2-oxo-3-pyrrolidinyl]-phenylmethyl ester, (2E,4S)-	1934276-50-2		3C protease inhibitor ⁵⁶	anti-Enterovirus 71 agent ⁵⁶	No	Antiviral Dataset
Benzenepropanamide, α-[[[(2E)-3-(3-bromophenyl)-1-oxo-2-propen-1-yl]amino]-N-[(1S)-1-formyl-2-[(3S)-2-oxo-3-pyrrolidinyl]ethyl]-, (αS)-	2229818-46-4		3C protease inhibitor ⁵⁷	anti-MERS-CoV and anti-SARS-CoV-1 agent ⁵⁷	No	Antiviral Dataset

The model was then also applied to the CAS COVID-19 Antiviral Candidate Compounds Dataset, which contains 49,437 compounds with potential antiviral activity identified by CAS scientists. The model predicted that 970 of these chemical compounds are likely to be active against 3CLpro of the coronavirus. From each of these applications, a few selected molecules with the highest inhibition probability are shown in Table 4. A full list of chemical structures predicted to be active is shown in Table S3 of Supporting Information. As expected, the model identified several well-known HIV-1 protease inhibitors (ritonavir and lopinavir)⁵⁴ and identified substances (RNs 2243743-58-8, 1934276-50-2, and 2229818-46-4) that target 3C protease/3CLpro and was shown to inhibit Enterovirus, MERS-CoV, and SARS-CoV-1 when tested in bioassays.⁵⁵⁻⁵⁷ These could represent new lead candidates as therapeutic agents for COVID-19 or other viral infections. The model further identified substances against host proteins involved in cellular processes, including diltiazem hydrochloride and leflunomide. Leflunomide is a dihydroorotate dehydrogenase inhibitor and involved in nucleotide synthesis.⁵⁸ In a recent multiomics study, dihydroorotate dehydrogenase was identified as a possible target

for SARS-CoV-2 infection.⁵⁹ In addition, leflunomide has been shown to inhibit SARS-CoV-2 activity.⁶⁰ Diltiazem hydrochloride is a calcium channel blocker that helps to increase blood flow and variably decrease the heart rate *via* strong depression of the A-V node. It has also been reported to be an inhibitor of the NF-κB signaling pathway. As hypertension is one of the common comorbidities seen in COVID-19 infected patients, and as diltiazem has dual effects (lowering high blood pressure and inhibiting NF-κB signaling), it may be an attractive substance to use in place of ACE inhibitors.^{61,62}

Furthermore, the 3CLpro model has been applied to screen-predicted active molecules from a data set that contains over 113 K substances with CAS-assigned pharmacological activity or a therapeutic role indexed in SARS-, MERS-, and COVID-19-related documents published since 2003. Beyond these substances known to be active from the training set, this process identified about 2500 additional substances predicted to be active against 3CLpro. Many of these compounds with predicted activities were previously shown as inhibitors of coronavirus 3CLpro, serving as a validating mechanism of the model that several substances predicted by the model are currently being

Table 5. Sample of Predicted Active Substances with Experimentally Validated Inhibitory Effect on Viral 3CLpro

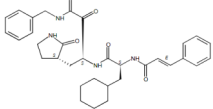
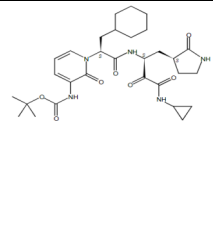
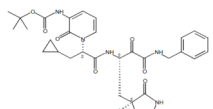
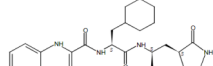
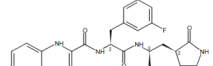
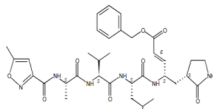
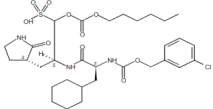
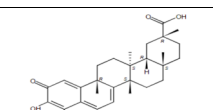
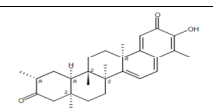
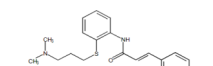
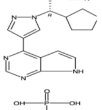
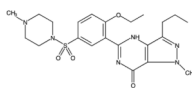
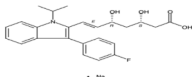
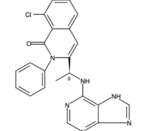
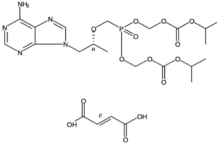
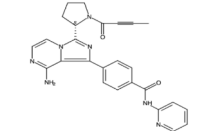
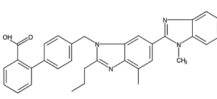
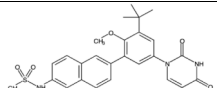
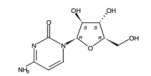
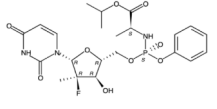
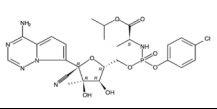
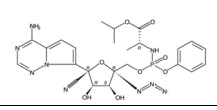
Substance	CAS Registry Number	Structure	Inhibition Values (μM)	Reference
β -[[[(2S)-3-Cyclohexyl-1-oxo-2-[(2E)-1-oxo-3-phenyl-2-propen-1-yl]amino]propyl]amino]- α ,2-dioxo-N-(phenylmethyl)- (βS, 3S)- 3-pyrrolidinebutanamide	2411412-50-3		SARS-CoV-2 IC ₅₀ = 2.39 EC ₅₀ = 2.07 MERS-CoV EC ₅₀ = 0.0004	Science (Washington, DC, United States) (2020), 368(6489), 409-412 Bioorganic & Medicinal Chemistry Letters (2020), 30(17)
Carbamic acid, N-[1-[(1S)-1-(cyclohexylmethyl)-2-[[[(1S)-3-(cyclopropylamino)-2,3-dioxo-1-[[[(3S)-2-oxo-3-pyrrolidinyl]methyl]propyl]amino]-2-oxoethyl]-1,2-dihydro-2-oxo-3-pyridinyl]-1,1-dimethylethyl ester	2412965-58-1		SARS-CoV-2 IC ₅₀ = 2.39	Bioorganic & Medicinal Chemistry Letters (2020), 30(17)
1,1-Dimethylethyl N-[1-[(1S)-1-(cyclopropylmethyl)-2-[[[(1S)-2,3-dioxo-1-[[[(3S)-2-oxo-3-pyrrolidinyl]methyl]-3-[[phenyl(methyl)amino]propyl]amino]-2-oxoethyl]-1,2-dihydro-2-oxo-3-pyridinyl]carbamate	2412965-59-2		SARS-CoV IC ₅₀ = 0.9 EC ₅₀ = 1.7 MERS-CoV IC ₅₀ = 0.58 SARS-CoV-2 IC ₅₀ = 0.67 EC ₅₀ = 1.75	Science (Washington, DC, United States) (2020), 368(6489), 409-412 Bioorganic & Medicinal Chemistry Letters (2020), 30(17)
N-[(1S)-1-(Cyclohexylmethyl)-2-[[[(1S)-1-formyl-2-[[[(3S)-2-oxo-3-pyrrolidinyl]ethyl]amino]-2-oxoethyl]-1H-indole-2-carboxamide	2103278-86-8		SARS-CoV-2 IC ₅₀ = 0.05 EC ₅₀ = 0.53	Bioorganic & Medicinal Chemistry Letters (2020), 30(17)
1H-Indole-2-carboxamide, N-[(1S)-1-[(3-fluorophenyl)methyl]-2-[[[(1S)-1-formyl-2-[[[(3S)-2-oxo-3-pyrrolidinyl]ethyl]amino]-2-oxoethyl]-	2413716-71-7		SARS-CoV-2 IC ₅₀ = 0.04 EC ₅₀ = 0.72	Bioorganic & Medicinal Chemistry Letters (2020), 30(17)
N-[(5-Methyl-3-isoxazolyl)carbonyl]-L-alanyl-L-valyl-N-[[[(1S,2E)-4-oxo-1-[[[(3S)-2-oxo-3-pyrrolidinyl]methyl]-4-(phenylmethoxy)-2-buten-1-yl]-L-leucinamide	884650-98-0		SARS-CoV-2 EC ₅₀ = 16.77 MERS-CoV EC ₅₀ = 0.28	Bioorganic & Medicinal Chemistry Letters (2020), 30(17)
8,10-Dioxo-2,5-diazahexadecanoic acid, 3-(cyclohexylmethyl)-4,9-dioxo-6-[[[(3S)-2-oxo-3-pyrrolidinyl]methyl]-7-sulfo-, 1-[(3-chlorophenyl)methyl] ester, sodium salt (1:1), (3S,6S)-	2122228-55-9		SARS-CoV IC ₅₀ = 0.8 EC ₅₀ = 0.025	WO2018023054
Celastrol	34157-83-0		SARS-CoV IC ₅₀ = 2.6	KR2011068191; Journal of Medicinal Chemistry (2016), 59(14), 6595-6628
Tingenone	50802-21-6		SARS-CoV IC ₅₀ = 5.5	KR2011068191; Journal of Medicinal Chemistry (2016), 59(14), 6595-6628
Cinanserin	1166-34-3		SARS-CoV IC ₅₀ = 4.92	CN 1472336; Journal of Virology (2005), 79(11), 7095-7103

Table 6. Examples of Substances Predicted to be Active Using the RdRp Model^{64,77–88a}

Substance	CAS Registry Number	Structure	Mode of action	Current Use	In COVID-19 clinical trials ^{65–67}	Data Set ^{34,35}
Ruxolitinib phosphate (Jakafi)	1092939-17-7		protein tyrosine kinases JAK 1 and 2 inhibitor ⁷⁷	treat intermediate or high-risk myelofibrosis and polycythemia vera; binding of ruxolitinib to kinases may lead to reduction of inflammation and inhibition of cellular proliferation ⁷⁷	Yes	FDA
Sildenafil (Revatio)	139755-83-2		phosphodiesterase type 5 inhibitor ⁸⁰	treat male erectile dysfunction; improve ability to exercise in adults with pulmonary arterial hypertension by relaxing blood vessels in lungs ⁸⁰	Yes	FDA
Fluvastatin sodium (Lescol XL)	93957-55-2		HMG-CoA reductase inhibitor ⁸¹	reduce risk of heart attack and stroke by reducing LDL cholesterol and triglycerides and increasing HDL cholesterol in blood ⁸¹	No	FDA
Duvelisib (Copiktra)	1201438-56-3		phosphoinositide 3-kinase (PI3K) inhibitor, specifically α and γ isoforms of PI3K ⁸²	treat chronic lymphocytic leukemia or small lymphocytic lymphoma ⁸²	Yes	FDA
Tenofovir disoproxil fumarate (Viread)	202138-50-9		reverse transcriptase inhibitor; analog of adenosine ⁸³	slow progression of disease caused by hepatitis B virus or human immunodeficiency virus ⁸³	Yes	FDA
Acalabrutinib (Calquence)	1420477-60-6		Bcr/tor's agammaglobulinemia tyrosine kinase inhibitor ⁸⁴	treat mantle cell lymphoma ⁸⁴	Yes	FDA
Telmisartan (Micardis)	144701-48-4		angiotensin II receptor blocker with high affinity for angiotensin II receptor AT1; peroxisome proliferator-activated receptor gamma modulator ⁸⁵	treat high blood pressure, heart failure, and diabetic kidney disease ⁸⁵	Yes	FDA
Dasabuvir (ABT333)	1132935-63-7		RNA-dependent RNA polymerase NS5B inhibitor ⁸⁶	treat hepatitis C; often used together with combination medication dasabuvir/ombitasvir/paritaprevir/ritonavir (Viekira pak) ⁸⁶	No	FDA
Cytarabine (Cytosine arabinoside)	147-94-4		DNA and RNA polymerase and nucleotide reductase enzyme inhibitor; interferes with DNA synthesis ⁸⁷	treat acute myeloid leukemia, acute lymphocytic leukemia, chronic myelogenous leukemia, and non-Hodgkin's lymphoma ⁸⁷	No	FDA
Sofosbuvir (Sovaldi)	1190307-88-0		hepatitis C virus RNA-dependent RNA polymerase NS5B inhibitor ⁸⁸	treat hepatitis C in combination with other medications including ribavirin, peginterferon- α , simeprevir, ledipasvir, daclatasvir, or velpatasvir ⁸⁸	Yes	FDA
L-Alanine, N-(4-chlorophenoxy)hydroxyphosphinyl-, 1-methylethyl ester, 6-ester with 2-C-(4-aminopyrrolo[2,1-fj][1,2,4]triazin-7-yl)-2,5-anhydro-3-C-methyl-D-altreronitriole	1191237-79-2		polymerase inhibitor ⁸⁹	nucleotide analog used as antiviral agent	No	Antiviral Dataset
D-Alanine, N-(hydroxyphenoxyphosphinyl)-, 1-methylethyl ester, 6-ester with 2-C-(4-aminopyrrolo[2,1-fj][1,2,4]triazin-7-yl)-2,5-anhydro-5-C-azido-D-altreronitriole	2294043-57-3		polymerase inhibitor ⁸⁹	nucleotide analog used as antiviral agent	No	Antiviral Dataset

^aSee Table S6 for additional substances predicted to be active.

pursued for repurposing against COVID-19. Table 5 shows several predicted substances with experimentally validated 3CLpro inhibition activity (IC_{50}) and inhibition of RNA replication (EC_{50}) for SARS-CoV-1, MERS-CoV, and/or SARS-CoV-2.

3.2. RdRp Target—Model and Results. In order to identify potential drug candidates of RdRp, we built machine learning models that can be used to predict the ability of a small molecule to inhibit the polymerase function of RdRp. The abovementioned training set molecules of RdRp were used as

training materials for this model. Each molecule was represented as a 42-component vector using MQNs that were used previously for organic molecule classification tasks.³⁸

A “Radial Kernel Support Vector Machine” model (SVM) was determined to be the most accurate and efficient model with a ROC-AUC value of 0.99 (Figure 2). Cumulative gain and lift analysis for this model is shown in Figures S2 and S3 and Tables S8 and S9 in Supporting Information. We applied this model for screening of RdRp inhibitors in the same three data sets screened above for 3CLpro: a data set of FDA-approved drugs, the CAS COVID-19 Antiviral data set, and a larger list of substances indexed by CAS from SARS, MERS, and COVID-19 published literature.

With a probability threshold of 0.50, we found more than 21,000 active candidates with diverse structural features when we applied the SVM model to the three data sets described above. Previously, CAS framework identifiers were used to evaluate the diversity of the CAS REGISTRY.⁶⁸ We found more than 2000 unique framework IDs within the predicted active candidates, indicating a high structural diversity of the screened active compound data sets. The relative importance of several individual features can be seen in more detail in Figure S1 of Supporting Information.

An important class of the screened active compounds for RdRp that we found were nucleotide analogues with structural similarity to remdesivir. As it was recently demonstrated, cyano-substituted adenosines⁶⁹ not only effectively compete with natural RdRp substrates but also inhibit RNA synthesis via steric effect. Several interesting nucleotide analogues including cyano-substituted adenosines that were found are listed in Table 6.

We used CAS REGISTRY information such as framework and ring identifiers to highlight the structural similarity between candidates. In addition to cyano-substituted compounds, we found several azido-substituted nucleotide analogues. Azido-substituted nucleotide analogues are known for their ability to generate reactive intermediates that are often used to modify enzyme active sites.⁷⁰

Beyond providing the molecular structures of these identified substances, Table 6 also provides information on their mode of action, current use, and inclusion in ongoing or upcoming clinical trials. As expected, the model identified several RdRp or polymerase inhibitors (dasabuvir, cytarabine, and sofosbuvir) and also identified several substances that inhibit enzymes or receptors involved in the host immune response (ruxolitinib phosphate, duvelisib, acalabrutinib, and telmisartan) and in cholesterol synthesis (fluvastatin sodium). In addition to the suggestion by the model that these substances could be useful in modulating RdRp, an overview of their current uses indicates they might have an ancillary benefit in treating COVID-19 patients. Of the substances inhibiting the host immune response, ruxolitinib phosphate and telmisartan are being studied in numerous current or upcoming COVID-19 clinical trials, including NCT04362137, NCT04348071, and NCT04360551.^{71–73}

Ruxolitinib phosphate is an inhibitor of protein kinase JAK1 and JAK2 shown to be part of the severe immune overreaction called a cytokine storm that can lead to life-threatening respiratory complications in some patients with COVID-19.⁷⁴ Telmisartan, an antagonist of angiotensin II receptor type 1 (AT1) and a modulator of peroxisome proliferator-activated receptor gamma, could be used to reduce the negative inflammatory response seen in COVID-19 patients.⁶³ Duvelisib, a phosphoinositide 3-kinase (PI3K) inhibitor, is being studied in

NCT04372602 to determine if it can reduce the aberrant hyperactivation of the innate immune system, preferentially polarize macrophages, reduce pulmonary inflammation, and limit viral persistence.⁷⁵ Fluvastatin sodium, a HMG-CoA reductase inhibitor and cholesterol-lowering drug, is being tested in clinical trial NCT00664742 that includes patients with metabolic syndrome.⁷⁶ As cardiovascular disease is a risk factor for COVID-19, this clinical study may provide some assistance for COVID-19 patients.

4. SUMMARY AND PERSPECTIVES

The COVID-19 pandemic caused by the new coronavirus SARS-CoV-2 represents a serious threat to the global health system. In order to aid development of efficient therapeutics against the disease, we use machine learning-based predictive modeling to identify novel drug candidates for the viral targets 3CLpro and RdRp. Chemist-curated training sets of substances were assembled from CAS data collections and integrated with curated bioassay data. Using suitable binary classifiers, we were able to screen a set of FDA-approved drugs, nearly 50,000 substances in the CAS COVID-19 Antiviral Candidate Compounds Dataset, and a list of 113,000 substances with CAS-assigned pharmacological activity or a therapeutic role indexed in SARS-, MERS-, and COVID-19-related documents. Through these screenings, we identified many potential inhibitors with potential activity against 3CLpro or RdRp. Many of these predicted active substances also have ancillary activities such as cardiovascular effects [diltiazem hydrochloride (cardizem)], cholesterol-reduction [fluvastatin sodium (lescol XL)], dihydroorotate dehydrogenase inhibition [leflunomide (arava)] and signal transduction inhibition [ruxolitinib phosphate (jakafi)], which may have other beneficial effects in treating COVID-19. For example, as noted above, heart disease is a known risk factor for COVID-19, so a candidate substance for treating COVID-19, which has known effectiveness treating heart disease, could potentially offer a dual benefit in certain cases. Dihydroorotate dehydrogenase inhibitors have been shown to have anti-SARS-CoV-2 activity, and Janus kinase inhibitors are being studied for inhibiting the cytokine storm process. Additionally, as mentioned above in this report, many of the predicted active substances have FDA approval for specific treatments, and several are included in current or future COVID-19 clinical trials, which could assist or accelerate finding therapeutic agents to use in COVID-19 treatment. In summary, it is hoped that the information provided in this study will be of value in the ongoing search for anticoronaviral therapeutic agents.

While this paper focused on identifying potential therapeutic compounds for use in the current COVID-19 crisis, it is not certain but possible that there will be additional pandemics in the years to come. Many of these will be of viral origin. Because viral transmission and spread can occur very fast, while the typical drug discovery process can take a decade or longer of costly development, it is urgent that preparation for future outbreaks begins now and that the current focus on antiviral agent research continues into the future. The ongoing development of computer-based drug discovery methods such as the machine learning procedures described here and elsewhere, molecular docking and virtual screening of potential therapeutics, and other artificial intelligence methods will be of central importance. Facilitating this is the ongoing increase in computer processing power, continued development of docking and structure prediction algorithms, and protein crystal

structure determination. Additionally, the use of high-throughput screening, omics technologies, and the repurposing of already developed drugs will continue and increase in importance as these can shave years off the development time for antiviral drugs. These new research methods will not replace human-based laboratory research but will instead complement it. Finally, because different types of viruses can cause epidemics (e.g., coronavirus, influenza viruses, Ebola viruses, and retroviruses), the development of broad-spectrum antiviral agents or vaccines would be of great value. This would also be useful in that despite the use of faster drug discovery methods, safety and effectiveness testing in humans still takes time. Organ-on-a-chip methods will help but will not eliminate this time-consuming step. Therefore, research on broad-spectrum antivirals will be essential. We hope the current work combining human data curation and machine learning-based predictive models to identify potential small molecule drug candidates for COVID-19 will contribute in some small way to the ongoing efforts in antiviral research.

■ ASSOCIATED CONTENT

SI Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/acsomega.0c03682>.

3CLpro data set, AUCs of evaluated 3CLpro models, chemical structures predicted to be active by the 3CLpro model, RdRp data set, AUCs of evaluated RdRp models, examples of substances predicted active by the RdRp SVM model, RdRp validation tables and figures, RdRp cumulative gain analysis, RdRp cumulative lift analysis, MQN feature importance analysis for the RdRp model, RdRp cumulative gain chart, and RdRp cumulative lift chart (XLSX)

■ AUTHOR INFORMATION

Corresponding Author

Julian Ivanov – CAS, A Division of the American Chemical Society, Columbus, Ohio 43210-3012, United States;
✉ jivanov@cas.org; orcid.org/0000-0002-3519-359X

Authors

Dmitrii Polshakov – CAS, A Division of the American Chemical Society, Columbus, Ohio 43210-3012, United States

Junko Kato-Weinstein – CAS, A Division of the American Chemical Society, Columbus, Ohio 43210-3012, United States

Qiongqiong Zhou – CAS, A Division of the American Chemical Society, Columbus, Ohio 43210-3012, United States;
✉ qzou@cas.org; orcid.org/0000-0001-6711-369X

Yingzhu Li – CAS, A Division of the American Chemical Society, Columbus, Ohio 43210-3012, United States

Roger Granet – CAS, A Division of the American Chemical Society, Columbus, Ohio 43210-3012, United States

Linda Garner – CAS, A Division of the American Chemical Society, Columbus, Ohio 43210-3012, United States

Yi Deng – CAS, A Division of the American Chemical Society, Columbus, Ohio 43210-3012, United States

Cynthia Liu – CAS, A Division of the American Chemical Society, Columbus, Ohio 43210-3012, United States; ✉ cliu@cas.org; orcid.org/0000-0003-3858-1501

Dana Albaiu – CAS, A Division of the American Chemical Society, Columbus, Ohio 43210-3012, United States

Jeffrey Wilson – CAS, A Division of the American Chemical Society, Columbus, Ohio 43210-3012, United States

Christopher Aultman – CAS, A Division of the American Chemical Society, Columbus, Ohio 43210-3012, United States

Complete contact information is available at:
<https://pubs.acs.org/doi/10.1021/acsomega.0c03682>

Author Contributions

†J.I. and D.P. contributed equally to this paper.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

We sincerely appreciate Don Swartwout, Haitao Chen, and Chris Gessner for their technical assistance during preparation of this paper. We are also very grateful to Manuel Guzman, Gilles George, and Mark Grabau for their support. We thank Steven P. Watkins and Cristina Tomeo for their assistance in the preparation.

■ REFERENCES

- (1) Gordon, D. E.; Jang, G. M.; Bouhaddou, M.; Xu, J.; Obernier, K.; White, K. M.; O'Meara, M. J.; Rezelj, V. V.; Guo, J. Z.; Swaney, D. L.; et al. A SARS-CoV-2 protein interaction map reveals targets for drug repurposing. *Nature* **2020**, *583*, 459.
- (2) Dai, W.; Zhang, B.; Jiang, X.-M.; Su, H.; Li, J.; Zhao, Y.; Xie, X.; Jin, Z.; Peng, J.; Liu, F.; et al. Structure-Based Design, Synthesis and Biological Evaluation of Peptidomimetic Aldehydes as a Novel Series of Antiviral Drug Candidates Targeting the SARS-CoV-2 Main Protease. *Science* **2020**, *368*, 1331–1335.
- (3) Zhang, L.; Lin, D.; Sun, X.; Rox, K.; Hilgenfeld, R. X-ray structure of main protease of the novel coronavirus SARS-CoV-2 enables design of α -ketoamide inhibitors, bioRxiv, doi: <https://doi.org/10.1101/2020.02.17.952879>.
- (4) Anand, K.; Ziebuhr, J.; Wadhwani, P.; Mesters, J. R.; Hilgenfeld, R. Coronavirus Main Proteinase (3CLpro) Structure: Basis for Design of Anti-SARS Drugs. *Science* **2003**, *300*, 1763–1767.
- (5) Tahir ul Qamar, M.; Alqahtani, S. M.; Alamri, M. A.; Chen, L. L. Structural basis of SARS-CoV-2 3CLpro and anti-COVID-19 drug discovery from medicinal plants. *J. Pharm. Anal.* **2020**, *10*, 313.
- (6) te Velhuis, A. J. W. Common and unique features of viral RNA-dependent Polymerases. *Cell. Mol. Life Sci.* **2014**, *71*, 4403–4420.
- (7) Gao, Y.; Yan, L.; Huang, Y.; Liu, F.; Zhao, Y.; Cao, L.; Wang, T.; Sun, Q.; Ming, Z.; Zhang, L.; et al. Structure of the RNA-dependent RNA polymerase from COVID-19 virus. *Science* **2020**, *368*, 779–782.
- (8) Elfiky, A. A. Ribavirin, Remdesivir, Sofosbuvir, Galidesivir, and Tenofovir against SARS-CoV-2 RNA dependent RNA polymerase (RdRp): A molecular docking study. *Life Sci.* **2020**, *253*, 117592.
- (9) (a) Vamathevan, J.; Clark, D.; Czodrowski, P.; Dunham, I.; Ferran, E.; Lee, G.; Li, B.; Madabhushi, A.; Shah, P.; Spitzer, M.; Zhao, S. Applications of machine learning in drug discovery and development. *Nat. Rev. Drug Discov.* **2019**, *18*, 463–477. (b) Ghosh, K.; Amin, S. A.; Gayen, S.; Jha, T. Chemical-informatics approach to COVID-19 drug discovery: Exploration of important fragments and data mining based prediction of some hits from natural origins as main protease (Mpro) inhibitors. *J. Mol. Struct.* **2021**, *1224*, 129026.
- (10) Islam, R.; Parves, M. R.; Paul, A. S.; Uddin, N.; Rahman, M. S.; Mamun, A. A.; Hossain, M. N.; Ali, M. A.; Halim, M. A. A molecular modeling approach to identify effective antiviral phytochemicals against the main protease of SARS-CoV-2. *J. Biomol. Struct. Dyn.* **2020**, *1–12*.
- (11) Kumar, V.; Roy, K. Development of a simple, interpretable and easily transferable QSAR model for quick screening antiviral databases in search of novel 3C-like protease (3CLpro) enzyme inhibitors against SARS-CoV diseases. *SAR QSAR Environ. Res.* **2020**, *31*, 511–526.
- (12) (a) Rahman, M. M.; Saha, T.; Islam, K. J.; Suman, R. H.; Biswas, S.; Rahat, E. U.; Hossen, M. R.; Islam, R.; Hossain, M. N.; Mamun, A.

- A.; et al. Virtual screening, molecular dynamics and structure–activity relationship studies to identify potent approved drugs for Covid-19 treatment. *J. Biomol. Struct. Dyn.* **2020**, 1–11. (b) Amin, S. A.; Ghosh, K.; Gayen, S.; Jha, T. Chemical-informatics approach to COVID-19 drug discovery: Monte Carlo based QSAR, virtual screening and molecular docking study of some in-house molecules as papain-like protease (PLpro) inhibitors. *J. Biomol. Struct. Dyn.* **2020**, 1.
- (13) Hansch, C.; Maloney, P. P.; Fujita, T.; Muir, R. M. Correlation of Biological Activity of Phenoxyacetic Acids with Hammett Substituent Constants and Partition Coefficients. *Nature* **1962**, *194*, 178–180.
- (14) Hansch, C.; Muir, R. M.; Fujita, T.; Maloney, P. P.; Geiger, F.; Streich, M. The Correlation of Biological Activity of Plant Growth-Regulators and Chloromycetin Derivatives with Hammett Constants and Partition Coefficients. *J. Am. Chem. Soc.* **1963**, *85*, 2817–2824.
- (15) Hansch, C.; Fujita, T. p - σ - π analysis—A method for the correlation of biological activity and chemical structure. *J. Am. Chem. Soc.* **1964**, *86*, 1616–1626.
- (16) Gutman, I.; Trinajstić, N. Graph theory and molecular orbitals. Total φ -electron energy of alternant hydrocarbons. *Chem. Phys. Lett.* **1972**, *17*, 535–538.
- (17) Hosoya, H. Topological Index. A Newly Proposed Quantity Characterizing the Topological Nature of Structural Isomers of Saturated Hydrocarbons. *Bull. Chem. Soc. Jpn.* **1971**, *44*, 2332–2339.
- (18) Randić, M. Characterization of molecular branching. *J. Am. Chem. Soc.* **1975**, *97*, 6609–6615.
- (19) Balaban, A. T. Highly discriminating distance-based topological index. *Chem. Phys. Lett.* **1982**, *89*, 399–404.
- (20) Balaban, A. T. Topological Indices Based on Topological Distances in Molecular Graph. *Pure Appl. Chem.* **1988**, *55*, 199–206.
- (21) Gasteiger, J.; Sadowski, J.; Schuur, J.; Selzer, P.; Steinhauer, L.; Steinhauer, V. Chemical Information in 3D Space. *J. Chem. Inf. Comput. Sci.* **1996**, *36*, 1030–1037.
- (22) Yan, A.; Gasteiger, J. Prediction of Aqueous Solubility of Organic Compounds Based on a 3D Structure Representation. *J. Chem. Inf. Comput. Sci.* **2003**, *43*, 429–434.
- (23) Mekenyan, O.; Peitchev, D.; Bonchev, D.; Trinajstić, N.; Bangov, I. P. Modeling the interaction of small organic-molecules with biomacromolecules. *Drug Res.* **1986**, *36*, 176–183.
- (24) Ivanov, J.; Karabunarliev, S.; Mekenyan, O. 3DGEN: A system for exhaustive 3D molecular design proceeding from molecular topology. *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 234–243.
- (25) Debnath, A. K.; Lopez de Compadre, R. L.; Debnath, G.; Shusterman, A. J.; Hansch, C. Corwin Hansch, Structure-activity relationship of mutagenic aromatic and heteroaromatic nitro compounds. Correlation with molecular orbital energies and hydrophobicity. *J. Med. Chem.* **1991**, *34*, 786–797.
- (26) Clark, T. Quantum Cheminformatics: An Oxymoron? (Part 1). In *Chemical Data Analysis in the Large: The Challenge of the Automation Age*; Hicks, M. G., Ed.; Proceedings of the Beilstein-Institut Workshop May 22nd - 26th; Logos Verlag Berlin: Bozen, Italy, 2000.
- (27) Clark, T. Quantum Cheminformatics: An Oxymoron? (Part 2). In *Rational Approaches to Drug Design*; Höltje, H.-D., Sippl, W., Eds.; Prous Science: Barcelona, 2001.
- (28) Hopfinger, A. J. A QSAR investigation of dihydrofolate reductase inhibition by Baker triazines based upon molecular shape analysis. *J. Am. Chem. Soc.* **1980**, *102*, 7196–7206.
- (29) Klopman, G. Artificial Intelligence Approach to Structure–Activity Studies. Computer Automated Structure Evaluation of Biological Activity of Organic Molecules. *J. Am. Chem. Soc.* **1984**, *106*, 7315–7321.
- (30) Klopman, G.; Wang, S. A Computer Automated Structure Evaluation (CASE) Approach to Calculation of Partition Coefficient. *J. Comput. Chem.* **1991**, *12*, 1025–1032.
- (31) Chakravarti, S.; Klopman, G.; Ivanov, J.; Saiakhov, R. *MC4PC—An Artificial Intelligence Approach to the Discovery of Quantitative Structure–Toxic Activity Relationships*; Taylor & Francis Group, 2005.
- (32) Cramer, R. D., III; Patterson, D. E.; Bunce, J. D. Comparative Molecular Field Analysis (CoMFA). 1. Effect of Shape on Binding of Steroids to Carrier Proteins. *J. Am. Chem. Soc.* **1988**, *110*, 5959–5967.
- (33) Cramer, R. D., III; Bunce, J. D.; Patterson, D. E.; Frank, I. E. Crossvalidation, Bootstrapping, and Partial Least Squares Compared with Multiple Regression in Conventional QSAR Studies. *Quant. Struct.-Act. Relat.* **1988**, *7*, 18.
- (34) Drugs@FDA: FDA-Approved Drugs. <https://www.accessdata.fda.gov/scripts/cder/daf/> (accessed May 26, 2020).
- (35) (a) <https://www.cas.org/covid-19-antiviral-compounds-dataset> (accessed May 2020). (b) Deng, Y.-H.; Wang, N.-N.; Zou, Z.-X.; Zhang, L.; Xu, K.-P.; Chen, A. F.; Cao, D.-S.; Tan, G.-S. Multi-Target Screening and Experimental Validation of Natural Products from Selaginella Plants against Alzheimer's Disease. *Front. Pharmacol.* **2017**, *8*, 539. (c) Zhou, Y.; Hou, Y.; Shen, J.; Huang, Y.; Martin, W.; Cheng, F. Network-based drug repurposing for novel coronavirus 2019-nCoV/SARS-CoV-2. *Cell Discovery* **2020**, *6*, 14.
- (36) Fawcett, T. An introduction to ROC analysis. *Pattern Recogn. Lett.* **2006**, *27*, 861–874.
- (37) RDKit, Open-Source Cheminformatics. <http://www.rdkit.org> (accessed June 2020).
- (38) Nguyen, K. T.; Blum, L. C.; van Deursen, R.; Reymond, J.-L. Classification of Organic Molecules by Molecular Quantum Numbers. *ChemMedChem* **2009**, *4*, 1803–1805.
- (39) Ogura, K.; Sato, T.; Yuki, H.; Honma, T. Support Vector Machine model for hERG inhibitory activities based on the integrated hERG database using descriptor selection by NSGA-II. *Sci. Rep.* **2019**, *9*, 12220.
- (40) Lee, A. A.; Yang, Q.; Bassyouni, A.; Butler, C. R.; Houh, X.; Jenkinson, S.; Price, D. A. Ligand biological activity predicted by cleaning positive and negative chemical correlations. *Proc. Natl. Acad. Sci.* **2019**, *116*, 3373.
- (41) <https://scikit-learn.org/stable/modules/svm.html#svm-kernels> (accessed May 2020).
- (42) Chen, T.; Guestrin, C. XGBoost: A Scalable Tree Boosting System. In *22nd SIGKDD Conference on Knowledge Discovery and Data Mining*, 2016.
- (43) Chen, T.; Tong, H. Higgs Boson Discovery with Boosted Trees. *JMLR: Workshop and Conference Proceedings*, 2015; Vol. 42, pp 69–80.
- (44) Mustapha, I. B.; Saeed, F. Bioactive Molecule Prediction Using Extreme Gradient Boosting. *Molecules* **2016**, *21*, 983.
- (45) Blondel, M.; Seki, K.; Uehara, K. Block coordinate descent algorithms for large-scale sparse multiclass classification. *Mach. Learn.* **2013**, *93*, 31–52.
- (46) Bell, R.; Koren, Y. Lessons from the Netflix prize challenge. *ACM SIGKDD Explor. Newsl.* **2007**, *9*, 75–79.
- (47) Zhang, Q.; Hughes-Oliver, J. M.; Ng, R. T. “A Model-Based Ensembling Approach for Developing QSARs”. *J. Chem. Inf. Model.* **2009**, *49*, 1857–1865.
- (48) Rogers, D.; Hahn, M. Extended-Connectivity Fingerprints. *J. Chem. Inf. Model.* **2010**, *50*, 742–754.
- (49) RDKit SMARTS-based implementation of the 166 public MACCS keys.
- (50) Carhart, R. E.; Smith, D. H.; Venkataraghavan, R. Atom Pairs as Molecular Features in Structure-Activity Studies: Definition and Applications. *J. Chem. Inf. Comput. Sci.* **1985**, *25*, 64–73.
- (51) Nilakantan, R.; Bauman, N.; Dixon, J. S.; Venkataraghavan, R. Topological Torsion: A New Molecular Descriptor for SAR Applications. Comparison with Other Descriptors. *J. Chem. Inf. Comput. Sci.* **1987**, *27*, 82–85.
- (52) Wildman, S. A.; Crippen, G. M. Prediction of Physicochemical Parameters by Atomic Contributions. *J. Chem. Inf. Comput. Sci.* **1999**, *39*, 868–873.
- (53) MOE-type descriptors using partial charges and surface area contributions. <http://www.chemcomp.com/journal/vsadesc.htm> (accessed May 2020).
- (54) Cvetkovic, R. S.; Goa, K. L. Lopinavir/Ritonavir. *Drugs* **2003**, *63*, 769–802.
- (55) Nie, Q.; Xu, X.; Zhang, Q.; Ma, Y.; Yin, Z.; Shang, L. 3D-quantitative structure-activity relationship study for the design of novel enterovirus A71 3C protease inhibitors. *Chem. Biol. Drug Des.* **2018**, *92*, 1750–1762.

- (56) Li, P.; Yang, B.; Hao, F.; Wang, P.; He, H.; Huang, L.; Zhang, X.; Zhang, S.; Peng, X.; Yin, K.; et al. Design, synthesis, and biological evaluation of anti-EV71 agents. *Bioorg. Med. Chem. Lett.* **2016**, *26*, 3346–3350.
- (57) Kumar, V.; Shin, J. S.; Shie, J.-J.; Ku, K. B.; Kim, C.; Go, Y. Y.; Huang, K.-F.; Kim, M.; Liang, P.-H. Identification and evaluation of potent Middle East respiratory syndrome coronavirus (MERS-CoV) 3CLPro inhibitors. *Antiviral Res.* **2017**, *141*, 101–106.
- (58) Teschner, S.; Burst, V. Leflunomide: a drug with a potential beyond rheumatology. *Immunotherapy* **2010**, *2*, 637–650.
- (59) Zheng, J.; Zhang, Y.; Liu, Y.; Baird, D.; Liu, X.; Wang, L.; Zhang, H.; Davey Smith, G.; Gaunt, T. Multi-omics study revealing tissue-dependent putative mechanisms of SARS-CoV-2 drug targets on viral infections and complex diseases, medRxiv, **2020**, DOI: 10.1101/2020.05.07.20093286.
- (60) Xiong, R.; Zhang, L.; Li, S.; Sun, Y.; Ding, M.; Wang, Y.; Zhao, Y.; Wu, Y.; Shang, W.; Jiang, X.; et al. Novel and potent inhibitors targeting DHODH, a rate-limiting enzyme in de novo pyrimidine biosynthesis, are broad-spectrum antiviral against RNA viruses including newly emerged coronavirus SARS-CoV-2, bioRxiv, **2020**, DOI: 10.1101/2020.03.11.983056.
- (61) O'Connor, S. E.; Grosset, A.; Janiak, P. The pharmacological basis and pathophysiological significance of the heart rate-lowering property of diltiazem. *Fundam. Clin. Pharmacol.* **1999**, *13*, 145–153.
- (62) Gilmore, T. D.; Herscovitch, M. Inhibitors of NF- κ B signaling: 785 and counting. *Oncogene* **2006**, *25*, 6887–6899.
- (63) Rothlin, R. P.; Vetulli, H. M.; Duarte, M.; Pelorosso, F. G. Telmisartan as tentative angiotensin receptor blocker therapeutic for COVID-19. *Drug Dev. Res.* **2020**, DOI: 10.1002/ddr.21679.
- (64) NIH, U.S. National Library of Medicine, druginfo.nlm.nih.gov/drugportal/ (accessed on May 26, 2020).
- (65) ClinicalTrials.gov, clinicaltrials.gov/ct2/home (accessed on July 29, 2020).
- (66) The ISRCTN Registry, www.isrctn.com (accessed on July 29, 2020).
- (67) EU Clinical Trials Register, www.clinicaltrialsregister.eu (accessed on July 29, 2020).
- (68) Lipkus, A. H.; Yuan, Q.; Lucas, K. A.; Funk, S. A.; Bartelt, W. F., III; Schenck, R. J.; Trippie, A. J. Structural Diversity of Organic Chemistry. A Scaffold Analysis of the CAS Registry. *J. Org. Chem.* **2008**, *73*, 4443–4451.
- (69) Gordon, C. J.; Tchesnokov, E. P.; Woolner, E.; Perry, J. K.; Feng, J. Y.; Porter, D. P.; Götte, M. Remdesivir is a direct-acting antiviral that inhibits RNA-dependent RNA polymerase from severe acute respiratory syndrome coronavirus 2 with high potency. *J. Biol. Chem.* **2020**, *295*, 6785–6797.
- (70) Polshakov, D.; Rai, S.; Wilson, R. M.; Mack, E. T.; Vogel, M.; Krause, J. A.; Burdzinski, G.; Platz, M. S. Photoaffinity Labeling with 8-Azidoadenosine and Its Derivatives: Chemistry of Closed and Opened Adenosine Diazaquinodimethanes. *Biochemistry* **2005**, *44*, 11241–11253.
- (71) Phase 3 Randomized, Double-blind, Placebo-controlled Multi-center Study to Assess the Efficacy and Safety of Ruxolitinib in Patients With COVID-19 Associated Cytokine Storm (RUXCOVID), ClinicalTrials.gov Identifier NCT04362137 clinicaltrials.gov/ct2/show/NCT04362137 (accessed on May 30, 2020).
- (72) Safety and Efficacy of Ruxolitinib for COVID-19, ClinicalTrials.gov Identifier NCT04348071 clinicaltrials.gov/ct2/show/NCT04348071 (accessed on May 30, 2020).
- (73) Pilot Clinical Trial of the Safety and Efficacy of Telmisartan for the Mitigation of Pulmonary and Cardiac Complications in COVID-19 patients, ClinicalTrials.gov Identifier NCT04360551 clinicaltrials.gov/ct2/show/NCT04360551 (accessed on May 30, 2020).
- (74) Novartis announces plan to initiate clinical study of Jakavi in severe COVID-19 patients and establish international compassionate use program, Novartis, News release. April 2, 2020, novartis.com/news/media-releases/novartis-announces-plan-initiate-clinical-study-jakavi-severe-covid-19-patients-and-establish-international-compassionate-use-program.
- (75) Duvelisib to Combat COVID-19, ClinicalTrials.gov Identifier NCT04372602 clinicaltrials.gov/ct2/show/NCT04372602 (accessed on May 30, 2020).
- (76) The Effect of Fluvastatin XL Treatment in Patients with Metabolic Syndrome, ClinicalTrials.gov Identifier NCT00664742 clinicaltrials.gov/ct2/show/NCT00664742 (accessed on May 30, 2020).
- (77) U.S. Food & Drug Administration, JAKAFI (ruxolitinib)-FDA, www.accessdata.fda.gov/drugsatfda_docs/label/2017/202192s0151lbl.pdf (accessed on May 22, 2020).
- (78) Butler, T.; Cho, A.; Kim, C. U.; Saunders, O. L.; Zhang, L. 1'-substituted carba-nucleoside analogs as antiviral agents. WO2009132135A1; World Intellectual Property Organization, 2009.
- (79) Beigelman, L.; Deval, J.; Prhvac, M. Substituted nucleosides, nucleotides and analogs thereof. WO2019053696A1; World Intellectual Property Organization, 2019.
- (80) Rust, K. Y.; Wilkens, H.; Kaiser, R.; Bregel, D.; Wilske, J.; Kraemer, T. Detection and Validated Quantification of the Phosphodiesterase Type 5 Inhibitors Sildenafil, Vardenafil, Tadalafil, and 2 of Their Metabolites in Human Blood Plasma by LC-MS/MS-Application to Forensic and Therapeutic Drug Monitoring Cases. *Ther. Drug Monit.* **2012**, *34*, 729–735.
- (81) Adams, S. P.; Sekhon, S. S.; Tsang, M.; Wright, J. M. Fluvastatin for lowering lipids. *Cochrane Database Syst. Rev.* **2018**, *3*, CD012282.
- (82) Flinn, I. W.; Cherry, M. A.; Maris, M. B.; Matous, J. V.; Berdeja, J. G.; Patel, M. Combination Trial of Duvelisib (IPI-145) With Rituximab or Bendamustine/Rituximab in Patients With non-Hodgkin Lymphoma or Chronic Lymphocytic Leukemia. *Am. J. Hematol.* **2019**, *94*, 1325–1334.
- (83) Agarwal, K.; Brunetto, M.; Seto, W. K.; Lim, Y.-S.; Fung, S.; Marcellin, P.; Ahn, S. H.; Izumi, N.; Chuang, W. L.; Bae, H.; Sharma, M.; et al. 96 weeks treatment of tenofovir alafenamide vs. tenofovir disoproxil fumarate for hepatitis B virus infection. *J. Hepatol.* **2018**, *68*, 672–681.
- (84) Wu, J.; Zhang, M.; Liu, D. Acalabrutinib (ACP-196): A Selective Second-Generation BTK Inhibitor. *J. Hematol. Oncol.* **2016**, *9*, 21.
- (85) Benson, S. C.; Pershadsingh, H. A.; Ho, C. I.; Chittiboyina, A.; Desai, P.; Pravenec, M.; Qi, N.; Wang, J.; Avery, M. A.; Kurtz, T. W. Identification of telmisartan as a unique angiotensin II receptor antagonist with selective PPAR γ -modulating activity. *Hypertension* **2004**, *43*, 993–1002.
- (86) Ferenci, P.; Bernstein, D.; Lalezari, J.; Cohen, D.; Luo, Y.; Cooper, C.; Tam, E.; Marinho, R. T.; Tsai, N.; Nyberg, A.; Box, T. D.; et al. ABT-450/r-ombitasvir and dasabuvir with or without ribavirin for HCV. *N. Engl. J. Med.* **2014**, *370*, 1983–1992.
- (87) Drug Information Panel, Summary for Cytarabine. druginfo.nlm.nih.gov/drugportal/name/cytarabine (accessed on May 28, 2020).
- (88) Bourlière, M.; Gordon, S. C.; Flamm, S. L.; Cooper, C. L.; Ramji, A.; Tong, M.; Ravendhran, N.; Vierling, J. M.; Tran, T. T.; Pianko, S.; et al. Sofosbuvir, Velpatasvir, and Voxilaprevir for Previously Treated HCV Infection. *N. Engl. J. Med.* **2017**, *376*, 2134–2146.