This article is made available via the ACS COVID-19 subset for unrestricted RESEARCH re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for the duration of the World Health Organization (WHO) declaration of COVID-19 as a global pandemic.

ACS Pharmacology & Translational Science

pubs.acs.org/ptsci

Article

# Hybrid *In Silico* Approach Reveals Novel Inhibitors of Multiple SARS-CoV-2 Variants

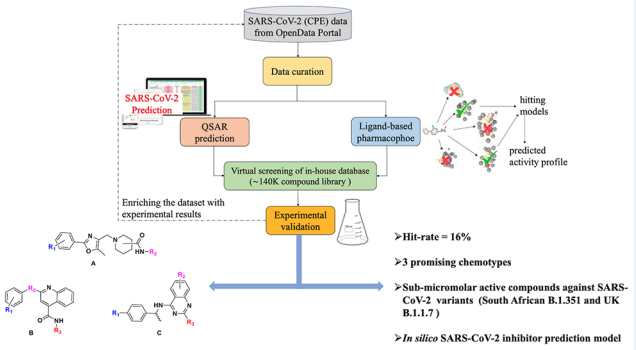
Sankalp Jain, Daniel C. Talley, Bolormaa Baljinnyam, Jun Choe, Quinlin Hanson, Wei Zhu, Miao Xu, Catherine Z. Chen, Wei Zheng, Xin Hu, Min Shen, Ganesha Rai, Matthew D. Hall, Anton Simeonov, and Alexey V. Zakharov*

Cite This: *ACS Pharmacol. Transl. Sci.* 2021, 4, 1675−1688

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

**ABSTRACT:** The National Center for Advancing Translational Sciences (NCATS) has been actively generating SARS-CoV-2 high-throughput screening data and disseminates it through the OpenData Portal (https://opendata.ncats.nih.gov/covid19/). Here, we provide a hybrid approach that utilizes NCATS screening data from the SARS-CoV-2 cytopathic effect reduction assay to build predictive models, using both machine learning and pharmacophore-based modeling. Optimized models were used to perform two iterative rounds of virtual screening to predict small molecules active against SARS-CoV-2. Experimental testing with live virus provided 100 (∼16% of predicted hits) active compounds (efficacy > 30%, $IC_{50} \leq 15\ \mu M$). Systematic clustering analysis of active compounds revealed three promising chemotypes which have not been previously identified as inhibitors of SARS-CoV-2 infection. Further investigation resulted in the identification of allosteric binders to host receptor angiotensin-converting enzyme 2; these compounds were then shown to inhibit the entry of pseudoparticles bearing spike protein of wild-type SARS-CoV-2, as well as South African B.1.351 and UK B.1.1.7 variants.




**KEYWORDS:** *COVID-19, SARS-CoV-2, virtual screening, machine learning, pharmacophore modeling*

In December 2019, a novel coronavirus strain SARS-CoV-2 began to spread in Wuhan, China[1] and eventually led to an alarming global pandemic. As of May 2021, the pandemic has reached over 154 million cases and the resulting complications have caused more than 30 million deaths worldwide.[2] Numerous strategies have been employed to find a reliable COVID-19 therapy including vaccine development, drug repurposing, and developing novel small-molecule SARS-CoV-2 inhibitors.[3−6] The FDA has now issued emergency use authorization for multiple vaccines; however, the outbreak is far from under control, mainly due to the emergence of SARS-CoV-2 variants. As per a recent CDC report, there are 13 variants, five of which are classified as variants of concern.[7,8]

At the beginning of the pandemic, the National Center for Advancing Translational Sciences (NCATS) started a COVID-19 drug repurposing campaign and created the OpenData Portal to make SARS-CoV-2-related assay data publicly accessible.[9] The COVID-19-targeted high-throughput screening (HTS) campaigns at NCATS apply a wide range of biochemical and cell-based assays, including the cytopathic effect assay (CPE) of live SARS-CoV-2 in Vero-E6 cells.[10] More recently, NCATS included data generated from testing of potential therapeutics against different SARS-CoV-2 variants (https://opendata.ncats.nih.gov/variant/assays).

Drug discovery is a time- and resource-intensive process; virtual screening (VS) to identify small-molecule protein modulators offers significant advantages, especially when used to complement traditional HTS methodology.[11,12] Multiple *in silico* studies related to SARS-CoV-2 have been reported, which employ virtual screening of small-molecule databases.[13−20] However, in most of these published communications, hit compounds were not experimentally validated in SARS-CoV-2 assays or were not counterscreened for cytotoxicity, rendering the results inconclusive.

While many efforts are focused on repurposing existing drugs,[21−23] we performed a hybrid virtual screening of two in-house libraries (∼140k compounds) in an effort to identify new chemotypes with antiviral activity and limited cytotoxicity, utilizing the NCATS publicly available screening data. This hybrid approach integrates a quantitative structure−activity

**Received:** July 15, 2021
**Published:** September 17, 2021

ACS Publications

Not subject to U.S. Copyright. Published 2021 by American Chemical Society

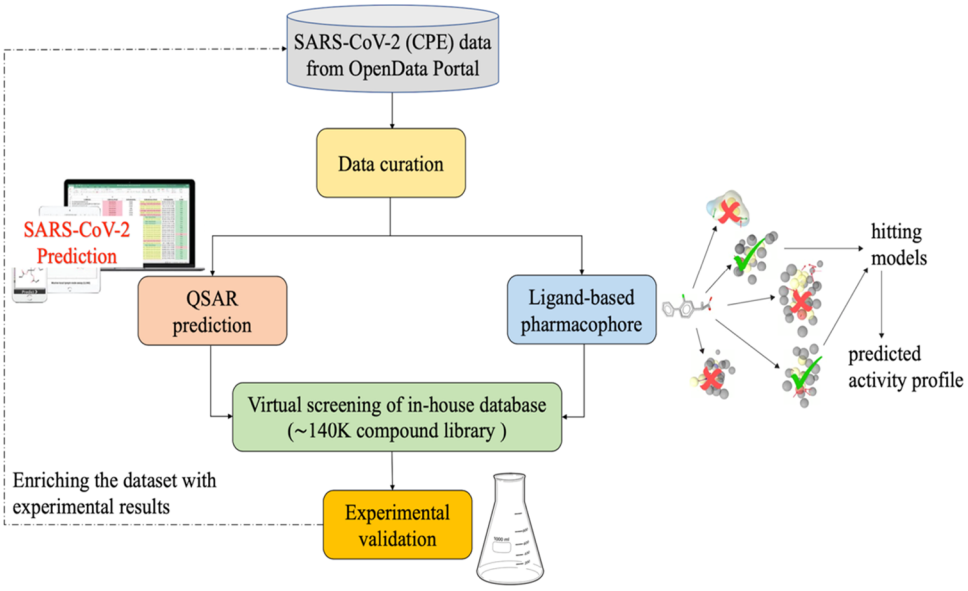



**Figure 1.** Flowchart of the virtual screening strategy used in this study.

relationship (QSAR) and ligand-based pharmacophore (LBP) modeling, followed by experimental testing of predicted hits in CPE and cytotoxicity assays. We executed two iterative rounds of virtual screening; hit compounds identified in the first round were experimentally tested, and these data were utilized to enrich the training data set for the proposed hybrid approach used in the second round (Figure 1).

These efforts resulted in a total of 100 compounds (out of 640 virtual screening hits; hit rate ∼16%) which showed inhibition (half-maximum inhibitory concentration, $IC_{50} \leq 15$ $\mu$M; the maximum inhibitory effect observed, efficacy[24] > 30%) in the CPE assay and minimal cytotoxicity ($IC_{50} > 30$ $\mu$M), where 68 of them had an efficacy greater than 70%. Interestingly, three novel antiviral chemotypes emerged with multiple (≥3) active structural analogues in each cluster. Some preliminary structure−activity relationships (SARs) were identified, which validates these chemotypes as candidates for further medicinal chemistry optimization as novel SARS-CoV-2 inhibitors.

In an effort to elucidate the mechanism of action, hit compounds/chemotypes were tested across several viral targets. Six novel SARS-CoV-2 CPE inhibitors were identified as allosteric ACE2 binders (using microscale thermophoresis, MST) and also blocked viral entry, as assessed by a pseudoparticle entry assay (PP assay). In most cases, ACE2 binding showed a direct correlation to activity in the PP assay. We further validated these six novel inhibitors in PP assays using both the South African and the UK SARS-CoV-2 variants; two compounds were identified with submicromolar activity against both variants.

To the best of our knowledge, this is the first study that identifies novel inhibitors of multiple SARS-CoV-2 variants with an elucidated mechanism of action. The curated data set and the optimized prediction models are publicly available *via* github (https://github.com/ncats/covid19_pred) as well as NCATS Predictor website (https://predictor.ncats.io/).

## ■ MATERIAL AND METHODS

The data used in this study were obtained from single-agent screening in both the SARS-CoV-2 cytopathic effect (CPE) and a host-toxicity counterscreen (CTG), CellTiter-Glo (CTG). These data are publicly accessible *via* the NCATS OpenData Portal (https://opendata.ncats.nih.gov/covid19/). As NCATS has been unceasingly performing screening campaigns, we combined these data with additional in-house quantitative HTS data, resulting in a data set of 9046 compounds.

**Data Set Curation.** The data set was curated following a protocol previously developed by Fourches et al.[25−27]

Briefly, the following steps were performed: (i) removal of inorganic compounds according to the chemical formula in MOE 2019.01;[28] (ii) removal of salts and compounds containing metals and/or rare or special atoms; (iii) standardization of chemical structures using Francis Atkinson's standardizer (https://github.com/flatkinson/standardiser); ans (iv) removal of duplicates and permanently charged compounds using MOE 2019.01.[28]

**Compound Labeling.** Compounds having an $IC_{50} < 30$ $\mu$M, curve class in the range of 1−3,[29] and a maximum response (MaxResponse) > 30% were considered active for the CPE reduction assay, whereas others were labeled as inactive. For the cytotoxicity counterscreen (CTG), compounds with an $IC_{50} < 30$ $\mu$M, curve class in the range of 1 to −3, and MaxResponse <−30% were considered active and others as inactive. In the combined data set, compounds active in the CPE reduction assay and inactive in the CTG counterscreen were considered as active. All others were labeled as inactive. While merging the data from multiple protocols, compounds with contradictory results in different experimental runs were removed from the study. For the first round of modeling, the data set was comprised of 8474 compounds (319 active and 8155 inactive). Enriching this data set with compounds identified in the first round of virtual screening and experimentally tested in the first round of screening, resulted in a data set of 9046 compounds (456 active and 8590 inactive).

**Descriptor Calculation.** Three different sets of descriptors were calculated for all data sets using RDKit (https://www.rdkit.org/).


https://doi.org/10.1021/acsptsci.1c00176
ACS Pharmacol. Transl. Sci. 2021, 4, 1675−1688

1. RDKit descriptors based on the two-dimensional structure (119 descriptors in total).
2. Morgan fingerprints (1024 bits).
3. Avalon fingerprints (1024 bits).

**Training and Test Set Selection.** From each class (active, inactive), 70% of the data were randomly selected and used as a training set. The remaining 30% of compounds were considered as the test set. Five-fold external cross-validation was omitted since pharmacophore modeling is computationally expensive and the selection of the best consensus model combining pharmacophore and machine learning approaches requires the established training and the test set. We emphasize that the selected consensus model was used in virtual screening and thus, prospectively validated. The composition of the resulting data sets is shown in Table 1.

**Table 1. Overview of the Data Sets Used in This Study**

| | total compounds | active | inactive | imbalance ratio (inactive/active) |
|---|---|---|---|---|
| | | First Round | | |
| full data set | 8474 | 319 | 8155 | 26:1 |
| training set | 5931 | 223 | 5708 | 26:1 |
| test set | 2543 | 96 | 2447 | 25:1 |
| | | Second Round | | |
| full data set | 9046 | 456 | 8590 | 19:1 |
| training set | 6332 | 319 | 6013 | 19:1 |
| test set | 2714 | 137 | 2577 | 19:1 |

**Virtual Screening Libraries.** To discover compounds active against SARS-CoV-2, we performed virtual screening using two of our internal libraries (∼140k compounds). These libraries contain a diverse collection of small molecules with an emphasis on medicinal chemistry-tractable scaffolds. The compound libraries were curated using the same protocol described in the Data set Curation section. These compounds were screened against the model and rank ordered based on the predicted activity score, which roughly corresponds to their probability of being active against SARS-CoV-2.

**Machine Learning: Stratified Bagging (SB).** Considering the high degree of data imbalance (active:inactive, 1:26), we used undersampling stratified bagging (SB); this method has been proven to be superior when dealing with imbalanced data sets.[30] SB is a machine learning technique based on an ensemble of models developed using multiple training data sets sampled from the original training set. It uses a traditional bagging approach (resampling with replacement) to create the training set of positive samples and randomly selects the same number of samples from the majority class. Thus, the total bagging training set size is double the minority class. Several models are then built and predictions are averaged to produce a final ensemble model output. Because of random sampling, about 37% of the compounds are left out in each run. These samples form "out-of-the-bag" sets, which are then used to test the final model. Although a small set of samples are selected each time, most compounds contribute to the overall bagging procedure since data sets were generated randomly. Random forest (RF) was used as a base classifier.[31] The number of trees was arbitrarily set to 100 (default) since it has been shown that the optimal number of trees is usually 64−128, while further increasing the number of trees does not necessarily improve the model's performance.[32]

Consensus QSAR modeling is another highly recommended approach that has been reported to outperform simple QSAR models.[33,34] In this study, we used a consensus approach that utilizes the consensus of the predictions from three different descriptors to predict anti-SARS-CoV-2 activity.

**Pharmacophore-Based Screening.** A pharmacophore describes the spatial arrangement of essential interactions of a drug with its respective binding site. Pharmacophore modeling and subsequent virtual screening (VS) is a well-established method utilized in drug discovery.[35,36] In this study, the generation of ligand-based pharmacophore models, their subsequent refinement, and virtual screening (VS) were performed with LigandScout 4.4 Advanced (Inte:Ligand GmbH). The conformational libraries for both pharmacophore modeling and the VS process were created with i:Con (max. 200 conformations per compound), a conformer generator implemented in LigandScout.[37] To design the ligand-based pharmacophore model (LBP), the most potent compounds were selected based on the $IC_{50}$ (<30 $\mu$M) and MaxResponse (>50.0%) values in the CPE assay. The molecules were clustered based on pharmacophore-based similarity (cluster distances 0.4, 0.6, 0.7, and 0.8, respectively). For each cluster obtained from different distance thresholds, merged-features pharmacophore (MFP) and shared-features pharmacophore (SFP) models were generated, which incorporate the features of selected compounds per cluster. A MFP merges all pharmacophore features for different molecules (in each cluster) into a single pharmacophore, interpolating the overlapping features. In comparison, an SFP contains a collection of overlapping pharmacophore features from different molecules (per cluster).[38] These features allow two or more similar bio-active molecules to bind to the macromolecule in a comparable way and trigger similar biological responses. Owing to the abstract nature of pharmacophore models, they represent an efficient approach for the virtual screening of large compound libraries.[39] To be more robust, we incorporated both MFP and SFP for our virtual screening. Furthermore, a good pharmacophore model should not only be able to estimate the activity of active compounds but also have the ability to identify active molecules from a database containing a large number of inactive compounds. To select the best models for screening, we applied these models on our complete data set (training and test set combined) and calculated the percentage of active and inactive that hit these pharmacophore models. The models that could identify 20% more active than inactive compounds were selected for the final virtual screening. The ligand-based pharmacophore models generated in the first and second rounds of screening are referred to as LBP-1 and LBP-2, respectively. The screening was performed using the iscreen module with default settings, with the maximum number of omitted features set to 2.

**First Round of *In Silico* Screening.** The complete collection of 138,749 compounds was tested against our stratified bagging models and ranked by the prediction score. This score takes a value between 0 and 1, indicates the probability of a compound to be active against SARS-CoV-2. In the first round of screening, we selected the top 300 predictions from each descriptor combination; 890 compounds were predicted to be in the top 300 prediction scores by at least two of the four models. We then retrieved all generated conformations of these 890 compounds prepared from the i:Con[37] and screened them against our pharmacophore models


https://doi.org/10.1021/acsptsci.1c00176
ACS Pharmacol. Transl. Sci. 2021, 4, 1675−1688

(LBP-1) (Table S1 in the Supporting Information). Finally, we ranked the compounds according to their pharmacophore-fit score. The top 320 compounds were selected for experimental validation in the SARS-CoV-2 CPE assay and CTG counter-screen.

**Second Round of *In Silico* Screening.** After obtaining results from the first round of modeling, followed by experimental validation, we updated our machine learning (SB) and ligand-based pharmacophore (LBP) model with the confirmed SARS-CoV-2 active and inactive compounds. We then removed the 320 compounds tested in the first round of experimental validation from our virtual screening library. The remaining compounds (138,429 compounds) were predicted using our updated stratified bagging models (SB-2) and ranked by the prediction score. In the second round of screening, we selected the top 500 predictions from each descriptor combination. This resulted in a subset of 1,325 compounds that were consistently present in the list of top predictions by at least two of the four models. We screened 138,429 compounds using the updated ligand-based pharmacophore models (LBP-2) shortlisted for screening (Table S2 in the Supporting Information). This gave us 65,952 compounds. We then ranked these compounds according to their pharmacophore fit and selected 320 compounds that were overlapping with 1,325 compounds for experimental validation.

**Model Performance Assessment.** Receiver operating characteristic area under the curve (ROC AUC) was used to assess the performance of the models. ROC AUC plots sensitivity (TP/(TP + FN)) against 1-specificity (TN/(TN + FP)). The higher the ROC AUC value, the better the model performs in distinguishing between active and inactive compounds. A ROC AUC value of 1.0 indicates a perfect classification model, whereas a value close to 0.5 indicates that the model provides random predictions. The ROC AUC score was calculated using the ROC curve (javascript) node in KNIME.[40] For the experimental validation results, the model performance was measured by the positive predicted value (PPV = TP/(TP + FP)). Based on earlier studies, we chose PPV and AUC as our model performance metrics.[41,42]

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

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

$$\text{balanced accuracy} = \frac{1}{2}\left(\frac{(\text{TP})}{(\text{TP} + \text{FN})} + \frac{(\text{TN})}{(\text{TN} + \text{FP})}\right)$$

$$\text{MCC} = \frac{\{(\text{TP} \times \text{TN}) - (\text{FP} \times \text{FN})\}}{\{(\text{TP} + \text{FP}) \times (\text{TP} + \text{FN}) \times (\text{TN} + \text{FP}) \times (\text{TN} + \text{FN})\}^{1/2}}$$

$$\text{precision} = \frac{\text{TP}}{(\text{TP} + \text{FP})}$$

where TP are true positives, TN are true negatives, FP are false positives, FN are false negatives, and MCC is Matthews correlation coefficient.

**Experimental Testing.** *SARS-CoV-2 Cytopathic Effect Assay.* The cytopathic effect (CPE) of SARS-CoV-2 was measured in Vero-E6 cells in a BSL-3 facility as described previously.[10] Cells were harvested, resuspended at 160 000 cells/mL in assay media (minimal essential medium, MEM, 2% (v/v) heat inactivated fetal bovine serum, FBS, 1% 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid, HEPES, 1% Pen/Strep/GlutaMax), and inoculated with SARS-CoV-2 (USA_WA1/2020) at a multiplicity of infection (MOI) of 0.002. Twenty-five microliters of the cell−virus mixture was dispensed per well of an assay-ready 384-well plate (Greiner, #781091). The assay-ready plates were prepared by adding 5 $\mu$L of assay media per well, prespotted with 60 nL of library compounds at a five-point serial dilution with concentrations ranging from 10 mM to 62 $\mu$M (final assay concentrations ranging from 20 $\mu$M to 124 nM) using an acoustic dispenser (Echo550, Labcyte, Inc.). Each plate contained two columns with 60 nL of dimethyl sulfoxide (DMSO) as negative (no inhibitor) control and 24 wells containing cells only (no virus) as a positive control. Plates were incubated for 72 h at 37 °C, 5% $CO_2$, 90% humidity. The cell viability was assessed by measuring the luminescence signal with Envision plate reader (PerkinElmer) after the addition of 30 $\mu$L/well of CellTiter-Glo reagent (Promega, Cat #G7573) and 10 min incubation at room temperature. The signal was normalized against negative (0% response) and positive control (100% response), and the resulting percent of inhibition data were fitted to a sigmoidal dose-response curve using the four-parameter Hill equation.

**CellTiter-Glo Counterscreen.** The assay was set up in the same way as in the CPE assay but omitting the addition of virus. DMSO and hyamine at 100 $\mu$M final concentration served as negative and positive controls, respectively. The obtained luminescence signal was normalized against negative control (0% response) and positive control (−100% response).

**SARS-CoV-2 Mpro Assay.** The ability of the compounds to inhibit the recombinant $M^{pro}$ activity was measured by a biochemical assay, SARS-CoV-2 $M^{pro}$ assay described previously.[43]

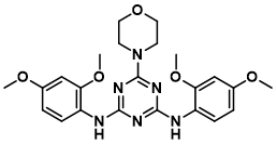
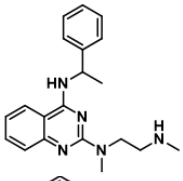
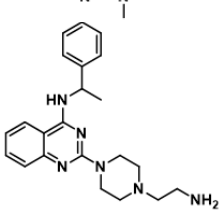
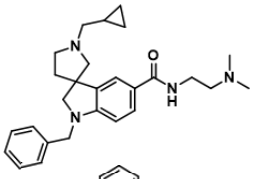
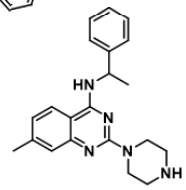
**ACE2-RBD AlphaLISA Proximity Assay.** The interaction of SARS-CoV-2 receptor-binding domain (RBD) with ACE2 was tested using an AlphaLISA proximity assay in combination with a corresponding TruHits counter assay as described elsewhere.[44]

**Microscale Thermophoresis Assay.** The binding of compounds to recombinant human ACE2 (Sino Biological, Cat #: 10108-H08H) was evaluated by microscale thermophoresis (MST). His-tagged ACE2 was labeled with RED-tris-NTA second generation dye (Nanotemper Technologies, Cat #: MO-L018) following manufacturer's protocol and diluted in MST buffer (10 mM HEPES pH 7.4, 150 mM NaCl, 10 mM $CaCl_2$, 0.01% Tween-20) to a final concentration of 3 nM. Hundred nanoliters of compounds in 2-fold dilution series were transferred to 384-well compound plate (Greiner, Cat #: 784201-1B) using an Echo 650 series acoustic dispenser (Labcyte Inc.), mixed with 10 $\mu$L of labeled protein and incubated for 15 min at room temperature (RT). MST traces were collected using a Monolith NT.Automated (Nanotemper Technologies) unit and a standard treated capillary chip (Nanotemper Technologies, Cat #: MO AK002) with following setting: 45% excitation power, medium MST power, and MST periods of 3 s/10 s/1 s. $K_d$ values were calculated by fitting the change in the normalized fluorescence signal of the thermograph using MO.Affinity analysis software.

**ACE2 Enzymatic Assay.** ACE2 enzyme activity was monitored in a fluorometric assay. Briefly, 25 nL of compounds were transferred to the 1536-well assay plate (Greiner, solid black medium-binding plates) using an Echo 650 (Labcyte Inc.) acoustic dispenser. Typically, 3 $\mu$L/well of 0.27 nM


https://doi.org/10.1021/acsptsci.1c00176
ACS Pharmacol. Transl. Sci. 2021, 4, 1675−1688

**Table 2. Five Most Potent and Efficacious Compounds Identified, along with *In Vitro*/Physicochemical ADME Data**

| ID | Structure | $IC_{50}$ (μM)[a] | Efficacy (%)[b] | $T_{1/2}$ (min)[c] | PAMPA (1e-6 cm/s)[d] | Solubility (μg/mL)[e] |
|---|---|---|---|---|---|---|
| 1 |  | 5.01 | 90 | 25 | ND | ND |
| 2 |  | 6.3 | 82 | 5.5 | 1110 | >50 |
| 3 |  | 7.94 | 100 | >30 | 60.6 | >56 |
| 4 |  | 7.94 | 98 | 9.9 | 64.9 | >96 |
| 5 |  | 10 | 89 | 4.4 | 1660 | 31 |

[a]$IC_{50}$: half-maximal inhibitory concentration value obtained from the CPE assay in eight-point dose response, measured in duplicate. [b]Efficacy: maximum inhibitory effect observed in CPE assay. [c]$T_{1/2}$: metabolic half-life measured in rat liver microsome fractions reported in minutes, with a minimum detectable half-life of 1 min.[47] [d]Parallel artificial membrane permeation assay (PAMPA) is reported as a metric of the passive permeability of the compounds ($1 \times 10^{-6}$ cm/s).[48,49] [e]Solubility—pION μSOL assay for kinetic aqueous solubility determination, pH 7.4.[50]

ACE2 (0.2 nM final concentration) suspension in assay buffer (PBS, pH 7.4, 0.01% Tween-20) was dispensed into assay plate with an Aurora Discovery BioRAPTR Dispenser (FRD; Beckton Dickenson) and incubated 15 min at room temperature (RT). Typically, 1 μL/well of 60 μM ACE2 substrate (AnaSpec, Cat #: AS-60757) was then added. The plate was centrifuged at 1000 rpm for 15 s and the fluorescence was detected with the PHERAstar plate reader (BMG Labtech) equipped with Module 340/440 at $t_1$ = 0 min and $t_2$ = 15 min at RT. Data was normalized to enzyme activity in the presence of DMSO, set as 0%, and in the presence of 6.2 μM MLN-4760, set as −100% inhibition. The resulting percent of inhibition data were fitted to a sigmoidal dose-response curve using a four-parameter Hill equation.

**Pseudotyped Particle (PP) Entry Assay.** Expi293F cells with stable expression of human ACE2 (HEK293-ACE2, Codex Biosolutions, Cat #: CB-97100-220) were seeded in white, solid bottom 384-well microplates (Greiner BioOne) at 6000 cells/well in 30 μL/well medium (DMEM, 10% FBS, 1× L-glutamine, 1× Pen/Strep, 1 μg/mL puromycin) and incubated at 37 °C with 5% $CO_2$ overnight (∼16 h). One hundred fifty nanoliters of compounds at 11-point titration, 1:3 dilution in DMSO, were dispensed via an Echo 650 (Labcyte Inc.) acoustic dispenser to assay plates. Cells were incubated with the test compounds for 1 h at 37 °C with 5% $CO_2$, before 2 μL/well of SARS-CoV-2-S PPs were added. PPs with the following spike variants were used: wild type (Codex Biosolutions, Cat #: CB-97100-154), South African variant

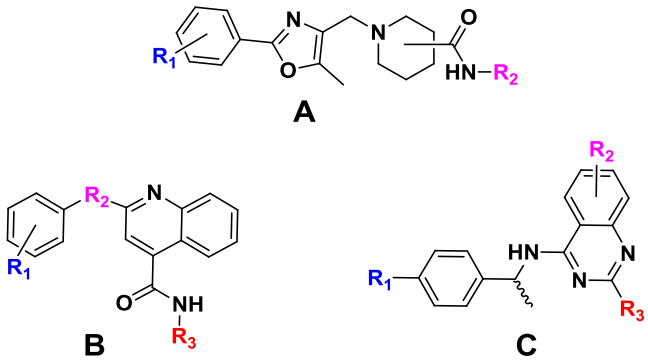
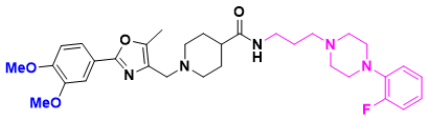
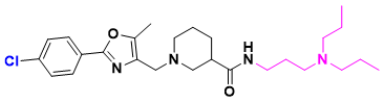
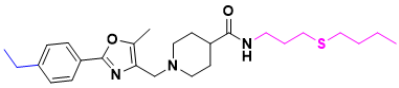



**Figure 2.** Three chemotypes (**A**−**C**) were identified as active in the CPE assay.


https://doi.org/10.1021/acsptsci.1c00176
ACS Pharmacol. Transl. Sci. 2021, 4, 1675−1688

Table 3. Notably Active Chemotype A Which Shows No Notable Cytotoxicity ($IC_{50} \leq 30\ \mu M$)

| ID | Structure | IC50 (μM)[a] | Efficacy (%)[c] | $T_{1/2}$ (min)[d] | PAMPA (1e-6 cm/s)[e] | Solubility (μg/mL)[f] |
|---|---|---|---|---|---|---|
| 6 |  | 8.91[b] | 80[b] | 8.3 | 266 | 56 |
| 7 |  | 12.6 | 88 | 4.3 | 1600 | 37 |
| 8 |  | 14.1 | 73 | 5.5 | 1470 | 4.4 |

[a]$IC_{50}$: half-maximal inhibitory concentration values obtained from the CPE assay in eight-point dose response, measured in duplicate. [b]Values represent data obtained from five-point dose response, measured in duplicate. [c]Efficacy: maximum inhibitory effect observed in CPE assay. [d]$T_{1/2}$: metabolic half-life measured in rat liver microsome fractions reported in minutes, the minimum detectable half-life of 1 min.[47] [e]Parallel artificial membrane permeation assay (PAMPA) is reported as a metric of the passive permeability of the compounds.[48,49] [f]Solubility—pION μSOL assay for kinetic aqueous solubility determination, pH 7.4.[50]

B.1.351 (Codex Biosolutions, Cat #: CB-97100-154), and U.K. variant B.1.1.7 (Codex Biosolutions, Cat #: CB-97100-153). The plates were spinoculated by centrifugation at 1500 rpm (453*g*) for 45 min at RT and incubated for 48 h at 37 °C with 5% $CO_2$ to allow cell entry of PPs and expression of a luciferase reporter. After the incubation, the supernatant was removed with gentle centrifugation using a Blue Washer (BlueCat Bio). Typically, 4 μL/well of the Bright-Glo Luciferase detection reagent (Promega) was added to assay plates and incubated for 5 min at RT. The luminescence signal was measured using a PHERAStar plate reader (BMG Labtech). Data were normalized against wells inoculated with SARS-CoV-2-S PPs as 100% entry and wells inoculated with "bald" PPs (containing no spike protein) as 0%.

An ATP content cytotoxicity assay was done as a counter assay. HEK293-ACE2 cells were seeded in white, solid bottom 384-well microplates (Greiner BioOne) and treated with compounds under the same experimental conditions as in the PP entry assay, omitting the inoculation step. After 48 h incubation, 4 μL/well of ATPLite (PerkinElmer) was added to assay plates and incubated for 15 min at RT. The luminescence signal was measured using a Viewlux plate reader (PerkinElmer). Data were normalized with wells containing cells as 100% and wells containing media only as 0%.

## RESULTS

**Hybrid Approach for *In Silico* Screening.** A combination of ligand- and structure-based methods has been used previously to discover small-molecule modulators of various targets,[35,36,45] since it improves the precision and reduces false positives.[46] In this study, we combined QSAR modeling with pharmacophore-based screening to identify novel chemotypes active against SARS-CoV-2. We used a consensus of the predictions based on the two approaches (Figure 1) to select compounds for experimental validation.

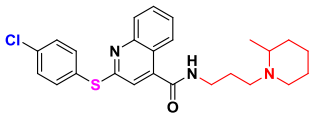
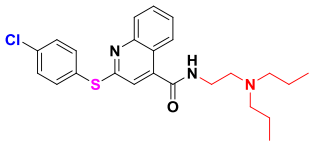
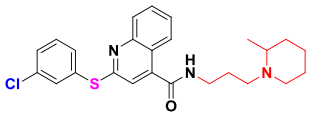
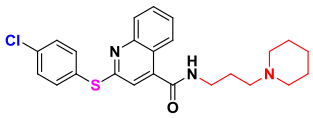
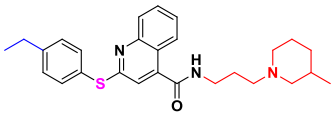
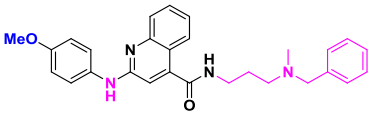
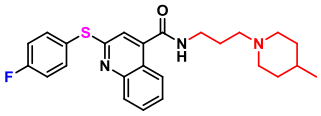
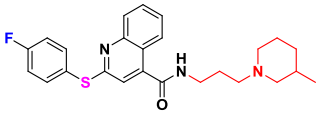
**Model Performance—Stratified Bagging.** The performance of QSAR models based on different combinations of descriptors and stratified bagging (SB) approaches is provided in Table S3 in the Supporting Information. It was measured by different metrics, including the receiver operating characteristic area under the curve (ROC AUC). All developed models showed ROC AUC values > 0.75. In the first round of modeling, the consensus of descriptors (RDKit, Morgan and Avalon) provided the best performance with ROC AUC = 0.80 on the test set. SB models generated in the first round of screening are referred to as SB-1. After obtaining the experimental results from the first round of virtual screening, we updated our SB model (referred to as SB-2). In the second round of modeling, the consensus of descriptors (RDKit, Morgan and Avalon) showed improved results with ROC AUC = 0.84 on the test set.

**Ligand-Based Pharmacophore Modeling.** For the first round of ligand-based pharmacophore modeling (LBP-1), we used 48 active compounds: clustering based on pharmacophore-based similarity (cluster distances of 0.4, 0.6, 0.7, and 0.8), followed by generation of ligand-based hypotheses led to a total of 44 pharmacophore hypotheses (merged-features pharmacophore (MFP) and shared-features pharmacophore (SFP)). Taking the computational constraints into account, 15 pharmacophore models that hit the majority (>20%) of active versus inactive (Table S1 in the Supporting Information) were selected for virtual screening. For the second round, referred to as LBP-2, we considered 53 actives and followed the same protocol as above. This resulted in 55 pharmacophore hypotheses (MFP and SFP). Pharmacophore models (20) were then selected for virtual screening. All pharmacophore hypotheses generated in this study are presented in the Supporting Information (Table S2 in the Supporting Information). In general, pharmacophoric sites such as hydrogen bond acceptor (HBA), hydrogen bond donor (HBD), aromatic ring, hydrophobic sites, and positive ionizable groups were prudently characterized.

**Experimental Testing of the First Round *In Silico* Screening Hits.** The 320 compounds selected from the first round of *in silico* screening (see the Materials and Methods section for details) were tested in the CPE assay in a five-point dilution series, with concentrations ranging from 20 μM to 124 nM. To exclude compounds with cytotoxic effects, the


https://doi.org/10.1021/acsptsci.1c00176
ACS Pharmacol. Transl. Sci. 2021, 4, 1675−1688

Table 4. Notably Active Chemotype B Which Shows No Notable Cytotoxicity ($IC_{50} \leq 30\ \mu M$)

| ID | Structure | $IC_{50}$ (μM)[a] | Efficacy (%)[c] | $T_{1/2}$ (min)[d] | PAMPA (1e-6 cm/s)[e] | Solubility (μg/mL)[f] |
|---|---|---|---|---|---|---|
| 9 |  | 6.31[b] | 100[b] | 3.3 | 815 | 31 |
| 10 |  | 10.0[b] | 97[b] | 2.2 | 564 | 1.9 |
| 11 |  | 10.0[b] | 100[b] | 2.0 | 1560 | 25 |
| 12 |  | 11.2 | 99 | 3.3 | Equilibrated | 30 |
| 13 |  | 12.6 | 87 | 2.3 | 1530 | >65 |
| 14 |  | 11.2[b] | 74[b] | 4.0 | Equilibrated | >65 |
| 15 |  | 14.1 | 66 | 3.3 | 1610 | >65 |
| 16 |  | 12.6 | 91 | 2.2 | 1840 | >65 |

[a]$IC_{50}$: half-maximal inhibitory concentration values obtained from the CPE assay in eight-point dose response, measured in duplicate. [b]Values represent data obtained from five-point dose response, measured in duplicate. [c]Efficacy: maximum inhibitory effect observed in CPE assay. [d]$T_{1/2}$: metabolic half-life measured in rat liver microsome fractions reported in minutes, minimum detectable half-life of 1 min.[47] [e]PAMPA (parallel artificial membrane permeation assay) is reported as a metric of the passive permeability of the compounds.[48,49] [f]Solubility—pION μSOL assay for kinetic aqueous solubility determination, pH 7.4.[50]

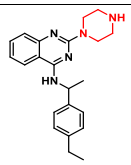
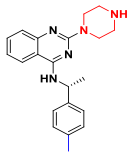
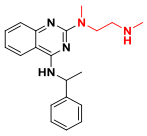
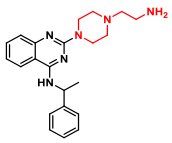
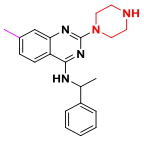
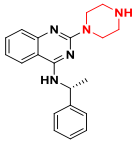
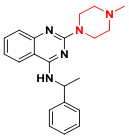
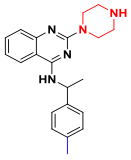
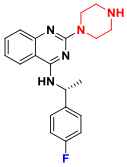
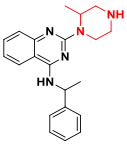
compounds were counterscreened in a cell viability assay. Out of the 320 compounds tested, 46 compounds showed a SARS-CoV-2 CPE inhibiting activity with a maximum response (MaxResponse) greater than 30% and $IC_{50}$ values of 3−15 μM. Of these 46, 42 compounds did not show any cytotoxicity or modest toxicity with an efficacy <25% (Table S4 in the Supporting Information). This provided a positive prediction value (PPV) (i.e., the fraction of model predicted positives that are experimentally confirmed) of 13%, which is 3-fold higher than the PPV calculated from the training set (PPV = 4%).

**Experimental Testing of the Second Round *In Silico* Screening Hits and Validation.** The selected 320 compounds from the second round of *in silico* screening (see the Materials and Methods section for details) were also tested in the CPE assay in a five-point dilution series, with concentrations ranging from 20 μM to 124 nM. For the second round testing, out of the 320 compounds, 65 compounds were identified with anti-SARS-CoV2-2 activity having a MaxResponse >30% and $IC_{50}$ values of 3−15 μM. Of these 65 compounds, 58 did not show any cytotoxicity or minimal toxicity (efficacy < 30%; Table S5 in the Supporting Information). Moreover, 27 of these 58 compounds exhibited an $IC_{50} \leq 10\ \mu M$ in the CPE assay. This further improved the PPV value to 18%.

For validation, 69 compounds were "cherry-picked" and retested in the CPE assay in duplicate as an eight-point


https://doi.org/10.1021/acsptsci.1c00176
ACS Pharmacol. Transl. Sci. 2021, 4, 1675−1688

Table 5. Notably Active Chemotype C Which Shows No Notable Cytotoxicity ($IC_{50} \leq 30\ \mu M$)

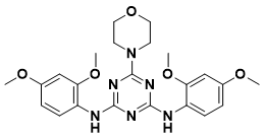
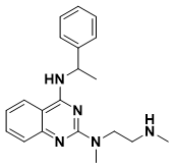
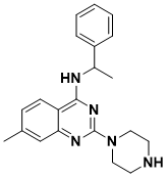
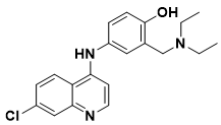
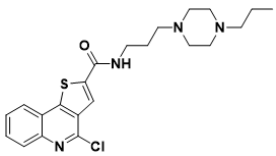
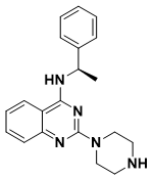
| ID | Structure | $IC_{50}$ (μM)[a] | Efficacy (%)[c] | $T_{1/2}$ (min)[d] | PAMPA (1e-6 cm/s)[e] | Solubility (μg/mL)[f] |
|---|---|---|---|---|---|---|
| **17** |  | 3.98[b] | 74[b] | 11 | Equilibrated | >53 |
| **18** |  | 4.47[b] | 91[b] | 5 | Equilibrated | 41 |
| **2** |  | 6.3 | 82 | 5.5 | 1110 | >50 |
| **4** |  | 7.94 | 100 | >30 | 60.6 | >56 |
| **5** |  | 10.0 | 89 | 4.4 | 1660 | 31 |
| **19** |  | 11.2[b] | 100[b] | <1 | 780 | >49 |
| **20** |  | 12.6[b] | 100[b] | 2.8 | 1570 | >51 |
| **21** |  | 12.6 | 92 | 7.7 | 772 | >51 |
| **22** |  | 14.1 | 81 | 2.8 | Equilibrated | 44.1 |
| **23** |  | 14.1 | 76 | 4.85 | 1273 | >51 |

[a]$IC_{50}$: half-maximal inhibitory concentration values obtained from the CPE assay in eight-point dose response, measured in duplicate. [b]Values represent data obtained from five-point dose response, measured in duplicate. [c]Efficacy: maximum inhibitory effect observed in CPE assay. [d]$T_{1/2}$: metabolic half-life measured in rat liver microsome fractions reported in minutes, a minimum detectable half-life of 1 min.[47] [e]PAMPA (parallel artificial membrane permeation assay) is reported as a metric of the passive permeability of the compounds.[48,49] [f]Solubility—pION μSOL assay for kinetic aqueous solubility determination, pH 7.4.[50]



https://doi.org/10.1021/acsptsci.1c00176
ACS Pharmacol. Transl. Sci. 2021, 4, 1675−1688

Table 6. Compounds Identified as ACE2 Binders and Inhibitors of Viral Entry in PP Assay

| ID | Structure | PP Assay, $IC_{50}$ (μM)[a] | CPE Assay, $IC_{50}$ (μM)[b] | ACE2 binding, $K_d$ (μM)[d] |
|---|---|---|---|---|
| 1 |  | 1.15 | 5.01 | 0.43 |
| 2 |  | 2.89 | 6.31 | 1.83 |
| 5 |  | 4.08 | 10 | 16.7 |
| 24 |  | 5.14 | 5.62[c] | 10.6 |
| 25 |  | 5.14 | 10 | 6.55 |
| 19 |  | 10.3 | 11.2[c] | 2.24 |

[a]Activity in the SARS-CoV-2 PP assay. [b]$IC_{50}$: half-maximal inhibitory concentration values obtained from the CPE assay in eight-point dose response, measured in duplicate. [c]Values represent data obtained from five-point dose response, measured in duplicate. [d]ACE2-binding affinity ($K_d$) measured by MST.

dilution series, with concentrations ranging from 20 μM to 78 nM. Fifty-three of the retested compounds were confirmed to have CPE inhibitory activity with an efficacy > 30% and $IC_{50}$ of 5−25 μM; five exhibited an $IC_{50} \leq 10$ μM with no notable cytotoxicity (Table S6 in the Supporting Information). The five most potent and efficacious compounds (**1**−**5**) from the follow-up CPE assay are shown in Table 2, along with associated *in vitro* ADME/physicochemical properties.

**Clustering and Preliminary SAR Analysis.** Hierarchical cluster analysis of the 100 active compounds revealed three promising chemotypes (Figure 2), where three or more active structural analogues were identified ($IC_{50} \leq 15$ μM and efficacy ≥ 70%) and no notable cytotoxicity ($IC_{50} \leq 30$ μM). Importantly, some preliminary structure−activity relationships (SAR) could be established for chemotypes, where the analogues analyzed (and present in in-house compound libraries) were structurally similar enough for direct comparison. The most promising analogues from each of the three chemotypes, activity in the CPE assay, as well as *in vitro* physicochemical properties are shown in Tables 3−5.

Within chemotype **A**, 26 analogues were tested from internal compound libraries and upon screening, three (compounds **6**−**8**) have $IC_{50}$ values ranging from 8.9 to 14.1 μM and efficacy ≥ 83% (Table 3). Although conclusive SAR trends were limited, the anti-SARS-CoV-2 activity (and cytotoxicity) is sensitive to substitutions on the phenyl ring attached to oxazole and both the position (3- vs 4-) and structure of the piperidinyl amide.


https://doi.org/10.1021/acsptsci.1c00176
ACS Pharmacol. Transl. Sci. 2021, 4, 1675−1688

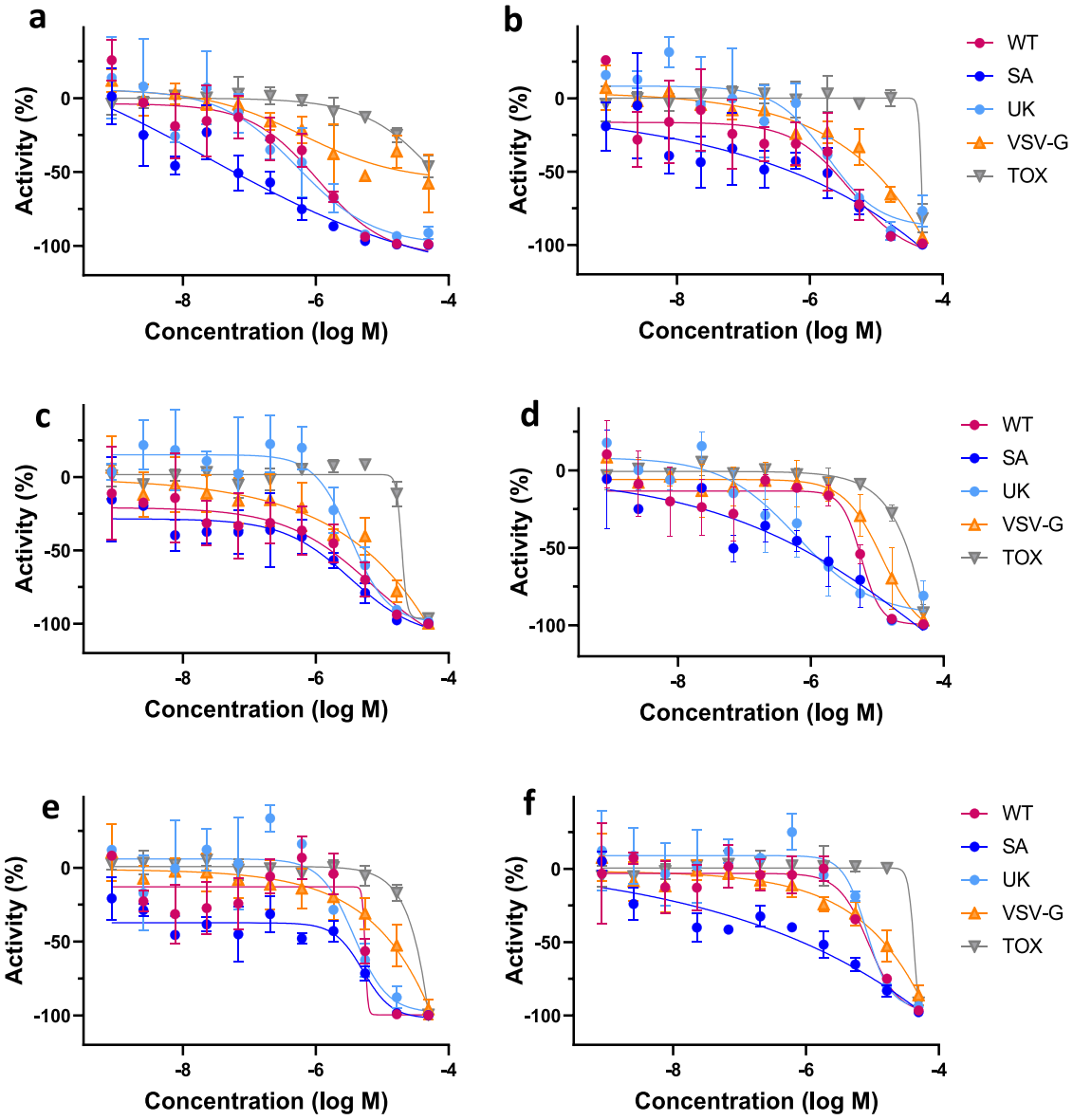



**Figure 3.** Dose-response curves of the six ACE2-binding compounds in PP and CTG assays. (a) Compound **1**, (b) compound **2**, (c) compound **5**, (d) compound **24**, (e) compound **25**, and (f) compound **19**. WT—wild-type SARS-CoV-2 variant assay; SA—South African B.1.351 variant assay; UK—UK B.1.1.7 variant assay; VSV-G—PP assay containing the G-protein of vesicular stomatitis virus; and Tox—cytotoxicity assay.

Within chemotype **B** were 18 in-house structural analogues, eight of which (compounds **9**−**16**) have promising $IC_{50}$ and efficacy values (Table 4). Analogues of this cluster were too structurally very diverse to analyze for conclusive SAR trends. With the exception of **10** which suffers from poor solubility, CPE-actives from this series have favorable solubility and permeability. Similar to chemotype **A**, all suffer from short metabolic half-time ($T_{1/2}$) in rat liver microsomes.

Quinazoline-containing chemotype **C** provided 10 promising analogues (compounds **2**, **4**, **5**, **17**, **18**, **20**−**23**), including three (**2**, **4**, **5**) of the most active compounds identified in the study (Table 5). Most of the notably active analogues contain variously substituted piperazines at the two position of the quinazoline core and 2-methyl-benzylamine at the four position. However, the open-chain (vs piperazine) analogue (**2**) is also quite active, suggesting that a two-position diamine with an ethylene spacer is perhaps part of the parent pharmacophore. Methyl- and ethyl-substitutions off the four position of the benzylamine phenyl ring (**18** and **17**, respectively) were well tolerated, while 4-fluoro (**22**) and 4-phenyl substituents reduced the activity. Similar to chemotypes **A** and **B**, this series has favorable solubility and permeability but suffers from poor metabolic $T_{1/2}$. However, it seems that the metabolic liability can be mitigated *via* the addition of an *N*-aminoethyl group off the piperazine ring (**4**; $T_{1/2}$ > 30 min).

**Mechanism of Action Studies.** In efforts to elucidate the mechanism of action against SARS-CoV-2, significantly active compounds were tested for their activity against some key events necessary for viral entry and replication. The SARS-CoV-2 main protease ($M^{pro}$) represents an attractive target for antiviral drug development because its inhibition prevents the formation of mature functional viral proteins and, thus, viral replication.[51] As such, active compounds were screened in the SARS-CoV-2 $M^{pro}$ enzymatic assay; however, no activity was observed.


https://doi.org/10.1021/acsptsci.1c00176
ACS Pharmacol. Transl. Sci. 2021, 4, 1675−1688

Compounds were also tested for their ability to interrupt the binding of the SARS-CoV-2 receptor-binding domain (RBD) of the spike protein to the host receptor ACE2, using an AlphaLisa proximity assay in combination with a counter assay to identify false-positive hits. All compounds tested showed activity in both RBD-ACE2 AlphaLisa and in the TruHit counterscreen (see the Methods section for details), rendering the results inconclusive.

Nonetheless, we determined whether compounds could bind ACE2 using microscale thermophoresis (MST); they were subsequently tested in a SARS-CoV-2 pseudoparticle (PP) entry assay to explore if an ACE2-binding compound could interfere with viral entry. In parallel, the compounds were tested in an ACE2 enzymatic assay. Six compounds were identified as ACE2 binders with an equilibrium dissociation constant ($K_d$) $\leq 20\ \mu$M. No inhibitory or agonistic activity was observed in the ACE2 enzymatic assay (Table 6).

All six ACE2-binding compounds were able to inhibit the PP entry into ACE2-overexpressing HEK293 cells, where the molecule with the strongest affinity to ACE2 showed the highest activity in PP entry inhibition (Table 6). Since these compounds do not bind S protein, we hypothesized that their activity should be independent of S protein sequence and, thus, active against different strains of SARS-CoV-2. Therefore, we tested them against other strains of SARS-CoV-2; compounds inhibited the entry of pseudoparticles bearing S proteins of South African B.1.351 and UK B.1.1.7 variants with the same or greater potency versus wild type (Figure 3).

## ■ DISCUSSION

A traditional QSAR modeling approach relies on the assumption that the biological activity of small molecules is correlated with their physicochemical properties or the so-called structural descriptors;[52−54] however, it does not consider the three-dimensional (3D) geometric features of the molecules. This results in an incomplete description of ligand−target interactions. Furthermore, QSAR models are also restricted to their applicability domain, i.e., the chemical space within which the models are originally trained.[55] To overcome these shortcomings, a hybrid approach was developed which combines QSAR models with pharmacophore-based screening that can retrieve ligands with structurally diverse scaffolds.

Utilization of the hybrid approach led to 4-fold improvement of the hit rate and revealed multiple novel scaffolds with activity against SARS-CoV-2. More importantly, 44 compounds experimentally confirmed as active in the CPE reduction assay did not show appreciable cytotoxicity.

Further analysis of active analogues revealed some preliminary SAR, although trends were limited due to significant structural differences within the set of analogues. This supports the hypothesis that these compounds are acting on a common target or *via* a shared mechanism to inhibit viral proliferation. Overall, the chemotypes identified showed good efficacy and potency as screening hits.

In an effort to elucidate the mechanism of action, active compounds were screened against some previously established SARS-CoV-2 targets that have been shown to mediate antiviral activity: SARS-CoV-2 $M^{pro}$ and RBD-ACE2 protein−protein interaction. None of the identified hits exhibited activity against these targets.

We identified six CPE-active compounds that are ACE2 binders and, likely as a direct result, are inhibitors of viral entry. Importantly, they were also able to inhibit the entry of pseudoparticles bearing spike protein from other variants of SARS-CoV-2 (South African B.1.351 and UK B.1.1.7) with similar or increased activity. As such, further development of these small molecules into drug candidates could provide therapeutic options less susceptible to common viral resistance mechanisms.

However, these compounds do not interrupt the RDB-ACE2 interaction. We assume allosteric binding to ACE2, as they do not inhibit ACE2 enzymatic activity; at the very least, they do not interfere with substrate binding. These inhibitors could interfere with the conformational change of S protein bound to ACE2 and/or influence the endosome environment, such as the pH decrease in the endosomal lumen, which triggers the conformational change of S protein. The compounds showed some reduced inhibitory activity, compared to SARS-CoV-2 PP, in the counterscreen experiments with PP containing the G-protein of vesicular stomatitis virus (VSV-G). VSV-G does not utilize ACE2 for host cell binding but requires low endosomal pH for a conformational change to induce membrane fusion. ACE2-overexpressing HEK293 cells were used for all PP assays. Consequently, these compounds could bind to ACE2, trap within the endosomes, and affect the VSV-G PP entry. Further experimentation is required to determine the exact mechanism by which these compounds disrupt viral entry.

This study demonstrates that our hybrid approach, combining machine learning with pharmacophore-based screening, increases the hit rate and allows for the discovery of novel scaffolds with activity against SARS-CoV-2. To date, no direct-acting antiviral small-molecule drugs have been approved by the FDA for the treatment of SARS-CoV-2. Additionally, multiple SARS-CoV-2 variants have emerged, some of which are classified as variants of concern by the CDC. Our VS approach showed a PPV of 18%, compared to a PPV of 4%, from the first experimental data set for modeling.

Among the SARS-CoV-2 inhibitors identified in this study are several novel chemotypes. Furthermore, two compounds identified have submicromolar inhibitory activity against the South African B.1.351 and UK B.1.1.7 variants. These preliminary results clearly warrant further investigation of each chemotype *via* medicinal chemistry efforts to thoroughly explore and establish the SAR for optimization of activity and also to improve upon physicochemical/ADME properties.

To accelerate further research on the finding of small molecules active against SARS-CoV-2, we provided the best-developed prediction models and modeling sets via github (https://github.com/ncats/covid19_pred) and through the NCATS Predictor website (https://predictor.ncats.io/).

## ■ ASSOCIATED CONTENT

### Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acsptsci.1c00176.

Additional information on the pharmacophore models, model performance, and the experimental results on the compounds screened in this study (PDF)

Pharmacophore_LBP-1; Pharmacophore_LBP-2 (ZIP)



https://doi.org/10.1021/acsptsci.1c00176
ACS Pharmacol. Transl. Sci. 2021, 4, 1675−1688

## AUTHOR INFORMATION


### Corresponding Author

**Alexey V. Zakharov** − *National Center for Advancing Translational Sciences (NCATS), National Institutes of Health, Rockville, Maryland 20850, United States;* orcid.org/0000-0003-2466-1711; Email: alexey.zakharov@nih.gov

### Authors

**Sankalp Jain** − *National Center for Advancing Translational Sciences (NCATS), National Institutes of Health, Rockville, Maryland 20850, United States*

**Daniel C. Talley** − *National Center for Advancing Translational Sciences (NCATS), National Institutes of Health, Rockville, Maryland 20850, United States*

**Bolormaa Baljinnyam** − *National Center for Advancing Translational Sciences (NCATS), National Institutes of Health, Rockville, Maryland 20850, United States*

**Jun Choe** − *National Center for Advancing Translational Sciences (NCATS), National Institutes of Health, Rockville, Maryland 20850, United States*

**Quinlin Hanson** − *National Center for Advancing Translational Sciences (NCATS), National Institutes of Health, Rockville, Maryland 20850, United States*

**Wei Zhu** − *National Center for Advancing Translational Sciences (NCATS), National Institutes of Health, Rockville, Maryland 20850, United States*

**Miao Xu** − *National Center for Advancing Translational Sciences (NCATS), National Institutes of Health, Rockville, Maryland 20850, United States*

**Catherine Z. Chen** − *National Center for Advancing Translational Sciences (NCATS), National Institutes of Health, Rockville, Maryland 20850, United States*

**Wei Zheng** − *National Center for Advancing Translational Sciences (NCATS), National Institutes of Health, Rockville, Maryland 20850, United States*

**Xin Hu** − *National Center for Advancing Translational Sciences (NCATS), National Institutes of Health, Rockville, Maryland 20850, United States*

**Min Shen** − *National Center for Advancing Translational Sciences (NCATS), National Institutes of Health, Rockville, Maryland 20850, United States*

**Ganesha Rai** − *National Center for Advancing Translational Sciences (NCATS), National Institutes of Health, Rockville, Maryland 20850, United States;* orcid.org/0000-0001-9763-9641

**Matthew D. Hall** − *National Center for Advancing Translational Sciences (NCATS), National Institutes of Health, Rockville, Maryland 20850, United States;* orcid.org/0000-0002-5073-442X

**Anton Simeonov** − *National Center for Advancing Translational Sciences (NCATS), National Institutes of Health, Rockville, Maryland 20850, United States*

Complete contact information is available at: https://pubs.acs.org/10.1021/acsptsci.1c00176


### Notes

The authors declare no competing financial interest.

## ACKNOWLEDGMENTS

The authors would like to thank Dac-Trung Nguyen for the technical help in setting up LigandScout on the AWS cluster and Pranav Shah for the *in vitro* ADME/physicochemical data on our most promising compounds. This research was supported by the Intramural Research Program of the National Center for Advancing Translational Sciences (NCATS), National Institutes of Health (NIH).

## ABBREVIATIONS

NCATS:National Center for Advancing Translational Sciences; FDA:Food and Drug Administration; CDC:Centers for Disease Control and Prevention; HTS:high-throughput screening; CPE:cytopathic effect; QSAR:quantitative structure−activity relationship; SAR:structure−activity relationship; PP:pseudoparticle; SB:stratified bagging; RF:random forest; VS:virtual screening; MFP:merged-features pharmacophore; SFP:shared-features pharmacophore; LBP:ligand-based pharmacophore; ROC AUC:receiver operating characteristic area under the curve; TP:true positives; TN:true negatives; FP:false positives; FN:false negatives; PPV:positive predicted value; MCC:Matthews correlation coefficient; RBD:receptor-binding domain; MST:microscale thermophoresis; HBA:hydrogen bond acceptor; HBD:hydrogen bond donor; $IC_{50}$:half-maximal inhibitory concentration; PAMPA:parallel artificial membrane permeation assay

## REFERENCES


(1) A Pneumonia Outbreak Associated with a New Coronavirus of Probable Bat Origin | Nature. https://www.nature.com/articles/s41586-020-2012-7 (accessed Aug 28, 2021).

(2) COVID-19 Data in Motion. https://coronavirus.jhu.edu/ (accessed May 5, 2021).

(3) Saul, S.; Einav, S. Old Drugs for a New Virus: Repurposed Approaches for Combating COVID-19. *ACS Infect. Dis.* **2020**, *6*, 2304−2318.

(4) Elfiky, A. A. Anti-HCV, Nucleotide Inhibitors, Repurposing against COVID-19. *Life Sci.* **2020**, *248*, No. 117477.

(5) Elfiky, E.; Ibrahim, N. S. *Anti-SARS and Anti-HCV Drugs Repurposing against the Papain-like Protease of the Newly Emerged Coronavirus (2019-NCoV)*; Research Square, 2020.

(6) De Savi, C.; Hughes, D. L.; Kvaerno, L. Quest for a COVID-19 Cure by Repurposing Small-Molecule Drugs: Mechanism of Action, Clinical Development, Synthesis at Scale, and Outlook for Supply. *Org. Process Res. Dev.* **2020**, *24*, 940−976.

(7) CDC. SARS-CoV-2 Variant Classifications and Definitions. https://www.cdc.gov/coronavirus/2019-ncov/cases-updates/variant-surveillance/variant-info.html (accessed May 5, 2021).

(8) Variant Therapeutic Data Summary. https://opendata.ncats.nih.gov/variant/summary (accessed May 5, 2021).

(9) Brimacombe, K. R.; Zhao, T.; Eastman, R. T.; Hu, X.; Wang, K.; Backus, M.; Baljinnyam, B.; Chen, C. Z.; Chen, L.; Eicher, T.; Ferrer, M.; Fu, Y.; Gorshkov, K.; Guo, H.; Hanson, Q. M.; Itkin, Z.; Kales, S. C.; Klumpp-Thomas, C.; Lee, E. M.; Michael, S.; Mierzwa, T.; Patt, A.; Pradhan, M.; Renn, A.; Shinn, P.; Shrimp, J. H.; Viraktamath, A.; Wilson, K. M.; Xu, M.; Zakharov, A. V.; Zhu, W.; Zheng, W.; Simeonov, A.; Mathé, E. A.; Lo, D. C.; Hall, M. D.; Shen, M. An OpenData Portal to Share COVID-19 Drug Repurposing Data in Real Time. *bioRxiv* **2020**, No. 135046.

(10) Chen, C. Z.; Shinn, P.; Itkin, Z.; Eastman, R. T.; Bostwick, R.; Rasmussen, L.; Huang, R.; Shen, M.; Hu, X.; Wilson, K. M.; Brooks, B. M.; Guo, H.; Zhao, T.; Klump-Thomas, C.; Simeonov, A.; Michael, S. G.; Lo, D. C.; Hall, M. D.; Zheng, W. Drug Repurposing Screen for Compounds Inhibiting the Cytopathic Effect of SARS-CoV-2. *Front. Pharmacol.* **2021**, *11*, No. 592737.

(11) Oprea, T. I. Virtual Screening in Lead Discovery: A Viewpoint. *Molecules* **2002**, *7*, 51−62.

(12) Good, A. 4.19—Virtual Screening. In *Comprehensive Medicinal Chemistry II*; Taylor, J. B.; Triggle, D. J., Eds.; Elsevier: Oxford, 2007; pp 459−494. DOI: 10.1016/B0-08-045044-X/00262-5.
(13) 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. *Mol. Inf.* **2020**, *39*, No. e2000028.
(14) Zhang, H.; Saravanan, K. M.; Yang, Y.; Hossain, M. T.; Li, J.; Ren, X.; Pan, Y.; Wei, Y. Deep Learning Based Drug Screening for Novel Coronavirus 2019-NCov. *Interdiscip. Sci.: Comput. Life Sci.* **2020**, *12*, 368−376.
(15) Berry, M.; Fielding, B. C.; Gamieldien, J. Potential Broad Spectrum Inhibitors of the Coronavirus 3CLpro: A Virtual Screening and Structure-Based Drug Design Study. *Viruses* **2015**, *7*, 6642−6660.
(16) Abuhammad, A.; Al-Aqtash, R. A.; Anson, B. J.; Mesecar, A. D.; Taha, M. O. Computational Modeling of the Bat HKU4 Coronavirus 3CLpro Inhibitors as a Tool for the Development of Antivirals against the Emerging Middle East Respiratory Syndrome (MERS) Coronavirus. *J. Mol. Recognit.* **2017**, *30*, No. e2644.
(17) Xu, C.; Ke, Z.; Liu, C.; Wang, Z.; Liu, D.; Zhang, L.; Wang, J.; He, W.; Xu, Z.; Li, Y.; Yang, Y.; Huang, Z.; Lv, P.; Wang, X.; Han, D.; Li, Y.; Qiao, N.; Liu, B. Systemic In Silico Screening in Drug Discovery for Coronavirus Disease (COVID-19) with an Online Interactive Web Server. *J. Chem. Inf. Model.* **2020**, *60*, 5735−5745.
(18) Singh, N.; Villoutreix, B. O. Resources and Computational Strategies to Advance Small Molecule SARS-CoV-2 Discovery: Lessons from the Pandemic and Preparing for Future Health Crises. *Comput. Struct. Biotechnol. J.* **2021**, *19*, 2537−2548.
(19) Alves, V. M.; Bobrowski, T.; Melo-Filho, C. C.; Korn, D.; Auerbach, S.; Schmitt, C.; Muratov, E. N.; Tropsha, A. QSAR Modeling of SARS-CoV Mpro Inhibitors Identifies Sufugolix, Cenicriviroc, Proglumetacin, and Other Drugs as Candidates for Repurposing against SARS-CoV-2. *Mol. Inf.* **2021**, *40*, No. 2000113.
(20) Muratov, E. N.; Amaro, R.; Andrade, C. H.; Brown, N.; Ekins, S.; Fourches, D.; Isayev, O.; Kozakov, D.; Medina-Franco, J. L.; Merz, K. M.; Oprea, T. I.; Poroikov, V.; Schneider, G.; Todd, M. H.; Varnek, A.; Winkler, D. A.; Zakharov, A. V.; Cherkasov, A.; Tropsha, A. A Critical Overview of Computational Approaches Employed for COVID-19 Drug Discovery. *Chem. Soc. Rev.* **2021**, *50*, 9121−9151.
(21) 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*, 1−18.
(22) Wang, Q.; Zhao, Y.; Chen, X.; Hong, A. Virtual Screening of Approved Clinic Drugs with Main Protease (3CLpro) Reveals Potential Inhibitory Effects on SARS-CoV-2. *J. Biomol. Struct. Dyn.* **2020**, 1−11.
(23) Bobrowski, T.; Chen, L.; Eastman, R. T.; Itkin, Z.; Shinn, P.; Chen, C. Z.; Guo, H.; Zheng, W.; Michael, S.; Simeonov, A.; Hall, M. D.; Zakharov, A. V.; Muratov, E. N. Synergistic and Antagonistic Drug Combinations against SARS-CoV-2. *Mol. Ther.* **2021**, *29*, 873−885.
(24) Inglese, J.; Auld, D. S.; Jadhav, A.; Johnson, R. L.; Simeonov, A.; Yasgar, A.; Zheng, W.; Austin, C. P. Quantitative High-Throughput Screening: A Titration-Based Approach That Efficiently Identifies Biological Activities in Large Chemical Libraries. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 11473−11478.
(25) Fourches, D.; Muratov, E.; Tropsha, A. Trust, but Verify: On the Importance of Chemical Structure Curation in Cheminformatics and QSAR Modeling Research. *J. Chem. Inf. Model.* **2010**, *50*, 1189−1204.
(26) Fourches, D.; Muratov, E.; Tropsha, A. Trust, but Verify II: A Practical Guide to Chemogenomics Data Curation. *J. Chem. Inf. Model.* **2016**, *56*, 1243−1252.
(27) Fourches, D.; Muratov, E.; Tropsha, A. Curation of Chemogenomics Data. *Nat. Chem. Biol.* **2015**, *11*, 535−535.
(28) *Molecular Operating Environment (MOE)*, 2019.01; Chemical Computing Group ULC, 1010 Sherbooke St. West, Suite #910, Montreal, QC, Canada, 2019.
(29) Wang, Y.; Jadhav, A.; Southal, N.; Huang, R.; Nguyen, D.-T. A Grid Algorithm for High Throughput Fitting of Dose-Response Curve Data. *Curr. Chem. Genomics* **2010**, *4*, 57−66.
(30) Jain, S.; Kotsampasakou, E.; Ecker, G. F. Comparing the Performance of Meta-Classifiers-a Case Study on Selected Imbalanced Data Sets Relevant for Prediction of Liver Toxicity. *J. Comput.-Aided Mol. Des.* **2018**, *32*, 583−590.
(31) Breiman, L. Random Forests. *Mach. Learn.* **2001**, *45*, 5−32.
(32) Oshiro, T. M.; Perez, P. S.; Baranauskas, J. A. How Many Trees in a Random Forest?. In *Machine Learning and Data Mining in Pattern Recognition; Lecture Notes in Computer Science*; Springer: Berlin, Heidelberg, 2012; pp 154−168. DOI: 10.1007/978-3-642-31537-4_13.
(33) Zakharov, A. V.; Varlamova, E. V.; Lagunin, A. A.; Dmitriev, A. V.; Muratov, E. N.; Fourches, D.; Kuz'min, V. E.; Poroikov, V. V.; Tropsha, A.; Nicklaus, M. C. QSAR Modeling and Prediction of Drug−Drug Interactions. *Mol. Pharmaceutics* **2016**, *13*, 545−556.
(34) Jain, S.; Siramshetty, V. B.; Alves, V. M.; Muratov, E. N.; Kleinstreuer, N.; Tropsha, A.; Nicklaus, M. C.; Simeonov, A.; Zakharov, A. V. Large-Scale Modeling of Multispecies Acute Toxicity End Points Using Consensus of Multitask Deep Learning Methods. *J. Chem. Inf. Model.* **2021**, DOI: 10.1021/acs.jcim.0c01164.
(35) Alamri, M. A.; Alamri, M. A. Pharmacophore and Docking-Based Sequential Virtual Screening for the Identification of Novel Sigma 1 Receptor Ligands. *Bioinformation* **2019**, *15*, 586−595.
(36) Vittorio, S.; Seidel, T.; Germanò, M. P.; Gitto, R.; Ielo, L.; Garon, A.; Rapisarda, A.; Pace, V.; Langer, T.; Luca, L. D. A Combination of Pharmacophore and Docking-Based Virtual Screening to Discover New Tyrosinase Inhibitors. *Mol. Inf.* **2020**, *39*, No. 1900054.
(37) Friedrich, N.-O.; de Bruyn Kops, C.; Flachsenberg, F.; Sommer, K.; Rarey, M.; Kirchmair, J. Benchmarking Commercial Conformer Ensemble Generators. *J. Chem. Inf. Model.* **2017**, *57*, 2719−2728.
(38) Wolber, G.; Dornhofer, A. A.; Langer, T. Efficient Overlay of Small Organic Molecules Using 3D Pharmacophores. *J. Comput.-Aided Mol. Des.* **2006**, *20*, 773−788.
(39) Langer, T.; Wolber, G. Pharmacophore Definition and 3D Searches. *Drug Discovery Today: Technol.* **2004**, *1*, 203−207.
(40) Berthold, M. R.; Cebron, N.; Dill, F.; Gabriel, T. R.; Kötter, T.; Meinl, T.; Ohl, P.; Thiel, K.; Wiswedel, B. KNIME—the Konstanz Information Miner: Version 2.0 and Beyond. *SIGKDD Explor. Newsl.* **2009**, *11*, 26−31.
(41) Huang, R.; Xu, M.; Zhu, H.; Chen, C. Z.; Zhu, W.; Lee, E. M.; He, S.; Zhang, L.; Zhao, J.; Shamim, K.; Bougie, D.; Huang, W.; Xia, M.; Hall, M. D.; Lo, D.; Simeonov, A.; Austin, C. P.; Qiu, X.; Tang, H.; Zheng, W. Biological Activity-Based Modeling Identifies Antiviral Leads against SARS-CoV-2. *Nat. Biotechnol.* **2021**, *39*, 747−753.
(42) Brown, J. B. Classifiers and Their Metrics Quantified. *Mol. Inf.* **2018**, *37*, No. 1700127.
(43) Zhu, W.; Xu, M.; Chen, C. Z.; Guo, H.; Shen, M.; Hu, X.; Shinn, P.; Klumpp-Thomas, C.; Michael, S. G.; Zheng, W. Identification of SARS-CoV-2 3CL Protease Inhibitors by a Quantitative High-Throughput Screening. *ACS Pharmacol. Transl. Sci.* **2020**, *3*, 1008−1016.
(44) Hanson, Q. M.; Wilson, K. M.; Shen, M.; Itkin, Z.; Eastman, R. T.; Shinn, P.; Hall, M. D. Targeting ACE2−RBD Interaction as a Platform for COVID-19 Therapeutics: Development and Drug-Repurposing Screen of an AlphaLISA Proximity Assay. *ACS Pharmacol. Transl. Sci.* **2020**, *3*, 1352−1360.
(45) Troger, F.; Delp, J.; Funke, M.; van der Stel, W.; Colas, C.; Leist, M.; van de Water, B.; Ecker, G. F. Identification of Mitochondrial Toxicants by Combined in Silico and in Vitro Studies—A Structure-Based View on the Adverse Outcome Pathway. *Comput. Toxicol.* **2020**, *14*, No. 100123.
(46) Jain, S.; Grandits, M.; Richter, L.; Ecker, G. F. Structure Based Classification for Bile Salt Export Pump (BSEP) Inhibitors Using Comparative Structural Modeling of Human BSEP. *J. Comput.-Aided Mol. Des.* **2017**, *31*, 507−521.

(47) Siramshetty, V. B.; Shah, P.; Kerns, E.; Nguyen, K.; Yu, K. R.; Kabir, M.; Williams, J.; Neyra, J.; Southall, N.; Nguyên, Đ.-T.; Xu, X. Retrospective Assessment of Rat Liver Microsomal Stability at NCATS: Data and QSAR Models. *Sci. Rep.* **2020**, *10*, No. 20713.
(48) Sun, H.; Nguyen, K.; Kerns, E.; Yan, Z.; Yu, K. R.; Shah, P.; Jadhav, A.; Xu, X. Highly Predictive and Interpretable Models for PAMPA Permeability. *Bioorg. Med. Chem.* **2017**, *25*, 1266−1276.
(49) Siramshetty, V.; Williams, J.; Nguyên, Đ.-T.; Neyra, J.; Southall, N.; Mathé, E.; Xu, X.; Shah, P. Validating ADME QSAR Models Using Marketed Drugs. *SLAS Discovery* **2021**, No. 24725552211017520.
(50) Sun, H.; Shah, P.; Nguyen, K.; Yu, K. R.; Kerns, E.; Kabir, M.; Wang, Y.; Xu, X. Predictive Models of Aqueous Solubility of Organic Compounds Built on A Large Dataset of High Integrity. *Bioorg. Med. Chem.* **2019**, *27*, 3110−3114.
(51) Ullrich, S.; Nitsche, C. The SARS-CoV-2 Main Protease as Drug Target. *Bioorg. Med. Chem. Lett.* **2020**, *30*, No. 127377.
(52) Recent Advances in QSAR Studies—Methods and Applications | Tomasz Puzyn | Springer. https://www.springer.com/gp/book/9781402097829 (accessed Oct 9, 2020).
(53) Kuz'min, V. E.; Artemenko, A. G.; Muratov, E. N.; Polischuk, P. G.; Ognichenko, L. N.; Liahovsky, A. V.; Hromov, A. I.; Varlamova, E. V. Virtual Screening and Molecular Design Based on Hierarchical Qsar Technology. In *Recent Advances in QSAR Studies: Methods and Applications*; Puzyn, T.; Leszczynski, J.; Cronin, M. T., Eds.; Challenges and Advances in Computational Chemistry and Physics; Springer: Netherlands, Dordrecht, 2010; pp 127−176. DOI: 10.1007/978-1-4020-9783-6_5.
(54) Alves, V. M.; Golbraikh, A.; Capuzzi, S. J.; Liu, K.; Lam, W. I.; Korn, D. R.; Pozefsky, D.; Andrade, C. H.; Muratov, E. N.; Tropsha, A. Multi-Descriptor Read Across (MuDRA): A Simple and Transparent Approach for Developing Accurate Quantitative Structure−Activity Relationship Models. *J. Chem. Inf. Model.* **2018**, *58*, 1214−1223.
(55) Roy, K.; Kar, S.; Das, R. N. Validation of QSAR Models. In *Understanding the Basics of QSAR for Applications in Pharmaceutical Sciences and Risk Assessment*; Roy, K.; Kar, S.; Das, R. N., Eds.; Academic Press: Boston, 2015; Chapter 7, pp 231−289. DOI: 10.1016/B978-0-12-801505-6.00007-7.


https://doi.org/10.1021/acsptsci.1c00176
*ACS Pharmacol. Transl. Sci.* 2021, 4, 1675−1688