

Allosteric Binders of ACE2 Are Promising Anti-SARS-CoV-2 Agents

Joshua E. Hochuli,^{||} Sankalp Jain,^{||} Cleber Melo-Filho, Zoe L. Sessions, Tesia Bobrowski, Jun Choe, Johnny Zheng, Richard Eastman, Daniel C. Talley, Ganesh Rai, Anton Simeonov, Alexander Tropsha,* Eugene N. Muratov,* Bolormaa Baljinnyam,* and Alexey V. Zakharov*Cite This: *ACS Pharmacol. Transl. Sci.* 2022, 5, 468–478

Read Online

ACCESS |



Metrics & More



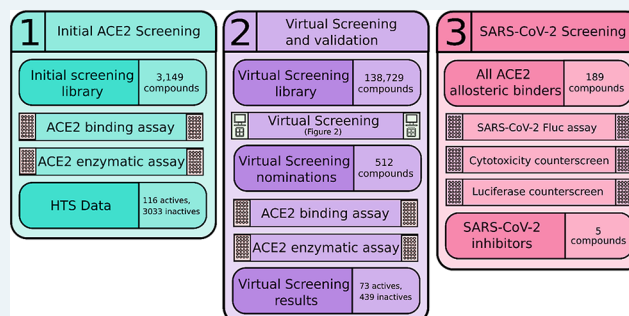
Article Recommendations



Supporting Information

ABSTRACT: The COVID-19 pandemic has had enormous health, economic, and social consequences. Vaccines have been successful in reducing rates of infection and hospitalization, but there is still a need for acute treatment of the disease. We investigate whether compounds that bind the human angiotensin-converting enzyme 2 (ACE2) protein can decrease SARS-CoV-2 replication without impacting ACE2's natural enzymatic function. Initial screening of a diversity library resulted in hit compounds active in an ACE2-binding assay, which showed little inhibition of ACE2 enzymatic activity (116 actives, success rate ~4%), suggesting they were allosteric binders. Subsequent application of *in silico* techniques boosted success rates to ~14% and resulted in 73 novel confirmed ACE2 binders with K_d values as low as 6 nM. A subsequent SARS-CoV-2 assay revealed that five of these compounds inhibit the viral life cycle in human cells. Further effort is required to completely elucidate the antiviral mechanism of these ACE2-binders, but they present a valuable starting point for both the development of acute treatments for COVID-19 and research into the host-directed therapy.

KEYWORDS: COVID-19, SARS-CoV-2, angiotensin-converting enzyme 2, ACE2, allosteric binder



The radical consequences of coronavirus disease 2019 (COVID-19) and the virus that causes it, severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), since its advent in late 2019, are evident in many facets of daily life. Since that time, many economies around the world have entered recessions, unemployment rates have dramatically increased, and the healthcare infrastructure in many countries is being pushed to its limits.^{1–3} More significantly, COVID-19 has killed almost 6 million people worldwide and has infected well over 400 million as of February 2022.⁴

Vaccine deployment has been a general success in reducing rates of infection⁵ and especially rates of hospitalization,^{5,6} though less so in viral transmission from a vaccinated individual.⁷ Importantly, there are breakthrough cases where vaccinated individuals contract the virus, which can still result in severe symptoms or death. Variants of the virus are actively developing and spreading;⁸ future viral mutants could evade the immune system detection established by a previous infection or vaccine. The tremendous effort the scientific community has dedicated to developing a vaccine is incredibly important and has truly altered the treatment landscape; however, future research into new and novel therapies for acute COVID-19 to alleviate the severity of infection is still required.

It is imperative that efforts be dedicated to the development of small-molecule inhibitors which target various parts of the viral life cycle, by direct action upon either viral proteins or host

factors required for viral replication. These therapies would not only be immediately applicable to COVID-19 but also be useful in any future coronavirus epidemics.⁹ There are currently several repurposed drugs granted emergency use authorization for the treatment of COVID-19. Remdesivir, a nucleoside analogue that inhibits SARS-CoV-2 RNA polymerase II, has taken a premier role in treatment since early 2020. Other drugs that are currently being used to treat COVID-19 cases are the rheumatoid arthritis drug baricitinib and the corticosteroid dexamethasone, which act to reduce the inflammation associated with severe infection.¹⁰ There are also at least two newly approved drugs for treatment of COVID-19: Paxlovid¹¹ and Molnupiravir,¹² but true efficacy in practice remains to be seen.

The National Center for Advancing Translational Sciences (NCATS) has put forth a significant effort to support and promote the development of anti-SARS-CoV-2 treatments. The NCATS OpenData Portal for COVID-19 drug repurposing¹³ allows researchers and public health officials to expedite the

Received: March 17, 2022

Published: June 22, 2022



development of SARS-CoV-2 interventions through open data sharing and analysis tools and also to prioritize antiviral discovery for further development in treating COVID-19. Furthermore, a number of researchers have been performing large-scale virtual screening (VS) of NCATS in-house compound libraries with the aim of identifying chemotypes with antiviral activity and limited host cell cytotoxicity.^{14,15} Various high-throughput screening (HTS) assays have been developed to expedite the process.^{16–18}

Although many small-molecule therapies for COVID-19 have been identified which target viral proteins, there is interest in discovering small-molecule modulators of host proteins integral to the SARS-CoV-2 life cycle. Drugs targeting the host cell machinery reduce the chance of viral resistance development and also have the potential to be applied to other viruses/indications which harness or utilize the same host targets.¹⁹ One such protein of particular interest serves as the primary entry receptor for SARS-CoV-2, angiotensin-converting enzyme 2 (ACE2).²⁰

ACE and ACE2 are key enzymes in the renin-angiotensin-aldosterone system (RAAS), which is implicated in renal, pulmonary, immune, and cardiovascular functions.^{21–23} Angiotensin I is cleaved into angiotensin II (Ang II) by ACE, and ACE2 is responsible for the conversion of angiotensin II to Ang1–7.^{24–27} Ang1–7 mediates vasodilatory, vasoprotective, anti-fibrotic, and anti-inflammatory effects through its binding to the G-protein coupled receptor Mas. Reduction of angiotensin II levels and the subsequent elevation of Ang1–7 by ACE2 enzymatic activity exerts beneficial actions not only on the heart and lungs but also on the brain, gastrointestinal system, and bone marrow.^{24,25}

Due to the complex role of the RAAS system, inhibition of ACE2 enzymatic activity will lead to increased inflammation, fibrosis, oxidative stress, and vasoconstriction.²⁵ ACE2 enzymatic inhibition, via either the SARS-CoV-2 spike protein interaction or small-molecule orthosteric binding,²⁶ results in downstream reduction of Ang1–7; this hinders anti-inflammatory compensation as well as allows for the over production and expression of Ang II, inciting further inflammation and fibrosis. As such, inhibition of the SARS-CoV-2 spike protein interaction with ACE2, without impacting ACE2 enzymatic activity, should be the goal for²⁹ COVID-19 therapeutic intervention.²⁷

Harnessing and properly utilizing computational tools^{28,29} have enabled many research groups to identify potential COVID-19 therapeutics, as well as to prevent the binding of SARS-CoV-2 to ACE2.³⁰ In some cases, molecular modeling has been used to virtually screen compounds³¹ and also to predict synergistic treatment³² that may aid in the treatment of COVID-19. We report here a hybrid discovery approach, where computational models are used in conjunction with high-throughput screening to establish a vast dataset of small molecules with desired binding properties (Figure 1).

The aims of this study are (i) to identify allosteric binders of ACE2 without enzyme inhibitory activity and (ii) to discover small molecules which allosterically bind ACE2 and are also active against SARS-CoV-2 in cell-based assays. To do so, we utilized assays for ACE2 binding and enzymatic activity to screen a large dataset of small molecules and also to develop initial data for *in silico* modeling. Independent modeling efforts were undertaken by both NCATS and University of North Carolina (UNC) teams to nominate a diverse set of molecules predicted to bind to ACE2 without significant interruption of enzymatic activity. Computationally nominated compounds

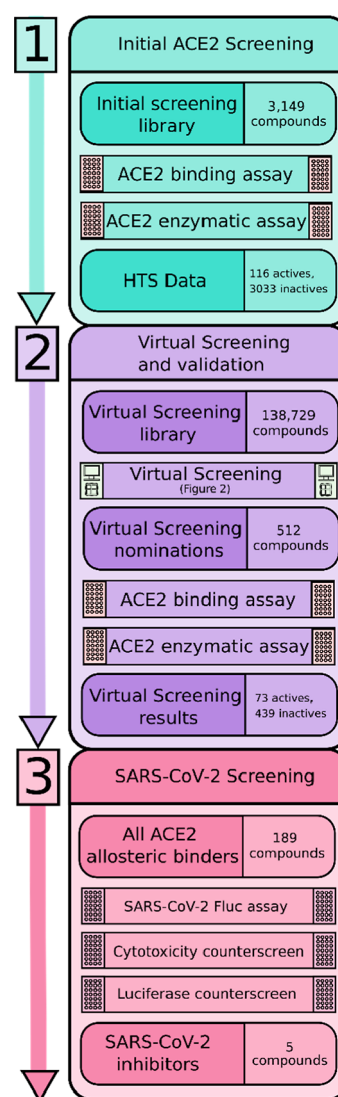


Figure 1. Overall study design.

were experimentally screened, resulting in 73 actives (out of 512 nominations; hit rate ~14%). Confirmed hits were subsequently screened in a SARS-CoV-2 replication assay in human cells. Appropriate counterscreens for cytotoxicity and luciferase assay interference were applied to all hits.

In total, five allosteric ACE2 binders with K_d ranging from 0.09 to 3 μM (inactive as ACE2 inhibitors) were identified as SARS-CoV-2 inhibitors with IC_{50} values ranging from 14 to 25 μM . Further efforts are required to elucidate the mechanism of action as SARS-CoV-2 inhibitors; however, hit compounds will serve as useful starting points for additional investigation via structure–activity relationship studies and medicinal chemistry optimization.

RESULTS AND DISCUSSION

Initial Screening. To identify small-molecule ACE2 binders which do not interfere with ACE2 enzymatic activity, we utilized microscale thermophoresis (MST) in combination with an enzymatic assay. Extended assay descriptions are included in the [Supporting Information](#).

Recombinant polyhistidine-tagged (His-tag) extracellular domain of ACE2 and the fluorogenic substrate MCA [(7-methoxycoumarin-4-acetic acid)—Ala-Pro-Lys (Dnp)-OH]

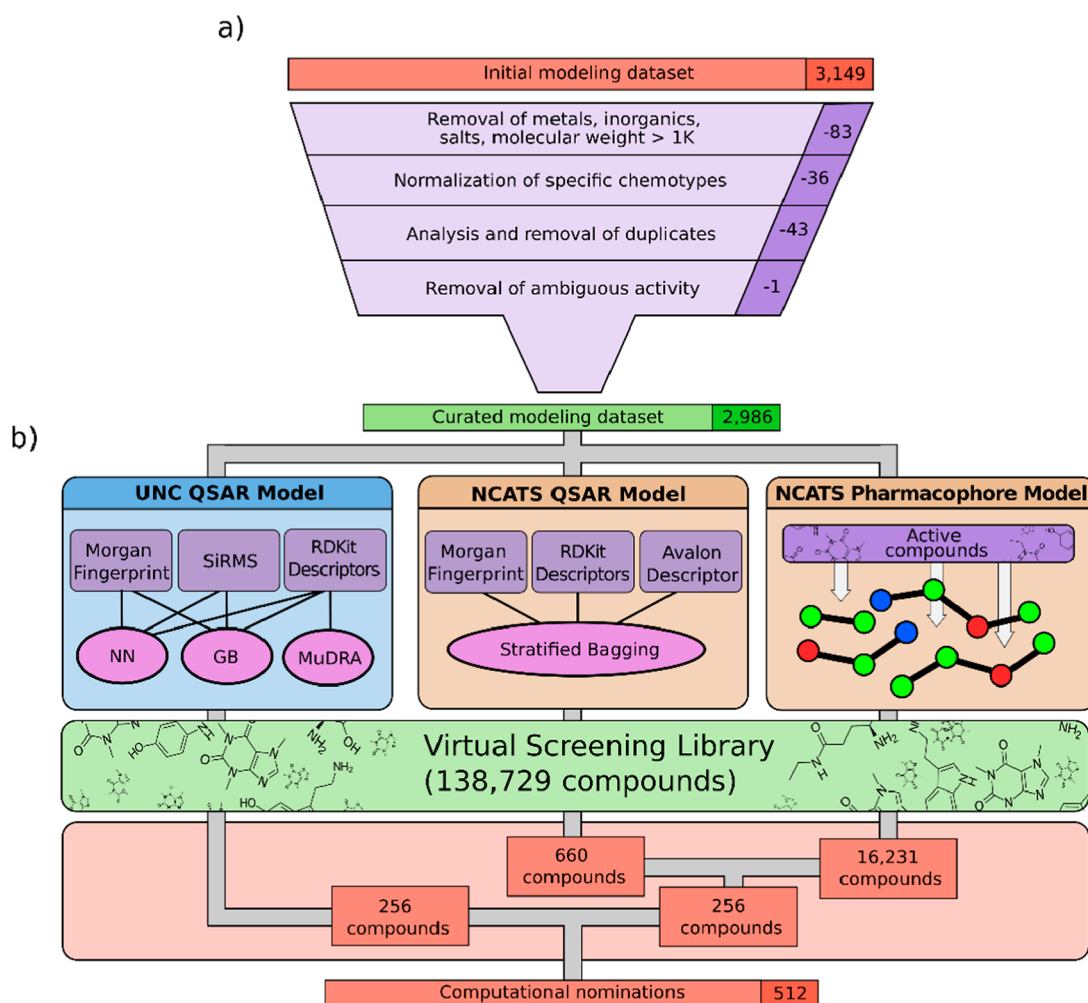


Figure 2. Computational workflow. (a) Curation protocol. UNC curation protocol^{34–36} shown and NCATS protocol is similar with minor deviations. (b) Modeling workflow. In total, three model types were used to screen compounds to produce 512 nominations for further experimental validation.

was used for the enzymatic assay. For MST, the recombinant ACE2 protein was labeled with a His-tag-specific fluorophore to monitor any binding events. MLN-4760, a known ACE2 inhibitor, was used to test and optimize the conditions of the enzymatic and MST assays. In the enzymatic assay, MLN-4760 showed a dose-dependent inhibition of ACE2 activity with a half-maximum inhibitory concentration (IC_{50}) of 1.50 nM (Figure S1A). The binding affinity of MLN-4760 to ACE2 measured by MST was 702 nM (Figure S1B). Consequently, MLN-4760 was used as a positive control in both assays to screen the compounds of the NCATS Pharmaceutical Collection and the anti-infective library.

A total of 3149 compounds were screened in the ACE2 enzymatic assay in a five-point dilution series with final compound concentrations ranging from 20 nM to 62 μ M in 1536-well-plate format. The Z' -factor for the assay had an average value of 0.72 ± 0.04 and a signal-to-background ratio of 13.82 ± 2.73 , indicating a robust assay performance. Compounds which showed ACE2 enzyme-modulating activity in the primary screen were cherry-picked and retested in a 11-point dilution series ranging from 0.4 nM to 123.5 μ M in duplicates. These compounds were tested in a counterscreen to check whether they interfere with the reporter fluorophore signal as well. Out of the 128 cherry-picked compounds, 112

compounds were confirmed, where 110 of them were inhibitors of the ACE2 enzyme and two activators.

The same small-molecule libraries were screened for ACE2 binding by MST in 96-capillary format at a single dose with final concentrations ranging from 39 to 392 μ M depending on the highest available concentration of the compounds in the library (based on solubility). Out of the 492 compounds (14.36%, Figure S1C) identified as potential hits, 405 unique compounds were selected for the affinity screening at seven-point dilution series. The compounds were counterscreened with a fluorophore-labeled His-peptide to identify compounds interacting with the His-tag of the recombinant ACE2, instead of the target protein. The hit compounds identified from the affinity screen with dissociation constant (K_d) values less than 30 μ M and not active in the ACE2 enzymatic assay were retested in a second round of the MST experiment. 116 compounds were validated as ACE2 putative allosteric binders.

Virtual Screening. UNC Modeling. Model validation statistics is reported in Table S2. Models in bold were selected for final predictive use as a consensus ensemble. According to standard metrics,³³ model performance was fairly poor (correct classification rates close to 0.5). However, we placed emphasis on the positive predictive value (PPV), which reflects the likelihood that any nominated molecule will be truly active in the experimental screening. We sought an increase in PPV over the

selection strategy for the initial screening set (3.7% active rate). Our consensus model performance predicted about a three fold increase in the rate of active compounds from less than four to slightly more than 10%. Compounds predicted as active by the highest number of models within the ensemble were selected for experimental validation.

NCATS Modeling. Detailed model statistics on the training set and the test set are provided in Table S3. All four models were selected for a final consensus ensemble, which showed AUC of ca. 65%. For the ligand-based pharmacophore (LBP) modeling, we used the 110 active compounds; clustering based on pharmacophore-based similarity (cluster distances of 0.4, 0.6, 0.7, and 0.8), followed by generation of 89 pharmacophore hypotheses: merged-feature pharmacophore (MFP) and shared-feature pharmacophore (SFP). Taking the computational constraints into account, 24 pharmacophore models that hit the majority (>20%) of active versus inactive were selected for VS (Table S1). In general, the pharmacophoric sites such as hydrogen bond acceptor, hydrogen bond donor, and aromatic ring were prudently characterized. The complete collection of 138,729 compounds was virtually screened by both stratified bagging (SB) and LBP consensus models. Out of 256 compounds selected for experimental evaluation, 58 compounds were selected by both SB and LBP models, 58 compounds were picked using the SB approach, and remaining 140 compounds were selected from our LBP models (Figure 2).

Postmodeling Screening and Follow-Up Experiments.

As described above, computational models were used to nominate a total of 512 compounds for experimental testing. Within the top 256 compounds, chosen from each institute, there were overlapping 11 compounds. Thus, we added additional 11 compounds from the UNC list. Then, all the molecules were tested for ACE2 binding by MST at a single concentration. Out of the 512 compounds, 130 (25.39%) were identified as binders and nine—as inconclusive (Figure 3). Next,

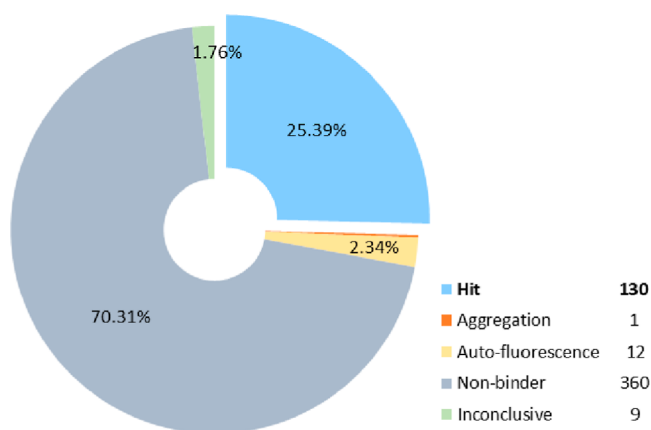


Figure 3. Experimental testing of the predicted molecules for ACE2 binding by MST at a single dose. Out of the tested 512 compounds, 130 (25.39%) were identified as hits, 360 as nonbinders, and 9 inconclusive. Twelve compounds were auto-fluorescent, and one compound caused aggregation.

these 139 potential hit compounds were measured by MST in dose–response at seven-point dilution series for ACE2 binding and in His-peptide counterscreen to test the binding specificity. These compounds were tested in the ACE2 enzymatic assay as well. Seventy five compounds were identified as ACE2 binders

with K_d values ranging from 6 nM to 562 μ M, where only two of them had moderate enzyme inhibitory activity.

To determine whether putative allosteric binders of ACE2 can affect SARS-CoV-2 infection, compounds with K_d below 10 μ M and showing no enzymatic inhibition were tested in the live SARS-CoV-2 Fluc assay (see Materials and Methods for details). The assay indirectly monitors the ability of compounds to inhibit viral replication and infection through various molecular mechanisms, including direct inhibition of viral entry or enzymatic processes, as well as acting on host pathways that modulate viral replication. The compounds were tested in the corresponding counterscreens for cytotoxicity and luciferase inhibitory activity.

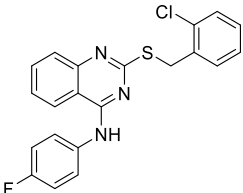
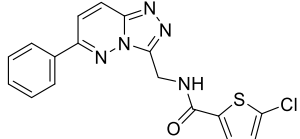
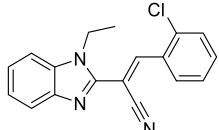
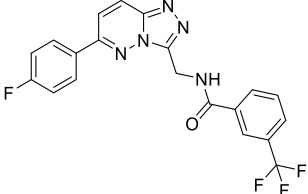
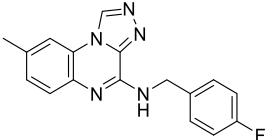
Overall, five compounds showed a SARS-CoV-2-inhibiting activity with an efficacy greater than 70% and half-maximal inhibitory concentration (IC_{50}) values of 14–25 μ M (Table 1). Of these compounds, only compound 1 showed a cytotoxic effect with $IC_{50} = 24.8 \mu$ M (Figure S2A), and compounds 3 and 5 had moderate luciferase-inhibiting activity (Figure S2B). The remaining compounds were inactive in the counterscreens and can be considered as true SARS-CoV-2 inhibitors. All five compounds were run through the ADMETlab³⁷ and PASS Online³⁸ web tools to predict various properties, which are reported in Tables S4 and S5, respectively.

Docking studies have been extensively employed in modern pharmaceutical sciences due to their ability to predict the conformation of small-molecule ligands interacting with the target-binding site. In order to understand the structural basis for allosteric inhibition, we docked our most potent ligand (NCGC00138760-01, $K_d = 0.092$) into three putative allosteric binding sites. The docking score (Glide XP score) for the three proposed binding sites led to the following results: allosteric site 1 (−5.691) < allosteric site 2 (−4.719) < allosteric site 3 (−3.898). Thus, allosteric binding site 1 is more preferred in comparison to sites 2 and 3.

As it can be seen in Figure 4, compound NCGC00138760-01 fits well in the hydrophobic binding pocket 1, showing H-bond interaction with ILE291, aromatic interaction with ILE29, ALA 431, ASP 367, and ASN 290, and pi-cation integration with LYS441. Further evidence could be derived from binding affinity score predictions obtained from LigandScout 4.4,^{39,40} which provided a score of −25.86 for allosteric binding site 1 versus −14.63 and −22.85 for allosteric binding sites 2 and 3, respectively. The proposed binding hypothesis can be further used to explore the structure–activity relationships of discovered chemotypes.

We used an MST assay (to measure binding to ACE2) and an enzymatic assay (to assess ACE2 activity) to discover putative allosteric binders of ACE2; it was presumed that a compound that shows strong binding to ACE2 with little to no inhibition of enzymatic activity is binding allosterically. Using the results of the initial binding and enzymatic activity assays (3,246 compounds), we developed QSAR models which were used to nominate a set of 512 virtual hits. Experimental validation of these compounds demonstrated that ca. 14% of them were allosteric ACE2 binders. This was a significant enrichment of hit rate over the prevalence of allosteric binders in the original assay, which was closer to 4%. All confirmed allosteric ACE2 binders are reported in Table S6. Interestingly, several ACE2-binding chemotypes were identified with >3 structural analogues included in the dataset as active. These hit chemotypes could serve as leads for additional investigation, where allosteric binding of ACE2 may be of interest.

Table 1. Compounds with the Capability to Bind to ACE2 and Inhibit SARS-CoV-2

ID	Structure	ACE2 binding, K_d (μM) ^a	IC_{50} (μM) ^b	Efficacy (%) ^c
1		0.092± 0.0001	15.67	140.19
2		0.382± 0.00006	17.58	124.61
3		0.623± 0.0001	13.96	177.42
4		0.727± 0.00008	19.72	100.73
5		3.31± 0.00013	24.83	70.61

^aACE2-binding affinity ($K_d \pm$ standard error) measured by MST. ^b IC_{50} : half-maximal inhibitory concentration values obtained in the SARS-CoV-2 Fluc assay, measured in triplicate. ^cEfficacy: maximum inhibitory effect observed in the SARS-CoV-2 Fluc assay.

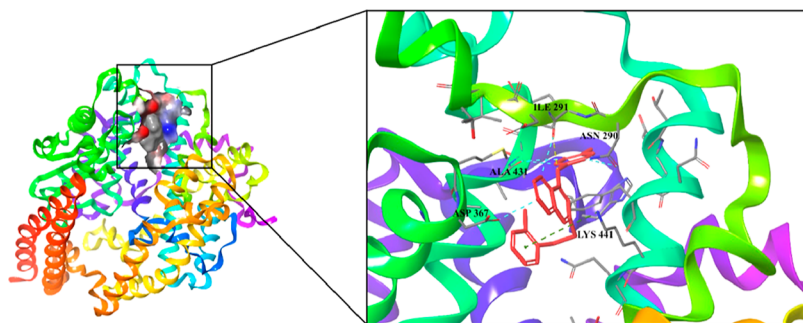


Figure 4. Surface model showing docking pose NCGC00138760-01 in the allosteric binding site 1 along with its interacting residues.

This set of compounds, along with hits from the original assay, were then subjected to a further experimental analysis to assess potential anti-SARS-CoV-2 activity. Thus, an assay for viral replication of a modified SARS-CoV-2 virus in human cells over-expressing ACE2 was applied to the set of allosteric ACE2 binders, along with appropriate counterscreens to rule out assay

artifacts. In total, five hit compounds reported here have significant binding to ACE2, no inhibition of ACE2 enzymatic activity, and significantly reduce SARS-CoV-2 replication in human cells. Thus, it is hypothesized that these compounds inhibit viral replication via ACE2 binding, thus preventing viral entry into the host cell. Further efforts are required to verify this

mechanism of action. To the best of our knowledge, this is the first successful attempt to discover antiviral agents against SARS-CoV-2 via ACE2 allosteric binding mechanism and we encourage further exploration of these hits for use in treatment of COVID-19.

MATERIALS AND METHODS

Experimental Section. MST Assay. Binding of the compounds to recombinant human ACE2 (Sino Biological, Cat #: 10108-H08H) was evaluated by MST. His-tagged ACE2 was labeled with RED-tris-NTA 2nd 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 CaCl₂, 0.01% Tween 20) to a final concentration of 3 nM. For the single-dose screen, 200 nL of library compounds were predispensed into the assay plate using Echo 650 series acoustic dispenser (Labcyte Inc.), mixed with 10 μ L of labeled protein, and incubated for 15 min at room temperature (RT). For dose–response experiments, 100 nL of compounds in twofold dilution series were transferred to a 384-well compound plate (Greiner, Cat #: 784201-1B). MST traces were collected using Monolith NT.Automated (Nanotemper Technologies) unit and standard treated capillary chip (Nanotemper Technologies, Cat #: MO AK002) with the 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 normalized fluorescence signal of the thermograph using MO.Affinity analysis software.

To identify false-positive binders, which could interact with the fluorophore-labeled His-tag instead of the target protein, all compounds tested in the dose–response experiments with ACE2 were counterscreened with poly-histidine control peptide (Nanotemper Technologies) under the same experimental conditions as with ACE2.

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. 3 μ L/well of 0.27 nM ACE2 (0.2 nM final concentration) suspension in assay buffer (PBS, pH 7.4, 0.01% Tween-20) was dispensed into an assay plate with Aurora Discovery BioRAPTR Dispenser (FRD; Beckton Dickenson) and incubated 15 min at RT. One μ L/well of 60 μ M ACE2 substrate MCA—Ala-Pro-Lys (Dnp)-OH (AnaSpec, Cat #: AS-60757) was then added. The substrate has the fluorophore MCA (7-methoxycoumarin-4-acetic acid) and Lys (Dnp, a dinitro-phenyl linked lysine) as a quencher and exhibits almost no fluorescence. ACE2 can recognize the site between Pro and Lys (Dnp) and cleave off the quencher, resulting in an increase of the fluorescence intensity. The plate was then 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 dimethyl sulfoxide (DMSO), set as 0%, and in the presence of 6.2 μ M MLN-4760, set as –100% inhibition. The resulting percent of inhibition data was fitted to a sigmoidal dose–response curve using the four-parameter Hill equation as described previously.⁴¹

Live SARS-CoV-2 Fluc Assay and Cytotoxicity Counterscreen. A live SARS-CoV-2 replication assay in A549-ACE2 host cells was used to measure the ability of compounds to perturb the replication of SARS-CoV-2. It employs an

engineered SARS-CoV-2 WA-1 lineage virus that has an integrated firefly luciferase reporter (Fluc, provided by Pei-Yong Shi, UTMB) and A549-ACE2 cells (generously provided by Pei-Yong Shi, UTMB), an adenocarcinoma human alveolar basal epithelial cell line stably overexpressing human ACE2.

Briefly, 20 nL/well of compounds in DMSO were spotted into 1536-well assay plates (Aurora E8, black clear bottom, tissue culture-treated plates) by acoustic dispensing. In parallel, 20 nL of DMSO was added to the first four columns of the plate, which serve as the no virus and neutral control wells. 4 μ L of A549-ACE2 cell suspension (4×10^5 cells/mL) was dispensed to all wells for a final density of 1,600 cells/well in Dulbecco's modified Eagle's medium (DMEM) with 2% fetal bovine serum (FBS). In addition, 1 μ L of media (DMEM, 2% FBS) was dispensed to columns 1 and 2. Thereafter, 1 μ L/well of SARS-CoV-2 (USA_WA1/2020) at multiplicity of infection of 0.2 suspended in media was dispensed to columns 3–48. This results in a DMSO final concentration of 0.4%. Assay plates were incubated for 48 h at 37 °C, 5% CO₂, and 90% humidity. After incubation, 2 μ L/well of One-Glo (Promega, Cat # E6120) detection reagent was added, and plates were incubated for 5 min at RT. Luminescence signal was measured on a BMG PHERAStar plate reader.

Raw data was normalized to the neutral control (cells infected with virus in the presence of DMSO, set as 0%) and positive control (cells without virus added, set as –100%) for each plate. The resulting percent of inhibition data was fitted to a sigmoidal dose–response curve using the four-parameter Hill equation.

In parallel, the compounds were tested in a cytotoxicity counterscreen against the A549-ACE2 cell line. The assay was set up in the same way as in the Fluc assay omitting the addition of virus. A549-ACE2 with DMSO solvent served as the negative control, whereas media and DMSO (no cells) were the positive control. The plates were incubated for 48 h at 37 °C, and one volume of CellTiter-Glo assay reagent (Promega, Madison, WI), which assesses viable cells (ATP content), was added using a BioRAPTR FRD (Beckman Coulter, Brea, CA). Cell viability was measured using a ViewLux μ HTS Microplate Imager (PerkinElmer, Waltham, MA). The obtained luminescence signal was normalized against negative control (0% response) and positive control (–100% response).

Firefly Luciferase Counterscreen. To identify false-positive hits, which could reduce the Fluc signal due to the inhibition of the luciferase enzyme rather than perturbing the viral infection, compounds were tested in a biochemical assay with recombinant luciferase from *Photinus pyralis* (firefly). Briefly, 25 nL/well compounds or DMSO as vehicle control (columns 1–4) was acoustically transferred to a white solid 1536-well plate (Greiner, Cat #: 789175-F). 3 μ L/well of 13.33 nM luciferase (10 nM final concentration) suspension in 50 mM Tris-acetate buffer, pH 7.6 was dispensed into the assay plate with Aurora Discovery BioRAPTR Dispenser (FRD; Beckton Dickenson) and incubated 15 min at RT. 3 μ L/well of buffer only was dispensed to columns 3–4 as no enzyme control. One μ L/well of 40 μ M D-luciferin in substrate buffer (50 mM Tris-acetate, pH 7.6, 10 mM Mg-acetate, 10 μ M ATP, 0.01% Tween-20, 0.05% BSA) was then added. The plate was centrifuged at 1000 rpm for 15 s, and the luminescence was detected with the PHERAstar plate reader (BMG Labtech). Data was normalized to enzyme activity in the presence of DMSO, set as 0%, and no enzyme control, set as –100% inhibition. The resulting percent of inhibition data was fitted to a sigmoidal dose–response curve using the four-parameter Hill equation.

Computational Details. Data Curation. All chemical structures and correspondent activity information were analyzed and prepared according to data curation protocols proposed by Fourches et al.^{34–36} In summary, specific chemotypes were normalized and explicit hydrogens were added. If present, polymers, salts, organometallic compounds, and mixtures were removed. Furthermore, we performed the analysis and exclusion of duplicates. The following criteria were applied for exclusion of duplicates: (i) if the reported activity of the duplicates was the same (i.e., in concordance), only one entry was kept in the dataset; (ii) if duplicates presented discordance in biological activity, both entries were excluded. 1 compound with ambiguous activity data was removed. In total, 163 compounds were removed in the curation process, six of which were actives.

Molecular Descriptors. UNC Protocol. Descriptors were chosen to cover a reasonable amount of descriptor types (whole-molecule descriptors, fragment descriptors, and topological descriptors) while maximizing ease-of-use. A total of seven different descriptors were generated for the dataset. Six of these (RDKit whole-molecule descriptors, Morgan fingerprint, Hashed Atom-Pair Fingerprint, Hashed Topological Torsion Fingerprint, and MACCS keys) were generated using the RDKit⁴² Python package.

The seventh descriptor type, simplex representation of molecular structures (SiRMS), was calculated using the Hit QSAR software.⁴³ At the 2D level, the connectivity of atoms in a simplex, the atom type, and bond nature (single, double, triple, and aromatic) were considered.⁴⁴ Bonded and nonbonded 2D simplexes were used. In addition to element and atom type, physicochemical characteristics of atoms, such as partial charge, lipophilicity, refraction, and the atom's ability to be a hydrogen-bond donor/acceptor, were used for atom differentiation in the simplexes. For the atom characteristics with real values (charge, lipophilicity, and refraction), a binning procedure was used to define discrete groups: (i) partial charge $A \leq -0.05 < B \leq 0 < C \leq 0.05 < D$, (ii) lipophilicity $A \leq -0.5 < B \leq 0 < C \leq 0.5 < D$, and (iii) refraction $A \leq 1.5 < B \leq 3 < C \leq 8 < D$. For hydrogen-bond characteristics, the atoms were also divided into three groups: A (acceptor of hydrogen in H-bond); D (donor of hydrogen in H-bond); and I (indifferent atom, i.e., atom that does not form H-bonds).⁴⁵

NCATS Protocol. We employed three different sets of descriptors: physicochemical descriptors (RDKit), Morgan fingerprints (1024 bits), and Avalon fingerprints (1024 bits), calculated using the RDKit toolkit.⁴² As consensus modeling approaches have been reported to outperform simple QSAR models,^{46–50} we also performed the consensus of descriptors (RDKit, Morgan, and Avalon).

Model Building. UNC QSAR Protocol. The models were developed using best practices as described in Cherkasov et al.⁵¹ Model types employed were gradient boosting implemented in scikit-learn,⁵² a simple neural network implemented in PyTorch,⁵³ and the Multiple Descriptor Read-Across⁵⁴ (MuDRA) model developed by some of the authors of this work. For models not based on decision trees, the descriptor matrix was normalized to prevent undesired weighting of certain descriptor dimensions. Descriptor dimensions with low variance (less than 0.0001) were eliminated due to being noninformative. Models were generated for most architecture-descriptor pairs. The MuDRA algorithm requires multiple descriptor spaces, so it was handled differently than the rest. The four descriptor types used for the MuDRA model were the Morgan fingerprint,

Hashed Atom-Pair Fingerprint, Hashed Topological Torsion Fingerprint, and MACCS keys.

NCATS QSAR Protocol. In order to overcome the problem of data imbalance, we used bagging with stratified under-sampling.⁵⁵ This method has proven to be among the best-performing methods for dealing with imbalanced datasets.⁵⁶ SB is a machine-learning technique that is based on an ensemble of models developed using multiple training datasets sampled from the original training set. It uses minority-class samples to create the training set of positive samples using a traditional bagging approach (resampling with replacement) and after that randomly selects the same number of samples from the majority class. Thus, the total bagging training set size was double the number of the minority-class molecules. Several models are then built and predictions averaged in order to produce a final ensemble model output. Because of random sampling, about 37% of the molecules are not selected and left out in each run. These samples create the “out-of-the-bag” sets, which are used for testing the performance of the final model.⁵⁶ Although a small set of samples are selected each time, the majority of molecules contributed to the overall bagging procedure since the datasets were generated randomly. Random forest (with default parameters) was used as a base-classifier. 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.

Pharmacophore-Based Modeling. In addition to QSAR, we also performed LBP modeling. A pharmacophore describes the spatial arrangement of essential interactions of a drug with its respective receptor-binding site. It is a well-established method VS in the early drug discovery process. In this study, the generation of LBP models, their subsequent refinement, and VS were performed with LigandScout 4.4 Advanced, available by Inte:Ligand GmbH. The conformational libraries for both pharmacophore modeling and the VS process were created with i:Con⁵⁷ (max. 200 conformations per compound), a conformer generator implemented in LigandScout.

To design the LBP models, the actives (from the training set) were clustered based on pharmacophore-based similarity (cluster distances 0.4, 0.6, 0.7, and 0.8, respectively). For each of the clusters obtained from different cluster distance thresholds, MFP and SFP models were generated that incorporate the features of selected compounds per cluster.⁵⁸ A good pharmacophore model should not only be able to estimate the activity of active compounds but also have the ability to identify the 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 dataset (training and test set combined) and calculated the percentage of active and inactive that hit these pharmacophore models. The models that hit 20% more active compounds versus inactive compounds were selected for the final VS. The screening was performed using iscreen module, with default settings with the maximum number of omitted features set to 2.

Model Validation. UNC Protocol. Models were evaluated using five-fold external cross-validation. The dataset was split randomly into five partitions. Each model was trained on four of the five partitions and tested on the fifth. This process was repeated five times so each partition was used once as a test set. Reported model statistics are an average of performance across each test set.

NCATS Protocol. From each class, 80% of the data was randomly selected and used as a training set. The remaining 20% of compounds were considered as the external validation set. For stratified bagging, since multiple training datasets were generated by selecting the molecules with replacement from the training set in a random fashion, this leaves out about 37% of the instances in each run. Therefore, these molecules that constitute the “out-of-the-bag” sets are later used for testing the performance of the final model.

Molecular Docking. To investigate the possible binding mode for the confirmed allosteric binders, we took the most potent ligand (NCGC00138760-01, $K_d = 0.092 \mu\text{M}$) and performed molecular docking in the three proposed allosteric binding sites.⁵⁹ The crystal structure of ACE2 (PDB ID: 6M0J, resolution = 2.5 Å)⁶⁰ was retrieved from Protein Data Bank database and was prepared for docking using the Protein Preparation Wizard of the Schrödinger Suite (2021).⁶¹ During the protein preparation, hydrogen atoms were added, water molecules were removed, and optimal protonation states and ASN/GLN/HIS flips were determined.

The three allosteric binding sites were defined as proposed by Dutta,⁵⁹ taking 20 Å around respective residues (Figure 5). The

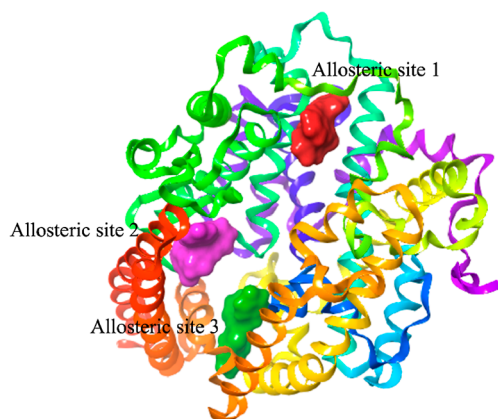


Figure 5. Structure of ACE2 (PDB ID: 6M0J). The allosteric sites 1, 2 and 3 are highlighted and depicted in red, pink, and green, respectively.

LigPrep module of Schrödinger Suite⁶² was used to generate the correct protonation states for the ligand, which were then used for the docking studies. The OPLS4 force field⁶³ was applied for the minimization of the structures and different ionization states were generated by adding or removing protons from the ligand at a target pH of 7.0 ± 2.0 using EPIK version 3.1.^{64,65} Tautomers were also generated for each ligand. For generation of stereoisomers, the information on chirality from the input file for each ligand was retained as is during the entire calculation. Docking was performed using the GlideXP scoring function^{61,66,67} implemented in Maestro (ver. 12.9).

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/acspsci.2c00049>.

Ligand-based pharmacophore models generated for screening, validation statistics for UNC models, model performance of NCATS stratified bagging models, activities for various models available in the PASS online web tool for all five hit compounds, activities for various models available in the ADMETlab 2.0 web tool for all five

hit compounds, activity of all confirmed allosteric ACE2 binding compounds, known ACE2 inhibitor MLN-4760 tested in ACE2 screening assays, results of the initial library screen and performance of the five ACE2 binders in the counterscreens (PDF)

■ AUTHOR INFORMATION

Corresponding Authors

Alexander Tropsha – Molecular Modeling Laboratory, UNC Eshelman School of Pharmacy, University of North Carolina, Chapel Hill, North Carolina 27599, United States; orcid.org/0000-0003-3802-8896; Email: tropsha@email.unc.edu

Eugene N. Muratov – Molecular Modeling Laboratory, UNC Eshelman School of Pharmacy, University of North Carolina, Chapel Hill, North Carolina 27599, United States; orcid.org/0000-0003-4616-7036; Email: murik@email.unc.edu

Bolormaa Baljinnyam – National Center for Advancing Translational Sciences, National Institutes of Health, Rockville, Maryland 20850, United States; Email: bolormaa.baljinnyam@nih.gov

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

Authors

Joshua E. Hochuli – Molecular Modeling Laboratory, UNC Eshelman School of Pharmacy and Curriculum in Bioinformatics and Computational Biology, University of North Carolina, Chapel Hill, North Carolina 27599, United States

Sankalp Jain – National Center for Advancing Translational Sciences, National Institutes of Health, Rockville, Maryland 20850, United States

Cleber Melo-Filho – Molecular Modeling Laboratory, UNC Eshelman School of Pharmacy, University of North Carolina, Chapel Hill, North Carolina 27599, United States

Zoe L. Sessions – Molecular Modeling Laboratory, UNC Eshelman School of Pharmacy, University of North Carolina, Chapel Hill, North Carolina 27599, United States

Tesia Bobrowski – Molecular Modeling Laboratory, UNC Eshelman School of Pharmacy, University of North Carolina, Chapel Hill, North Carolina 27599, United States

Jun Choe – National Center for Advancing Translational Sciences, National Institutes of Health, Rockville, Maryland 20850, United States; orcid.org/0000-0002-3389-5969

Johnny Zheng – National Center for Advancing Translational Sciences, National Institutes of Health, Rockville, Maryland 20850, United States

Richard Eastman – National Center for Advancing Translational Sciences, National Institutes of Health, Rockville, Maryland 20850, United States; orcid.org/0000-0003-3580-380X

Daniel C. Talley – National Center for Advancing Translational Sciences, National Institutes of Health, Rockville, Maryland 20850, United States

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

Anton Simeonov — National Center for Advancing Translational Sciences, National Institutes of Health, Rockville, Maryland 20850, United States

Complete contact information is available at:
<https://pubs.acs.org/10.1021/acspsci.2c00049>

Author Contributions

^{||}J.E.H. and S.J. contributed equally. Conceptualization, E.N.M., A.T., B.B., and A.V.Z.; methodology, J.E.H., S.J., R.E., E.N.M., and A.T.; validation, J.E.H. and C.M.-F.; assay conceptualization, A.S.; experimental validation, B.B. and R.E.; qHTS analysis and selection of chemotypes, G.R. and D.C.T.; formal analysis, L.C. and H.G.; investigation, T.B., Z.L.S., and E.N.M.; data curation, C.M.-F., T.B., Z.L.S., and S.J.; writing—original draft, J.E.H. and S.J.; writing—review and editing, all authors; visualization, J.E.H. and B.B. The senior authors assume the overall responsibility for the results reported herein.

Notes

The authors declare the following competing financial interest(s): A.T. and E.N.M. are co-founders of Predictive, LLC, which develops computational methodologies and software for toxicity prediction.

Data, workflows, and models produced in this study are available on GitHub (https://github.com/molecularmodelinglab/ACE2_allosteric_modeling) and on Chembench⁶⁸ (<https://chembench.mml.unc.edu/>). External software is available at the respective links: KNIME (<https://www.knime.com>), Atkinson standardizer (<https://github.com/flatkinson/standardiser>), and scikit-learn (<https://scikit-learn.org/stable/>).

ACKNOWLEDGMENTS

Authors from UNC-Chapel Hill were supported by National Institutes of Health (grants U19AI171292 and R01GM140154). J.E.H. was supported by National Institutes of Health Training grant (T32GM008570). This research was supported by the Intramural Research Program of the National Center for Advancing Translational Sciences (NCATS) and National Institutes of Health (NIH).

REFERENCES

- (1) Morgan, A. K.; Awafo, B. A.; Quartey, T. The Effects of COVID-19 on Global Economic Output and Sustainability: Evidence from around the World and Lessons for Redress. *Sustain. Sci. Pract. Pol.* **2021**, *17*, 76–80.
- (2) Milani, F. COVID-19 Outbreak, Social Response, and Early Economic Effects: A Global VAR Analysis of Cross-Country Interdependencies. *J. Popul. Econ.* **2021**, *34*, 223–252.
- (3) Smith Jervelund, S.; Eikemo, T. A. The Double Burden of COVID-19. *Scand. J. Publ. Health* **2021**, *49*, 1–4.
- (4) Coronavirus disease (COVID-19) pandemic. <https://www.who.int/emergencies/diseases/novel-coronavirus-2019> (accessed Jan 18, 2022).
- (5) Tartof, S. Y.; Slezak, J. M.; Fischer, H.; Hong, V.; Ackerson, B. K.; Ranasinghe, O. N.; Frankland, T. B.; Ogun, O. A.; Zamparo, J. M.; Gray, S.; Valluri, S. R.; Pan, K.; Angulo, F. J.; Jodar, L.; McLaughlin, J. M. Six-Month Effectiveness of BNT162B2 mRNA COVID-19 Vaccine in a Large US Integrated Health System: A Retrospective Cohort Study. *SSRN Electron. J.* **2021**, *398*, 1407–1416, DOI: 10.2139/ssrn.3909743.
- (6) Tenforde, M. W.; Self, W. H.; Adams, K.; Gaglani, M.; Ginde, A. A.; McNeal, T.; Ghamande, S.; Douin, D. J.; Talbot, H. K.; Casey, J. D.; Mohr, N. M.; Zepeski, A.; Shapiro, N. I.; Gibbs, K. W.; Files, D. C.; Hager, D. N.; Shehu, A.; Prekker, M. E.; Erickson, H. L.; Exline, M. C.; Gong, M. N.; Mohamed, A.; Henning, D. J.; Steingrub, J. S.; Peltan, I. D.; Brown, S. M.; Martin, E. T.; Monto, A. S.; Khan, A.; Hough, C. L.;

Busse, L. W.; ten Lohuis, C. C.; Duggal, A.; Wilson, J. G.; Gordon, A. J.; Qadir, N.; Chang, S. Y.; Mallow, C.; Rivas, C.; Babcock, H. M.; Kwon, J. H.; Halasa, N.; Chappell, J. D.; Luring, A. S.; Grijalva, C. G.; Rice, T. W.; Jones, I. D.; Stubblefield, W. B.; Baughman, A.; Womack, K. N.; Rhoads, J. P.; Lindsell, C. J.; Hart, K. W.; Zhu, Y.; Olson, S. M.; Kobayashi, M.; Verani, J. R.; Patel, M. M. Association Between mRNA Vaccination and COVID-19 Hospitalization and Disease Severity. *JAMA* **2021**, *326*, 2043.

(7) Singanayagam, A.; Hakki, S.; Dunning, J.; Madon, K. J.; Crone, M. A.; Koycheva, A.; Derqui-Fernandez, N.; Barnett, J. L.; Whitfield, M. G.; Varro, R.; Charlett, A.; Kundu, R.; Fenn, J.; Cutajar, J.; Quinn, V.; Conibear, E.; Barclay, W.; Freemont, P. S.; Taylor, G. P.; Ahmad, S.; Zambon, M.; Ferguson, N. M.; Lalvani, A.; Badhan, A.; Dustan, S.; Tejpal, C.; Ketkar, A. v.; Narean, J. S.; Hammett, S.; McDermott, E.; Pillay, T.; Houston, H.; Luca, C.; Samuel, J.; Bremang, S.; Evetts, S.; Poh, J.; Anderson, C.; Jackson, D.; Miah, S.; Ellis, J.; Lackenby, A. Community Transmission and Viral Load Kinetics of the SARS-CoV-2 Delta (B.1.617.2) Variant in Vaccinated and Unvaccinated Individuals in the UK: A Prospective, Longitudinal, Cohort Study. *Lancet Infect. Dis.* **2021**, *22*, 183.

(8) Walensky, R. P.; Walke, H. T.; Fauci, A. S. SARS-CoV-2 Variants of Concern in the United States—Challenges and Opportunities. *JAMA* **2021**, *325*, 1037.

(9) Bobrowski, T.; Melo-Filho, C. C.; Korn, D.; Alves, V. M.; Popov, K. I.; Auerbach, S.; Schmitt, C.; Moorman, N. J.; Muratov, E. N.; Tropsha, A. Learning from History: Do Not Flatten the Curve of Antiviral Research. *Drug Discovery Today* **2020**, *25*, 1604–1613.

(10) Marshall, W. F., III COVID-19 (Coronavirus) Drugs: Are There Any That Work?, 2021.

(11) *Emergency Use Authorization 105*. United States Food and Drug Administration, 2022.

(12) *Emergency Use Authorization 108*. United States Food and Drug Administration, 2022.

(13) Brimacombe, K. R.; Zhao, T.; Eastman, R. T.; Hu, X.; Wang, K.; Backus, M.; Baljinnam, 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**, *2020*, 06.04.135046.

(14) 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.

(15) Hu, X.; Shrimp, J. H.; Guo, H.; Xu, M.; Chen, C. Z.; Zhu, W.; Zakharov, A.; Jain, S.; Shinn, P.; Simeonov, A.; Hall, M. D.; Shen, M. Discovery of TMPRSS2 Inhibitors from Virtual Screening. *bioRxiv* **2021**, *2020*, 12.28.424413.

(16) 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.; 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. *Frontiers in Pharmacology* **2021**, *11*.

(17) 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.

(18) 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 COVID19 Therapeutics: Development and Drug Repurposing Screen of an AlphaLISA Proximity Assay. *ACS Pharmacol. Transl. Sci.* **2020**, *3*, 1352–1360.

(19) Nathan, C. Fresh Approaches to Anti-Infective Therapies. *Sci. Transl. Med.* **2012**, *4*, 140sr2.

- (20) Hussain, M.; Jabeen, N.; Raza, F.; Shabbir, S.; Baig, A. A.; Amanullah, A.; Aziz, B. Structural Variations in Human ACE2 May Influence Its Binding with SARS-CoV-2 Spike Protein. *J. Med. Virol.* **2020**, *92*, 1580–1586.
- (21) Pacurari, M.; Kafoury, R.; Tchounwou, P. B.; Ndebele, K. The Renin-Angiotensin-Aldosterone System in Vascular Inflammation and Remodeling. *Int. J. Inflammation* **2014**, *2014*, 1–13.
- (22) Maron, B. A.; Leopold, J. A. The Role of the Renin-Angiotensin-Aldosterone System in the Pathobiology of Pulmonary Arterial Hypertension (2013 Grover Conference Series). *Pulm. Circ.* **2014**, *4*, 200–210.
- (23) Suzuki, Y. Angiotensin II, the Immune System and Renal Diseases: Another Road for RAS? *Nephrol., Dial., Transplant.* **2003**, *18*, 1423–1426.
- (24) Patel, S.; Rauf, A.; Khan, H.; Abu-Izneid, T. Renin-Angiotensin-Aldosterone (RAAS): The Ubiquitous System for Homeostasis and Pathologies. *Biomed. Pharmacother.* **2017**, *94*, 317–325.
- (25) South, A. M.; Diz, D. I.; Chappell, M. C. COVID-19, ACE2, and the Cardiovascular Consequences. *Am. J. Physiol. Heart Circ. Physiol.* **2020**, *318*, H1084–H1090.
- (26) Gheblawi, M.; Wang, K.; Viveiros, A.; Nguyen, Q.; Zhong, J.-C.; Turner, A. J.; Raizada, M. K.; Grant, M. B.; Oudit, G. Y. Angiotensin-Converting Enzyme 2: SARS-CoV-2 Receptor and Regulator of the Renin-Angiotensin System. *Circ. Res.* **2020**, *126*, 1456–1474.
- (27) Gross, L. Z. F.; Sacerdoti, M.; Piiper, A.; Zeuzem, S.; Leroux, A. E.; Biondi, R. M. ACE2, the Receptor That Enables Infection by SARS-CoV-2: Biochemistry, Structure, Allostery and Evaluation of the Potential Development of ACE2 Modulators. *ChemMedChem* **2020**, *15*, 1682–1690.
- (28) Muratov, E. N.; Bajorath, J.; Sheridan, R. P.; Tetko, I. V.; Filimonov, D.; Poroikov, V.; Oprea, T. I.; Baskin, I. I.; Varnek, A.; Roitberg, A.; Isayev, O.; Curtalolo, S.; Fourches, D.; Cohen, Y.; Aspuru-Guzik, A.; Winkler, D. A.; Agrafiotis, D.; Cherkasov, A.; Tropsha, A. QSAR without Borders. *Chem. Soc. Rev.* **2020**, *49*, 3525–3564.
- (29) 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.
- (30) Havranek, B.; Chan, K. K.; Wu, A.; Procko, E.; Islam, S. M. Computationally Designed ACE2 Decoy Receptor Binds SARS-CoV-2 Spike (S) Protein with Tight Nanomolar Affinity. *J. Chem. Inf. Model.* **2021**, *61*, 4656–4669.
- (31) Terali, K.; Baddal, B.; Gülcan, H. O. Prioritizing Potential ACE2 Inhibitors in the COVID-19 Pandemic: Insights from a Molecular Mechanics-Assisted Structure-Based Virtual Screening Experiment. *J. Mol. Graphics Modell.* **2020**, *100*, 107697.
- (32) 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.
- (33) Cherkasov, A.; Muratov, E. N.; Fourches, D.; Varnek, A.; Baskin, I. I.; Cronin, M.; Dearden, J.; Gramatica, P.; Martin, Y. C.; Todeschini, R.; Consonni, V.; Kuz'min, V. E.; Cramer, R.; Benigni, R.; Yang, C.; Rathman, J.; Terflöth, L.; Gasteiger, J.; Richard, A.; Tropsha, A. QSAR Modeling: Where Have You Been? Where Are You Going To? *J. Med. Chem.* **2014**, *57*, 4977–5010.
- (34) 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.
- (35) 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.
- (36) Fourches, D.; Muratov, E.; Tropsha, A. Curation of Chemogenomics Data. *Nat. Chem. Biol.* **2015**, *11*, 535.
- (37) Xiong, G.; Wu, Z.; Yi, J.; Fu, L.; Yang, Z.; Hsieh, C.; Yin, M.; Zeng, X.; Wu, C.; Lu, A.; Chen, X.; Hou, T.; Cao, D. ADMETlab 2.0: An Integrated Online Platform for Accurate and Comprehensive Predictions of ADMET Properties. *Nucleic Acids Res.* **2021**, *49*, W5–W14.
- (38) Filimonov, D. A.; Lagunin, A. A.; Glorizova, T. A.; Rudik, A. v.; Drushilovskii, D. S.; Pogodin, P. v.; Poroikov, V. v. Prediction of the Biological Activity Spectra of Organic Compounds Using the Pass Online Web Resource. *Chem. Heterocycl. Compd.* **2014**, *50*, 444–457.
- (39) *Inte:Ligand Software Development and Consulting GmbH. LigandScout*, version 4.4: Vienna, Austria, 2019.
- (40) Wolber, G.; Langer, T. LigandScout: 3-D Pharmacophores Derived from Protein-Bound Ligands and Their Use as Virtual Screening Filters. *J. Chem. Inf. Model.* **2005**, *45*, 160–169.
- (41) 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.
- (42) RDKit: open-source cheminformatics.
- (43) Kuz'min, V.; Artemenko, A.; Ognichenko, L.; Hromov, A.; Kosinskaya, A.; Stelmakh, S.; Sessions, Z. L.; Muratov, E. N. Simplex Representation of Molecular Structure as Universal QSAR/QSPR Tool. *Struct. Chem.* **2021**, *32*, 1365–1392.
- (44) Kuz'min, V. E.; Muratov, E. N.; Artemenko, A. G.; Gorb, L.; Qasim, M.; Leszczynski, J. The Effects of Characteristics of Substituents on Toxicity of the Nitroaromatics: HiT QSAR Study. *J. Comput.-Aided Mol. Des.* **2008**, *22*, 747–759.
- (45) Kuz'min, V. E.; Muratov, E. N.; Artemenko, A. G.; Varlamova, E. V.; Gorb, L.; Wang, J.; Leszczynski, J. Consensus QSAR Modeling of Phosphor-Containing Chiral AChE Inhibitors. *QSAR Comb. Sci.* **2009**, *28*, 664–677.
- (46) Zhu, H.; Tropsha, A.; Fourches, D.; Varnek, A.; Papa, E.; Gramatica, P.; Öberg, T.; Dao, P.; Cherkasov, A.; Tetko, I. v. Combinatorial QSAR Modeling of Chemical Toxicants Tested against Tetrahymena Pyriformis. *J. Chem. Inf. Model.* **2008**, *48*, 766–784.
- (47) Zakharov, A. V.; Peach, M. L.; Sitzmann, M.; Nicklaus, M. C. QSAR Modeling of Imbalanced High-Throughput Screening Data in PubChem. *J. Chem. Inf. Model.* **2014**, *54*, 705–712.
- (48) Xu, Y.; Pei, J.; Lai, L. Deep Learning Based Regression and Multiclass Models for Acute Oral Toxicity Prediction with Automatic Chemical Feature Extraction. *J. Chem. Inf. Model.* **2017**, *57*, 2672–2685.
- (49) Zakharov, A. v.; Zhao, T.; Nguyen, D.-T.; Peryea, T.; Sheils, T.; Yasgar, A.; Huang, R.; Southall, N.; Simeonov, A. Novel Consensus Architecture To Improve Performance of Large-Scale Multitask Deep Learning QSAR Models. *J. Chem. Inf. Model.* **2019**, *59*, 4613–4624.
- (50) 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**, *61*, 653–663.
- (51) Cherkasov, A.; Muratov, E. N.; Fourches, D.; Varnek, A.; Baskin, I. I.; Cronin, M.; Dearden, J.; Gramatica, P.; Martin, Y. C.; Todeschini, R.; Consonni, V.; Kuz'min, V. E.; Cramer, R.; Benigni, R.; Yang, C.; Rathman, J.; Terflöth, L.; Gasteiger, J.; Richard, A.; Tropsha, A. QSAR Modeling: Where Have You Been? Where Are You Going To? *J. Med. Chem.* **2014**, *57*, 4977–5010.
- (52) Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; Vanderplas, J.; Passos, A.; Cournapeau, D.; Brucher, M.; Perrot, M.; Duchesnay, E. Scikit-Learn: Machine Learning in Python. *J. Mach. Learn. Res.* **2011**, *12*, 2825–2830.
- (53) Paszke, A.; Gross, S.; Massa, F.; Lerer, A.; Bradbury, J.; Chanan, G.; Killeen, T.; Lin, Z.; Gimelshein, N.; Antiga, L.; Desmaison, A.; Kopf, A.; Yang, E.; DeVito, Z.; Raison, M.; Tejani, A.; Chilamkurthy, S.; Steiner, B.; Fang, L.; Bai, J.; Chintala, S. PyTorch: An Imperative Style, High-Performance Deep Learning Library. In *Advances in Neural Information Processing Systems* 32; Wallach, H., Larochelle, H., Beygelzimer, A., d Alché-Buc, F., Fox, E., Garnett, R., Eds.; Curran Associates, Inc., 2019; pp 8024–8035.
- (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) Tetko, I. V.; Novotarskyi, S.; Sushko, I.; Ivanov, V.; Petrenko, A. E.; Dieden, R.; Lebon, F.; Mathieu, B. Development of Dimethyl Sulfoxide Solubility Models Using 163,000 Molecules: Using a Domain Applicability Metric to Select More Reliable Predictions. *J. Chem. Inf. Model.* **2013**, *53*, 1990–2000.

(56) 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.

(57) 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.

(58) Jain, S.; Talley, D. C.; Baljinnayam, B.; Choe, J.; Hanson, Q.; Zhu, W.; Xu, M.; Chen, C. Z.; Zheng, W.; Hu, X.; Shen, M.; Rai, G.; Hall, M. D.; Simeonov, A.; Zakharov, A. v. Hybrid In Silico Approach Reveals Novel Inhibitors of Multiple SARS-CoV-2 Variants. *ACS Pharmacol. Transl. Sci.* **2021**, *4*, 1675–1688.

(59) Dutta, K. Allosteric Site of ACE-2 as a Drug Target for COVID-19. *ACS Pharmacol. Transl. Sci.* **2022**, *5*, 179–182.

(60) Lan, J.; Ge, J.; Yu, J.; Shan, S.; Zhou, H.; Fan, S.; Zhang, Q.; Shi, X.; Wang, Q.; Zhang, L.; Wang, X. Structure of the SARS-CoV-2 Spike Receptor-Binding Domain Bound to the ACE2 Receptor. *Nature* **2020**, *581*, 215–220.

(61) Schrödinger LLC. *Schrödinger Release 2022-2*: New York, NY, 2021;.

(62) Schrödinger LLC. *Schrödinger Release 2015-1: LigPrep*. version 3.3: New York, NY, 2015;.

(63) Lu, C.; Wu, C.; Ghoreishi, D.; Chen, W.; Wang, L.; Damm, W.; Ross, G. A.; Dahlgren, M. K.; Russell, E.; von Bargen, C. D.; Abel, R.; Friesner, R. A.; Harder, E. D. OPLS4: Improving Force Field Accuracy on Challenging Regimes of Chemical Space. *J. Chem. Theory Comput.* **2021**, *17*, 4291–4300.

(64) Greenwood, J. R.; Calkins, D.; Sullivan, A. P.; Shelley, J. C. Towards the Comprehensive, Rapid, and Accurate Prediction of the Favorable Tautomeric States of Drug-like Molecules in Aqueous Solution. *J. Comput.-Aided Mol. Des.* **2010**, *24*, 591–604.

(65) Shelley, J. C.; Chollet, A.; Frye, L. L.; Greenwood, J. R.; Timlin, M. R.; Uchimaya, M. Epik: A Software Program for PK a Prediction and Protonation State Generation for Drug-like Molecules. *J. Comput.-Aided Mol. Des.* **2007**, *21*, 681–691.

(66) Friesner, R. A.; Murphy, R. B.; Repasky, M. P.; Frye, L. L.; Greenwood, J. R.; Halgren, T. A.; Sanschagrin, P. C.; Mainz, D. T. Extra Precision Glide: Docking and Scoring Incorporating a Model of Hydrophobic Enclosure for Protein–Ligand Complexes. *J. Med. Chem.* **2006**, *49*, 6177–6196.

(67) Halgren, T. A.; Murphy, R. B.; Friesner, R. A.; Beard, H. S.; Frye, L. L.; Pollard, W. T.; Banks, J. L. Glide: A New Approach for Rapid, Accurate Docking and Scoring. 2. Enrichment Factors in Database Screening. *J. Med. Chem.* **2004**, *47*, 1750–1759.

(68) Capuzzi, S. J.; Kim, I. S.-J.; Lam, W. I.; Thornton, T. E.; Muratov, E. N.; Pozefsky, D.; Tropsha, A. Chembench: A Publicly Accessible, Integrated Cheminformatics Portal. *J. Chem. Inf. Model.* **2017**, *57*, 105–108.