A 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, Alexey V. Zakharov*

National Center for Advancing Translational Sciences (NCATS), National Institutes of Health, 9800 Medical Center Drive, Rockville, MD, 20850, United States

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 cytophatic 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%, $IC_{50} \leq 15$ μM). Systematic clustering analysis of active compounds revealed three promising chemotypes which have not been previously identified as inhibitors of SARS-CoV-2 infection. Further analysis identified allosteric binders to host receptor angiotensin-converting enzyme 2, which were able 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, especially due to the emergence of several SARS-CoV-2 variants. As per a recent CDC report, there are 13 variants, 5 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 all data generated utilizing SARS-CoV-2 related assays 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 testing of 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 to identify small molecule protein modulators offers significant advantages, especially when used to complement

traditional HTS methodology[11,12]. Multiple *in silico* studies have been reported which employ virtual screening of small molecule databases[13–18]. However, in the majority of published COVID-19-related virtual screening 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[19,20], 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 quantitative structure-activity relationship (QSAR) and ligand-based pharmacophore modeling, followed by experimental testing of predicted hits in the 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 dataset for the proposed hybrid approach used in the second round (**Fig. 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, $IC_{50} \leq 15$ µM; Efficacy > 30%) in the CPE assay and minimal cytotoxicity ($IC_{50} > 30$ µM), where 68 of them had an efficacy greater than 70%. Interestingly, three novel antiviral chemotypes emerged with multiple (≥3) active structural analogs in each cluster. Some preliminary structure-activity relationships (SARs) were identified within each cluster, which further 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, identified as allosteric ACE2 binders (using microscale thermophoresis), 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 showing submicromolar activity against both variants.

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

Results

Hybrid approach for *in silico* screening

Combination of ligand- and structure-based methods into consensus approach has been used earlier to discover novel actives against various targets[21–23], since it improves the precision and reduces false positives[24]. 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 (**Fig. 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) approach is provided in Supplementary Table 1. It was measured by different

metrics, including operating characteristic area under the curve (ROC AUC). All developed models showed ROC AUC values > 0.75. In the 1st 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 1st round of screening are referred to as SB-1. After obtaining the experimental results from the 1st round of virtual screening, we updated our SB model (referred to as SB-2). In the 2nd 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 1st round of ligand-based pharmacophore (LBP) modeling, we used the 48 active compounds: clustering based on pharmacophore-based similarity (cluster distance 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 (Supplementary Table 2) were selected for virtual screening. The ligand-based pharmacophore models generated in the first round of screening are referred to as LBP-1. For the 2nd 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 (Supplementary Table 3). In general, the 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 1st Round *in silico* Screening Hits

The 320 compounds selected from the 1st round of *in silico* screening (see Methods for details) were tested in the CPE assay in a 5-point dilution series, with concentrations ranging from 20 µM to 124 nM. To exclude compounds with cytotoxic effects, the compounds were counter-screened 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% (Supplementary Table 4). This gives us a positive predicted 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 2nd Round *in silico* Screening Hits and Validation

The selected 320 compounds from the 2nd round of *in silico* screening (see Methods for details) were also tested in the CPE assay in a 5-point dilution series, with concentrations ranging from 20 µM to 124 nM. For the 2nd 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%; Supplementary Table 5). Moreover, 27 of these 58 compounds exhibited an $IC_{50} \leq 10$ µM in the CPE reduction assay. This further improved the PPV value to 18%.

For validation, 69 compounds were 'cherry-picked' and retested in the CPE reduction assay in duplicate as an 8-point dilution series, with concentrations ranging from 20 µM to 78 nM. 53 of the retested compounds were confirmed to have CPE inhibitory activity with an efficacy > 30%

and IC_{50} of 5-25 µM, and five of them exhibited an $IC_{50} \leq 10$ µM with no notable cytotoxicity (Supplementary Table 6). The five most potent and efficacious compounds (**1-5**) from the follow-up CPE assay are shown in **Fig. 2a**.

Clustering and Preliminary SAR Analysis

Hierarchical cluster analysis of the 100 CPE reducing compounds, identified in the two rounds of virtual screening, revealed three promising chemotypes (**Fig. 2b**), where three or more active structural analogs 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 analogs analyzed (and present in in-house compound libraries) were structurally similar enough for direct comparison. The most promising analogs from each of the three chemotypes, activity in the CPE assay, as well as *in vitro* physicochemical properties are shown in **Fig. 3** and **Table 1**.

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

Within chemotype **B** were 18 in-house structural analogs, eight of which (compounds **9-16**) have promising IC_{50} and efficacy values (**Fig. 3b**). Analogs 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 analogs (compounds **2, 4, 5, 17, 18, 20-23**), including three (**2, 4, 5**) of the most active compounds identified in the study (**Table 1**). Most of the notably active analogs contain variously substituted piperazines at the 2-position of the quinazoline core and 2-methyl-benzylamine at the 4-position. However, the open-chain (vs piperazine) analog (**2**) is also quite active, suggesting a 2-position diamine with an ethylene spacer is perhaps part of the parent pharmacophore. Methyl- and ethyl- substitutions off the 4-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 the metabolic liability can be mitigated *via* 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 of virtual screening hits active in the CPE reducing assay, 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 anti-viral drug development because its inhibition prevents the formation of mature functional viral proteins and thus, viral replication[25]. As such, active compounds were screened in the SARS-CoV-2 M^{pro} enzymatic assay, however no activity was observed.

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. The compounds showed activity in both RBD-ACE2 AlphaLisa and in the TruHit counterscreen (see Methods for details), rendering to inconclusive results.

Nonetheless, we tested the compounds for their binding to ACE2 using microscale thermophoresis (MST) followed by a SARS-CoV-2 pseudoparticle (PP) entry assay, to explore if a molecule interacting with the viral receptor could interfere with host cell 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 µM, with no inhibitory or agonistic activity in ACE2 enzymatic assay (**Table 2**).

All six 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 2**). Since these compounds do not bind S protein, we hypothesized 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. As result, we found that 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 compared to the wild type (**Fig. 4**).

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[26], however it does not consider the 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[27]. To overcome these shortcomings, a hybrid approach was developed which combines QSAR models with pharmacophore 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 analogs revealed some preliminary SAR, although trends were limited due to significant structural differences within the set of analogs. This supports the hypothesis that these compounds are acting on a common target or *via* a shared mechanism to inhibit viral proliferation, thus decreasing the viral cytopathic effect. Overall, the chemotypes identified showed good efficacy and potency as screening hits. These preliminary screening results clearly warrant further investigation of each chemotype *via* medicinal chemistry to thoroughly explore and establish the SAR, as well as optimize physicochemical properties.

In an effort to elucidate the mechanism of action, active compounds were screened against some previously established SARS-CoV-2 targets which 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.

Interestingly, we identified six CPE-active compounds which are ACE2 binders and inhibitors of viral entry. We assume these molecules are allosteric binders to ACE2, as they did not inhibit ACE2 enzymatic activity, meaning they did not at least interfere with the substrate binding. Although these compounds do not interrupt the RDB-ACE2 interaction, they were able to inhibit the entry of pseudoparticles resembling wild-type as well as other variants of SARS-CoV-2. These compounds might 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 do not utilize ACE2 for host cell binding but require 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 hinder viral entry.

To accelerate further research on 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/).

Figures and Tables
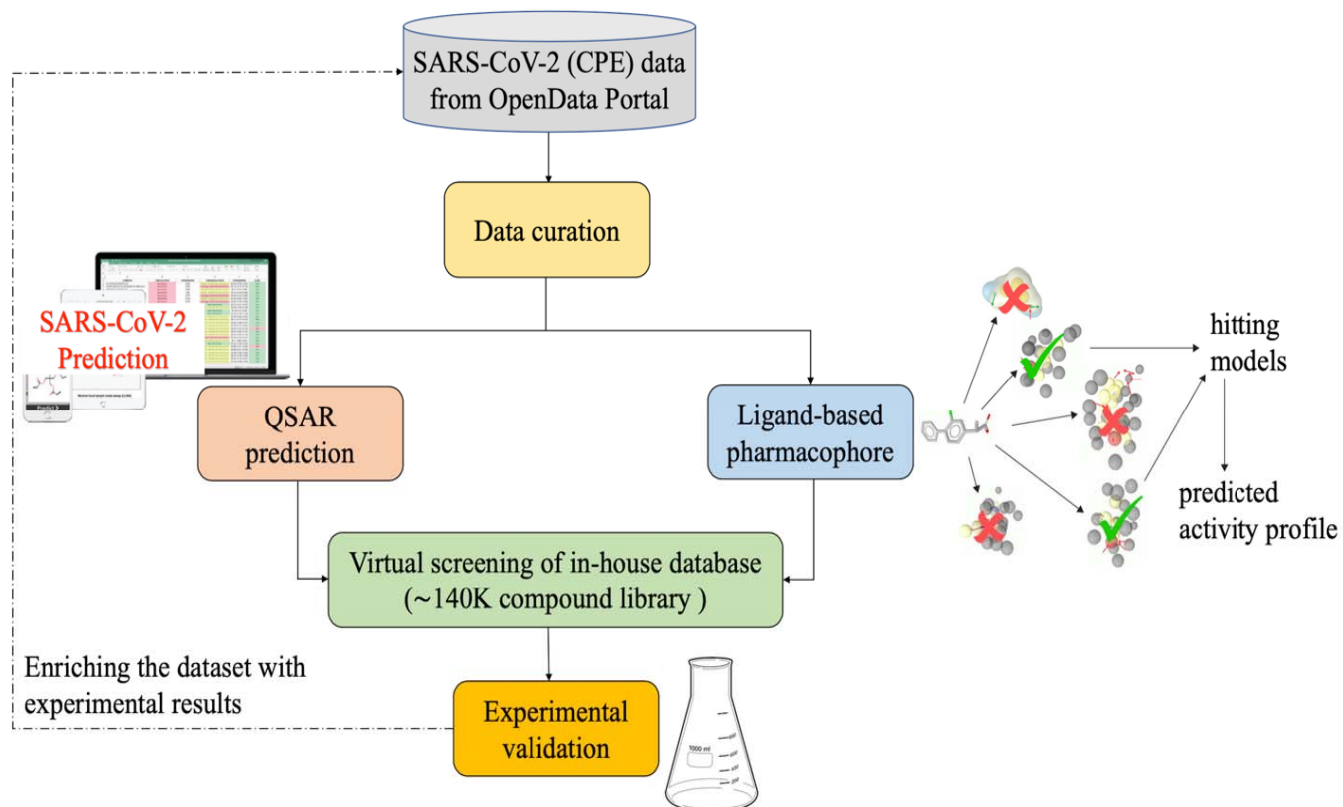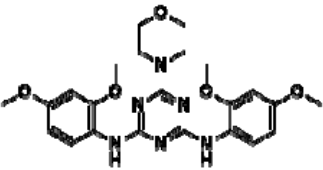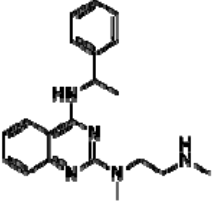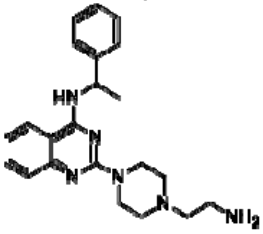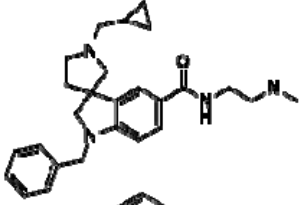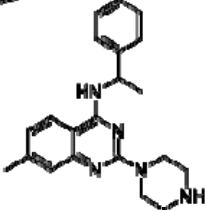
Fig. 1: Flowchart of the virtual screening strategy adopted in this study.

a

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

B

C
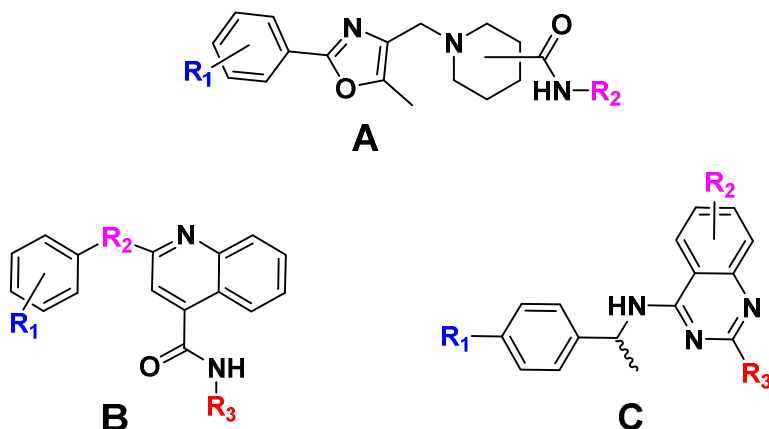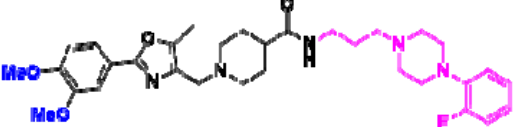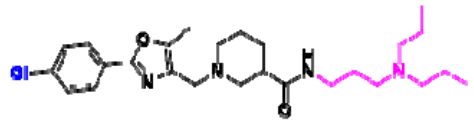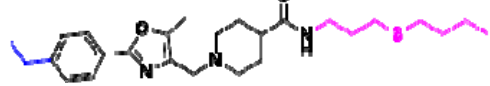
b

Fig. 2: a. Five most potent and efficacious compounds identified, along with *in vitro*/physico-chemical ADME data. [a]IC$_{50}$: half-maximal inhibitory concentration value obtained from the CPE assay in 8-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 lysates reported in minutes, with a minimum detectable half-life of 1 minute. [d]PAMPA (parallel artificial membrane permeation assay) is reported as a metric of the passive permeability of the compounds (1x10^{-6} cm/s). [e]Solubility – pION μSOL assay for kinetic aqueous solubility determination, pH 7.4. **b.** Three chemotypes (**A-C**) identified as active in the CPE assay.

a

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

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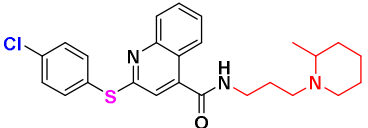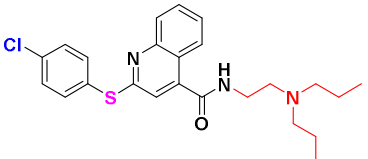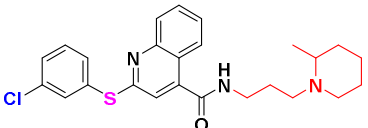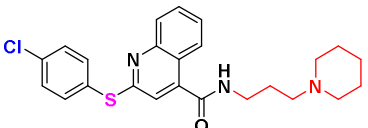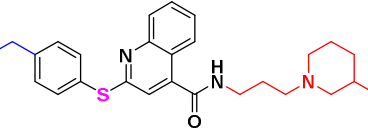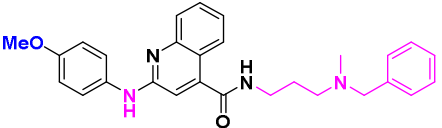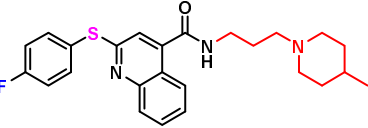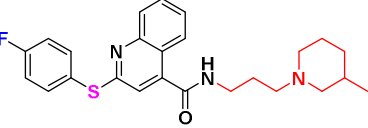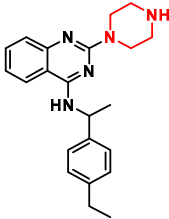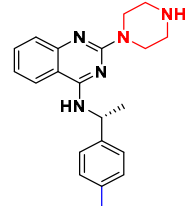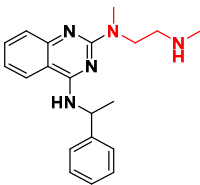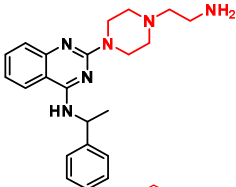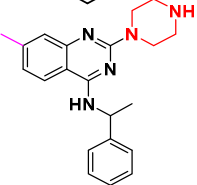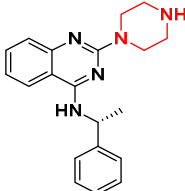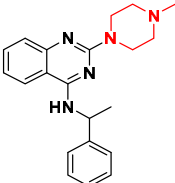
ID	Structure	IC_{50} $(\mu M)^a$	Efficacy $(\%)^c$	$T_{1/2}$ $(min)^d$	PAMPA $(1e\text{-}6\ cm/s)^e$	Solubility $(\mu g/mL)^f$
9		6.31^b	100^b	3.3	815	31
10		10.0^b	97^b	2.2	564	1.9
11		10.0^b	100^b	2.0	1560	25
12		11.2	99	3.3	Equilibrated	30
13		12.6	87	2.3	1530	>65
14		11.2^b	74^b	4.0	Equilibrated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

Fig. 3: Notably active chemotypes **(a) A** and **(b) B** which show no notable cytotoxicity ($IC_{50} \leq 30$ µM). [a]IC_{50}: half-maximal inhibitory concentration values obtained from the CPE assay in 8-point dose response, measured in duplicate. [b] Values represent data obtained from 5-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 lysates reported in minutes, minimum detectable half-life of 1 minute. [e]PAMPA (parallel artificial membrane permeation assay) is reported as a metric of the passive permeability of the compounds. [f]Solubility – pION µSOL assay for kinetic aqueous solubility determination, pH 7.4.

ID	Structure	IC$_{50}$ (µM)[a]	Efficacy (%)[c]	T$_{1/2}$ (min)[d]	PAMPA (1e-6 cm/s)[e]	Solubility (µ/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
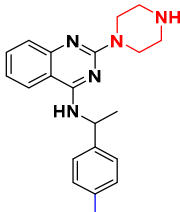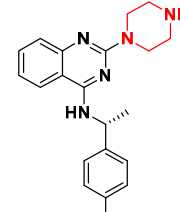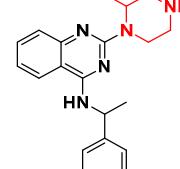
21		12.6	92	7.7	772	>51
22		14.1	81	2.8	Equilibrated	44.1
23		14.1	76	4.85	1273	>51

Table 1. Notably active chemotype **C** which show no notable cytotoxicity ($IC_{50} \leq 30$ µM). [a]IC_{50}: half-maximal inhibitory concentration values obtained from the CPE assay in 8-point dose response, measured in duplicate. [b] Values represent data obtained from 5-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 lysates reported in minutes, minimum detectable half-life of 1 minute. [e]PAMPA (parallel artificial membrane permeation assay) is reported as a metric of the passive permeability of the compounds. [f]Solubility – pION µSOL assay for kinetic aqueous solubility determination, pH 7.4.

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

Table 2. Compounds identified as ACE2 binders and inhibitors of viral entry in PP assay.

[a]Activity in the SARS-CoV-2 PP assay. [b]IC_{50}: half maximal inhibitory concentration values obtained from the CPE assay in 8-point dose response, measured in duplicate. [c]Values represent data obtained from 5-point dose response, measured in duplicate. [d]ACE2 binding affinity (K_d) measured by MST.

Fig. 4: Dose-response curves of the six ACE2 binding compounds in PP and cytotoxicity assays.

a, compound **1**; **b,** compound **2**; **c,** compound **5**; **d,** compound **24**; **e,** compound **25**; **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; Tox – cytotoxicity assay.
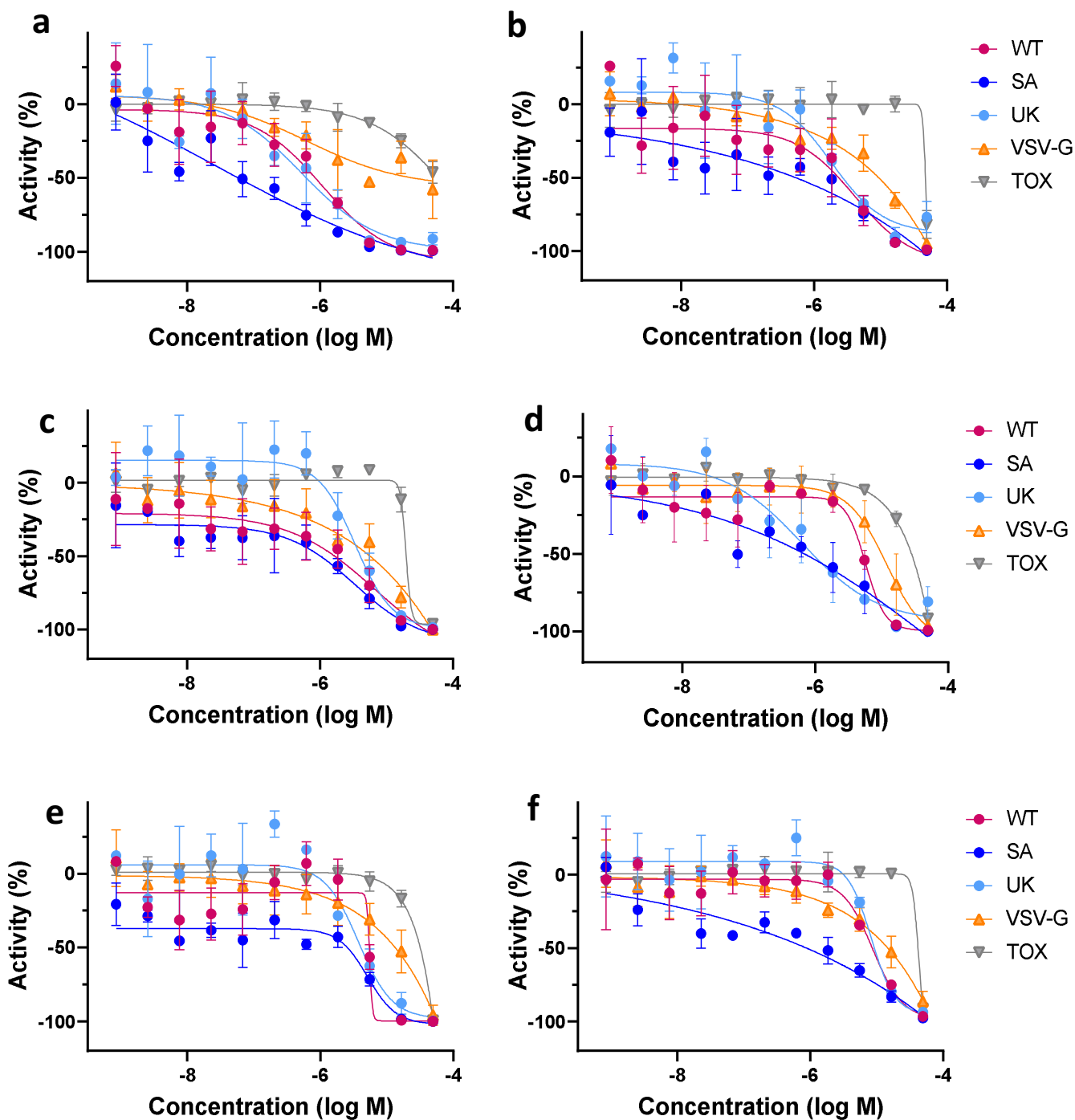
Supplementary Information

Supplementary Table 1-8

Supplementary Pharmacophore models

Acknowledgement

References

1. A pneumonia outbreak associated with a new coronavirus of probable bat origin | Nature. https://www.nature.com/articles/s41586-020-2012-7.

2. Home. *Johns Hopkins Coronavirus Resource Center* https://coronavirus.jhu.edu/.

3. Saul, S. & Einav, S. Old Drugs for a New Virus: Repurposed Approaches for Combating COVID-19. *ACS Infect Dis* (2020) doi:10.1021/acsinfecdis.0c00343.

4. Elfiky, A. A. Anti-HCV, nucleotide inhibitors, repurposing against COVID-19. *Life Sci.* **248**, 117477 (2020).

5. A, E. & Ns, I. Anti-SARS and anti-HCV drugs repurposing against the Papain-like protease of the newly emerged coronavirus (2019-nCoV). (2020) doi:10.21203/rs.2.23280/v1.

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.* **24**, 940–976 (2020).

7. CDC. SARS-CoV-2 Variant Classifications and Definitions. *Centers for Disease Control and Prevention* https://www.cdc.gov/coronavirus/2019-ncov/cases-updates/variant-surveillance/variant-info.html (2020).

8. Variant Therapeutic Data Summary. https://opendata.ncats.nih.gov/variant/summary.

9. Brimacombe, K. R. *et al.* An OpenData portal to share COVID-19 drug repurposing data in real time. *bioRxiv* 2020.06.04.135046 (2020) doi:10.1101/2020.06.04.135046.

10. Chen, C. Z. *et al.* Drug Repurposing Screen for Compounds Inhibiting the Cytopathic Effect of SARS-CoV-2. *Front. Pharmacol.* **11**, (2021).

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

12. Good, A. 4.19 - Virtual Screening. in *Comprehensive Medicinal Chemistry II* (eds. Taylor, J. B. & Triggle, D. J.) 459–494 (Elsevier, 2007). 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 Inform* **39**, e2000028 (2020).

14. Zhang, H. *et al.* Deep Learning Based Drug Screening for Novel Coronavirus 2019-nCov. *Interdiscip Sci* 1–9 (2020) doi:10.1007/s12539-020-00376-6.

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* **7**, 6642–6660 (2015).

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. *Journal of Molecular Recognition* **30**, e2644 (2017).

17. Xu, C. *et al.* Systemic In Silico Screening in Drug Discovery for Coronavirus Disease (COVID-19) with an Online Interactive Web Server. *J. Chem. Inf. Model.* (2020) doi:10.1021/acs.jcim.0c00821.

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. *Computational and Structural Biotechnology Journal* **19**, 2537–2548 (2021).

19. Zhou, Y. *et al.* Network-based drug repurposing for novel coronavirus 2019-nCoV/SARS-CoV-2. *Cell Discovery* **6**, 1–18 (2020).

20. 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. (2020).

21. Alamri, M. A. & Alamri, M. A. Pharmacophore and docking-based sequential virtual screening for the identification of novel Sigma 1 receptor ligands. *Bioinformation* **15**, 586–595 (2019).

22. Vittorio, S. *et al.* A Combination of Pharmacophore and Docking-based Virtual Screening to Discover new Tyrosinase Inhibitors. *Molecular Informatics* **39**, 1900054 (2020).

23. Troger, F. *et al.* Identification of mitochondrial toxicants by combined in silico and in vitro studies – A structure-based view on the adverse outcome pathway. *Computational Toxicology* **14**, 100123 (2020).

24. 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.* **31**, 507–521 (2017).

25. Ullrich, S. & Nitsche, C. The SARS-CoV-2 main protease as drug target. *Bioorganic & Medicinal Chemistry Letters* **30**, 127377 (2020).

26. Recent Advances in QSAR Studies - Methods and Applications | Tomasz Puzyn | Springer. https://www.springer.com/gp/book/9781402097829.

27. Roy, K., Kar, S. & Das, R. N. Chapter 7 - Validation of QSAR Models. in *Understanding the Basics of QSAR for Applications in Pharmaceutical Sciences and Risk Assessment* (eds. Roy, K., Kar, S. & Das, R. N.) 231–289 (Academic Press, 2015). doi:10.1016/B978-0-12-801505-6.00007-7.

28. 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* **4**, 57–66 (2010).

29. 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.* **32**, 583–590 (2018).

30. Breiman, L. Random Forests. *Machine Learning* **45**, 5–32 (2001).

31. 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* 154–168 (Springer, Berlin, Heidelberg, 2012). doi:10.1007/978-3-642-31537-4_13.

32. Zakharov, A. V. *et al.* QSAR Modeling and Prediction of Drug–Drug Interactions. *Mol. Pharmaceutics* **13**, 545–556 (2016).

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.04.447130; this version posted June 4, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. This article is a US Government work. It is not subject to copyright under 17 USC 105 and is also made available for use under a CC0 license.

33. Jain, S. *et al.* 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.

34. Friedrich, N.-O. *et al.* Benchmarking Commercial Conformer Ensemble Generators. *J Chem Inf Model* **57**, 2719–2728 (2017).

35. Berthold, M. R. *et al.* KNIME - the Konstanz Information Miner: Version 2.0 and Beyond. *SIGKDD Explor. Newsl.* **11**, 26–31 (2009).

36. Huang, R. *et al.* Biological activity-based modeling identifies antiviral leads against SARS-CoV-2. *Nature Biotechnology* 1–7 (2021) doi:10.1038/s41587-021-00839-1.

37. Brown, J. B. Classifiers and their Metrics Quantified. *Molecular Informatics* **37**, 1700127 (2018).

38. Zhu, W. *et al.* Identification of SARS-CoV-2 3CL Protease Inhibitors by a Quantitative High-Throughput Screening. *ACS Pharmacol. Transl. Sci.* **3**, 1008–1016 (2020).

39. Hanson, Q. M. *et al.* 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.* **3**, 1352–1360 (2020).