

# PolypharmDB, a deep learning-based resource, quickly identifies repurposed drug candidates for COVID-19

Dar'ya S. Redka<sup>1</sup>, Stephen S. MacKinnon<sup>1</sup>, Melissa Landon<sup>1</sup>, Andreas Windemuth<sup>1</sup>, Naheed Kurji<sup>1</sup>, Vijay Shahani<sup>1,2</sup>

<sup>1</sup>Applied Science, Cyclica Inc., Toronto, Canada

<sup>2</sup>Corresponding author: vijay.shahani@cyclicarx.com

## Abstract

There is an immediate need to discover treatments for COVID-19, the pandemic caused by the SARS-CoV-2 virus. Standard small molecule drug discovery workflows that start with library screens are an impractical path forward given the timelines to discover, develop, and test clinically. To accelerate the time to patient testing, here we explored the therapeutic potential of small molecule drugs that have been tested to some degree in a clinical environment, including approved medications, as possible therapeutic interventions for COVID-19. Motivating our process is a concept termed polypharmacology, i.e. off-target interactions that may hold therapeutic potential. In this work, we used Ligand Design, our deep learning drug design platform, to interrogate the polypharmacological profiles of an internal collection of small molecule drugs with federal approval or going through clinical trials, with the goal of identifying molecules predicted to modulate targets relevant for COVID-19 treatment. Resulting from our efforts is PolypharmDB, a resource of drugs and their predicted binding of protein targets across the human proteome. Mining PolypharmDB yielded molecules predicted to interact with human and viral drug targets for COVID-19, including host proteins linked to viral entry and proliferation and key viral proteins associated with the virus life-cycle. Further, we assembled a collection of prioritized approved drugs for two specific host-targets, TMPRSS2 and cathepsin B, whose joint inhibition was recently shown to block SARS-CoV-2 virus entry into host cells. Overall, we demonstrate that our approach facilitates rapid response, identifying 30 prioritized candidates for testing for their possible use as anti-COVID drugs. Using the PolypharmDB resource, it is possible to identify repurposed drug candidates for newly discovered targets within a single work day. We are making a complete list of the molecules we identified available at no cost to partners with the ability to test them for efficacy, in vitro and/or clinically.

**Keywords:** SARS-CoV-2 virus, COVID-19, coronavirus, TMPRSS2, Cathepsin B, host-target, polypharmacology, off-target interactions

### **Abbreviations:**

SARS-CoV-2: Severe acute respiratory syndrome-related coronavirus

COVID-19: Coronavirus Disease-2019

3CLpro: papain-like protease

PLpro: main protease

RdRp: nonstructural protein

ACE2: angiotensin converting enzyme 2

TMPRSS2: transmembrane protease serine 2

## Introduction

Severe acute respiratory syndrome-related coronavirus SARS-CoV-2 (formerly referred to as 2019-nCoV), causes the serious novel coronavirus disease COVID-19 and represents a major global health risk [1]. SARS-CoV-2 belongs to the coronavirus family found in both animals and humans. This family is responsible for infections ranging from the common cold to the Severe Acute Respiratory Syndrome (SARS) and Middle East Respiratory Syndrome (MERS) [2]. The death toll from the SARS-CoV-2 outbreak has quickly surpassed that of the 2003 SARS Epidemic (774), with the number of confirmed cases escalating worldwide at an unprecedented rate [3].

SARS-CoV-2 is a positive-sense single-stranded RNA virus within a membrane envelope [4]. Consistent with other coronaviruses, a distinguishing feature of SARS-CoV-2, is its crown of Spike proteins, a key unit that mediates viral entry through its interaction with the host protein ACE2 [5,6]. Other key non-structural proteins include viral proteases 3C-like protease (3CLpro) and papain-like protease (PLpro), responsible for cleaving viral polypeptides into functional units [7,8]. Further, PLpro has demonstrated deubiquitinase activity and has been linked to suppressing key immunogenic signalling molecules like NF-kappaB and interferon factor 3 [9,10]. Lastly, RNA-dependent RNA polymerase (RdRp) is critical for viral proliferation by replicating genomic RNA for nascent viruses [7].

Currently no treatments for SARS-CoV-2 exist. In the last few months a rapid global response aimed at the discovery of therapeutic agents has yielded proposals that include pre- and post-exposure prophylaxis, vaccines, targeted treatments, and medicines to ameliorate symptoms [11–13]. Unfortunately, though vaccine development has accelerated greatly through the herculean efforts of researchers, the time required to obtain safety and efficacy data diminishes their impact on the current pandemic. Developing novel drugs to interfere with SARS-CoV-2 pathogenesis directly is a desirable long-term endeavor, but would have limited impact on the current crisis. Therefore, treating COVID-19 in the near term must be accomplished with the methods and medicines at our current disposal, calling for a repurposing strategy.

Initial repurposing efforts have focused on existing antivirals. One of the most promising programs in clinical testing is Gilead's remdesivir, a 1'-cyano-substituted adenosine nucleotide analogue prodrug with established activity for ebola RdRp [14]. While the use of existing antivirals holds promise, there is no guarantee that these treatments will be effective, a point illustrated by HIV drugs lopinavir and ritonavir, a joint treatment that proved ineffective in a clinical trial with 199 SARS-CoV-2 patients [15]. Other molecules have been identified, including chloroquine and hydroxychloroquine, both antimalarials that have been suggested as capable of reducing symptoms and decreasing recovery time, though its mechanism of action is poorly understood [13,16]. Given the uncertainty of the therapeutic potential of repurposed antivirals for COVID-19, discovery of additional effective medicines is of paramount importance. SARS-CoV-2 has several viral proteins, including the Spike protein, 3CLpro, PLPro, and RdRp, identified as putative drug targets. Importantly, a number of host proteins attenuate viral progression and may serve as suitable drug targets in their own right, as discussed below.

While most drugs are designed to target a single protein, they necessarily interact with other proteins (e.g., transports, xenobiotic metabolizers, carriers, etc.). In fact, it is anticipated that typical xenobiotics interact with 30 to 300 proteins [17]; this multiple activity is termed *polypharmacology*. In the context of this work we exploit the notion of polypharmacology to uncover interactions of existing approved and investigational medications with either SARS-CoV-2 proteins or host protein targets. Based on this

strategy, herein, we outline a rapid approach to prioritizing possible interventions for COVID-19, with the flexibility to adapt to additional data as they become available. To execute this strategy three major elements are required: 1) a curated dataset of molecules with associated safety data 2) a set of putative targets and their structures, and 3) a discriminator in the form of a deep learning engine.

Our curated dataset of small molecules includes approved medicines and a collection of experimental medications with clinical safety data. In our initial efforts, and in the absence of direct SARS-CoV-2 targets and structures, we assembled two distinct panels of protein targets: 1) a panel of potential host human proteins that have been shown to be involved for either SARS-CoV or MERS-CoV infections, and 2) a panel of SARS-CoV-2 viral protein target, based in the homology models of SARS-CoV. Our discriminator is Cyclica Inc.'s Ligand Design platform, a deep learning approach capable of supporting both repurposing efforts as well as *de novo* drug discovery. As a result, we produced a collection of drug-target interactions, termed PolypharmDB that is ready to be mined for molecules with predicted activities against specific targets relevant to SARS-CoV-2, both current targets and those yet to be validated experimentally.

Notably, a recent study identified two desirable host targets for SARS-CoV-2, which were already present in our panel: TMPRSS2 and cathepsin B. The authors highlighted the host-target, TMPRSS2, as a possible standalone target, but also alluded to the potential of a combination approach with molecules targeted at cathepsin B [18]. The ability to find possible candidates through PolypharmDB highlights its utility in a situation where information is updating rapidly.

## Methods

### Target selection

#### Host targets

In the absence of the experimental data specific to SARS-CoV-2 human targets, we conducted a literature review of human host proteins involved in SARS-CoV or MERS-CoV viral infections. This effort revealed a collection of 15 potential therapeutic targets for the SARS-CoV-2 viral infection (Table 1). These targets served as initial priority targets.

Of the selected host targets, three are members of the transmembrane serine protease family (TMPRSS2, TMPRSS11A, and TMPRSS4), which prime the Spike protein for its interaction with ACE2, a process required for viral entry [19–23]. Specifically, studies suggest TMPRSS2 activity is crucial for the spread of SARS-CoV and MERS-CoV in the airways of murine models [23]. Studies suggest that endosomal cathepsins and furin are also capable of priming the Spike protein of SARS-like viruses [24], and therefore we included three endosomal cathepsins (cathepsin L, B, and S) and furin in our target panel.

Other targets include the TCDD-inducible poly(ADP-ribose) polymerase (TIPARP) and its regulator Aryl Hydrocarbon Receptor (AHR) AHR, a nuclear receptor/transcription factor. AHR has been shown to be activated in response to coronavirus infection, leading to upregulation of its downstream effector, TIPARP [25]. TIPARP in turn facilitates viral replication and reduces interferon expression, suggesting that therapeutic inhibitors of either AHR and or TIPARP may inhibit viral life cycle [25].

Other proteases that were shown to be involved in priming SARS-CoV or MERS-CoV viral spike-driven entry into host cells, and therefore included in targets panel are Aminopeptidase N (ANPEP) [26],

Dipeptidyl Peptidase 4 (DPP4, also known as CD26) [27], and its scaffolding protein tetraspanin CD9 [28], and ADAM metallopeptidase domain 17 (ADAM17; also known as tumor necrosis factor- $\alpha$ -converting enzyme (TACE) [29].

**Table 1:** We credit the scientific community for enabling our target selection who, following the SARS-CoV and MERS-CoV outbreaks, used a variety of approaches to identify potential human and non-human therapeutic targets. Resulting from these studies are several promising human target candidates.

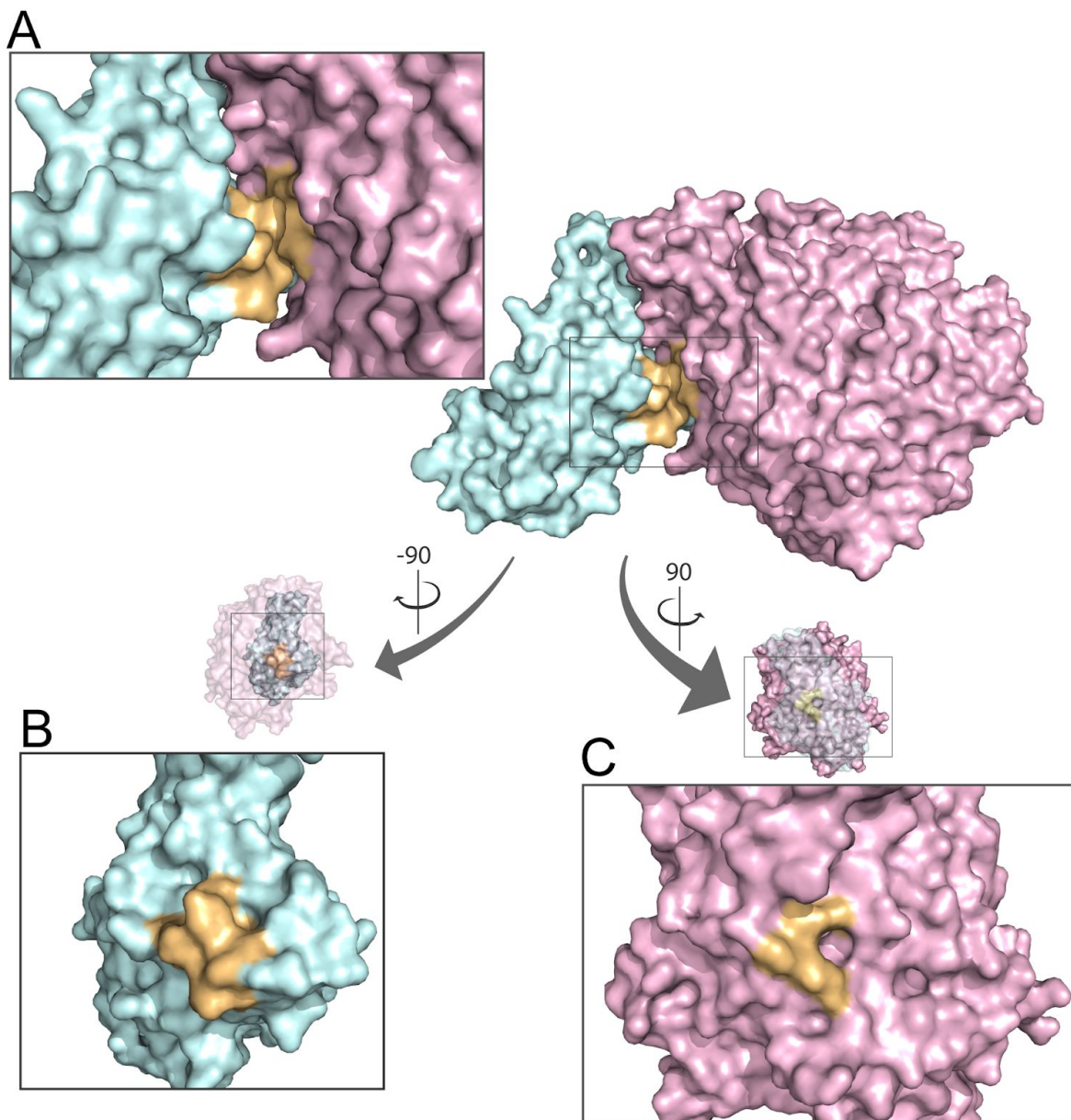
Human Targets	Uniprot name	Involvement	Reference
TCDD-inducible poly(ADP-ribose) polymerase (TIPARP) and Aryl Hydrocarbon Receptor (AHR)	AHR_HUMAN, PARPT_HUMAN	viral replication	[23]
Dipeptidyl Peptidase 4 and its scaffolding protein tetraspanin CD9	DPP4_HUMAN, CD9_HUMAN	facilitation of viral entry	[27] [28]
Transmembrane serine proteases	TMPS2_HUMAN, TMPS4_HUMAN, TM11A_HUMAN	facilitation of viral entry	[23] [22]
Endosomal cathepsins and furin proteases	CATL1_HUMAN, CATL2_HUMAN, CATB_HUMAN, CATS_HUMAN, FURIN_HUMAN	facilitation of viral entry	[24]
Aminopeptidase N	AMPN_HUMAN	viral entry receptor	[26]
ADAM metallopeptidase domain 17 (ADAM17), also known as tumor necrosis factor- $\alpha$ -converting enzyme (TACE)	ADA17_HUMAN	facilitation viral entry	[29]
Phospholipase A2 $\alpha$	PA24A_HUMAN	viral replication	[26]

## Viral targets and interactions with host protein ACE2

To facilitate predictions with Cyclica's structure based approach, we modelled SARS-CoV-2 protein structures from SARS-CoV and templates of other related viruses. While SARS-CoV-2 genome codes for more than a dozen proteins, we focused on 3CLpro and Spike protein for small molecule intervention given the existence of high quality template structures.

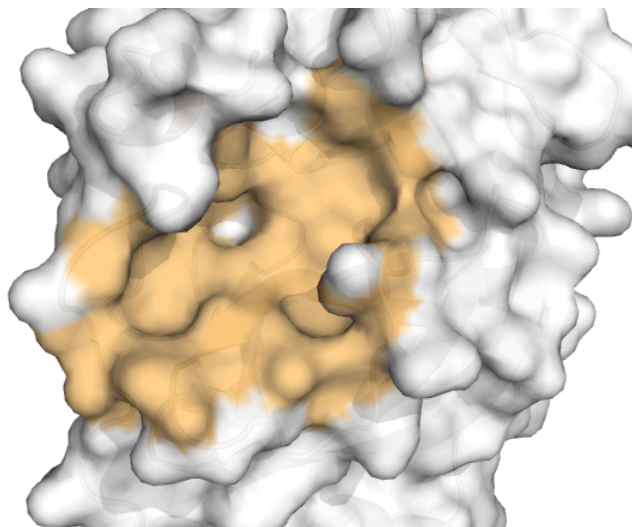
We used SwissModel [30] to model the Spike protein of SARS-CoV-2 in complex with the human ACE2 receptor (Figure 1), based on a known structure of the same complex for Spike-SARS-CoV [31] ([PDB ID: 2AJF](#)), which has a 77.8% sequence identity to Spike-SARS-CoV-2. Amino acid sequences of the SARS-CoV-1 Spike protein and human ACE2 were extracted from the PDB record (2AJF, chains E, F and A, B, respectively) and used to query (via BLASTP) the betacoronavirus database on the NCBI BLAST server on Feb 6, 2020. The top matches belonging to SARS-CoV-2 were then used to create a 2-chain model with PDB entry 2ajf as the template, using the SwissModel protein modeling service. The crystal structure of the Spike-SARS-CoV-2 in complex with ACE2 has been released recently ([PDB ID: 6LZG](#); to be published), but it was not available at the time of this work. We evaluated the similarity of our homology model of the ACE2-Spike-SARS-CoV-2 and the crystal structure within the 10 Å radius of the predicted binding pockets, and found the RMSD of 0.35 Å over all Ca atoms and the RMSD of 0.79 Å for all heavy atoms between the two structures.

The resulting structural model was divided into 3 different pocket models, a required input for Ligand Design. A pocket model consists of a protein structure combined with a model ligand. The latter can be any chemical structure and serves only to indicate the approximate position of the hypothesized binding location. Here, for simplicity, we use one of the amino acids located at the protein/protein binding interface as the model ligand. The three pocket models investigated are: Spike protein (E-chain) site defined by ACE2 LYS-353 (A chain), ACE2 (A chain) site defined by S TYR-491 (E chain), and the S - ACE2 complex (A+E chains) site defined by ACE2 LYS-353 (A chain). These correspond to three different binding hypotheses for a potential ligand, where the ligand binds to either of the two proteins separately, or to the common interface.



**Figure 1:** The interaction (A) between Spike-SARS-CoV-2 (blue, B) and ACE2 (red, C). The area highlighted in yellow shows a 5 Å radius surrounding the intended binding site of the drug.

In the case of 3CLpro from SARS-Cov-2, we used the available homology model published recently by another group (Figure 2) [32]. That model was based on the structure of 3CLpro SARS-CoV (PDB ID: 2a5i), which has 96% sequence identity to the 3CLpro SARS-CoV-2. For comparison, we included four structures of 3CLpro SARS-CoV in our screen (PDB IDs: 2D2D, 3TIU, 3TNSns, 4OW0), and found a similar list of top binding molecules for those targets (data not shown).



**Figure 2:** The active site of 3CLpro-SARS-CoV-2 (homology model). The area in yellow highlights the designated binding site of the drug.

## Screening Procedure

Ligand Design (LD) is a flexible deep learning platform for identifying molecules that best satisfy a defined objective function. Briefly, LD traverses a chemical space iteratively by repeatedly deriving new molecules (children) from previous selections (parents) and then filtering them down according to the objective function. The fit to the objective function improves with each generation, until convergence, when no better molecules can be found. Chemical spaces available for searching range from fixed libraries of thousands or millions of molecules (notably, when operating in this fashion LD only explores a single generation consisting of all molecules in the fixed library), to fully generative chemical spaces of  $10^{30}$ ,  $10^{60}$ , or  $10^{180}$  molecules depending on the method used to enumerate chemical space [33].

Objective functions can be simple computed physicochemical properties, ligand-based predictive models (such as for ADME or toxicity) and predicted binding to one or more target or anti-target proteins. In Ligand Design protein binding predictions are performed with MatchMaker, a deep learning model trained on the entire human proteome that combines structural and experimental data to predict the binding of small molecules to protein “pockets”, which are defined by the protein structure (experimental or modeled) and a location at which ligands may bind. MatchMaker has been found to be far more accurate than molecular docking, and is also several orders of magnitudes more computationally efficient [34,35]. Target binding objective functions do not focus on just one protein, but take the entire polypharmacological profile of a molecule into account. In the simplest case, as in this study, the objective function can simply be for the molecule to bind one target and not bind all others. The hosts and

viral targets identified were designated as desirable binders to be assessed against all molecules in our fixed library.

## PolypharmDB Assembly

In order to expedite time to the patient, we performed a LD fixed library screen of 10,224 drugs in the DrugBank database against ~8700 proteins, including our human and viral targets of interest. The resulting polypharmacological database, PolypharmDB, contained the predicted binding profiles of 2118 drugs that were approved (and not withdrawn) for use in the market, 2242 drugs in clinical trials (but not yet approved), and 5547 molecules that are in pre-clinical exploration (listed as “Experimental” in the DrugBank database) or are nutraceuticals. A distinguishing feature of PolypharmDB is that each molecule is linked to its polypharmacology, enabling the analysis of both on- and off-target interactions. PolypharmDB can be leveraged to concentrate on the approved small molecule drugs only (see *Results*), or to explore all of the small molecules drugs present in the DrugBank database. Further, desired protein targets can be added or removed on the fly, allowing rapid identification of molecules for novel therapeutic targets.

## Re-prioritized Host-targets

Since our initial analysis of this panel, TMPRSS2 has been shown to be essential for the SARS-CoV-2 viral entry by activating the spike-protein [18], and has been highlighted as an important potential therapeutic target for COVID-19 [36]. TMPRSS11A has been suggested to have more localized expression in the respiratory tract [22], and therefore may also be a worthwhile target to consider for SARS-CoV-2. Interestingly, since our initial investigation, cathepsins L and B were shown to complement the activity of TMPRSS2 in facilitating host cell entry for SARS-CoV-2, with full suppression of viral entry accomplished with their dual suppression [18].

## Results

### Viral Targets: Top Repurposing Candidates

We attempted to predict molecules that would hinder the functionality of two SARS-CoV-2 proteins, 3CLpro and Spike protein, in several ways. First, we mined PolypharmDB for molecules that interacted with the active sites of 3CLpro, based on previously reported structures of the active site of SARS-CoV-1 (e.g., PBD IDs: 2D2D, 3TIU, 3TNS, 4OW0) (Figure 2). Second, we predicted molecules that would compete with binding of Spike protein to the human ACE2 protein by exploring: 1) the Spike-binding site on ACE2 (Figure 1B), 2) the ACE2-binding site on Spike protein (Figure 1C), and 3) the interface of ACE2 and Spike proteins (Figure 1A).

Table 2 shows the top five approved drugs that we predict to interact with the active site of 3Cpro. Notably, 3 out of those drugs belong to the mTOR-signalling pathway (temsirolimus, sirolimus, everolimus). The extended list of top scoring drugs for 3CLpro, including investigational molecules in addition to approved (see Supplemental Data S1).

Table 2 also shows the top scoring approved drugs for the site of interaction between ACE2 and the Spike protein. Interestingly, both the screen for the Spike-binding site of ACE2 alone (Figure 1B), and the screen for the site that included structural information for both ACE2 and Spike-protein (Figure 1C), produced a



list of top binders populated with cephalosporin-like and penicillin-like antibiotics. Screening of the ACE2-binding site of the Spike protein (Figure 1A) produced a smaller number of high scoring molecules, and therefore only top 3 approved drugs are shown in Table 2. The extended lists of top predicted binders including non-approved drugs are shown in Supplemental Data S1.

**Table 2:** Top hits (approved) for 3CLpro and the interaction of ACE2 and Spike (SARS-CoV-2).

Drug Name	Clinical Description
<b>3CLpro-SARS-CoV-2</b>	
Temsirolimus	A antineoplastic agent used in the treatment of renal cell carcinoma (RCC) that works by inhibiting mTOR.
Sirolimus	An immunosuppressant used to prevent organ transplant rejections and to treat lymphangioleiomyomatosis.
Bremelanotide	A small peptide used to treat hypoactive sexual desire disorder in premenopausal women
Everolimus	A mammalian target of rapamycin (mTOR) kinase inhibitor used to treat various types of malignancies.
Deutetrabenazine	A vesicular monoamine transporter 2 inhibitor used for the symptomatic treatment of chorea in Huntington's disease and tardive dyskinesia.
<b>ACE2 (anticipated Spike protein binding site)</b>	
Latamoxef	A broad-spectrum beta-lactam antibiotic similar in structure to the cephalosporins except for the substitution of an oxazabicyclo moiety for the thiazabicyclo moiety of certain cephalosporins.
Cefazolin	A broad-spectrum cephalosporin antibiotic mainly used for the treatment of skin bacterial infections and other moderate to severe bacterial infections in the lung, bone, joint, stomach, blood, heart valve, and urinary tract.
Cefoxitin	A semi-synthetic, broad-spectrum antibiotic for parenteral administration used for the treatment of serious bacterial infections, such as urinary tract infection, blood infection, bone and joint infection, and lower respiratory tract infection.
Bacampicillin	A prodrug of ampicillin and is microbiologically inactive. It is absorbed following oral administration
Pheneticillin	A penicillin antibiotic which is approved for use internationally.
<b>Spike-SARS-CoV-2</b>	
Enoxacin	A broad-spectrum 6-fluoronaphthyridinone antibacterial agent (fluoroquinolones) structurally related to nalidixic acid.
Linezolid	An oxazolidinone antibiotic used to treat susceptible strains of multidrug resistant bacterial infections.
Colchicine	A medication used to treat gout and Behçet's disease
<b>ACE2-Spike-SARS-CoV-2 Interaction</b>	
Pheneticillin	A penicillin antibiotic which is approved for use internationally.
Ticarcillin	A beta lactam antibiotic used to treat a variety of infections in the body.
Cefuroxime	A cephalosporin indicated for the treatment of a variety of infections including acute bacterial otitis media, several upper respiratory tract infections, skin infections, urinary tract infections, gonorrhea, early Lyme disease, and impetigo.
Mezlocillin	A semisynthetic ampicillin-derived acylureido penicillin
Meropenem	A carbapenem antibiotic used to treat a wide variety of infections in the body.

## Human Targets: Top Repurposing Candidates

As described in *Methods*, from PolypharmDB we explored molecules that are predicted to interact with any of the 15 human targets that have been shown to be involved in SARS-CoV and MERS-CoV infections. Out of 15 human targets we initially selected, 13 targets produced molecules with a significantly high rank



( $p < 0.03$ ). The top approved binders for 11 out of those targets are shown in the Supplemental Data S2, and the extended list of top hits can be found in the Supplemental Data S3. Since the completion of our initial PolypharmDB analysis, two of the human targets, TMPRSS2 and cathepsin B, that we have selected based on their involvement in SARS-CoV and MERS-CoV [23,24], were shown to be essential for the infectivity of SARS-CoV-2 [18]. Those two targets are highlighted in the next section, and are excluded from the Supplemental Data S2 and S3.

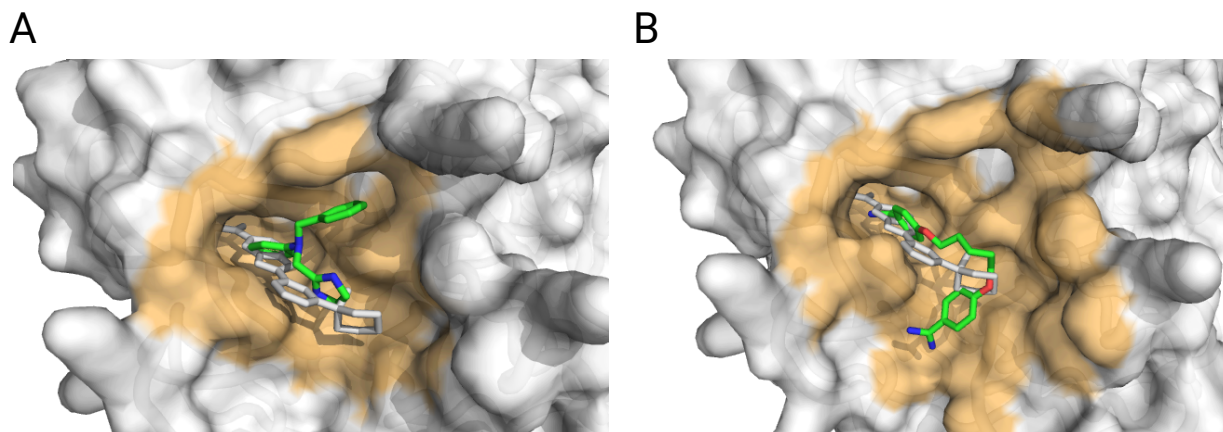
### TMPRSS2 + Cathepsin B: Top Repurposing Candidates

A recent publication in Cell showed that TMPRSS2 (Transmembrane serine protease), with support from the endosomal cathepsin B is essential for priming of SARS-CoV-2 for the host-cell entry [18]. TMPRSS2 was also highlighted as an essential host target for SARS-CoV-2 in a recent In Focus article by ACS Central Science [36]. The Cell publication revealed that small molecules were capable of complete suppression of viral entry *in vitro*. Specifically, the authors found that along with inhibition of TMPRSS2 with camostat mesylate, residual viral entry can be abolished with the addition of a cathepsin B inhibitor, E-64d. Unfortunately, given the low potency of camostat mesylate, and that E-64d lacks safety data, the combination is unsuitable for human use.

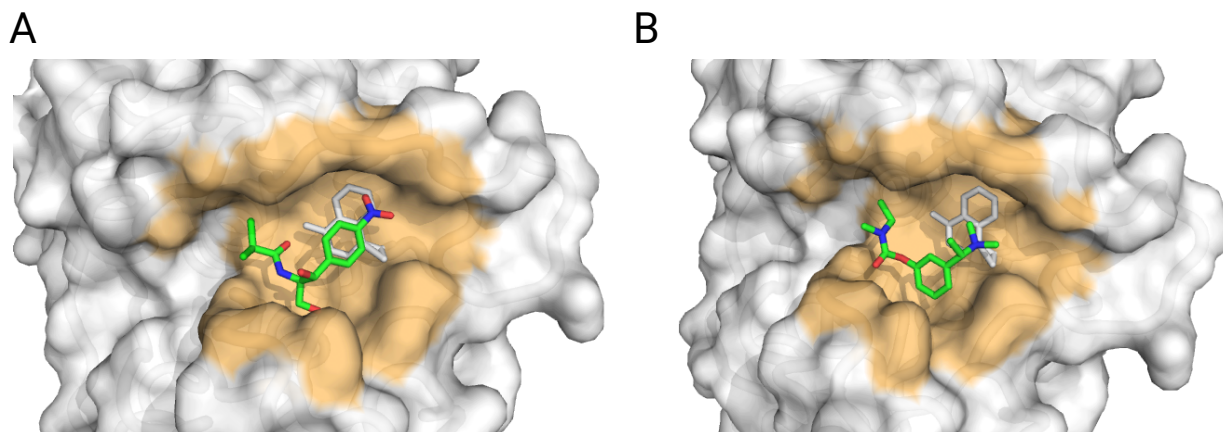
We hypothesized that our deep learning method can discover new combinations of safe medicines to inhibit these targets. Turning back to the original results of mining PolypharmDB, Table 3 highlights possible small molecules that are likely binders to both TMPRSS2 and cathepsin B. Furthermore, we explored PolypharmDB for any candidates that may bind both targets, so called dual-targeting agents. To find these molecules, we used a scoring method we refer to as the Expectation Value Score (EV-score), which uses the rank of a drug-protein interaction to produce a combined score for a selected panel of targets. The molecules identified for the individual targets and using EV-score can be tested individually, and in the case of the TMPRSS2 and cathepsin B targets, drugs can be tested in combination for their capacity in preventing viral entry. Table 3 summarizes these results for the top 10 scoring approved drugs. Notably, the top 10 drugs in each case cover a wide range of classes of drug. The extended list of top scoring molecules, both approved and investigational, can be found in Supplemental Data S4.

One of the top scoring molecules for TMPRSS2, antazoline, a histamine H1 receptors antagonist, is displayed docked to its target in Figure 3A. Pentamidine, an anti-parasitic and anti-fungal agent used to treat trypanosomiasis, leishmaniasis, and some fungal infections, particularly pneumocystis pneumonia in HIV-infected patients [37], also ranked highly for TMPRSS2 and is shown docked in Figure 3B. Chloramphenicol, a broad spectrum antibiotic, which appears both in the top hits list for cathepsin B and the top EV-score list for both targets, is shown in Figure 4A docked to its predicted binding site of cathepsin B. A parasymphomimetic agent, rivastigmine, used in treatment of Alzheimer's related dementia, also appears in the top hits list for both cathepsin B and EV-score for both targets, and its displayed docked to its predicted binding site on cathepsin B in Figure 4B.

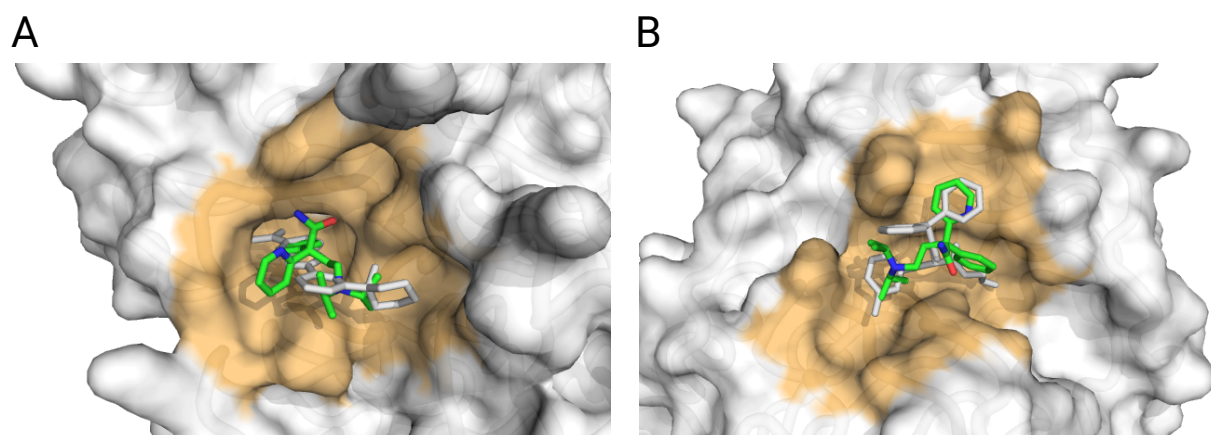
One drug, disopyramide, an anti-arrhythmic agent with anticholinergic and anesthetic properties is shared by all three lists (*i.e.*, TMPRSS2, cathepsin B, and the combination of the two), and is shown docked to both targets in Figure 5. Disopyramide has been reported to inhibit multiple human targets, such as the muscarinic acetylcholine receptors [38], the voltage-gated sodium channels [39], and the voltage-gated potassium channels [40]. Notably, disopyramide is approved as a long term therapy with tolerable anticholinergic side effects [41].



**Figure 3:** Predicted poses of antazoline (A) and pentamidine (B) within the TMPRSS2 proposed binding site (homology model). The area highlighted in yellow shows a 5 Å radius surrounding the docked molecule. The reference ligand that was used to define the binding pocket, based on the homology model is displayed in gray. Docked poses are for illustrative purposes only and do not influence the rank of the molecule.



**Figure 4:** Chloramphenicol (A) and rivastigmine (B) structures posed within the cathepsin B site. The area highlighted in yellow shows a 5 Å radius surrounding the docked molecule. The reference ligand that was used to define the binding pocket, based on the crystal structure is displayed in gray. Docked poses are for illustrative purposes only and do not influence the rank of the molecule.



**Figure 5:** Disopyramide with predicted poses within the TMPRSS2 binding site (A; homology model of the protein) and to cathepsin B binding site (B). The area highlighted in yellow shows a 5 Å radius surrounding the docked molecule. The reference ligand that was used to define the binding pocket, based either on the homology model (TMPRSS2) or the crystal structure (cathepsin B) is displayed in gray. Docked poses are for illustrative purposes only and do not influence the rank of the molecule.

**Table 3:** Top 10 approved drugs for TMPRSS2, cathepsin B, and combination of the two.

TMPRSS2	
Drug Name	Clinical Description
Antazoline	An antihistamine agent used for the symptomatic treatment of nasal congestion and allergic conjunctivitis.
Pentamidine	An antifungal agent used to treat <i>Pneumocystis pneumonia</i> in patients infected with HIV.
Tolazoline	A vasodilator used to treat pulmonary artery anomalies.
Dihydralazine	An antihypertensive agent; a peripheral vasodilator
Histrelin	A GnRH agonist found in subcutaneous implants used for the treatment of pediatric patients with central precocious puberty and the palliative treatment of advanced prostate cancer.
Naphazoline	A sympathomimetic vasoconstrictor used for the symptomatic relief of redness and itching of the eye, and nasal congestion.
Disopyramide	A class 1A antiarrhythmic agent used to treat life-threatening ventricular arrhythmias.
Hydroxystilbamidine	Hydroxystilbamidine isethionate is used in the therapy of some patients with nonprogressive blastomycosis of the skin, and pulmonary or systemic blastomycosis in children, with fewer side effects than amphotericin B. Hydroxystilbamidine isethionate is also used in pathology for diagnostic purposes.
Anastrozole	A competitive, selective, non-steroidal aromatase inhibitor used as adjuvant therapy for the treatment of hormone receptor-positive breast cancer in postmenopausal women.
Amlodipine	A calcium channel blocker used to treat hypertension and angina.
Cathepsin B	
Drug Name	Clinical Description
Isocarboxazid	A monoamine oxidase inhibitor used to treat enduring and debilitating symptoms of depression following inadequate clinical response to other antidepressant drugs.
Chloramphenicol	A broad spectrum antibiotic that is effective against a variety of susceptible and serious bacterial infections but is not frequently used because of its high risk of bone marrow toxicity.
Phenacemide	Used to control certain seizures in the treatment of epilepsy. This medicine acts on the central nervous system (CNS) to reduce the number and severity of seizures.
Rivastigmine	A cholinesterase inhibitor used to treat mild to moderate dementia in Alzheimer's and Parkinson's.

Disopyramide	A class 1A antiarrhythmic agent used to treat life-threatening ventricular arrhythmias.
Cefaclor	A second generation cephalosporin used to treat a wide variety of infections in the body.
Phenoxybenzamine	An alpha adrenergic agonist used to treat pheochromocytoma and episodes of hypertension and sweating.
Cephalexin	A first generation cephalosporin used to treat certain susceptible bacterial infections.
Selegiline	A monoamine oxidase inhibitor used to treat major depressive disorder and Parkinson's.
Phenelzine	A monoamine oxidase inhibitor used to treat atypical, nonendogenous, or neurotic depression.
<b>Polypharm panel: TMPRSS2 and Cathepsin B</b>	
<b>Drug Name</b>	<b>Clinical Description</b>
Disopyramide	A class 1A antiarrhythmic agent used to treat life-threatening ventricular arrhythmias.
Isocarboxazid	A monoamine oxidase inhibitor used to treat enduring and debilitating symptoms of depression following inadequate clinical response to other antidepressant drugs.
Chloramphenicol	A broad spectrum antibiotic that is effective against a variety of susceptible and serious bacterial infections but is not frequently used because of its high risk of bone marrow toxicity.
Phenacemide	Used to control certain seizures in the treatment of epilepsy. This medicine acts on the central nervous system (CNS) to reduce the number and severity of seizures.
Rivastigmine	A cholinesterase inhibitor used to treat mild to moderate dementia in Alzheimer's and Parkinson's.
Cefaclor	A second generation cephalosporin used to treat a wide variety of infections in the body.
Selegiline	A monoamine oxidase inhibitor used to treat major depressive disorder and Parkinson's.
Cephalexin	A first generation cephalosporin used to treat certain susceptible bacterial infections.
Phenelzine	A monoamine oxidase inhibitor used to treat atypical, nonendogenous, or neurotic depression.
Phenoxybenzamine	An alpha adrenergic agonist used to treat pheochromocytoma and episodes of hypertension and sweating.

## Discussion

In this study, we applied our deep learning platform, Ligand Design, to uncover the intrinsic polypharmacology of approved drugs as well as clinically tested drug candidates in order to identify putative therapeutic agents for the ongoing COVID-19 pandemic. Cyclica's MatchMaker™ technology was used to rapidly generate PolypharmDB, a collection of drug molecules and a ranked list of their predicted off-target protein interactions including SARS-CoV-2 viral proteins. Therapeutic drug candidates can be mined from this resource by supplying putative therapeutic targets from either the host or SARS-CoV-2.

While these results are encouraging, notable limitations to our approach exist. MatchMaker was trained on human drug-target interactions, including small molecule structures and resolved 3D-structure for proteins. To the extent that MatchMaker is generalizing beyond the protein structure on which it was trained, signal should exist for detecting potential binders; however, while we have verified this generalization by cross-validation, we have not performed prospective validation. Adding to the uncertainty is the utilization of modelled structures, and the identification of previously undrugged locations, for example the proposed Spike protein-ACE2 interface. Applications of MatchMaker to new species and undocumented binding sites is a topic of ongoing development within Cyclica. This lowers our confidence that MatchMaker correctly predicts small molecule binders for viral proteins. Recently, reliable crystal structures of key SARS-CoV-2 proteins have been solved by the community. We re-evaluated the similarity of the homology model of 3CLpro SARS-CoV-2 to its recently published crystal

structure ([PDB ID: 6lu7](#)), and found the RMSD of 0.55 Å over all Ca atoms and the RMSD of 0.97 Å for all heavy atoms between the two structures within a 10 Å radius of the predicted binding pocket. Results generated by the homology model and the crystal structure have significant overlap (88.7% for the top 100 molecules), lending confidence to the results generated for other modelled structures.

While we are confident that MatchMaker can predict binders as we have prospective validation in other therapeutic areas [42], our initial set of human therapeutic targets for COVID-19 treatment are not validated clinically. Given the uncertainty around therapeutic relevance, we erred on the side of inclusion, allowing researchers to discriminate amongst targets based on their evaluation of supporting literature and their own leading hypotheses. This resulted in a fairly long list of possible targets and repurposing candidates that can be explored by researchers.

The utility of PolypharmDB is its ability to rapidly respond to the growing wealth of research surrounding SARS-CoV-2. The recent identification of TMPRSS2 and cathepsin B is a prime example of how PolypharmDB can be deployed for the rapid identification of drug candidates. To expedite the path to potential treatment, we applied further measures to reduce the candidate pool to approved medicines. Also, we applied the notion of polypharmacology to select drug candidates that may interact with either target, so called “dual-targeted” molecules. However, while this approach is unlikely to discover a potent dual-targeted molecule, it did help uncover additional molecules that may have been overlooked with a single target focus. These targets and the proposed drug molecules represent a unique opportunity to apply therapeutics individually or in combination.

Recently, the Krogan group, in a terrific example of the collaborative efforts underway, reported on experimentally identified human-viral protein-protein interactions [43]. The publication used a combination of viral protein expression and used affinity purification mass spectrometry to identify SARS-CoV-2-human protein protein interactions. Future exploration using PolypharmDB will be based on high priority targets stemming from aforementioned study. Further, we commit to provide candidate repurposed drugs to researchers who approach us with their own priority targets.

Last, recognizing that there is immediate need for treatments, this approach has focused on the repurposing capabilities of Ligand Design through the generation of PolypharmDB. Ligand Design is capable of exploring larger chemical space in the form of fixed libraries, or enumerated libraries like Enamine’s Real Space. Another initiative will look at generating small molecule matter directed at SARS-CoV-2. As with PolypharmDB, we will continue to support long term drug discovery efforts for SARS-CoV-2 and COVID-19.

## Acknowledgments

We would like to express our gratitude to Dr. Shane Climie, for sharing his vast experience and helping us frame our thinking around this project. We would also like to thank Dr. Robert Vernon for helping us assess the homology models used in this paper, Dr. David Kuter for his help assembling our protein structures to support the creation of our figures, and Steve Constable for facilitating our computational workload. And last but certainly not least, a major thank you to all the members of our incredible R&D team who, thanks to their efforts in building our technology, made this methodology a possibility.

## References

- 1 Kupferschmidt, K. and Cohen, J. (2020) Will novel virus go pandemic or be contained? *Science, American Association for the Advancement of Science* **367**, 610–611.
- 2 Su, S., Wong, G., Shi, W., Liu, J., Lai, A. C. K., Zhou, J., Liu, W., Bi, Y. and Gao, G. F. (2016) Epidemiology, Genetic Recombination, and Pathogenesis of Coronaviruses. *Trends Microbiol.* **24**, 490–502.
- 3 Coronavirus disease (COVID-19) Situation Dashboard.
- 4 Masters, P. S. (2006) The Molecular Biology of Coronaviruses. In *Advances in Virus Research*, pp 193–292, Academic Press.
- 5 Du, L., He, Y., Zhou, Y., Liu, S., Zheng, B.-J. and Jiang, S. (2009) The spike protein of SARS-CoV – a target for vaccine and therapeutic development. *Nat. Rev. Microbiol., Nature Publishing Group* **7**, 226–236.
- 6 Wrapp, D., Wang, N., Corbett, K. S., Goldsmith, J. A., Hsieh, C.-L., Abiona, O., Graham, B. S. and McLellan, J. S. (2020) Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation. *Science, American Association for the Advancement of Science* **367**, 1260–1263.
- 7 Ziebuhr, J., Snijder, E. J. and Gorbalenya, A. E. (2000) Virus-encoded proteinases and proteolytic processing in the Nidovirales. *J. Gen. Virol., Microbiology Society*, **81**, 853–879.
- 8 Báez-Santos, Y. M., St John, S. E. and Mesecar, A. D. (2015) The SARS-coronavirus papain-like protease: structure, function and inhibition by designed antiviral compounds. *Antiviral Res.* **115**, 21–38.
- 9 Frieman, M., Ratia, K., Johnston, R. E., Mesecar, A. D. and Baric, R. S. (2009) Severe Acute Respiratory Syndrome Coronavirus Papain-Like Protease Ubiquitin-Like Domain and Catalytic Domain Regulate Antagonism of IRF3 and NF- $\kappa$ B Signaling. *J. Virol., American Society for Microbiology Journals* **83**, 6689–6705.
- 10 Clementz, M. A., Chen, Z., Banach, B. S., Wang, Y., Sun, L., Ratia, K., Baez-Santos, Y. M., Wang, J., Takayama, J., Ghosh, A. K., et al. (2010) Deubiquitinating and Interferon Antagonism Activities of Coronavirus Papain-Like Proteases. *J. Virol., American Society for Microbiology Journals* **84**, 4619–4629.
- 11 Atlanta Site Added to NIH Clinical Trial of a Vaccine for COVID-19 | NIH: National Institute of Allergy and Infectious Diseases.
- 12 Post-exposure Prophylaxis / Preemptive Therapy for SARS-Coronavirus-2 - Full Text View - ClinicalTrials.gov.
- 13 Wang, M., Cao, R., Zhang, L., Yang, X., Liu, J., Xu, M., Shi, Z., Hu, Z., Zhong, W. and Xiao, G. (2020) Remdesivir and chloroquine effectively inhibit the recently emerged novel coronavirus (2019-nCoV) in vitro. *Cell Res., Nature Publishing Group* **30**, 269–271.
- 14 Tchesnokov, E. P., Feng, J. Y., Porter, D. P. and Götte, M. (2019) Mechanism of Inhibition of Ebola Virus RNA-Dependent RNA Polymerase by Remdesivir. *Viruses* **11**.
- 15 Cao, B., Wang, Y., Wen, D., Liu, W., Wang, J., Fan, G., Ruan, L., Song, B., Cai, Y., Wei, M., et al. (2020) A Trial of Lopinavir–Ritonavir in Adults Hospitalized with Severe Covid-19. *N. Engl. J. Med., Massachusetts Medical Society* **0**, null.
- 16 Gao, J., Tian, Z. and Yang, X. (2020) Breakthrough: Chloroquine phosphate has shown apparent efficacy in treatment of COVID-19 associated pneumonia in clinical studies. *Biosci. Trends* **14**, 72–73.
- 17 Zhou, H., Gao, M. and Skolnick, J. (2015) Comprehensive prediction of drug-protein interactions and side effects for the human proteome. *Sci. Rep., Nature Publishing Group* **5**, 1–13.
- 18 Hoffmann, M., Kleine-Weber, H., Schroeder, S., Krüger, N., Herrler, T., Erichsen, S., Schiergens, T. S., Herrler, G., Wu, N.-H., Nitsche, A., et al. (2020) SARS-CoV-2 Cell Entry Depends on ACE2 and TMPRSS2 and Is Blocked by a Clinically Proven Protease Inhibitor. *Cell, Elsevier* **0**.
- 19 Matsuyama, S., Nagata, N., Shirato, K., Kawase, M., Takeda, M. and Taguchi, F. (2010) Efficient Activation of the Severe Acute Respiratory Syndrome Coronavirus Spike Protein by the Transmembrane Protease TMPRSS2. *J. Virol., American Society for Microbiology Journals* **84**, 12658–12664.

- 20 Glowacka, I., Bertram, S., Müller, M. A., Allen, P., Soilleux, E., Pfefferle, S., Steffen, I., Tsegaye, T. S., He, Y., Gnirss, K., et al. (2011) Evidence that TMPRSS2 Activates the Severe Acute Respiratory Syndrome Coronavirus Spike Protein for Membrane Fusion and Reduces Viral Control by the Humoral Immune Response. *J. Virol., American Society for Microbiology Journals* **85**, 4122–4134.
- 21 Shulla, A., Heald-Sargent, T., Subramanya, G., Zhao, J., Perlman, S. and Gallagher, T. (2011) A Transmembrane Serine Protease Is Linked to the Severe Acute Respiratory Syndrome Coronavirus Receptor and Activates Virus Entry. *J. Virol., American Society for Microbiology Journals* **85**, 873–882.
- 22 Zmora, P., Hoffmann, M., Kollmus, H., Moldenhauer, A.-S., Danov, O., Braun, A., Winkler, M., Schughart, K. and Pöhlmann, S. (2018) TMPRSS11A activates the influenza A virus hemagglutinin and the MERS coronavirus spike protein and is insensitive against blockade by HAI-1. *J. Biol. Chem., American Society for Biochemistry and Molecular Biology* **293**, 13863–13873.
- 23 Iwata-Yoshikawa, N., Okamura, T., Shimizu, Y., Hasegawa, H., Takeda, M. and Nagata, N. (2019) TMPRSS2 Contributes to Virus Spread and Immunopathology in the Airways of Murine Models after Coronavirus Infection. *J. Virol., American Society for Microbiology Journals* **93**.
- 24 Najjar, F. E., Lampe, L., Baker, M. L., Wang, L.-F. and Dutch, R. E. (2015) Analysis of Cathepsin and Furin Proteolytic Enzymes Involved in Viral Fusion Protein Activation in Cells of the Bat Reservoir Host. *PLOS ONE, Public Library of Science* **10**, e0115736.
- 25 Grunewald, M. E., Shaban, M. G., Mackin, S. R., Fehr, A. R. and Perlman, S. (2020) Murine Coronavirus Infection Activates the Aryl Hydrocarbon Receptor in an Indoleamine 2,3-Dioxygenase-Independent Manner, Contributing to Cytokine Modulation and Proviral TCDD-Inducible-PARP Expression. *J. Virol., American Society for Microbiology Journals* **94**.
- 26 Santiago, C., Mudgal, G., Reguera, J., Recacha, R., Albrecht, S., Enjuanes, L. and Casasnovas, J. M. (2017) Allosteric inhibition of aminopeptidase N functions related to tumor growth and virus infection. *Sci. Rep., Nature Publishing Group* **7**, 1–14.
- 27 Kleine-Weber, H., Schroeder, S., Krüger, N., Prokscha, A., Naim, H. Y., Müller, M. A., Drosten, C., Pöhlmann, S. and Hoffmann, M. (2020) Polymorphisms in dipeptidyl peptidase 4 reduce host cell entry of Middle East respiratory syndrome coronavirus. *Emerg. Microbes Infect., Taylor & Francis* **9**, 155–168.
- 28 Earnest, J. T., Hantak, M. P., Li, K., Jr, P. B. M., Perlman, S. and Gallagher, T. (2017) The tetraspanin CD9 facilitates MERS-coronavirus entry by scaffolding host cell receptors and proteases. *PLOS Pathog., Public Library of Science* **13**, e1006546.
- 29 Haga, S., Nagata, N., Okamura, T., Yamamoto, N., Sata, T., Yamamoto, N., Sasazuki, T. and Ishizaka, Y. (2010) TACE antagonists blocking ACE2 shedding caused by the spike protein of SARS-CoV are candidate antiviral compounds. *Antiviral Res.* **85**, 551–555.
- 30 Waterhouse, A., Bertoni, M., Bienert, S., Studer, G., Tauriello, G., Gumienny, R., Heer, F. T., de Beer, T. A. P., Rempfer, C., Bordoli, L., et al. (2018) SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Res., Oxford Academic* **46**, W296–W303.
- 31 Li, F., Li, W., Farzan, M. and Harrison, S. C. (2005) Structure of SARS Coronavirus Spike Receptor-Binding Domain Complexed with Receptor. *Science, American Association for the Advancement of Science* **309**, 1864–1868.
- 32 Stoermer, M. (2020) Homology Models of Wuhan Coronavirus 3CLpro Protease, ChemRxiv.
- 33 Gorse, A.-D. (2005, December 31) Diversity in Medicinal Chemistry Space. *Curr. Top. Med. Chem.*
- 34 (2019) Cyclica Validation Note: Proteome screening with Matchmaker achieves performance comparable to experimental techniques.
- 35 (2019) Cyclica Validation Note: MatchMaker: a leap forward in proteome screening beyond molecular docking.
- 36 Liu, C., Zhou, Q., Li, Y., Garner, L. V., Watkins, S. P., Carter, L. J., Smoot, J., Gregg, A. C., Daniels, A. D., Jervey, S., et al. (2020) Research and Development on Therapeutic Agents and Vaccines for COVID-19 and Related Human Coronavirus Diseases. *ACS Cent. Sci., American Chemical Society* **6**, 315–331.
- 37 Huang, Y.-S., Yang, J.-J., Lee, N.-Y., Chen, G.-J., Ko, W.-C., Sun, H.-Y. and Hung, C.-C. (2017) Treatment of *Pneumocystis jirovecii* pneumonia in HIV-infected patients: a review. *Expert Rev. Anti Infect. Ther., Taylor & Francis* **15**, 873–892.



- 38 Yamamoto, N., Ozaki, T., Keida, Y., Ohtsuka, M. and Goto, T. (1999) A comparison of the binding characteristics of class I antiarrhythmic agents for human muscarinic m1-m3 receptors. *J. Cardiovasc. Pharmacol.* **34**, 53–59.
- 39 Sugao, M., Fujiki, A., Nishida, K., Sakabe, M., Tsuneda, T., Iwamoto, J., Mizumaki, K. and Inoue, H. (2005) Repolarization dynamics in patients with idiopathic ventricular fibrillation: pharmacological therapy with bepridil and disopyramide. *J. Cardiovasc. Pharmacol.* **45**, 545–549.
- 40 Casis, O. and Sánchez-Chapula, J. A. (1998) Disopyramide, imipramine, and amitriptyline bind to a common site on the transient outward K<sup>+</sup> channel. *J. Cardiovasc. Pharmacol.* **32**, 521–526.
- 41 Bauman, J. L., Gallastegui, J., Strasberg, B., Swiryn, S., Hoff, J., Welch, W. J. and Bauernfeind, R. A. (1986) Long-term therapy with disopyramide phosphate: Side effects and effectiveness. *Am. Heart J.* **111**, 654–660.
- 42 (2020) Cyclica Case Study: Rapid discovery of a Sirtuin 2 inhibitor using Ligand Design.
- 43 Gordon, D. E., Jang, G. M., Bouhaddou, M., Xu, J., Obernier, K., O’Meara, M. J., Guo, J. Z., Swaney, D. L., Tummino, T. A., Huettenhain, R., et al. (2020) A SARS-CoV-2-Human Protein-Protein Interaction Map Reveals Drug Targets and Potential Drug-Repurposing. *bioRxiv* 2020.03.22.002386.

## Appendix/Supplemental

See Supplemental Data S1-S4.