# Open Science Discovery of Oral Non-Covalent SARS-CoV-2 Main Protease Inhibitor Therapeutics

The COVID Moonshot Consortium*

## Abstract

The COVID-19 pandemic is a stark reminder that a barren global antiviral pipeline has grave humanitarian consequences. Future pandemics could be prevented by accessible, easily deployable broad-spectrum oral antivirals and open knowledge bases that derisk and accelerate novel antiviral discovery and development. Here, we report the results of the **COVID Moonshot**, a fully open-science structure-enabled drug discovery campaign targeting the SARS-CoV-2 main protease. We discovered a novel chemical scaffold that is differentiated from current clinical candidates in terms of toxicity, resistance, and pharmacokinetics liabilities, and developed it into noncovalent orally-bioavailable nanomolar inhibitors with clinical potential. Our approach leveraged crowdsourcing, high-throughput structural biology, machine learning, and exascale molecular simulations. In the process, we generated a detailed map of the structural plasticity of the main protease, extensive structure-activity relationships for multiple chemotypes, and a wealth of biochemical activity data. In a first for a structure-based drug discovery campaign, all compound designs (>18,000 designs), crystallographic data (>500 ligand-bound X-ray structures), assay data (>10,000 measurements), and synthesized molecules (>2,400 compounds) for this campaign were shared rapidly and openly, creating a rich open and IP-free knowledgebase for future anti-coronavirus drug discovery.

## Introduction

The development of broad-spectrum oral antivirals is a critical but underexplored aspect of COVID-19 response and pandemic preparedness. Despite rapid progress in vaccine development, COVID-19 will likely become endemic[1], continuing to cause a significant number of deaths, especially in the Global South, unless there is an accessible treatment[2]. Antiviral therapeutics are a necessary and complementary strategy to vaccination in order to control COVID-19[3]. COVID-19 is not an isolated event, but the latest exemplar of a series of significant threats to human health caused by beta-coronaviruses also responsible for the SARS (2003) and MERS (2010) pandemics[4]. Open knowledge bases and technology infrastructures for antiviral drug discovery will enable pandemic preparedness by kindling the currently barren global antivirals pipeline and providing multiple starting points for the development of therapeutics. Here, we report the open science discovery of a novel oral antiviral and a roadmap for the development of future SARS-CoV-2 and pan-coronavirus antivirals.

The SARS-CoV-2 main protease (Mpro; or 3CL-protease) is an attractive target for antiviral development due to its essential role in viral replication, a large degree of conservation across coronaviruses, and dissimilarity to human proteases[5] (**Supplementary FIgure 1**). Pioneering studies during and after the 2003 SARS pandemic established the linkage between Mpro inhibition and antiviral activity in cell culture[6]. This is corroborated by recent in vitro and in vivo studies for SARS-CoV-2[7,8] and the recent clinical success of nirmatrelvir (the Mpro inhibitor component of Paxlovid)[9].

To warrant early use in the course of disease or even prophylactically among at-risk populations, an antiviral drug would need to be orally available with an excellent safety profile. Given the historical difficulties in developing peptidomimetic compounds into oral drugs, and the risk of downstream idiosyncratic hazards of covalent inhibition, we chose to pursue novel non-covalent non-peptidomimetic scaffolds. While first-generation oral Mpro inhibitors have now demonstrated clinical efficacy[10,11], the need for CYP3A4 inhibitor co-dosing (ritonavir, in the case of Paxlovid) to achieve sufficient human exposure may significantly limit use in at-risk populations due to potentially significant drug-drug interactions[12]. There remains a significant need for chemically differentiated oral antiviral protease inhibitors with the potential to rapidly enter clinical development.

*Authors are ordered alphabetically. For correspondence: john.chodera@choderalab.org (John D. Chodera); alpha.lee@postera.ai (Alpha Lee); nir.london@weizmann.ac.il (Nir London); annette.vondelft@cmd.ox.ac.uk (Annette von Delft); frank.von-delft@diamond.ac.uk (Frank von Delft)

**Crowdsourced progression of X-ray fragment hits rapidly generated potent lead compounds with diverse chemotypes**

The COVID Moonshot is an open science drug discovery campaign targeting SARS-CoV-2 Mpro[13,14], building off a rapid crystallographic fragment screening campaign that assessed 1495 fragment-soaked crystals screened within weeks to identify 78 hits that densely populated the active site (**Figure 1A**)[15]. This dataset was posted online on 18 Mar 2020[16], just days after the screen was completed[16]. The non-covalent fragment hits did not show detectable inhibition in a fluorescence based enzyme activity assay (assay dynamic range $IC_{50}$ < 100 μM). However, they provided a high-resolution map of key interactions that optimized compounds may exploit to inhibit Mpro[17].

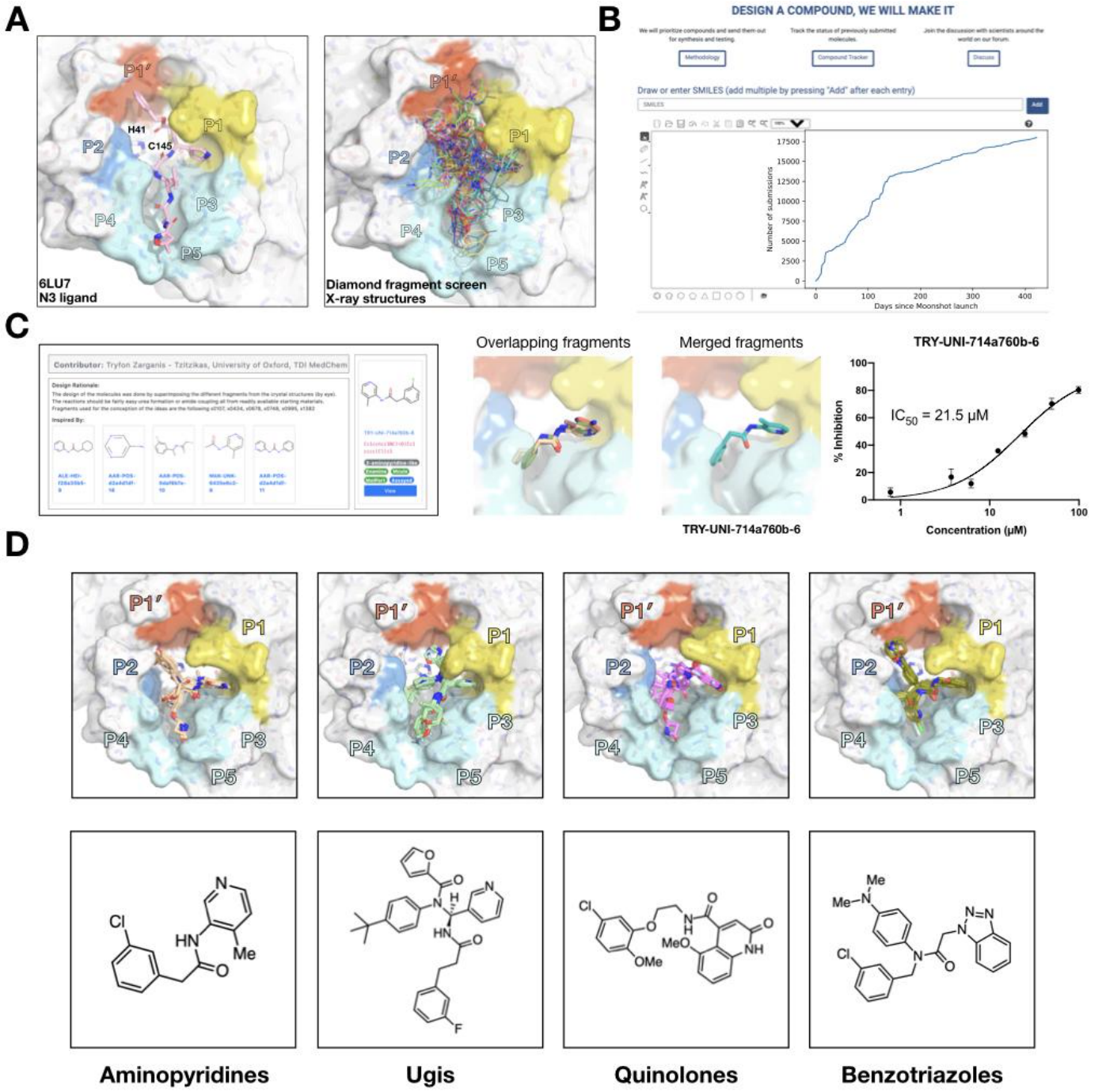
Numerous approaches have been proposed to advance from fragments to lead compounds[18,19]. One strategy, *fragment merging*, aims to combine multiple fragments into a single more potent molecule, whereas *fragment expansion* elaborates a fragment to engage neighboring interactions. While these strategies are usually applied to a single fragment or a handful of fragments, our large-scale fragment screen produced a dense ensemble of hits, providing a unique opportunity for rapid lead generation by combining chemotypes from multiple fragments. Nonetheless, this requires heuristic chemical reasoning that accounts for the spatial orientation of fragments in the binding site---a feat that can challenge algorithms but is potentially also solvable by humans. Building on successes in crowdsourced protein[20] and RNA[21] design campaigns, we hypothesized that crowdsourced human intuition and algorithmic strategies could accelerate the generation of potent lead compounds and furnish diverse chemical matter, as different chemists would employ different approaches and reasoning strategies.

We launched an online crowdsourcing platform [http://postera.ai/covid] on 18 March 2020 (**Figure 1B**), soliciting participants to submit compounds designed based on the fragment hits[14]. Data from biochemical assays and X-ray crystallography were released rapidly on the same platform, enabling contributing designers to build on all available data, as well as designs contributed by others. To ensure there would be no delays in ultimately delivering potential drug candidates straight to generics manufacture due to IP licensing issues, all designers were asked to contribute their designs directly into the public domain, with every design and all related experimental data immediately disclosed online, made openly available explicitly free of IP restrictions. This aggressive open science policy enabled contributors from multiple fields in both academia and industry to freely share their ideas. Within the first week, we received over 2,000 submissions, representing a diverse set of design strategies.

Gratifyingly, many submissions exploited spatially overlapping fragment hits. For example, the submission **TRY-UNI-714a760b-6** was inspired by five overlapping fragments, furnishing a noncovalent inhibitor with an SARS-CoV-2 Mpro enzymatic $IC_{50}$ of 21 μM (**Figure 1C**). This compound seeded the "aminopyridine" series, whose pursuit toward a preclinical candidate is described in detail below. Apart from the aminopyridine series, our campaign identified three major additional chemically distinct lead series with measurable potencies against SARS-CoV-2 Mpro inspired by reported SARS-CoV-1 inhibitors (**Figure 1D**). Those compounds span the same binding pocket but feature different chemotypes, and the significant quantity of SAR subsequently generated for these series furnishes multiple backup series with different risk profiles. Other groups have subsequently further elaborated on the Ugi[22,23] and the benzotriazole series we generated[24].

## Figure 1: Crowdsourcing rapidly identified novel chemotype scaffolds by merging fragment hits.

**A:** The Diamond / XChem fragment screen that initiated this SARS-CoV-2 Mpro inhibitor discovery campaign generated 58 hits that completely cover the Mpro active site, with a variety of chemotypes engaging each pocket; 1495 X-ray datasets were collected and 78 solved structures for hits were publicly posted 18 Mar 2020[15].
The peptidomimetic N3 ligand is shown at left for comparison to indicate natural substrate engagement in the binding site, defining the peptide sidechain numbering scheme used throughout this work. The catalytic Cys145 cleaves the scissile peptide bond between P1 and P1', with His41-Cys145 forming a catalytic dyad whose coupled charge states that shuttle between zwitterionic and neutral states[25].
**B:** On March 18th 2020, the COVID Moonshot set up a crowdsourcing website to capture how to progress hits with measurable potency from fragments. We received over 2000 submissions in the first week.
**C:** Many submissions, such as TRY-UNI-714a760b-6, exploited spatially overlapping fragment hits to design potent leads that are also synthetically facile (see Figure 3).
**D:** Crowdsourcing efforts also identified three additional major lead series inspired by reported SARS-CoV-1 inhibitors, with measurable biochemical activity, and well-resolved engagement modes from high throughput crystallography.



**Machine learning and free energy perturbation enabled rapid optimization cycles**

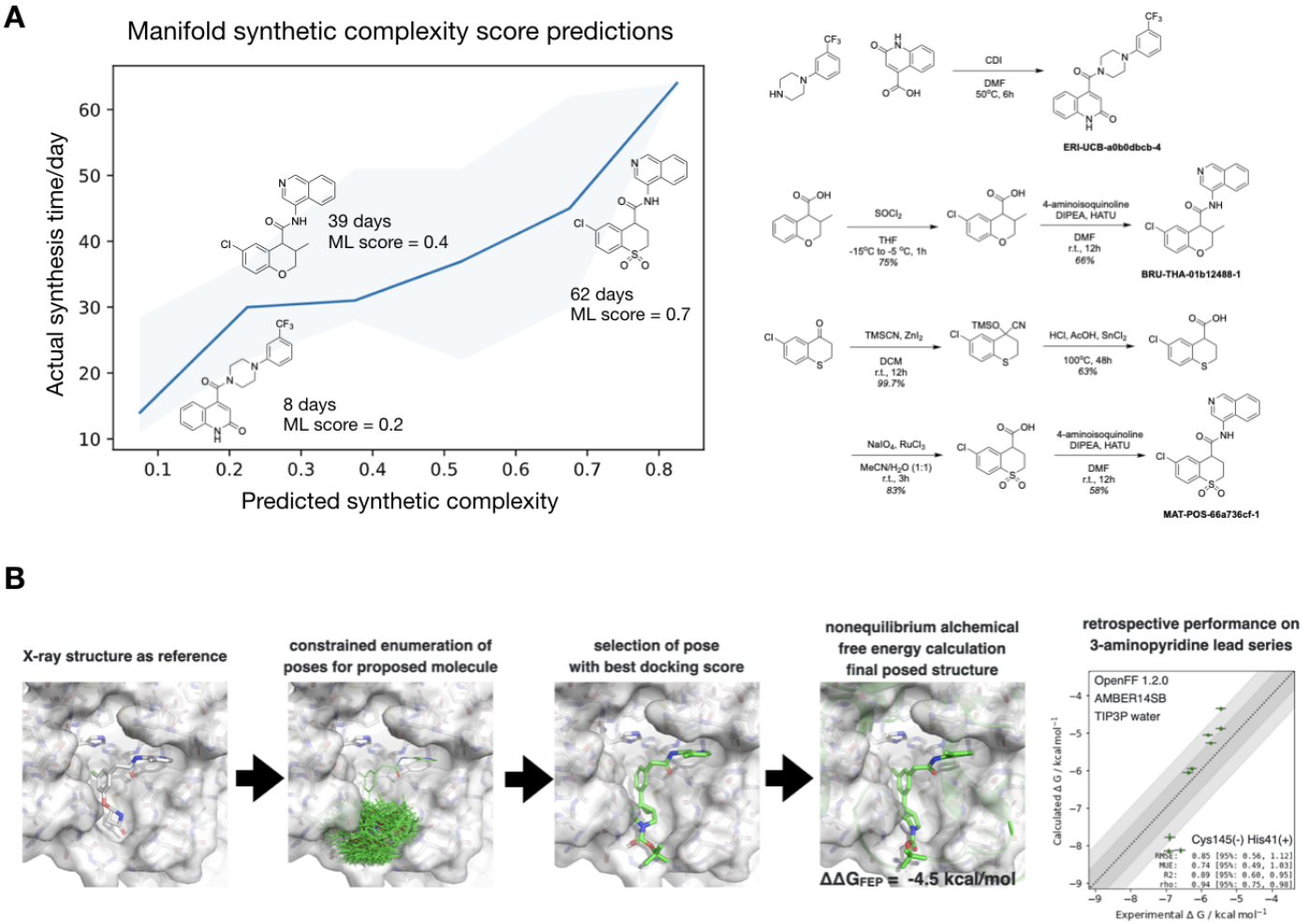
With a growing number of chemically diverse submissions aiming to progress multiple lead series, we next turned to computational methods to aid in triaging with the goal of increasing potency. To execute a rapid fragment-to-lead campaign, we used models to plan synthetic routes, enumerate synthetically accessible virtual libraries, and estimate potencies to prioritize which compounds to target for synthesis.

We used a major synthetic contract research organization (CRO), Enamine, to carry out rapid synthesis of progressed compound designs. The large number of building blocks and advanced intermediates available from Enamine – totalling to over 10 million – presents a unique opportunity for accelerating chemistry. However, the sheer number of building blocks, as well as the frequency in which building blocks go in and out of stock, presented the need for a rapid algorithmic approach to identifying compounds that could be quickly synthesized from available building blocks in a few reliable synthetic steps by chemists at Enamine. To address this challenge, we used a machine learning approach to predict synthetic tractability[26,27] that plans efficient retrosynthetic routes in seconds per compound. We automatically computed synthetic routes for all crowdsourced submissions utilizing Enamine's in-stock building block inventories. From the computed routes, synthetic complexity was estimated based on the number of steps and the probability of success of each step. The synthetic accessibility score, as well as the predicted synthetic route, were then used to aid medicinal chemistry decision making. **Figure 2A** (left) shows that our predicted synthetic complexity correlates with the actual time taken to synthesize target compounds. **Figure 2A** (right) demonstrates how the algorithm was able to pick out advanced intermediates as starting materials.

We estimated potency using alchemical free energy calculations[28–30], an accurate physical modeling technique that has hitherto not been deployed in a high throughput setup due to its prohibitive computational cost. We employed Folding@Home[31]---a worldwide distributed computing network where hundreds of thousands of volunteers around the world contributed computing power to create the world's first exascale computing resource[32]---to compute the free energy of binding of all 20,000+ crowdsourced submissions using the Open Force Field Initiative "Parsley" small molecule force fields[33] and nonequilibrium switching with the open source perses alchemical free energy toolkit[34] based on the GPU-accelerated OpenMM framework[35], culminating in over 1 ms of simulation time[32]. We first performed a small retrospective study using bioactivity data generated from the first week of crowdsourced compound designs, triaged solely using synthetic accessibility; **Figure 2B** shows that the results of these free energy calculations showed good correlation with experimentally-measured affinities. Henceforth, results from alchemical free energy calculations were published live, and used to guide compound selection and iterative design (see **Data Availability**).

**Figure 2**: **Machine learning route synthesis prediction enables rapid and inexpensive synthesis of compounds, while alchemical free energy calculations prioritizes potent compounds from large virtual synthetic libraries.**

**A:** Machine learning forecasts experimental synthesis time (left) and returns efficient routes that leverage over 10 million in-stock advanced intermediates (right). Our algorithm predicts the probability of each step being successful, and predicts synthetic accessibility by taking the product of the probabilities along the whole route. We analysed all compounds made in COVID Moonshot from 2020-09-01 to 2021-05-14. The right panel exemplifies the experimental execution of the predicted routes, demonstrating the ability of the algorithm to build on functionalized intermediates to shorten synthesis.
**B:** Applying alchemical free energy calculations at scale enables us to estimate the potency of compounds. The figure shows our automated free energy calculation workflow and retrospective evaluation on the first month of The COVID Moonshot data, which inspired confidence for large scale deployment during this campaign.



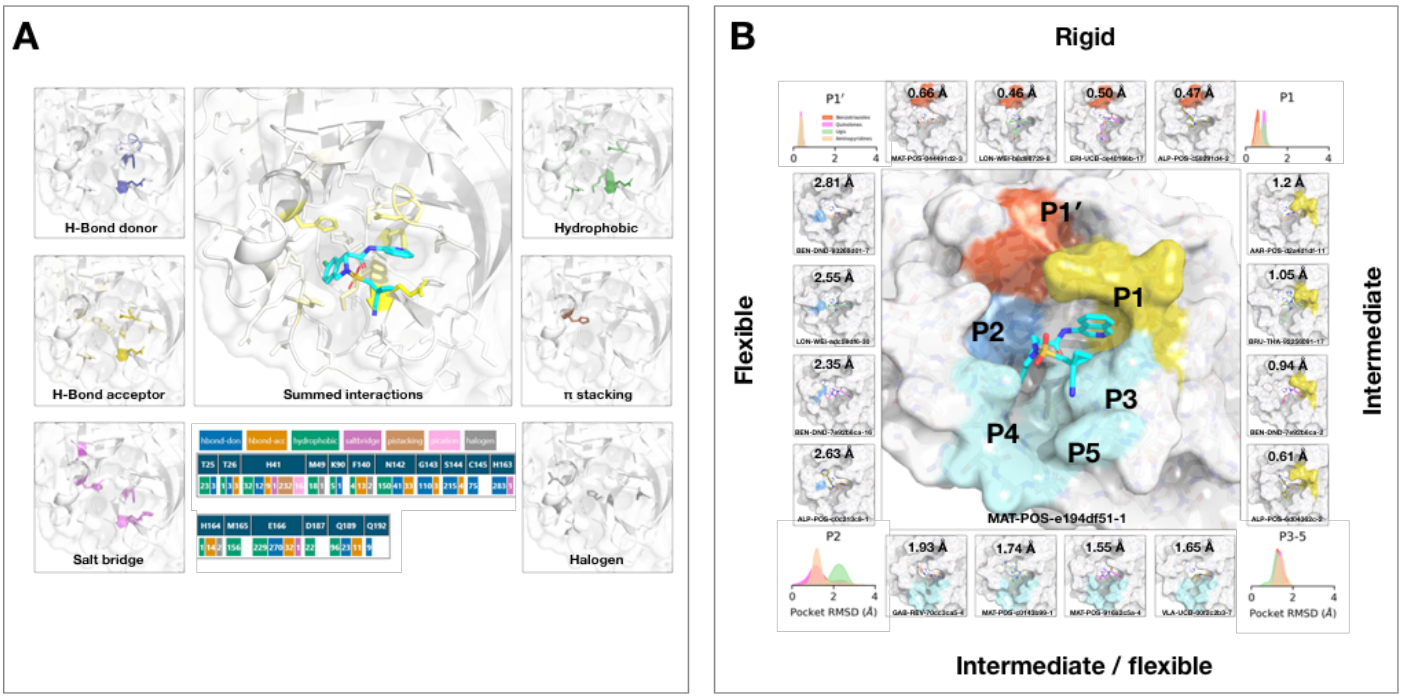
**High-throughput structural biology uncovered binding modes and interactions underlying potency**

Compounds selected based on synthetic accessibility and alchemical free energy calculations were synthesized and tested. We profiled every compound through crystal soaking and X-ray diffraction. Analysis of this large trove of structural data, totalling over 587 crystal structures (as of 25 Jan 2022), reveals the hotspots for ligand engagement and plasticity of each binding pocket. **Figure 3A** highlights the statistics of intermolecular interactions between the residues and our ligands across a preprocessed set of 416 complexes. The P1 and P2 pockets are the hotspots of interactions, yet the interaction patterns are starkly different. The salient interactions sampled by our ligands in the P1 pocket are H163 (H-bond donor), E166 (H-bond acceptor), S144 (H-bond donor), and N142 (hydrophobic interactions). Whereas P2 is dominated by π-stacking interactions with H41 and hydrophobic interactions M165. The P1' and P3/4/5 pockets are sparingly sampled by our ligands; the former can be targeted via hydrophobic interactions (T25), whereas the latter via H-bonds (Q192).

This pattern of intermolecular interactions is reflected in the plasticity of the different subpockets. The dominance of directional interactions in P1 renders it significantly more rigid than P2 (**Figure 3B**). Interestingly, this degree of rigidity is also dependent on the chemical series, with the Ugi and benzotriazole series being able to significantly deform the P2 pocket. Those series comprise more heavy atoms and span a larger region of the binding site, thus changes in P2 pocket interactions could be better tolerated.

**Figure 3: High-throughput crystallography reveals hotspots for ligand engagement and structural plasticity of the binding sites.**

**A:** The five subpockets exhibit different preferences for intermolecular interactions. The figure highlights the location of different types of interactions, with the shading indicating the frequency. The bottom column tallies the number of times each interaction was seen in our structures for different residues. The interaction map was generated using PLIPify and summarizes the interactions witnessed across 416 complexes
**B:** The subpockets have different degrees of plasticity, which is also dependent on the chemical series. The corners of the figure shows the distribution of side chain RMSD deviations from the structure of PET-UNK-29afea89-2 (middle panel). The boxes exemplify ligands that significantly deform the pockets.



**Design of an orally bioavailable inhibitor with potent antiviral activity**

Our medicinal chemistry strategy was driven by the design of potent ligand-efficient and geometrically compact inhibitors that fit tightly in the substrate binding pocket. The former strategy aimed to increase the probability of achieving oral bioavailability, while the latter heuristic was motivated by the substrate envelope hypothesis for avoiding viral resistance[36]. **Figure 4A** shows our medicinal chemistry strategy towards the lead compound.

Starting from the fragment hit, we explored the P1 pocket, which admits a steep structure-activity relationship (SAR), perhaps unsurprising given its rigidity and preference for directional H-bond interactions (**Figure 3A**). A significant increase in potency was unlocked by replacing pyridine with isoquinoline, which picks up additional hydrophobic interactions with N142. The SAR around the P2 pocket is considerably more tolerant to modifications, and broadly favours hydrophobic moieties. A step-change in potency was achieved by rigidifying the scaffold: We introduced the tetrahydropyran ring to transform the P2 substituent into a chromane moiety (compound **MAT-POS-b3e365b9-1**; the racemic mixture **VLA-UCB-1dbca3b4-15**, which was initially synthesized, has a $IC_{50}$ of 360 nM; **Figure 4A**), chosen because of building block availability. Despite possessing a degree of molecular complexity, **MAT-POS-b3e365b9-1** is still a one-step amide coupling (**Figure 2A**). We then further explored the P2 pocket with a library chemistry strategy in mind. Thus we first morphed the chromane into a tetrahydroisoquinoline to introduce a functionalisable handle (**MAT-POS-3ccb8ef6-1, Figure 4A**), which gratifyingly maintained potency. Finally, we constructed a focused library realised via sulphonamide Schotten-Baumann coupling, furnishing a surprising increase in both enzymatic inhibition and antiviral efficacy, leading to the lead compound presented here (**MAT-POS-e194df51-1, Figure 4A**)

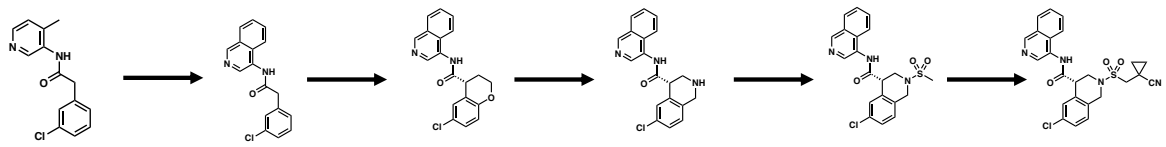
Our compound **MAT-POS-e194df51-1** was subsequently profiled in SARS-CoV-2 antiviral assays across multiple cell lines, exhibiting $EC_{50}$ of 64 nM in A549-ACE2-TMPRSS2 cells and 126 nM in HelaAce2, without measurable cytotoxicity (**Figure 4B,** data available via GitHub link). This is in line with overall cellular efficacy for the chemical series; of 150 compounds with enzyme assay IC50s <500nM assessed in A549-ACE2-TMPRSS2 cellular CPE assays, 15 compounds showed lower IC50s values compared to the internal control Nirmatrelvir (Paxlovid) that was measured at an $IC_{50}$ of 218 nM in this assay (**Figure 4C**). This is in line with good antiviral activity of our lead compound across “crowdsourced” antiviral assays across different laboratories and cell lines, including assays performed with and without p-gp inhibitors and with Nirmatrelvir as an internal control (**Figure 4D**). We also observed good cross-reactivity of our lead compound **MAT-POS-e194df51-1** and analogue **MAT-POS-a54ce14d-2** against known SARS-CoV-2 variants alpha, beta, delta and omicron variants (**Figure 4E**). Additionally supporting the antiviral activity of our lead series in more complex organoid models, closely related molecules **PET-UNK-29afea89-2** and **MAT-POS-932d1078-3** with $IC_{50}$s in HelaAce2 CPE assays of 240 nM and 331 nM and 657 nM and 2.57 μM in A549-ACE2-TMPRSS2 CPE assays, respectively (**Suppl Figure 2A** and **B**) show an over 100-fold reduction of intracellular viral RNA and infectious virus secretion into the apical compartment of human induced pluripotent stem cell-derived kidney organoids (**Suppl Figure 2D** and **E**), an accessible model for the human kidney, an organ that is infected in COVID-19 patients, as reported previously for earlier analogues of the same series[37].

Further, **MAT-POS-e194df51-1** exhibits properties required for an orally bioavailable therapeutic: using the rat pharmacokinetics (**Figure 4F**) and in vitro human metabolism data **(Figure 4G)**, projections suggest an oral human dose (see **Methods**) in the range 450-2000 mg BID (two times a day) or 100-350mg TID (three times a day). Further, preliminary profiling on this compound has not revealed safety liabilities. This compound is currently undergoing further evaluation as part of the preclinical development pipeline.

**Figure 4: Iterative medicinal chemistry furnished an orally bioavailable inhibitor.**

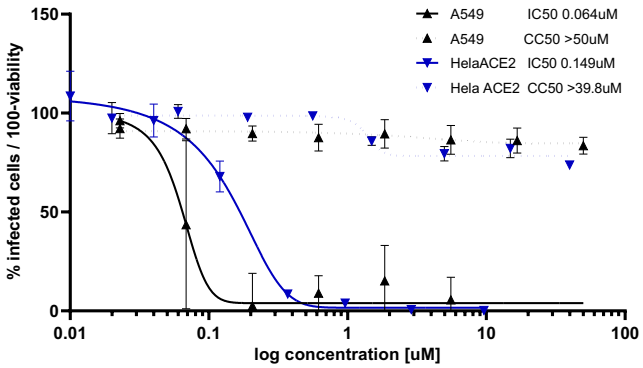
**A:** Summary of the salient structure-activity relationship around the P1 and P2 binding pocket that improved potency in Mpro biochemical assay by 2.5 orders of magnitude.
**B:** Antiviral activity of **MAT-POS-e194df51-1** cellular antiviral assays, with an $IC_{50}$ of 64 nM in A549-ACE2-TMPRSS2 cells assessing cytopathic effect (CPE, black, plotted as 100 - % viability), and 126 nM in HelaAce2 assays (blue, plotted as % infected cells). Both assays were performed with p-gp inhibitors.
**C**: Histogram comparing antiviral efficacy of all COVID Moonshot compounds measured to date in an A549-ACE2-TMPRSS2 CPE cellular antiviral assay.
**D:** Detailed cellular antiviral assessment of key compounds comprising the synthetic strategy (Fig 4A) across different cell lines and assay techniques, with and without p-gp inhibitors, demonstrating efficacy of **MAT-POS-e194df51-1** in various set-ups and laboratories.
**E**: **MAT-POS-e194df51-1** shows good cross-reactivity against known circulating variants of SARS-CoV-2 in antiviral cellular assays in a CPE assay in HelaACE2 cells.
**F**: PK profile of development compound **MAT-POS-e194df51-1** in rats with a 2mg/kg intravenous and 10mg/kg oral dosing with good oral availability.
**G**: ADME characteristics of **MAT-POS-e194df51-1** demonstrate translational potential of this compound.

## A

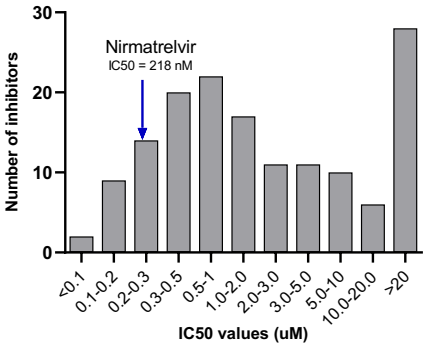


| | TRY-UNI-714a760b-6 | ADA-UCB-6c2cb422-1 | MAT-POS-b3e365b9-1 | MAT-POS-3ccb8ef6-1 | MAT-POS-e194df51-1 | MAT-POS-e194df51-1 |
|---|---|---|---|---|---|---|
| $IC_{50}$(Mpro)/uM | 25 | 0.73 | 0.21 | 0.28 | 0.141 | 0.037 |
| $EC_{50}$(SARS-CoV-2, A549)/uM | n.d. | 4.5 | 7.0 | 1.9 | 1.65 | 0.064 |

## B




## C



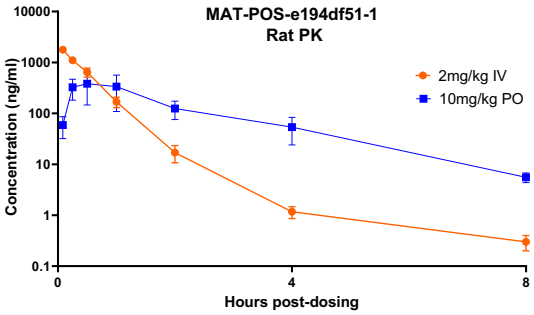

## D

| | A549 (+ p-gp) CPE | | | HelaAce2 (+ p-gp) CPE | | | Calu-3 (no p-gp) FFU | | | VeroE6 (no p-gp) CPE | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | IC50 | IC90 | CC50 | IC50 | IC90 | CC50 | IC50 | IC90 | CC50 | IC50 | IC90 |
| Nirmatrelvir | 0.218 | 0.336 | >50 | 0.0604 | 0.12 | > 39.8 | 2.01 | 6.08 | >100 | 2.71 | 3.71 |
| MAT-POS-e194df51-1 | 0.0638 | 0.126 | >50 | 0.149 | 0.365 | > 39.8 | 1.15 | 4.68 | >100 | | |
| MAT-POS-24589f88-3 | 1.65 | 21.2 | >50 | 0.428 | 1.4 | > 39.8 | 1.02 | 1.31 | >100 | 0.213 | 0.391 |
| MAT-POS-3ccb8ef6-1 | 1.87 | 10.2 | >50 | | | | 2.57 | 6.96 | >100 | 0.513 | 2.16 |
| MAT-POS-b3e365b9-1 | 6.97 | >50 | >50 | 0.481 | 1.29 | > 39.8 | 1.06 | 3.43 | >100 | 1.57 | 3.48 |
| ADA-UCB-6c2cb422-1 | 4.49 | 7.06 | >50 | | | | >10 | | >100 | 3.24 | 6.41 |
| TRY-UNI-714a760b-6 | | | | | | | | | | | |

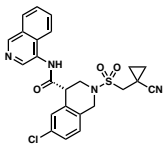
## E

| | MAT-POS-e194df51-1 | | Nirmatrelvir | |
|---|---|---|---|---|
| | IC50 | CC50 | IC50 | CC50 |
| Alpha variant (B.1.1.7. | 0.38 | >20 | 0.12 | >10 |
| Beta variant (B.1.351) | 1.48 | >20 | 0.21 | >10 |
| Delta variant (B.1.617.2) | 1.52 | >20 | 0.21 | >10 |
| Omicron variant (B.1.529) | 0.29 | >20 | 0.07 | >10 |
| MA-SARS-CoV-2/WA1 | 0.43 | >20 | 0.14 | >10 |

## F




## G



### MAT-POS-e194df51-1

| | | | | |
|---|---|---|---|---|
| **Antiviral efficacy** | | | | |
| Mpro IC50 /uM | | 0.037 | | |
| A549 IC50 /uM | | 0.064 | | |
| **In vitro ADME** | | | | |
| LogD [measured] | | 2.5 | | |
| MDCK-LE FA (%) | | 92.9 | | |
| | **Rat** | **Dog** | **Minipig** | **Human** |
| Liver microsomes Cl ul/min/kg | 604 | 164 | 542 | 152 |
| Liver microsomes t ½ (min) | 2.4 | 8.5 | 2.6 | 9.1 |
| Heps Cl ul/min/kg | 67.6 | 61.4 | 65.9 | 10.3 |
| Heps t ½ (min) | 10.3 | 11.3 | 10.5 | 67.5 |
| PPB free fraction (%) | 5.4 | | | 10.1 |
| **Safety / Drug-drug interactions** | | | | |
| Cyp450 (uM) 2C9/2D6/3A4 | | 25/9.4/10.3 | | |
| PXR risk | | Low | | |
| Herg (uM) | | >30 | | |
| **In vivo pharmacokinetics** | | | | |
| Rat IV Vd (l/kg) | 1.05 | | | |
| Rat IV CL | 34.8 | | | |
| Rat t ½ IV/PO (h) | 0.448 / 1.4 | | | |
| Rat Bioavailability (%) | 18 | | | |

**Open Science drug discovery presents a viable route to patient impact**

Herein, we report the success of an open science patent-free antiviral discovery program in rapidly developing an orally bioavailable antiviral in response to emergent pandemic threat in producing compounds entering preclinical studies. As a result of the open science policy, a large number of collaborators (now **The COVID Moonshot Consortium**) were eager and able to provide in-kind support, providing synthesis, assays and in vitro/vivo experiments. By making all data immediately available, and all compounds purchasable from Enamine, we aim to accelerate research globally along parallel tracks following up on our initial work.

The COVID Moonshot and its lead compounds for COVID-19 have been adopted into the drug development portfolio of the Drugs for Neglected Diseases Initiative (DNDi), funded by the Wellcome Trust through the COVID-19 Therapeutics Accelerator (ACT-A) to enable accelerated Phase I-enabling studies in preparation for clinical trials[38]. Downstream, we envisage a direct-to-generics production and distribution model where the eventual drug price will immediately reflect the cost of producing the compounds alongside a reasonable margin, rather than the monopoly rent associated with recovering substantial development costs and based on restrictive intellectual property positions on the compound. Working with globally-focused development and manufacturing organisations and public health funding organisations united in our goal of global equitable access, we are building a streamlined, economically viable downstream development plan designed to deliver affordable treatment options from this chemical series, with a primary focus on patients in Low-and-middle-income countries (LMICs).

Open science efforts have transformed many areas of biosciences, with illustrious examples such as the Human Genome Project[39], the Structural Genomics Consortium[40], and the RAS Initiative[41]. Yet its inroad in therapeutic discovery is slow because of the perceived need for commercial return. We hope The COVID Moonshot is an exemplar for open science drug discovery[42] that paves the way towards a new paradigm for infectious diseases drug discovery - a disease area of grave public importance but chronically underfunded by the private sector[42].

**Materials and Methods**

The Supplementary Material contains detailed methods and protocols for all the experiments reported in this manuscript.

**Data availability**

- **All compound submissions and experimental data are available via GitHub:** https://github.com/postera-ai/COVID_moonshot_submissions
- **All compound designs, datasets, and X-ray structures are indexed on the COVID Moonshot website:** https://postera.ai/covid
- **Bioactivity data can be interactively browsed at the COVID Moonshot website:** https://covid.postera.ai/covid/activity_data
- **Alchemical free energy calculations code and datasets are indexed on GitHub:** https://github.com/foldingathome/covid-moonshot
- **All X-ray structures are available for interactive viewing or bulk download on Fragalysis:** https://fragalysis.diamond.ac.uk/viewer/react/preview/target/Mpro

**Resource availability**

- **Synthesized compounds**: We have made all compounds assayed here are now available from the current Enamine catalogue, and readily available for purchase from Enamine (and other suppliers)

via the Manifold platform accessible for each compound page on the COVID Moonshot website: https://covid.postera.ai/covid/activity_data

**Acknowledgements:**

The COVID Moonshot acknowledges funding by the Wellcome Trust on behalf of the Covid-19 Therapeutics Accelerator. The COVID Moonshot project is particularly grateful to UCB Pharma Ltd and UCB S.A for the support from their Medicinal and Computational Chemistry groups, to Novartis Institute for Biomedical Research for generous in-kind ADME and PK contributions, Takeda for in-kind contribution of antiviral assays/pan-corona biochemical assays, and Nanosyn for protease panel assays. We thank CDD Vault and OpenEye Scientific for their in-kind contribution allowing the consortium to use their software. We also thank the numerous volunteers that contributed compound designs to the COVID moonshot and to the citizen scientist volunteers of Folding@home for donating their computing resources, and Amazon Web Services for key support of Folding@home infrastructure .

**Funding acknowledgments and Financial Disclosures**
Funding acknowledgement and Disclosures statements for each author is listed in the consortium spreadsheet (The COVID Moonshot Consortium).

**References**

1. Phillips, N. The coronavirus is here to stay - here's what that means. *Nature* **590**, 382–384 (2021).
2. Winchester, N. Covid-19 vaccinations: is the Global South falling behind? *UK House of Lords library* https://lordslibrary.parliament.uk/covid-19-vaccinations-is-the-global-south-falling-behind/ (2021).
3. Zhao, T. *et al.* Warnings regarding the potential coronavirus disease 2019 (COVID-19) transmission risk: Vaccination is not enough. *Infect. Control Hosp. Epidemiol.* 1 (2021).
4. de Wit, E., van Doremalen, N., Falzarano, D. & Munster, V. J. SARS and MERS: recent insights into emerging coronaviruses. *Nat. Rev. Microbiol.* **14**, 523–534 (2016).
5. Ullrich, S. & Nitsche, C. The SARS-CoV-2 main protease as drug target. *Bioorg. Med. Chem. Lett.* **30**, 127377 (2020).
6. Pillaiyar, T., Manickam, M., Namasivayam, V., Hayashi, Y. & Jung, S.-H. An Overview of Severe Acute Respiratory Syndrome-Coronavirus (SARS-CoV) 3CL Protease Inhibitors: Peptidomimetics and Small Molecule Chemotherapy. *J. Med. Chem.* **59**, 6595–6628 (2016).
7. Zhang, L. *et al.* Crystal structure of SARS-CoV-2 main protease provides a basis for design of improved α-ketoamide inhibitors. *Science* **368**, 409–412 (2020).
8. Jin, Z. *et al.* Structure of Mpro from SARS-CoV-2 and discovery of its inhibitors. *Nature* **582**, 289–293 (2020).
9. Owen, D. R. *et al.* An oral SARS-CoV-2 mpro inhibitor clinical candidate for the treatment of COVID-

19. *bioRxiv* (2021) doi:10.1101/2021.07.28.21261232.

10. Study of pf-07321332 in healthy participants. https://clinicaltrials.gov/ct2/show/NCT04756531.
11. Study of PBI-0451 in Healthy Subjects. https://clinicaltrials.gov/ct2/show/NCT05011812.
12. Heskin, J. *et al.* Caution required with use of ritonavir-boosted PF-07321332 in COVID-19 management. *Lancet* **399**, 21–22 (2022).
13. von Delft, F. *et al.* A white-knuckle ride of open COVID drug discovery. *Nature* **594**, 330–332 (2021).
14. Chodera, J., Lee, A. A., London, N. & von Delft, F. Crowdsourcing drug discovery for pandemics. *Nat. Chem.* **12**, 581 (2020).
15. Douangamath, A. *et al.* Crystallographic and electrophilic fragment screening of the SARS-CoV-2 main protease. *Nat. Commun.* **11**, 5047 (2020).
16. Main protease structure and XChem fragment screen. *Diamond Light Source* https://www.diamond.ac.uk/covid-19/for-scientists/Main-protease-structure-and-XChem.html (2020).
17. XChem @ Diamond. https://fragalysis.diamond.ac.uk/viewer/react/preview/target/Mpro.
18. Rees, D. C., Congreve, M., Murray, C. W. & Carr, R. Fragment-based lead discovery. *Nat. Rev. Drug Discov.* **3**, 660–672 (2004).
19. Erlanson, D. A., Fesik, S. W., Hubbard, R. E., Jahnke, W. & Jhoti, H. Twenty years on: the impact of fragments on drug discovery. *Nat. Rev. Drug Discov.* **15**, 605–619 (2016).
20. Cooper, S. *et al.* Predicting protein structures with a multiplayer online game. *Nature* **466**, 756–760 (2010).
21. Lee, J. *et al.* RNA design rules from a massive open laboratory. *Proc. Natl. Acad. Sci. U. S. A.* **111**, 2122–2127 (2014).
22. Kitamura, N. *et al.* Expedited Approach toward the Rational Design of Noncovalent SARS-CoV-2 Main Protease Inhibitors. *J. Med. Chem.* (2021) doi:10.1021/acs.jmedchem.1c00509.
23. Zaidman, D. *et al.* An automatic pipeline for the design of irreversible derivatives identifies a potent SARS-CoV-2 Mpro inhibitor. *Cell Chem Biol* In Press (2021) doi:10.1101/2020.09.21.299776.
24. Han, S. H. *et al.* Structure-Based Optimization of ML300-Derived, Noncovalent Inhibitors Targeting the Severe Acute Respiratory Syndrome Coronavirus 3CL Protease (SARS-CoV-2 3CLpro). *J. Med. Chem.* (2021) doi:10.1021/acs.jmedchem.1c00598.
25. Kneller, D. W. *et al.* Structural plasticity of SARS-CoV-2 3CL Mpro active site cavity revealed by room

temperature X-ray crystallography. *Nat. Commun.* **11**, 3202 (2020).

26. Schwaller, P. *et al.* Molecular Transformer: A Model for Uncertainty-Calibrated Chemical Reaction Prediction. *ACS Cent Sci* **5**, 1572–1583 (2019).
27. Morris, A. *et al.* Discovery of SARS-CoV-2 main protease inhibitors using a synthesis-directed de novo design model. *Chem. Commun.* **57**, 5909–5912 (2021).
28. Mey, A. S. J. *et al.* Best Practices for Alchemical Free Energy Calculations. *arXiv [q-bio.BM]* (2020).
29. Abel, R., Wang, L., Mobley, D. L. & Friesner, R. A. A Critical Review of Validation, Blind Testing, and Real- World Use of Alchemical Protein-Ligand Binding Free Energy Calculations. *Curr. Top. Med. Chem.* **17**, 2577–2585 (2017).
30. Armacost, K. A., Riniker, S. & Cournia, Z. Exploring Novel Directions in Free Energy Calculations. *J. Chem. Inf. Model.* **60**, 5283–5286 (2020).
31. Shirts, M. & Pande, V. S. COMPUTING: Screen Savers of the World Unite! *Science* **290**, 1903–1904 (2000).
32. Zimmerman, M. I. *et al.* SARS-CoV-2 simulations go exascale to predict dramatic spike opening and cryptic pockets across the proteome. *Nat. Chem.* **13**, 651–659 (2021).
33. Qiu, Y. *et al.* Development and Benchmarking of Open Force Field v1.0.0, the Parsley Small Molecule Force Field. *ChemRxiv* (2020) doi:10.26434/chemrxiv.13082561.v2.
34. Rufa, D. A. *et al.* Towards chemical accuracy for alchemical free energy calculations with hybrid physics-based machine learning / molecular mechanics potentials. *bioRxiv* 2020.07.29.227959 (2020) doi:10.1101/2020.07.29.227959.
35. Eastman, P. *et al.* OpenMM 7: Rapid development of high performance algorithms for molecular dynamics. *PLoS Comput. Biol.* **13**, e1005659 (2017).
36. Nalam, M. N. L. *et al.* Evaluating the substrate-envelope hypothesis: structural analysis of novel HIV-1 protease inhibitors designed to be robust against drug resistance. *J. Virol.* **84**, 5368–5378 (2010).
37. Jansen, J. *et al.* SARS-CoV-2 infects the human kidney and drives fibrosis in kidney organoids. *Cell Stem Cell* (2021) doi:10.1016/j.stem.2021.12.010.
38. Wellcome Trust Grant ref: 224021/Z/21/Z - COVID Moonshot funded by COVID-19 Therapeutics Accelerator to rapidly develop a safe, globally accessible and affordable antiviral pill. *Drugs for Neglected Diseases initiative (DNDi)* https://dndi.org/press-releases/2021/covid-moonshot-funded-by-

wellcome-to-rapidly-develop-safe-globally-accessible-affordable-antiviral-pill/ (2021).

39. Lander, E. S. *et al.* Initial sequencing and analysis of the human genome. *Nature* **409**, 860–921 (2001).
40. Williamson, A. R. Creating a structural genomics consortium. *Nat. Struct. Biol.* **7 Suppl**, 953 (2000).
41. U.S. Cancer Institute 'Megaproject' to Target Common Cancer-Driving Protein. *Science* https://www.science.org/content/article/us-cancer-institute-megaproject-target-common-cancer-driving-protein.
42. Lee, A., Chodera, J. & von Delft, F. Why we are developing a patent-free Covid antiviral therapy. *Knowable mag.* (2021) doi:10.1146/knowable-092721-1.
43. Malla, T. R. *et al.* Mass spectrometry reveals potential of β-lactams as SARS-CoV-2 Mpro inhibitors. *Chem. Commun.* **57**, 1430–1433 (2021).
44. Noske, G. D. *et al.* A Crystallographic Snapshot of SARS-CoV-2 Main Protease Maturation Process. *J. Mol. Biol.* **433**, 167118 (2021).
45. Winter, G. *et al.* DIALS: implementation and evaluation of a new integration package. *Acta Crystallogr D Struct Biol* **74**, 85–97 (2018).
46. Winter, G. *et al.* How best to use photons. *Acta Crystallogr D Struct Biol* **75**, 242–261 (2019).
47. Winter, G. & McAuley, K. E. Automated data collection for macromolecular crystallography. *Methods* **55**, 81–93 (2011).
48. Kabsch, W. Integration, scaling, space-group assignment and post-refinement. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 133–144 (2010).
49. Winter, G., Lobley, C. M. C. & Prince, S. M. Decision making in xia2. *Acta Crystallogr. D Biol. Crystallogr.* **69**, 1260–1273 (2013).
50. Vonrhein, C. *et al.* Data processing and analysis with the autoPROC toolbox. *Acta Crystallogr. D Biol. Crystallogr.* **67**, 293–302 (2011).
51. Krojer, T. *et al.* The XChemExplorer graphical workflow tool for routine or large-scale protein-ligand structure determination. *Acta Crystallogr D Struct Biol* **73**, 267–278 (2017).
52. Pearce, N. M. *et al.* A multi-crystal method for extracting obscured crystallographic states from conventionally uninterpretable electron density. *Nat. Commun.* **8**, 15123 (2017).
53. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 486–501 (2010).

54. Long, F. *et al.* AceDRG: a stereochemical description generator for ligands. *Acta Crystallogr D Struct Biol* **73**, 112–122 (2017).

55. Bruno, I. J. *et al.* Retrieval of crystallographically-derived molecular geometry information. *J. Chem. Inf. Comput. Sci.* **44**, 2133–2144 (2004).

56. Cottrell, S. J., Olsson, T. S. G., Taylor, R., Cole, J. C. & Liebeschuetz, J. W. Validating and understanding ring conformations using small molecule crystallographic data. *J. Chem. Inf. Model.* **52**, 956–962 (2012).

57. Corman, V. M. *et al.* Detection of 2019 novel coronavirus (2019-nCoV) by real-time RT-PCR. *Euro Surveill.* **25**, (2020).

58. Varghese, F. S. *et al.* Berberine and Obatoclax Inhibit SARS-Cov-2 Replication in Primary Human Nasal Epithelial Cells In Vitro. *Viruses* **13**, (2021).

59. Jochmans, D., Leyssen, P. & Neyts, J. A novel method for high-throughput screening to quantify antiviral activity against viruses that induce limited CPE. *J. Virol. Methods* **183**, 176–179 (2012).

60. Rosales, R. *et al.* Nirmatrelvir, Molnupiravir, and Remdesivir maintain potent in vitro activity against the SARS-CoV-2 Omicron variant. *bioRxiv* 2022.01.17.476685 (2022) doi:10.1101/2022.01.17.476685.

61. Gapsys, V. *et al.* Large scale relative protein ligand binding affinities using non-equilibrium alchemy. *Chem. Sci.* **11**, 1140–1152 (2019).

62. Maier, J. A. *et al.* ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *J. Chem. Theory Comput.* **11**, 3696–3713 (2015).

63. Crooks, G. E. Path-ensemble averages in systems driven far from equilibrium. *Phys. Rev. E* **61**, 2361–2366 (2000).

64. Bennett, C. H. Efficient estimation of free energy differences from Monte Carlo data. *J. Comput. Phys.* **22**, 245–268 (1976).

65. Salentin, S., Schreiber, S., Haupt, V. J., Adasme, M. F. & Schroeder, M. PLIP: fully automated protein-ligand interaction profiler. *Nucleic Acids Res.* **43**, W443–7 (2015).

## Materials and Methods

### 0. Compound registration and data flow process

All compound designs from the internal medicinal chemistry team, collaborators, and external submitters were captured through the online compound design submission platform [https://postera.ai/covid] along with submitter identity, institution, design rationale, and any inspiration fragments. A forum thread was created to discuss these designs and attached to the compound design. Each submitted batch of related designs received a unique ID including the first three letters of the submitter name and submitter institution, and each compound design submitted received a unique ID ("PostEra ID") that appended a unique molecule sequence ID within the submission batch ID. Internally, compound designs, synthesized compounds, and compounds with experimental data were tracked with corresponding records in a CDD Vault (Collaborative Drug Discovery Inc.).

Stereochemistry: While the design platform enabled submitters to register compounds with specific defined or uncertain stereochemistry, compounds were initially synthesized and biochemically assayed as racemates, and if active, chirally separated compounds were assayed separately. To track this, compound designs with unspecified stereochemistry were submitted to generate new molecule IDs if the original designs contained specific stereochemistry, and data for the assayed racemates attached to the compound record with unspecified stereochemistry. Compounds with specified stereochemistry were then registered for stereochemically resolved components of active compounds (if not already present). Because the absolute stereochemical identity of the enantiopure compounds was unknown at time of receipt, assay data was attached to compound records with specified *relative* stereochemistry, rather than absolute stereochemistry. For compounds where sufficient data was available from a variety of sources to propose the absolute stereochemistry (e.g. X-ray data for the compound or a close analogue), the "suspected_SMILES" record was updated, along with an articulated rationale in the "why_suspected_SMILES" field. As a result, caution must be exercised when using data for enantiopure compounds for downstream uses (e.g. whole-dataset machine learning) without verifying the absolute stereochemistry is known with confidence.

### 1. Experimental methods

#### 1.1 Fluorescence MPro inhibition assay

Compounds were seeded into assay-ready plates (Greiner 384 low volume, cat 784900) using an Echo 555 acoustic dispenser, and DMSO was back-filled for a uniform concentration in assay plates (DMSO concentration maximum 1%) Screening assays were performed in duplicate at 20μM and 50μM. Hits of greater than 50% inhibition at 50μM were confirmed by dose response assays. Dose response assays were performed in 12 point dilutions of 2-fold, typically beginning at 100μM. Highly active compounds were repeated in a similar fashion at lower concentrations beginning at 10μM or 1μM. Reagents for Mpro assay were dispensed into the assay plate in 10μl volumes for a final volume of 20μL.

Final reaction concentrations were 20mM HEPES pH7.3, 1.0mM TCEP, 50mM NaCl, 0.01% Tween-20, 10% glycerol, 5nM Mpro, 375nM fluorogenic peptide substrate ([5-FAM]-AVLQSGFR-[Lys(Dabcyl)]-K-amide). Mpro was pre-incubated for 15 minutes at room temperature with compound before addition of substrate and a further 30 minute incubation. Protease reaction was measured in a BMG Pherastar FS with a 480/520 ex/em filter set. Raw data was mapped and normalized to high (Protease with DMSO) and low (No Protease) controls using Genedata Screener software. Normalized data was then uploaded to CDD Vault (Collaborative Drug Discovery). Dose response curves were generated for IC50 using nonlinear regression with the Levenberg–Marquardt algorithm with minimum inhibition = 0% and maximum inhibition = 100%.

### 1.2 RapidFire MPro inhibition assay

The assay was performed according to the published procedure[43]. Briefly, compounds were seeded into assay-ready plates (Greiner 384PP, cat# 781280) using an ECHO 650T dispenser and DMSO was back-filled for a uniform concentration in assay plates (DMSO concentration < 1%, final volume = 500 nL.). A 15 µM enzyme stock solution is prepared in 20 mM HEPES, pH 7.5 and 300 mM NaCl, and subsequently diluted to a working solution of 300 nM Mpro in assay buffer (20 mM HEPES, pH 7.5 and 50 mM NaCl) before the addition of 25 µL to each well using a Multidrop Combi (Thermo Scientific). After a quick centrifugation step (1000 rpm, 15 s) the plate is incubated for 15 min at room temperature. The reaction is initiated with the addition of 25 µL of 4 µM 11-mer (TSAVLQSGFRK-NH2, initially custom synthesized by the Schofield group, GLBiochem, used until March 2021), or 10 µM 37-mer (ALNDFSNSGSDVLYQPPQTSITSAVLQSGFRKMAFPS-NH2, GLBiochem, used after March 2021), dissolved in assay buffer. After centrifugation (1000 rpm, 14 s) the reaction is incubated for 10 min (11-mer) or 5 min (37-mer) at room temperature before quenching with 10 % formic acid. The reactions are analysed with MS using RapidFire (RF) 365 high-throughput sampling robot (Agilent) connected to an iFunnel Agilent 6550 accurate mass quadrupole time-of-flight (Q-TOF) mass spectrometer using electrospray. All compounds are triaged by testing the % inhibition at 5 and 50 µM final concentration. Dose response curves uses an 11-point range of 100--0.0017 µM inhibitor concentrations. RapidFire integrator software (Agilent) was used to extract the charged states from the total ion chromatogram data followed by peak integration. For the 11-mer peptide the m/z (+1) charge states of both the substrate (1191.67 Da) and cleaved N-terminal product TSAVLQ (617.34 Da) were used and the 37-mer peptide the m/z (+2) charge states of the substrate (3960.94 Da) and m/z (+1) of the cleaved C-terminal product SGFRKMAFPS (1125.57 Da). Percentage conversion (product peak integral / (product peak integral + substrate peak integral))*100) and percentage inhibitions were calculated and normalised against DMSO control with deduction of any background signal in Microsoft Excel. IC50s were calculated using Levenberg–Marquardt algorithm used to fit a restrained Hill equation to the dose-response data with both GraphPad PRISM and CDD.

### 1.2 High throughput X-ray crystallography

Purified protein[15] at 24 mg/ml in 20 mM HEPES pH 7.5, 50 mM NaCl buffer was diluted to 12 mg/ml with 20 mM HEPES pH 7.5, 50 mM NaCl before performing crystallization using the sitting-drop vapour diffusion method with a reservoir solution containing 11% PEG 4 K, 5% DMSO, 0.1 M MES pH 6.5. Crystals of Mpro in the monoclinic crystal form (C2), with a single monomer in the asymmetric unit, were grown with drop ratios of 0.15 µl protein, 0.3 µl reservoir solution and 0.05 µl seeds prepared from previously produced crystals of the same crystal form[15]. Crystals in the orthorhombic crystal form ($P2_12_12_1$), with the Mpro dimer present in the asymmetric unit, were grown with drop ratios of 0.15 µl protein, 0.15 µl reservoir solution and 0.05 µl seeds prepared from crystals of an immature Mpro mutant in the same crystal form[44].

Compounds were soaked into crystals by adding compound stock solutions directly to the crystallisation drops using an ECHO liquid handler. In brief, 40-90 nl of DMSO solutions (between 20 and 100 mM) were transferred directly to crystallisation drops using giving a final compound concentration of 2-20 mM and DMSO concentration of 10-20%. Drops were incubated at room temperature for approx. 1-3 h prior to mounting and flash cooling in liquid nitrogen without the addition of further cryoprotectant.

Data was collected at Diamond Light Source on the beamline I04-1 at 100 K and processed with the fully automated pipelines at Diamond[45–47], which include XDS [48], xia2[49], autoPROC[50] and DIALS[45]. Further analysis was performed using XChemExplorer[51] with electron density maps generated using DIMPLE (http://ccp4.github.io/dimple/). Ligand-binding events were identified using PanDDA[52] (https://github.com/ConorFWild/pandda) and ligands were manually modelled into PanDDA-calculated event maps or electron density maps using Coot[53]. Ligand restraints were calculated with ACEDRG[54] or GRADE

(grade v. 1.2.19 (Global Phasing Ltd., Cambridge, United Kingdom, 2010)) and structures refined with Buster (Buster v. 2.10.13 (Cambridge, United Kingdom, 2017)). Models and quality annotations were reviewed using XChemReview (citation?), Buster-Report (Buster v. 2.10.13 (Cambridge, United Kingdom, 2017)) and Mogul[55,56].

Coordinates, structure factors and PanDDA event maps for all data sets are available on Fragalysis (https://fragalysis.diamond.ac.uk/viewer/react/preview/target/Mpro).

### 1.3 Viral screening assays

A variety of antiviral replication assays were performed in collaborating laboratories, including cytopathic effect (CPE) inhibition assays at the IIBR, Israel, and Katholieke Universiteit Leuven, RT-qPCR for viral RNA at Radboud University Medical Center, Netherlands, immunofluorescence assays at University of Nebraska Medical centre, USA, and plaque assays and focus forming unit assays at University of Oxford, UK.

#### 1.3.1 Antiviral Cytopathic Effect Assay, VeroE6 (IIBR, Ness-Ziona, Israel)

SARS-CoV-2 (GISAID accession EPI_ISL_406862) was kindly provided by Bundeswehr Institute of Microbiology, Munich, Germany. Virus stocks were propagated (4 passages) and tittered on Vero E6 cells. Handling and working with SARS-CoV-2 virus was conducted in a BSL3 facility in accordance with the biosafety guidelines of the Israel Institute for Biological Research (IIBR). Vero E6 were plated in 96-well plates and treated with compounds in medium containing 2 % fetal bovine serum. The assay plates containing compound dilutions and cells were incubated for 1 hour at 37℃ temperature prior to adding Multiplicity of infection (MOI) 0.01 of viruses. Viruses were added to the entire plate, including virus control wells that did not contain test compound and Remdesivir drug used as positive control. After 72h incubation viral cytopathic effect (CPE) inhibition assay was measured with XTT reagent. Three replicate plates were used.

#### 1.3.2 Antiviral Immunoflourescence assay, VeroE6 (Pathology and Microbiology, University of Nebraska Medical Centre, USA, St Patrick Reid)

Vero E6 cells were pretreated with 20 uM of the Moonshot compounds for around 2h. Cells were then infected with SARS-CoV-2 at a MOI of 0.1 for 24h. Virus infection was terminated by 4 % PFA fixation. Cells were stained using a Rabbit SARS-CoV-2 antibody (Sino Biological 40150-R007) as a primary antibody, and Alexa-488, Hoechst and Cell Mask (Thermo Fisher) as a secondary antibody. Images were collected on the Operetta system imaging system, and analysed using the Harmony software.

#### 1.3.3 Antiviral Focus Forming Unit Assay, Calu-3 (University of Oxford, UK)

***Cell culture***. The African green monkey Vero E6 cell line (ATCC CRL-1586) was cultured in Dulbecco's modified Eagle medium (DMEM) with Glutamax supplemented with 100 µg/mL streptomycin, 100 U/mL penicillin, and 10 % heat-inactivated fetal calf serum (FCS). The human lung cancer cell line Calu-3 (Anderson Ryan, Department of Oncology, Medical Science Division, University of Oxford) was cultured in a 1:1 mixture of DMEM with Glutamax and Ham's F-12 medium supplemented with 100 µg/mL streptomycin, 100 U/mL penicillin, and 10% heat-inactivated FCS. All cells were maintained as mycoplasma free, with regular verifications by polymerase chain reaction (PCR).

**Virus propagation**. SARS-CoV-2 England/2/2020 was provided at passage 1 from Public Health England, Collindale. Passage 2 submaster and passage 3 working stocks were produced by infecting Vero E6 cells at a multiplicity of infection of 0.01 in virus propagation medium (DMEM with Glutamax supplemented with 2 % FCS) and incubating until cytopathic effect was visible. The cell supernatant was then centrifuged at 500 g

for 5 minutes, aliquoted and stored at -80°C. The titre of viral stocks was determined by plaque assay. All subsequent assays were performed using a passage 3 stock.

**Cell viability**. Cell viability was was measured using the CellTiter 96 R AQueous One Solution Cell Proliferation MTA (3-(4,5-dimethylthiazol-2-yl)-5-(3-carboxymethoxyphenyl)-2-(4-sulfophenyl)-2H - 15 tetrazolium, inner salt) Assay (Promega) according to the manufacturer's instruction after treatment with compound. Briefly, Calu 3 cells were treated with compounds in quadruplicate for 3 days. Wells with 200 µL growth medium with and without cells were included as controls in quadruplicate. After the incubation, 100 µL of growth medium was removed and 20 µL of MTS reagent was added to the remaining medium in each well. After a further one to two hour incubation, the absorbance at 490 nm was measured on a Molecular Devices SpectraMax M5 microplate reader.

***Antiviral assays.*** For Focus forming unit assays, a SARS-CoV-2 Microneutralization assay from the W James lab (Dunn School of Pathology, University of Oxford) was adapted for use as a FFU assay. Briefly, 3 half log dilutions of each supernatant to be analyzed were prepared in virus propagation medium. 20µL of each dilution was inoculated into wells of a 96-well plate in quadruplicate followed by 100 µL Vero E6 cells at 4.5 x 10^5 cells/mL in virus propagation medium. The plates were incubated for 2 hours prior to the addition of 100 µL of 1.8 % CMC overlay, and then incubated for a further 24 hours. After 24 hours the overlay was carefully removed and the cells washed once with PBS before fixing with 50µL of 4 % paraformaldehyde, after 30 minutes the paraformaldehyde was removed and replaced with 100µL of 1 % ethanolamine in PBS. The cells were permeabilized by replacing the ethanolamine with 2 % Triton X100 in PBS and incubating at 37°C for 30 minutes. The plates were then washed 3 times with wash buffer (0.1 % Tween 20 in PBS) inverted and gently tapped onto tissue to dry before the addition of 50 µl of EY2A anti-N human mAb (Arthur Huang (Taiwan)/Alain Townsend (Weatherall Institute of Molecular Medicine, University of Oxford)) at 10 pmol in wash buffer. The plates were rocked at room temperature for 1 hour, washed and incubated with 100µl of secondary antibody Anti-Human IgG (Fc-specific)-peroxidase-conjugate produced in Goat diluted 1:5000 at room temperature for 1 hour. 50µL of TrueBlue peroxidase substrate was added to the wells and incubated at RT for 10 min on the rocker, after 10 minutes the substrate was removed and the plates washed with ddH2O for 10 minutes. The H2O was removed and the plates allowed to air dry. The foci were then counted using an ELISPOT classic reader system (AID GmbH).

### 1.3.4 Antiviral qPCR assay, VeroE6 and kidney organoids (Radboud University Medical Center, Nijmegen, Netherlands)

***Cell culture*** African green monkey Vero E6 kidney cells (ATCC CRL-1586) and Vero FM kidney cells (ATCC CCL-81) were cultured in Dulbecco's modified Eagle medium (DMEM) with 4.5 g/L glucose and L-glutamine (Gibco), supplemented with 10% fetal calf serum (FCS, Sigma Aldrich), 100 µg/ml streptomycin and 100 U/ml penicillin (Gibco). Cells were maintained at 37 °C with 5% $CO_2$. Human induced pluripotent stem cell-derived kidney organoids were prepared as previously described [37].

***Virus propagation.*** SARS-CoV-2 (isolate BetaCoV/Munich/BavPat1/2020) was kindly provided by Prof. C. Drosten (Charité-Universitätsmedizin Berlin, Institute of Virology, Berlin, Germany) and was initially cultured in Vero E6 cells up to three passages in the laboratory of Prof. Bart Haagmans (Viroscience Department, Erasmus Medical Center, Rotterdam, The Netherlands). Vero FM cells were infected with passage 3 stock at an MOI of 0.01 in infection medium (DMEM containing L-glutamine, 2% FCS, 20 mM HEPES buffer, 100 µg/ml streptomycin and 100 U/ml penicillin). Cell culture supernatant containing virus was harvested at 48 hours post-infection (hpi), centrifuged to remove cellular debris, filtered using a 0.2 µm syringe filter (Whatman), and stored in 100 µl aliquots at -80°C.

***Virus titration.*** Vero E6 cells were seeded in 12-well plates at a density of 500,000 cells/well. Cell culture medium was discarded at 24 h post-seeding, cells were washed twice with PBS and infected with 10-fold dilutions of the virus stock in unsupplemented DMEM. At 1 hpi, cells were washed with PBS and replaced

with overlay medium, consisting of Minimum Essential medium (Gibco), 2% FCS, 20 mM HEPES buffer, 100 µg/ml streptomycin, 100 U/ml penicillin and 0.75% carboxymethyl cellulose (Sigma Aldrich). At 72 hpi, the overlay medium was discarded, cells were washed with PBS and stained with 0.25% crystal violet solution containing 4% formaldehyde for 30 minutes. Afterwards, staining solution was discarded and plates were washed with PBS, dried and plaques were counted.

***Antiviral assay.*** Vero E6 cells were seeded onto 24-well plates at a density of 150,000 cells/well. At 24 h post-seeding, cell culture medium was discarded, cells were washed twice with PBS and infected with SARS-CoV-2 at an MOI of 0.01 in the presence of six concentrations of the inhibitors (25 µM – 0.06 µM). At 1 hpi, the inoculum was discarded, cells were washed with PBS, and infection medium containing the same concentration of the inhibitors was added to the wells. SARS-CoV-2 infection in the presence of 0.1% DMSO was used as a negative control. At 24 hpi, 100 µl of the cell culture supernatant was added to RNA-Solv reagent (Omega Bio-Tek) and RNA was isolated and precipitated in the presence of glycogen according to manufacturer's instructions. TaqMan Reverse Transcription reagent and random hexamers (Applied Biosystems) were used for cDNA synthesis. Semi-quantitative real-time PCR was performed using GoTaq qPCR (Promega) BRYT Green Dye-based kit using primers targeting the SARS-CoV-2 E protein gene[57] (Forward primer, 5′-ACAGGTACGTTAATAGTTAATAGCGT-3'; Reverse primer, 5′-ACAGGTACGTTAATAGTTAATAGCGT-3'). A standard curve of a plasmid containing the E gene qPCR amplicon was used to convert Ct values relative genome copy numbers. For viability assays, Vero E6 cells were seeded in 96-well white-bottom culture plates (Perkin Elmer) at a density of 30,000 cells per well. At 24 h post-seeding, cells were treated with the same concentrations of compounds as used for the antiviral assay. Cells treated with 0.1 % DMSO were used as a negative control. At 24 h post-treatment, cell viability was assessed using the Cell Titer Glo 2.0 kit (Promega) using the Victor Multilabel Plate Reader (Perkin Elmer) to measure luminescence signal.

**Antiviral assays in organoids**. Human iPSC-derived kidney organoids cultured in transwell filters (Corning) were infected with SARS-CoV-2 in the presence of 1 and 10 µM of MAT-POS-932d1078-3, PET-UNK-29afea89-2 or 0.1% DMSO using an MOI of 1.0 in Essential 6 medium (Gibco) at 37°C and 5 % $CO_2$, exposing the cells both basolaterally and apically to the inoculum. After 24 h, medium containing the inoculum was removed and fresh Essential 6 medium containing the same concentration of inhibitor was added to the basolateral compartment and cells were cultured for an additional 24 h. At 48 hpi, organoids were washed in PBS, and the apical surface was exposed to Essential 6 medium for 10 min at 37°C, which was collected and used for viral titration. Individual organoids were harvested for RNA isolation using the PureLink RNA mini kit (Thermo Fisher) according to manufacturer's instructions. Viral RNA copies were analyzed by RT-qPCR on the SARS-CoV E gene, as described previously[58].

### 1.3.5 Cytopathic Effect Assay, hACE2-TMPRSS2 cells (Katholieke Universiteit Leuven)

***Virus isolation and virus stocks*** All virus-related work was conducted in the high-containment BSL3 facilities of the KU Leuven Rega Institute (3CAPS) under licenses AMV 30112018 SBB 219 2018 0892 and AMV 23102017 SBB 219 2017 0589 according to institutional guidelines. The Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) strain used for this study was the Alpha variant of Concern (derived from hCoV-19/Belgium/rega-12211513/2020; EPI_ISL_791333, 2020-12-21). Virus sample was originally isolated in-house from nasopharyngeal swabs taken from travellers returning to Belgium (baseline surveillance) and were subjected to sequencing on a MinION platform (Oxford Nanopore) directly from the nasopharyngeal swabs. Virus stocks were then grown on Vero E6 cells in (DMEM 2% FBS medium) and passaged one time on A549-ACE2TMPRSS2 cells. Median tissue culture infectious doses (TCID50) was defined by end-point titration.

***A549-ACE2-TMPRSS2 assay*** A549-Dual™ hACE2-TMPRSS2 cells obtained by Invitrogen (Cat. a549d-cov2r) were cultured in DMEM 10% FCS (Hyclone) supplemented with 10 µg/ml blasticidin (Invivogen, ant-

bl-05), 100 µg/ml hygromycin (Invivogen, ant-hg-1), 0.5 µg/ml puromycin (Invivogen, ant-pr-1) and 100 µg/ml zeocin (Invivogen, ant-zn-05). For antiviral assay, cells were seeded in assay medium (DMEM 2%) at a density of 15,000 cells/well. One day after, compounds were serially diluted in assay medium (DMEM supplemented with 2% v/v FCS) and cells were infected with their respective SARS-CoV-2 strain at a MOI of approximately 0.003 TCID50/ml. On day 4 pi., differences in cell viability caused by virus-induced CPE or by compound-specific side effects were analyzed using MTS as described previously [59]. Cytotoxic effects caused by compound treatment alone were monitored in parallel plates containing mock-infected cells.

### 1.3.6 Immunofluorescence SARS-CoV-2 antiviral screening assay, Hela-ACE2 (Mount Sinai)

Assessment of cross-reactivity against SARS-CoV-2 variant strains and cytotoxicity assays were performed as previously described[60]. In brief, two thousand HeLa-ACE2 cells (BPS Bioscience) were seeded into 96-well plates in DMEM (10% FBS) and incubated for 24 hours at 37°C, 5% CO2. Two hours before infection, the medium was replaced with 100 µL of DMEM (2% FBS) containing the compound of interest at concentrations 50% greater than those indicated, including a DMSO control. Plates were then transferred into the BSL3 facility and 100 PFU (MOI = 0.025) was added in 50 µL of DMEM (2% FBS), bringing the final compound concentration to those indicated. Plates were then incubated for 48 hours at 37°C. After infection, supernatants were removed and cells were fixed with 4% formaldehyde for 24 hours prior to being removed from the BSL3 facility. The cells were then immunostained for the viral N protein (an inhouse mAb 1C7, provided by Dr. Thomas Moran, Thomas.Moran@mssm.edu) with a DAPI counterstain. Infected cells (488 nm) and total cells (DAPI) were quantified using the Celigo (Nexcelcom) imaging cytometer. Infectivity was measured by the accumulation of viral N protein (fluorescence accumulation). Percent infection was quantified as ((Infected cells/Total cells) - Background) *100 and the DMSO control was then set to 100% infection for analysis. Data was fit using nonlinear regression and IC50s for each experiment were determined using GraphPad Prism version 8.0.0 (San Diego, CA). Cytotoxicity was also performed using the MTT assay (Roche), according to the manufacturer's instructions. Cytotoxicity was performed in uninfected cells with same compound dilutions and concurrent with viral replication assay. All assays were performed in biologically independent triplicates.

## 2. Computational methods

### 2.1 Synthetic route planning

We employ an approach based on the Molecular Transformer technology[26]. Our algorithm uses natural language processing to predict the outcomes of chemical reactions and design retrosynthetic routes starting from commercially available building blocks. This proprietary platform is provided free of charge by PostEra Inc (http://postera.ai). Additionally, Manifold (https://postera.ai/manifold) was built by PostEra Inc. during the project to search the entire space of purchasable molecules, and automatically find the optimal building blocks.

### 2.2 Alchemical free energy calculations

Large-scale alchemical free energy calculations were conducted in "Sprints" in which each set of calculations aimed to prioritize compounds that could be produced from a common synthetic intermediate using Enamine's extensive building block library, resulting in synthetic libraries of hundreds to tens of thousands. Virtual synthetic libraries were organized into a star map, where all transformations were made with respect to a single reference X-ray structure and compound with experimentally measured bioactivity. X-ray structures were prepared using the OpenEye Toolkit SpruceTK with manually controlled protonation states for the key His61:Cys145 catalytic dyad (variously using zwitterionic or uncharged states). Initial poses of

target compounds were were generated via constrained conformer enumeration to identify minimally-clashing poses using Omega (from the OpenEye Toolkit) using a strategy that closely follows an exercise described in a blog post by Pat Walters (http://practicalcheminformatics.blogspot.com/2020/03/building-on-fragments-from-diamondxchem_30.html). Alchemical free energy calculations were then prepared using the open source perses relative alchemical free energy toolkit[34] (https://github.com/choderalab/perses), and nonequilibrium switching alchemical free energy calculations[61] were run on Folding@home using the OpenMM compute core[35]. Nonequilibrium switching calculations used 1 ns nonequilibrium alchemical trajectories, where most calculations were performed with 1 fs timesteps without constraints to hydrogen due to technical limitations that have been resolved in calculations employing OpenMM 7.5.1 and later. We used the Open Force Field Initiative OpenFF "Parsley" small molecule force fields[33] (multiple generations were released and used as the project evolved) and the AMBER14SB protein force field[62] with recommended ion parameters, and TIP3P water. Calculations were analyzed using the fah-xchem dashboard (https://github.com/choderalab/fah-xchem) using the Bennett acceptance ratio[63,64] (https://threeplusone.com/pubs/gecthesis) and posted online in real time for the medicinal chemistry team to consult in making decisions about which compounds to prioritize.

Scripts for performing and analyzing alchemical free energy calculations, as well as an index of computed datasets and dashboards are available at https://github.com/foldingathome/covid-moonshot

Code for generating the COVID Moonshot alchemical free energy calculation web dashboards is available here: https://github.com/choderalab/fah-xchem

### 2.3 Structural flexibility and interactions analysis

Protein-ligand interactions are the driving forces for molecular recognition. In this work, the *plipify* repo (https://github.com/volkamerlab/plipify) is used to detect shared interaction hot spots within the different MPro structures. plipify is a python wrapper built on top of PLIP[65], a tool that enables automatic generation of protein-ligand interaction profiles for single complexes, to allow combining these profiles for multiple structures.

To generate the hotspots (depicted in **Figure 3A**), the fragalysis data was downloaded (as of January 2022, https://fragalysis.diamond.ac.uk/api/targets/?format=json&title=Mpro). The respective 587 pre-aligned complex structures were further investigated (found under data/{target}/aligned/{crystal_name}/{crystal_name}_bound.pdb). Only one chain per structure is kept, and the structures are protonated using Amber's *reduce* function. Plipify is invoked and structures are excluded from further analysis if they do not contain exactly one binding site (i.e. PLIP detects either zero or more than 1 binding sites), the sequence contains gaps ('-') or the sequence length differs more than a standard deviation from the average length across all investigated structures.

This resulted in a final set of 416 complex structures, used to generate the interaction fingerprints. Note for this study, only hbond-donor, hbond-acceptor, salt bridge, hydrophobic, pistacking, and halogen interactions are inspected. Interacting residues are only included if the summed interaction count per residue over all investigated structures is greater than five. The resulting structural depiction (**Figure 3A**) were generated using pymol, and structure Mpro-P1788_0A_bound_chainA (protonated) is displayed (scripts available at https://github.com/volkamerlab/plipify/tree/master/projects/01/fragalysis.ipynb).

Scripts used to generate structural **Figure 1A**, **Figure 3B**, and **Supplementary FIgure 1B** are available at https://github.com/choderalab/covid-moonshot-manuscript/

## 3 Human dose prediction

At this early stage, using screening in vivo rat PK data and the in vitro EC50 and EC90 data, we predicted human dose based on Caldwell and Tang methods for Cl and Caldwell for Vd,ss. Estimating the rate of absorption (ka, based on rat PK data) as 0.4-0.6 $hr^{-1}$ gives a human dose range when incorporating EC50 and EC90 concentration targets.

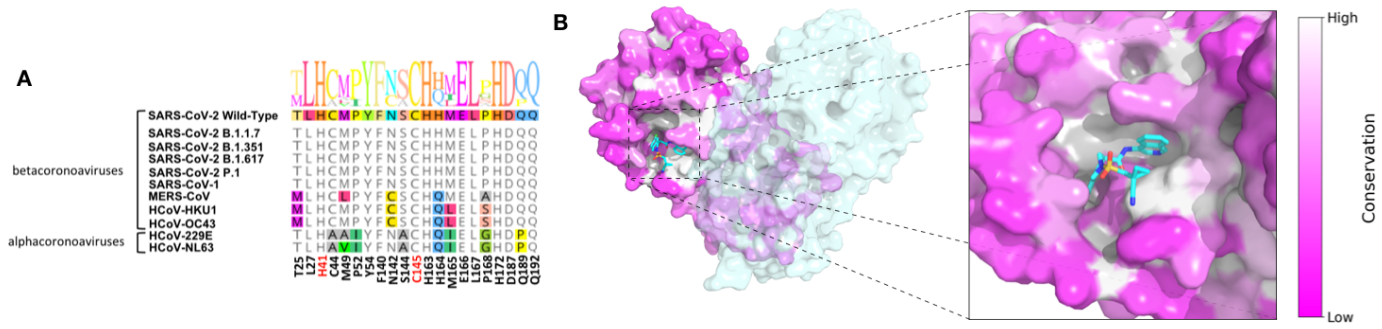
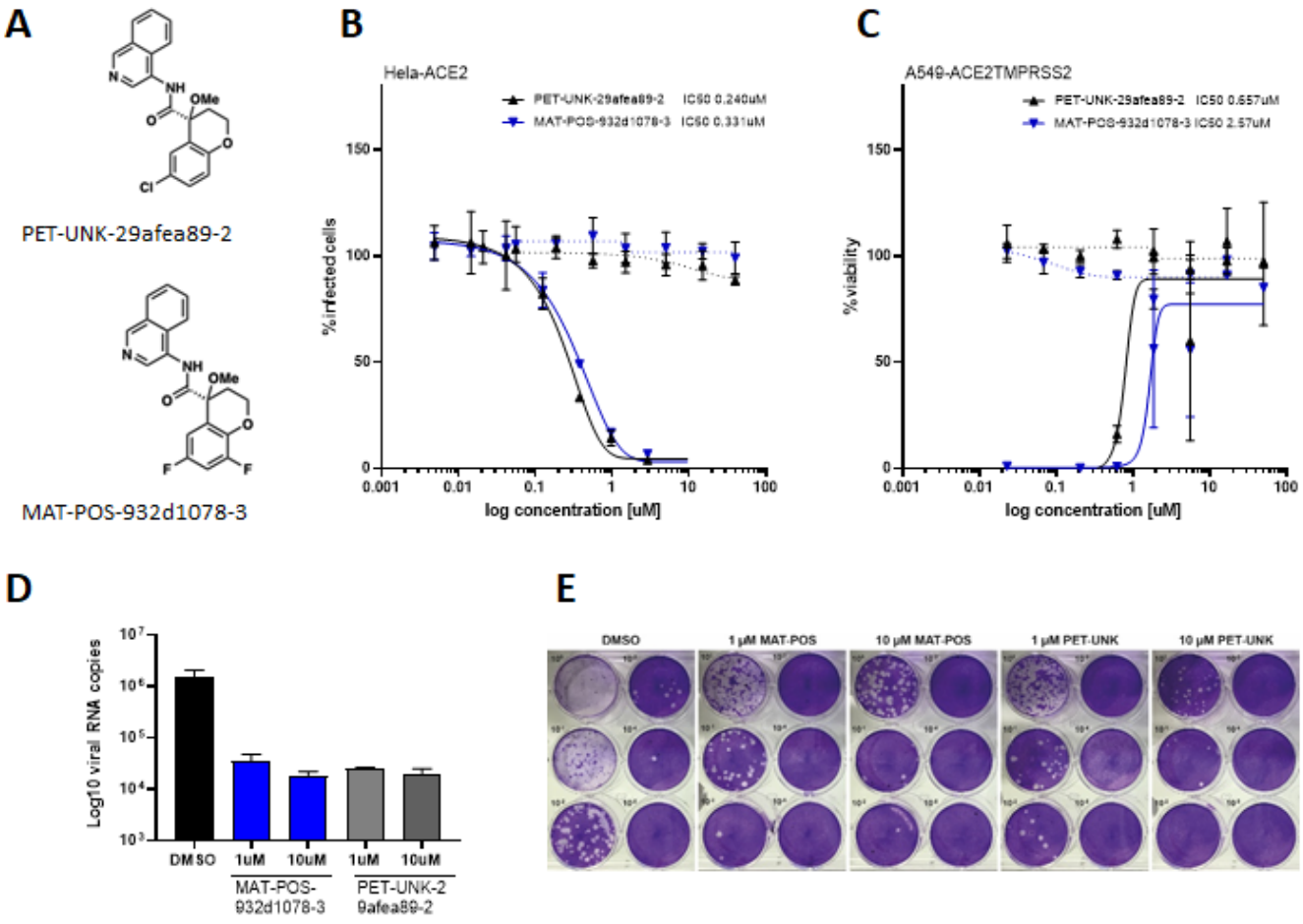
| Rate of Absorption Estimated from rat | BID (mg) EC50 (EC90) | TID (mg) EC50 (EC90) |
|---|---|---|
| Ka = 0.6 | 988 (1941) | 169 (332) |
| Ka = 0.4 | 434 (853) | 108 (212) |

## Supplementary Figures

**Supplementary Figure 1: The SARS-CoV-2 main viral protease (Mpro) is highly conserved across coronaviruses.**

**A:** Mpro sequences across coronaviruses are highly conserved due to their requirement to cleave viral polyproteins in numerous locations, showing very little variation in residues lining the active site near the scissile bond.

**B:** Available structural data for Mpro from multiple coronaviruses shows a high degree of structural conservation, especially in the vicinity of the active site.



**Supplementary Figure 2: Closely related analogues of the lead compound, PET-UNK-29afea89-2 and MAT-POS-932d1078-3, demonstrate antiviral activity across different cellular antiviral assays and a kidney organoid model.** **A**: shows the chemical structure of PET-UNK-29afea89-2 and MAT-POS-932d1078-3. **B:** dose-response curves of both compounds in Immunofluorescence assays in Hela-ACE2 cells, and **C**: Cytopathic Effect assays in A549-ACE2-TMPRSS2 cells. The curves also show the cytotoxicity data (dotted lines), demonstrating the lack of cytotoxic activity across all three cell lines. **D-E**: Antiviral activity of MAT-POS-932d1078-3 and PET-UNK-29afea89-2 in kidney organoids infected with SARS-CoV-2 in the presence of 1 μM and 10 μM of compounds or DMSO as a control. **D:** Intracellular viral RNA measured by qPCRand **E**: infectious viral titers released from the apical side of the organoids at 48 hpi, measured by plaque assay on Vero E6 cells. Data in D are mean and SD of 2 biological replicates from a representative experiment of 2 independent experiments. Intracellular viral RNA levels in **D** were normalized to expression of the β-actin housekeeping gene.