Open Science Discovery of Potent Non-Covalent SARS-CoV-2 Main Protease Inhibitors

Melissa L. Boby[1,2,3,*], Daren Fearon[4,5,*], Matteo Ferla[6,*], Mihajlo Filep[7,*], Lizbé Koekemoer[8,*], Matthew C. Robinson[9,*], The COVID Moonshot Consortium, John D. Chodera[3,#], Alpha A Lee[9,#], Nir London[7,#], Annette von Delft[6,8,10,#], Frank von Delft[4,5,8,10,11,#]

[1] Pharmacology Graduate Program, Weill Cornell Graduate School of Medical Sciences, New York, NY 10065, USA
[2] Program in Chemical Biology, Sloan Kettering Institute, Memorial Sloan Kettering Cancer Center, New York, NY 10065, USA
[3] Program in Computational and Systems Biology, Sloan Kettering Institute, Memorial Sloan Kettering Cancer Center, New York, NY 10065, USA
[4] Diamond Light Source Ltd., Harwell Science and Innovation Campus, Didcot, OX11 0QX, UK
[5] Research Complex at Harwell, Harwell Science and Innovation Campus, Didcot OX11 0FA, United Kingdom
[6] Oxford Biomedical Research Centre, National Institute for Health Research, University of Oxford, Oxford, UK
[7] Department of Chemical and Structural Biology, The Weizmann Institute of Science, Rehovot, Israel, 7610001
[8] Centre for Medicines Discovery, Nuffield Department of Medicine, University of Oxford, Oxford, UK
[9] PostEra Inc., 1 Broadway, 14th Floor, Cambridge, MA 02142, USA
[10] Structural Genomics Consortium, Nuffield Department of Medicine, University of Oxford, Oxford, UK
[11] Department of Biochemistry, University of Johannesburg, Auckland Park, Johannesburg 2006, South Africa

[*] Equal contribution ; 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)

The COVID Moonshot Consortium Membership:
https://docs.google.com/spreadsheets/d/1OnPzj5IvYtD5uu6tqmlhvOX47__WgN1VrbzMH_kqns0

Abstract

We report the results of the **COVID Moonshot**, a fully open-science, crowd sourced, structure-enabled drug discovery campaign targeting the SARS-CoV-2 main protease. We discovered a non-covalent, non-peptidic inhibitor scaffold with lead-like properties that is differentiated from current main protease inhibitors. Our approach leveraged crowdsourcing, machine learning, exascale molecular simulations, and high-throughput structural biology and chemistry. We generated a detailed map of the structural plasticity of the SARS-CoV-2 main protease, extensive structure-activity relationships for multiple chemotypes, and a wealth of biochemical activity data. All compound designs (>18,000 designs), crystallographic data (>840 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

Despite rapid progress in vaccine development, the global failure to abate coronavirus disease 2019 (COVID-19) culminating in more than 690 million confirmed cases worldwide by July 2023, will likely cause the virus to become endemic (*1*), and continue 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*). Several directly acting oral antivirals are now approved for usage against severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) infection, including ritonavir-boosted nirmatrelvir(*4*), ensitrelvir (Japan)(*5*), and molnupiravir(*6*).

COVID-19 is not an isolated event, but the latest exemplar of a series of threats to human health caused by beta-coronaviruses also responsible for the SARS (2003) and Middle East respiratory syndrome (MERS) (2010) pandemics(*7*). Open knowledge bases and technology infrastructures for antiviral drug discovery will enable pandemic preparedness by refreshing the antivirals pipeline and providing multiple starting points for the development of therapeutics. Here, we report the open science discovery of a potent SARS-CoV-2 antiviral lead compound and a roadmap for the potential 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 of its structure and substrate profile to human proteases(*8*) (**Fig S1**). Pioneering studies during and after the 2003 SARS pandemic established the linkage between Mpro inhibition and antiviral activity in cell culture(*9*). This work has been corroborated by in vitro and in vivo studies for SARS-CoV-2(*10*, *11*) and the clinical success of nirmatrelvir (the Mpro inhibitor component of Paxlovid)(*12*) and ensitrevir (Xocova)(*13*, *14*).

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 non-covalent, non-peptidomimetic scaffolds. First-generation oral Mpro inhibitors have now demonstrated clinical efficacy(*15*, *16*), but the need for cytochrome P450 3A4 (CYP3A4) inhibitor co-dosing (ritonavir, in the case of Paxlovid) to achieve sufficient human exposure may substantially limit use in at-risk populations due to drug-drug interactions(*17*). There remains a need for chemically differentiated oral antiviral protease inhibitors with the potential to enter clinical development.

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(*18*, *19*), 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**)(*20*). This dataset was posted online on 18 Mar 2020(*21*), days after the screen was completed(*21*). 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(*22*).

Numerous approaches have been proposed to advance from fragments to lead compounds(*23*, *24*). 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 an opportunity for rapid lead generation by combining chemotypes from multiple fragments. Nonetheless, this approach 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(*25*) and ribonucleic acid (RNA) (*26*) design campaigns, we hypothesized that crowdsourced human analysis 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(*19*). Compounds selected for synthesis were evaluated by biochemical assays (**Data S1**) and x-ray crystallography and the results were released rapidly on the same platform, enabling contributing designers to build on all available data, as well as designs contributed by others. To facilitate transparency and maximal speed, and to avoid delays around intellectual property, 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 (**Data S2**).

Many submissions exploited spatially overlapping fragment hits. For example, the submission **TRY-UNI-714a760b-6** was inspired by five overlapping fragments, furnishing a non-covalent inhibitor with an SARS-CoV-2 Mpro enzymatic IC_{50} of 23.7 μM (**Figure 1C**). This compound seeded the "aminopyridine" series, whose optimization is described in detail below. We should note that only 11 of the 998 fragments in the DSi-poised library(*27*, *28*) contained a 3-amino pyridine, yet four of them were successfully identified in the structures and were consequently picked-up for merging by the designers. Apart from the aminopyridine series, our campaign identified three other major chemically distinct lead series with measurable potencies against SARS-CoV-2 Mpro inspired by reported SARS-CoV-1 inhibitors (**Fig S2**). Those compounds span the same binding pocket but feature different chemotypes, and the large quantity of structure-activity relationship (SAR) subsequently generated for these series furnishes multiple backup series with different risk profiles. Other groups have subsequently further elaborated on the Ugi(*29*, *30*) and the benzotriazole series we generated(*31*).

Analysis of the submissions provides some hints to the utility of crowdsourcing as a general strategy for hit-discovery or hit-to-lead campaigns. A qualitative assessment of the textual description of submitted designs (**Fig S3**) hints that many of the designers used tools like docking to assess fragment 'linking', 'merging' or 'combination'. When trying to more thoroughly categorize submissions it doesn't appear that hypothesis-driven designs perform better than docking driven designs, however 'predicting' historical SARS inhibitors is the best performing strategy (**Fig S4**; **Figure 1D**). Throughout the campaign, designs were contributed by both the core group of labs and medicinal chemists leading this project, as well as by the community. One could hypothesize that the core group being committed to the project, as well as thoroughly invested in the campaign details would contribute more potent designs. However, there is no obvious difference in the distributions of designs produced by the core group vs. the community, in the early stages of the campaign

(**Figure 1D**) nor were the designs contributed by the community less synthetically accessible (**Figure 1E**). Later in the campaign, (lead optimization stage) the number of submissions from the community decreased, and comparing potency became irrelevant as other attributes of the molecules were being optimized. It is important to mention that several key compounds along the optimization trajectory (**Figure 5A**) of our lead were contributed by the community and not core group members including: **TRY-UNI-714a760b-6**, **ADA-UCB-6c2cb422-1** and **VLA-UCB-1dbca3b4-15** (the racemic mixture of **MAT-POS-b3e365b9-1**). Although anecdotal, this example demonstrates the potential power of crowdsourcing as a strategy to drive fragment-to-lead campaigns.

Technologies to support rapid optimization cycles

With a growing number of chemically diverse submissions, we relied on a team of experienced medicinal chemists supported by computational methods to aid in triaging design proposals 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 did not use an "autonomous" approach - expert judgment is used to make decisions given all the model predictions. Further, in the context of a fast-moving campaign, we prioritized making progress over granular "human-versus-machine" evaluations.

Synthetic route predictions guided decision making to accelerate design-make-test-analyze cycles
We used an established synthetic contract research organization (CRO), Enamine, to carry out rapid synthesis of progressed compound designs. To take full advantage of the available building block collection, we used a machine learning approach that plans efficient retrosynthetic routes to predict synthetic tractability(*32*, *33*). 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. Our predicted synthetic complexity correlated with the actual time taken to synthesize target compounds and the algorithm was able to pick out advanced intermediates as starting materials (**Figure 2A**).

Alchemical free energy calculations prioritized potent compounds for synthesis
We estimated potency of proposed designs and virtual synthetic libraries of analogues using alchemical free energy calculations(*34–36*), 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(*37*)—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(*38*)—to compute the free energy of binding of all 20,000+ crowdsourced and internal design submissions using the Open Force Field Initiative "Parsley" small molecule force fields(*39*) and nonequilibrium switching with the open source PERSES alchemical free energy toolkit(*40–42*) based on the GPU-accelerated OpenMM framework(*43*)(*38*) (see **Detailed Methods**). Comprehensive sampling was prioritized over efficiency of computation given the abundant compute resources available on Folding@home.

We first performed a small retrospective study using bioactivity data generated from the first week of crowdsourced compound designs, triaged solely using synthetic accessibility. The results of these free energy calculations showed good correlation with experimentally-measured affinities (**Figure 2B**). Henceforth, alchemical free energy calculations were used as an additional (though not sole) criteria to guide compound selection and iterative design (see **Data Availability**). During the campaign, distinct objectives were solicited from submitters to address medicinal chemistry problems, and free energy calculations were used to assess these submissions based on predicted potency. **Figure 2C** shows that predicted pIC_{50} tracks experimental measurements across 3 chronologically distinct design campaigns: decoration of the benzopyran ring, replacement of the benzopyran system, and replacement of the isoquinoline system. Note that some design ideas with low predicted pIC_{50} were synthesized since the medicinal chemistry team

balanced between gaining insights on structure-activity and structure-property relationship and potency optimization. The champion compounds from each design campaign are highlighted on the right panel of **Figure 2C**. Although free energy calculations identified multiple potency-improving transformations, the strategically useful one was the swap from pyran to a piperidine sulfonamide system, which is on the critical path to the lead compound (**Figure 5A)**. On average, 80 GPU hours/compound was used across the three panels (see **Detailed Methods**).

A major strength of alchemical free energy calculations proved to be their ability to select potent analogues from virtual synthetic libraries from which the medicinal chemistry team had already selected compounds sharing a common intermediate, as well as highlighting submitted designs predicted to be highly potent but where major synthetic effort would be required. Our design team prioritized for synthesis small libraries suggested by the aforementioned computational approaches. Chemically-related groups of outliers frequently provided new chemical insight that informed modeling choices (**Fig S5**). The approach was not without drawbacks, including the need to focus on a single reference compound and structure to design transformation networks (rather than leveraging the abundant structural data), the requirement that protonation states be fixed for the entire calculation (requiring the entire transformation network to be recomputed to assess a different protonation state), and the relatively large computational cost required to handle large replacements (see **Detailed Methods**). The method is also not uniformly accurate across all chemical transformations, and accurately estimating its accuracy beforehand is challenging. For example, isoquinoline replacements show lower correlation between calculated and predicted free energy (Panel 3 of **Figure 2B**) compared to the benzopyran replacements (Panel 2 of **Figure 2B**).

Nanomole-scale high-throughput chemistry enabled rapid evaluation of structure-activity relationship (SAR)
A complementary method for rapid SAR evaluation was the use of nanomole scale high-throughput chemistry(*44*, *45*) (HTC), coupled with a 'direct to biology'(*46–48*) biochemical screening. Two examples include the optimization of the Chan-Lam reaction(*49*) to extend molecule **ADA-UCB-6c2cb422-1** and amide coupling to extend **MAT-POS-4223bc15-21** (**Figure 2D**). In both cases, we determined the co-crystal structures of the parent compounds (**Fig S6**) and suggested vectors that could target the P4 pocket of Mpro. Optimization of the reaction conditions was performed for the starting building block with model amines (**Figs S7-S8**) and the optimal conditions were applied to HTC with a library of 300 amine building blocks (**Data S3**). Yield estimation was performed in both cases and showed for the Chan-Lam library only 29 of the library yielded >30% of the desired product, in comparison to 151 for the amide coupling. Nevertheless, the crude mixtures were subjected to a biochemical assay against Mpro (**Data S3**). Seven compounds were selected for resynthesis from the Chan-Lam series and 20 from the amide series (**Fig S9**). In parallel to synthesis, the crude reaction mixtures were subjected to soaking and x-ray crystallography. The structures verified the extended compounds indeed adopt a similar binding mode to the parent. Chan-Lam extended compounds indeed occupied P4, whereas the amides extended towards P3/P5, in both cases forming new interactions with Mpro (**Figure 2D**). Upon resynthesis, one of the Chan-Lam compounds was able to slightly improve over the parent compound IC_{50}. Several of the amide-coupling series were able to improve by up to 300-fold on the parent acid-compound (up-to 3-fold on the corresponding methyl-amide) with the best inhibitor exhibiting an IC_{50} of 28nM against Mpro.

Covalent targeting strategies
Another approach that was attempted in order to rapidly gain potency was the use of electrophiles to covalently target the catalytic Cys145. The original fragment screen(*20*) that launched this effort, included electrophiles(*50*) and resulted in 48 structures of covalently bound fragments, the vast majority of which were chloroacetamides. Some of the earliest, and most potent, fragment merges explored by both the core group as well as the community were of chloroacetamide (**Figure 1D**) and further optimization improved chloroacetamide fragments' IC_{50} to as low as 300nM (**Fig S10**). Chloroacetamides, however, are not considered suitable for therapeutics and therefore we aimed to move towards acrylamides by derivatizing potent reversible hits(*30*) (**Fig S11**). Ultimately, we focused on a non-covalent series, but the chlorophenyl

moiety that remained throughout the series (**Figure 5A**) was adopted from a chloroacetamide hit fragment (**AAR-POS-0daf6b7e-10; Figure 1C**).

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

We selected compounds based on synthetic tractability and alchemical free energy calculations. We profiled every compound through crystal soaking and x-ray diffraction totaling in 584 structures (see **Table S1 and Fig S12** for average statistics; **Data S4** for Crystallographic and refinement statistics and **Fig S13** for ligand density for the structures highlighted in this manuscript). Analysis of a subset of this large trove of structural data, (n=367, up to July 2021), reveals the hotspots for ligand engagement and plasticity of each binding pocket. **Figure 3** highlights the statistics of intermolecular interactions between the residues and our ligands. 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), and N142 (hydrophobic interactions). Whereas P2 interactions are 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 more rigid than P2 (**Figure 4**). The rigidity is also dependent on the chemical series (**Fig S2**), with the Ugi and benzotriazole series being able to 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.

Design of a SARS-CoV-2 Mpro inhibitor lead series 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(*51*). **Figure 5A** outlines the critical intermediates on the path towards an optimized lead compound.

Starting from the fragment hit, we explored the P1 pocket, which admits a steep SAR, perhaps unsurprising given its rigidity and preference for directional H-bond interactions (**Figure 3A**). An 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 favors hydrophobic moieties. A step-change in potency was achieved by rigidifying the scaffold: We introduced a 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 5A**), chosen because of building block availability. Despite possessing a degree of molecular complexity, **MAT-POS-b3e365b9-1** is obtained via a one-step amide coupling (**Figure 2A**). We then further explored the P2 pocket with a library chemistry strategy in mind. Thus, guided by free energy calculations (**Figure 2C**), we first substituted the chromane for a tetrahydroisoquinoline to introduce a functionalizable handle (**MAT-POS-3ccb8ef6-1, Figure 5A**), which maintained potency. Finally, we constructed a focused library realized via sulphonamide Schotten-Baumann coupling (**Fig. S14**), furnishing an increase in both enzymatic inhibition and cellular antiviral efficacy. This work led to a potent antiviral chemical series (**Figure 5A)** with a favorable safety profile, low brain penetrance (**Fig S15; Data S5**), improved oral bioavailability, but moderate in vitro-in vivo correlation in clearance (**Fig S16; Data S5**; all measured cellular antiviral data is available in **Data S6**).

As an example for the aminopyridine lead series, we discuss antiviral efficacy, absorption, distribution, metabolism and excretion (ADME) and pharmacokinetic (PK) characteristics of compound **MAT-POS-e194df51-1:**. **MAT-POS-e194df51-1** was 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 5B**). This is in line with overall cellular efficacy for the chemical series; of 150 compounds with enzyme assay IC_{50} <500 nM assessed in A549-ACE2-TMPRSS2 cellular cythopatic effect (CPE) assays, 15 compounds showed lower EC_{50} values compared to the internal control nirmatrelvir that was measured at an EC_{50} of 218 nM in this assay (**Figure 5C**). Similarly, good antiviral activity was measured across "crowdsourced" antiviral assays across different laboratories and cell lines, including assays performed with and without p-gp inhibitors and using nirmatrelvir as an internal control (**Figure 5D**). We also observed good cross-reactivity of our lead compound **MAT-POS-e194df51-1** against known SARS-CoV-2 variants Alpha, Beta, Delta and Omicron (**Figure 5E**). Closely related molecules **PET-UNK-29afea89-2** and **MAT-POS-932d1078-3** with EC_{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 (**Fig S17A** and **B**) show a >100-fold reduction of intracellular viral RNA and infectious virus secretion into the apical compartment of human induced pluripotent stem cell-derived kidney organoids (**Fig S16D** 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(*52*). **MAT-POS-e194df51-1** exhibits favorable properties required for an orally bioavailable inhibitor (**Figure 5F,G**). In addition, crystallographic studies reveal that the interaction pattern of **MAT-POS-e194df51-1** with the Mpro binding site is distinct to approved Mpro inhibitors nirmatrelvir and ensitrelvir (S-217622) (**Fig S18)**, thus potentially offering complementary resistance profiles and justifying further development.

Open Science presents a viable route to early drug discovery

The results presented here reflect the success of an open science, patent-free antiviral discovery program in rapidly developing a differentiated optimized lead in response to an emerging pandemic threat. As a result of the open science policy, a large number of collaborators (now **The COVID Moonshot Consortium**) were 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. As a striking example for the impact of open-science, the Shionogi clinical candidate S-217622 (which has now received emergency approval in Japan as Xocova [ensitrelvir]) was identified in-part based on crystallographic data from The COVID Moonshot Consortium(*53*).

Despite our optimization and characterization efforts, considerable gaps from reporting a clinical candidate remain: The series requires further PK and pharmacodynamic (PD) optimization; in particular it displays high clearance and low bioavailability. As it stands, it would likely not be able to achieve therapeutic exposure without a pharmacokinetic booster (such as ritonavir). To move forward, additional in-depth safety data is required, as well as additional PK data from a second species to enable accurate human dose prediction. The COVID Moonshot and its lead series for COVID-19 have been adopted into the drug development portfolio of the Drugs for Neglected Diseases Initiative (DNDi) for further lead optimization and downstream pre-clinical development. This work is funded by a $10M Wellcome Trust through the COVID-19 Therapeutics Accelerator (ACT-A), of which results will be reported upon filing Clinical Trials Authorization (CTA) (*54*). To reach Phase II readiness we expect a further $7.5M will be required in addition to process route development costs(*55*).

Open science efforts have transformed many areas of biosciences, with examples such as the Human Genome Project(*56*), the Structural Genomics Consortium(*57*), and the RAS Initiative(*58*). The COVID Moonshot provides an example of open science drug discovery leading to advances in infectious diseases drug discovery—a research area of grave public importance but one which is chronically underfunded by the private sector(*59*).

Methods Summary

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 registered and assayed separately. Because the absolute stereochemical identity of 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.

Submission analysis: the submitter names were standardised by removing affiliations and expansion of first name abbreviations, the submissions by two users who submitted large batches of compounds in an automated way in contravention of the goal of the project were removed. The word cloud was generated by filtered against 1,000 most common words and removing grammatical inflections and generating an image with an online word cloud generator. The classification of the methodology was done by presence of keywords determined by a simple keyword classifier with manually determined words (circa 100 training, 100 test) wherein 'dock', 'seesar', 'vina', 'autodock', 'screen', 'drug-hunter' were typical of docking, while 'by-eye', 'merg[ing]', 'link[ing]', 'coupl[ing]' were typical of hypothesis driven methods. A large fraction could not be accurately classified due to paucity of information. SAScore was calculated with Postera Manifold under the retrosynthesis route.

1. Experimental methods
1.1 Protease activity assays
1.1.1 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%.

The assay was calibrated at different enzyme concentrations to confirm linearity and response of protease activity, as well as optimization of buffer components for most stable and reproducible assay conditions. Substrate concentration was chosen after titration to minimize saturation of signal in the plate reader while obtaining a satisfactory and robust dynamic range of typically 5-6 fold over control without enzyme. We used low substrate concentrations of the bright FRET peptide to avoid "inner filter effect"(*60*), and to bias towards detection of competitive inhibitors(*61*). As positive control, under our assay condition Nirmatrelvir has IC50 of 2.6nM.

1.1.2 RapidFire MPro inhibition assay

The assay was performed according to the published procedure(*62*). 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(*20*) 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(*20*). 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(*63*).

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(*64–66*), which include XDS(*67*), xia2(*68*), autoPROC(*69*) and DIALS(*64*). Further analysis was performed using XChemExplorer(*70*) with electron density maps generated using DIMPLE (http://ccp4.github.io/dimple/). Ligand-binding events were identified using PanDDA(*71*) (https://github.com/ConorFWild/pandda) and ligands were manually modelled into PanDDA-calculated event maps or electron density maps using Coot(*72*). Ligand restraints were calculated with ACEDRG(*73*) 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(*74*), Buster-Report (Buster v. 2.10.13 (Cambridge, United Kingdom, 2017)) and Mogul(*75*, *76*).

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(*52*).

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(*77*) (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(*78*).

1.3.5 High-content SARS-CoV-2 antiviral screening assay, Hela-ACE2 (Takeda via Calibr/TSRI)

SARS-CoV-2/HeLa-ACE2 high-content screening assay. Compounds are acoustically transferred into 384-well µclear-bottom plates (Greiner, Part. No. 781090-2B) and HeLa-ACE2 cells are seeded in the plates in 2% FBS at a density of 1.0×10^3 cells per well. Plated cells are transported to the BSL3 facility where SARS-CoV-2 (strain USA-WA1/2020 propagated in Vero E6 cells) diluted in assay media is added to achieve ~30 – 50% infected cells. Plates are incubated for 24 h at 34℃ 5% CO_2, and then fixed with 8% formaldehyde. Fixed cells are stained with human polyclonal sera as the primary antibody, goat anti-human H+L conjugated Alexa 488 (Thermo Fisher Scientific A11013) as the secondary antibody, and antifade-46-diamidino-2-

phenylindole (DAPI; Thermo Fisher Scientific D1306) to stain DNA, with PBS 0.05% Tween 20 washes in between fixation and subsequent primary and secondary antibody staining. Plates are imaged using the ImageXpress Micro Confocal High-Content Imaging System (Molecular Devices) with a 10× objective, with 4 fields imaged per well. Images are analyzed using the Multi-Wavelength Cell Scoring Application Module (MetaXpress), with DAPI staining identifying the host-cell nuclei (the total number of cells in the images) and the SARS-CoV-2 immunofluorescence signal leading to identification of infected cells.

Uninfected host cell cytotoxicity counter screen. Compounds are acoustically transferred into 1,536-well plates (Corning No. 9006BC). HeLa-ACE2 cells are maintained as described for the infection assay and seeded in the assay-ready plates at 400 cells/well in DMEM with 2% FBS. Plates are incubated for 24 hours at 37℃ 5% CO_2. To assess cell viability, 2 mL of 50% Cell-Titer Glo (Promega No G7573) diluted in water is added to the cells and luminescence measured on an EnVision Plate Reader (Perkin Elmer).

Data analysis. Primary in vitro screen and the host cell cytotoxicity counter screen data are uploaded to Genedata Screener, Version 16.0. Data are normalized to neutral (DMSO) minus inhibitor controls (2.5 µM remdesivir for antiviral effect and 10 µM puromycin for infected host cell toxicity). For the uninfected host cell cytotoxicity counter screen 40 µM puromycin (Sigma) is used as the positive control. For dose response experiments compounds are tested in technical triplicates on different assay plates and dose curves are fitted with the four parameter Hill Equation.

1.3.6 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(*79*). 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(*80*). 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(*32*). 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 batches ("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 His41:Cys145 catalytic dyad (variously using zwitterionic or uncharged states) and His163 in P1 (which interacts with the 3-aminopyridine or isoquinoline nitrogen in our primary lead series). As the most relevant protonation states were uncertain, when computational resources afforded, calculations were carried out using multiple protonation state variants (His41:Cys145 either neutral or zwitterionic; His163 neutral or protonated) and the most predictive model on available retrospective data for that scaffold selected for nominating prospective predictions for that batch. 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(*40*) (https://github.com/choderalab/perses), and nonequilibrium switching alchemical free energy calculations(*81*) were run on Folding@home using the OpenMM compute core(*43*). 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(*39*) (multiple generations between 1.1.1 and 1.3.1 were released and used as the project evolved) and the AMBER14SB protein force field(*82*) with recommended ion parameters(*83*, *84*), and TIP3P water(*85*). As many assayed compounds as possible were included in each batch of transformations to enable continual retrospective assessment and to leverage

existing measured affinities in extrapolating predicted affinities. Analysis of free energy calculations used the maximum likelihood estimator [https://doi.org/10.1021/acs.jcim.9b00528] to reconstruct the optimal predicted absolute free energy (and hence pIC50) estimate from available experimental measurements. Calculations were analyzed using the fah-xchem dashboard (https://github.com/choderalab/fah-xchem) using the Bennett acceptance ratio(*86*, *87*) (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.

We note that our primary aim was computing estimates of relative binding free energies for large alchemical transformations using abundant computing resources (which exceeded 1 exaFLOP/s(*38*)) rather than aggressive optimization of the cost/transformation. Batches of transformations used between 100-200 parallel 4 ns nonequilibrium cycles per transformation, selected based on the number of atoms modified in the transformation, resulting in 100-200 ns/transformation in aggregate. A Tesla V100 achieves ~200 ns/day for our solvated Mpro complex, meaning ~2-4 GPU-days/transformation was consumed on a V100 equivalent GPU. To give typical scales, **Figure 2C Panel 1** ran 6319 transformations of 140 cycles, resulting in ~3.5 ms of simulation time or ~424K GPU-hours; **Figure 2C Panel 2** ran 5077 transformations of ~200 cycles, resulting in ~4 ms simulation time, or ~480K GPU-hours; **Figure 2C Panel 3** ran 686 transformations of ~200 cycles, resulting in ~548 μs of simulation time, or ~66K GPU-hours.

Scripts for setting up and analysing the perses alchemical free energy calculations on Folding@home, as well as an index of computed datasets and dashboards are available at https://github.com/foldingathome/covid-moonshot

Code used for generating the COVID Moonshot alchemical free energy calculation web dashboards is available here: https://github.com/choderalab/fah-xchem
Retrospective calculations for transformations in the main synthetic series shown in **Figure 5A** were performed with an early release of perses 0.10.2 constructed as a simplified example that anyone can run to illustrate how these calculations work on standard GPU workstations, and use standard alchemical replica exchange protocols of 5 ns/replica (which just take a few hours on standard workstations, as opposed to the expensive nonequilibrium protocols used in the Sprints). Input scripts for this calculation are available in the perses distribution under `examples/moonshot-mainseries/` [https://github.com/choderalab/perses/tree/main/examples/moonshot-mainseries].

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(*88*), 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 July 2021, https://fragalysis.diamond.ac.uk/api/targets/?format=json&title=Mpro). The respective 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 367 complex structures, used to generate the interaction fingerprints. Note for this study, only hbond-donor, hbond-acceptor, salt bridge, hydrophobic, pi-stacking, and halogen interactions

are inspected. Additional code was added to *PLIPify* to split the hbond-donor and hbond-acceptor interactions into backbone and sidechain interactions (https://github.com/volkamerlab/plipify/pull/18). Interacting residues are only included if the summed interaction count per residue over all investigated structures is greater than five. Careful examination of examples of the interactions led us to filter out the S144 interactions from the final report as none of the interactions were convincing (24 hbond-don-bb, 168 hbond-don-sc, and 4 hbond-acc-sc interactions). 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). Finally, structures containing compounds exhibiting some of the major interactions identified were used to generate the figures in **Figure 3B**.

4. Chemical methods

4.1 HTC library synthesis

4.1.1 Chan-Lam reaction

The arylamine library was made by reacting the boronic acid (**Fig S7D**), under the optimized reaction conditions (1 eq. amine; 0.2 eq. CuI; 0.8 eq. DMAP; 2 eq. Hex_3N; DMSO; under air; RT; 2 days) with 296 amines (200 aromatic, 48 primary, and 48 secondary aliphatic amines; **Data S2**). For library production, we used Echo LDV plates and an Echo 555 acoustic dispenser for liquid handling. After the allotted reaction time, plate copies were made after diluting the reaction mixture with 4.6 µL DMSO and transferring 1 µL of the obtained solution to an 384-well plate, for either biochemical assay or yield estimation.

4.1.2 Amide coupling

The amide library was made by reacting the carboxylic acid (**Fig S8E**) under the optimized reaction conditions (2 eq. amine; 2 eq. EDC; 2 eq. HOAt; 5 eq. DIPEA; DMSO; RT; 24h) with 300 amines (202 aromatics, 49 primary, and 49 secondary aliphatic amines Data S2). For library production, we used Echo LDV plates and an Echo 555 acoustic dispenser for liquid handling. Plate copies were made after diluting the reaction mixture with 4 µL DMSO. For yield estimation, 1 µL of the diluted library was transferred to an LC/MS-ready 384-well plate, followed by dilution with 20% ACN in water to the final volume of 50 µL. The desired product was identified in 60% of wells.

4.2 General compounds synthesis and characterization

All compounds were directly purchased from Enamine Inc., following Enamine's standard QC for compound collections. In addition, in Supplementary Chemistry we discuss the synthesis procedure, as well as LCMS and 1H NMR characterisation of compounds which were discussed in the manuscript with associated bioactivity data.

All COVID Moonshot compounds are publicly available as a screening collection that can be ordered in bulk or as singleton through Enamine. The compound identifiers of the COVID Moonshot collection are in the Supplementary Data, together with Enamine's internal QC data comprising LCMS spectra for all compounds and NMR spectra for selected compounds.

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 acknowledgements and Disclosures statements for each author are listed in the consortium spreadsheet (**Data S7**). This research was funded in part, by the Wellcome Trust [Grant number 209407/Z/17/Z]. For the purpose of open access, the author has applied a CC-BY public copyright license to any Author Accepted Manuscript version arising from this submission.

Data availability

- **All compound designs, datasets, and X-ray structures can be browsed on the COVID Moonshot website**: https://postera.ai/covid ; The compound submissions and experimental data are available via GitHub https://github.com/postera-ai/COVID_moonshot_submissions and the bioactivity data can be interactively browsed https://covid.postera.ai/covid/activity_data ; All of the data is also available in a permanent archive(*89*)
- **Alchemical free energy calculations code and datasets are indexed on GitHub and stored in a permanent archive**(*90*) https://github.com/foldingathome/covid-moonshot
- **All X-ray structures are available for interactive viewing and comparison or bulk download via Fragalysis:** https://fragalysis.diamond.ac.uk/viewer/react/preview/target/Mpro ; Structures were deposited to the PDB (**Data S4**) and are also available in a permanent archive.(*91*)

Resource availability

- **Synthesized compounds**: We have made all compounds assayed here available from the current Enamine catalog, 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

Supplementary Materials

Materials and Methods
Supplementary Chemistry
Figs. S1 to S18
Data S1-S7
References 93-94

References

1. N. Phillips, The coronavirus is here to stay - here's what that means. *Nature*. **590**, 382–384 (2021).

2. N. Winchester, Covid-19 vaccinations: is the Global South falling behind? *UK House of Lords library* (2021), (available at https://lordslibrary.parliament.uk/covid-19-vaccinations-is-the-global-south-falling-behind/).

3. T. Zhao, C. Hu, M. Ayaz Ahmed, C. Cheng, Y. Chen, C. Sun, Warnings regarding the potential coronavirus disease 2019 (COVID-19) transmission risk: Vaccination is not enough. *Infect. Control Hosp. Epidemiol.*, 1 (2021).

4. J. Hammond, H. Leister-Tebbe, A. Gardner, P. Abreu, W. Bao, W. Wisemandle, M. Baniecki, V. M. Hendrick, B. Damle, A. Simón-Campos, R. Pypstra, J. M. Rusnak, EPIC-HR Investigators, Oral Nirmatrelvir for High-Risk, Nonhospitalized Adults with Covid-19. *N. Engl. J. Med.* **386**, 1397–1408 (2022).

5. M. W. McCarthy, Ensitrelvir as a potential treatment for COVID-19. *Expert Opin. Pharmacother.* **23**, 1995–1998 (2022).

6. A. Jayk Bernal, M. M. Gomes da Silva, D. B. Musungaie, E. Kovalchuk, A. Gonzalez, V. Delos Reyes, A. Martín-Quirós, Y. Caraco, A. Williams-Diaz, M. L. Brown, J. Du, A. Pedley, C. Assaid, J. Strizki, J. A. Grobler, H. H. Shamsuddin, R. Tipping, H. Wan, A. Paschke, J. R. Butterton, M. G. Johnson, C. De Anda, MOVe-OUT Study Group, Molnupiravir for Oral Treatment of Covid-19 in Nonhospitalized Patients. *N. Engl. J. Med.* **386**, 509–520 (2022).

7. E. de Wit, N. van Doremalen, D. Falzarano, V. J. Munster, SARS and MERS: recent insights into emerging coronaviruses. *Nat. Rev. Microbiol.* **14**, 523–534 (2016).

8. S. Ullrich, C. Nitsche, The SARS-CoV-2 main protease as drug target. *Bioorg. Med. Chem. Lett.* **30**, 127377 (2020).

9. T. Pillaiyar, M. Manickam, V. Namasivayam, Y. Hayashi, S.-H. Jung, 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).

10. L. Zhang, D. Lin, X. Sun, U. Curth, C. Drosten, L. Sauerhering, S. Becker, K. Rox, R. Hilgenfeld, Crystal structure of SARS-CoV-2 main protease provides a basis for design of improved α-ketoamide inhibitors. *Science*. **368**, 409–412 (2020).

11. Z. Jin, X. Du, Y. Xu, Y. Deng, M. Liu, Y. Zhao, B. Zhang, X. Li, L. Zhang, C. Peng, Y. Duan, J. Yu, L. Wang, K. Yang, F. Liu, R. Jiang, X. Yang, T. You, X. Liu, X. Yang, F. Bai, H. Liu, X. Liu, L. W. Guddat, W. Xu, G. Xiao, C. Qin, Z. Shi, H. Jiang, Z. Rao, H. Yang, Structure of Mpro from SARS-CoV-2 and discovery of its inhibitors. *Nature*. **582**, 289–293 (2020).

12. D. R. Owen, C. M. N. Allerton, A. S. Anderson, L. Aschenbrenner, M. Avery, S. Berritt, B. Boras, R. D. Cardin, A. Carlo, K. Coffman, A. Dantonio, L. Di, H. Eng, R. Ferre, K. S. Gajiwala, S. A. Gibson, S. E. Greasley, B. L. Hurst, E. P. Kadar, A. S. Kalgutkar, J. C. Lee, J. Lee, W. Liu, S. W. Mason, S. Noell, J. J. Novak, R. S. Obach, K. Ogilvie, N. C. Patel, M. Pettersson, D. K. Rai, M. R. Reese, M. F. Sammons, J. G. Sathish, R. S. P. Singh, C. M. Steppan, A. E. Stewart, J. B. Tuttle, L. Updyke, P. R. Verhoest, L. Wei, Q. Yang, Y. Zhu, An oral SARS-CoV-2 mpro inhibitor clinical candidate for the treatment of COVID-19. *bioRxiv* (2021), , doi:10.1101/2021.07.28.21261232.

13. H. Mukae, H. Yotsuyanagi, N. Ohmagari, Y. Doi, T. Imamura, T. Sonoyama, T. Fukuhara, G. Ichihashi, T. Sanaki, K. Baba, Y. Takeda, Y. Tsuge, T. Uehara, A Randomized Phase 2/3 Study of Ensitrelvir, a Novel Oral SARS-CoV-2 3C-Like Protease Inhibitor, in Japanese Patients with Mild-to-Moderate COVID-19 or Asymptomatic SARS-CoV-2 Infection: Results of the Phase 2a Part. *Antimicrob. Agents Chemother.* **66**, e0069722 (2022).

14. H. Mukae, H. Yotsuyanagi, N. Ohmagari, Y. Doi, H. Sakaguchi, T. Sonoyama, G. Ichihashi, T. Sanaki, K. Baba, Y. Tsuge, T. Uehara, Efficacy and safety of ensitrelvir in patients with mild-to-moderate

COVID-19: the phase 2b part of a randomized, placebo-controlled, phase 2/3 study. *Clin. Infect. Dis.* (2022), doi:10.1093/cid/ciac933.

15. Study of pf-07321332 in healthy participants, (available at https://clinicaltrials.gov/ct2/show/NCT04756531).

16. Study of PBI-0451 in Healthy Subjects, (available at https://clinicaltrials.gov/ct2/show/NCT05011812).

17. J. Heskin, S. J. C. Pallett, N. Mughal, G. W. Davies, L. S. P. Moore, M. Rayment, R. Jones, Caution required with use of ritonavir-boosted PF-07321332 in COVID-19 management. *Lancet*. **399**, 21–22 (2022).

18. F. von Delft, M. Calmiano, J. Chodera, E. Griffen, A. Lee, N. London, T. Matviuk, B. Perry, M. Robinson, A. von Delft, A white-knuckle ride of open COVID drug discovery. *Nature*. **594**, 330–332 (2021).

19. J. Chodera, A. A. Lee, N. London, F. von Delft, Crowdsourcing drug discovery for pandemics. *Nat. Chem.* **12**, 581 (2020).

20. A. Douangamath, D. Fearon, P. Gehrtz, T. Krojer, P. Lukacik, C. D. Owen, E. Resnick, C. Strain-Damerell, A. Aimon, P. Ábrányi-Balogh, J. Brandão-Neto, A. Carbery, G. Davison, A. Dias, T. D. Downes, L. Dunnett, M. Fairhead, J. D. Firth, S. P. Jones, A. Keeley, G. M. Keserü, H. F. Klein, M. P. Martin, M. E. M. Noble, P. O'Brien, A. Powell, R. N. Reddi, R. Skyner, M. Snee, M. J. Waring, C. Wild, N. London, F. von Delft, M. A. Walsh, Crystallographic and electrophilic fragment screening of the SARS-CoV-2 main protease. *Nat. Commun.* **11**, 5047 (2020).

21. Main protease structure and XChem fragment screen. *Diamond Light Source* (2020), (available at https://www.diamond.ac.uk/covid-19/for-scientists/Main-protease-structure-and-XChem.html).

22. XChem @ Diamond, (available at https://fragalysis.diamond.ac.uk/viewer/react/preview/target/Mpro).

23. D. C. Rees, M. Congreve, C. W. Murray, R. Carr, Fragment-based lead discovery. *Nat. Rev. Drug Discov.* **3**, 660–672 (2004).

24. D. A. Erlanson, S. W. Fesik, R. E. Hubbard, W. Jahnke, H. Jhoti, Twenty years on: the impact of fragments on drug discovery. *Nat. Rev. Drug Discov.* **15**, 605–619 (2016).

25. S. Cooper, F. Khatib, A. Treuille, J. Barbero, J. Lee, M. Beenen, A. Leaver-Fay, D. Baker, Z. Popović, F. Players, Predicting protein structures with a multiplayer online game. *Nature*. **466**, 756–760 (2010).

26. J. Lee, W. Kladwang, M. Lee, D. Cantu, M. Azizyan, H. Kim, A. Limpaecher, S. Yoon, A. Treuille, R. Das, EteRNA Participants, RNA design rules from a massive open laboratory. *Proc. Natl. Acad. Sci. U. S. A.* **111**, 2122–2127 (2014).

27. O. B. Cox, T. Krojer, P. Collins, O. Monteiro, R. Talon, A. Bradley, O. Fedorov, J. Amin, B. D. Marsden, J. Spencer, F. von Delft, P. E. Brennan, A poised fragment library enables rapid synthetic expansion yielding the first reported inhibitors of PHIP(2), an atypical bromodomain. *Chem. Sci.* **7**, 2322–2330 (2016).

28. A. Carbery, R. Skyner, F. von Delft, C. M. Deane, Fragment Libraries Designed to Be Functionally Diverse Recover Protein Binding Information More Efficiently Than Standard Structurally Diverse Libraries. *J. Med. Chem.* **65**, 11404–11413 (2022).

29. N. Kitamura, M. D. Sacco, C. Ma, Y. Hu, J. A. Townsend, X. Meng, F. Zhang, X. Zhang, M. Ba, T. Szeto, A. Kukuljac, M. T. Marty, D. Schultz, S. Cherry, Y. Xiang, Y. Chen, J. Wang, Expedited Approach toward the Rational Design of Noncovalent SARS-CoV-2 Main Protease Inhibitors. *J. Med. Chem.* (2021), doi:10.1021/acs.jmedchem.1c00509.

30. D. Zaidman, P. Gehrtz, M. Filep, D. Fearon, R. Gabizon, A. Douangamath, J. Prilusky, S. Duberstein, G. Cohen, C. D. Owen, E. Resnick, C. Strain-Damerell, P. Lukacik, Covid-Moonshot Consortium, H. Barr, M. A. Walsh, F. von Delft, N. London, An automatic pipeline for the design of irreversible

derivatives identifies a potent SARS-CoV-2 Mpro inhibitor. *Cell Chem Biol*. **28**, 1795–1806.e5 (2021).

31. S. H. Han, C. M. Goins, T. Arya, W.-J. Shin, J. Maw, A. Hooper, D. P. Sonawane, M. R. Porter, B. E. Bannister, R. D. Crouch, A. A. Lindsey, G. Lakatos, S. R. Martinez, J. Alvarado, W. S. Akers, N. S. Wang, J. U. Jung, J. D. Macdonald, S. R. Stauffer, 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.

32. P. Schwaller, T. Laino, T. Gaudin, P. Bolgar, C. A. Hunter, C. Bekas, A. A. Lee, Molecular Transformer: A Model for Uncertainty-Calibrated Chemical Reaction Prediction. *ACS Cent Sci*. **5**, 1572–1583 (2019).

33. A. Morris, W. McCorkindale, T. C. M. Consortium, N. Drayman, J. D. Chodera, S. Tay, N. London, A. A. Lee, Discovery of SARS-CoV-2 main protease inhibitors using a synthesis-directed de novo design model. *Chem. Commun.* . **57**, 5909–5912 (2021).

34. A. S. J. Mey, B. Allen, H. E. Bruce Macdonald, J. D. Chodera, M. Kuhn, J. Michel, D. L. Mobley, L. N. Naden, S. Prasad, A. Rizzi, J. Scheen, M. R. Shirts, G. Tresadern, H. Xu, Best Practices for Alchemical Free Energy Calculations. *arXiv [q-bio.BM]* (2020), (available at http://arxiv.org/abs/2008.03067).

35. R. Abel, L. Wang, D. L. Mobley, R. A. Friesner, 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).

36. K. A. Armacost, S. Riniker, Z. Cournia, Exploring Novel Directions in Free Energy Calculations. *J. Chem. Inf. Model.* **60**, 5283–5286 (2020).

37. M. Shirts, V. S. Pande, COMPUTING: Screen Savers of the World Unite! *Science*. **290**, 1903–1904 (2000).

38. M. I. Zimmerman, J. R. Porter, M. D. Ward, S. Singh, N. Vithani, A. Meller, U. L. Mallimadugula, C. E. Kuhn, J. H. Borowsky, R. P. Wiewiora, M. F. D. Hurley, A. M. Harbison, C. A. Fogarty, J. E. Coffland, E. Fadda, V. A. Voelz, J. D. Chodera, G. R. Bowman, SARS-CoV-2 simulations go exascale to predict dramatic spike opening and cryptic pockets across the proteome. *Nat. Chem.* **13**, 651–659 (2021).

39. Y. Qiu, D. Smith, S. Boothroyd, H. Jang, J. Wagner, C. C. Bannan, T. Gokey, V. T. Lim, C. Stern, A. Rizzi, X. Lucas, B. Tjanaka, M. R. Shirts, M. Gilson, J. Chodera, C. I. Bayly, D. Mobley, L.-P. Wang, Development and Benchmarking of Open Force Field v1.0.0, the Parsley Small Molecule Force Field. *ChemRxiv* (2020), doi:10.26434/chemrxiv.13082561.v2.

40. D. A. Rufa, H. E. Bruce Macdonald, J. Fass, M. Wieder, P. B. Grinaway, A. E. Roitberg, O. Isayev, J. D. Chodera, Towards chemical accuracy for alchemical free energy calculations with hybrid physics-based machine learning / molecular mechanics potentials. *bioRxiv* (2020), p. 2020.07.29.227959.

41. D. A. Rufa, I. Zhang, H. E. Bruce Macdonald, P. B. Grinaway, I. Pulido, M. M. Henry, J. Rodríguez-Guerra, M. Wittmann, S. K. Albanese, W. G. Glass, A. Silveira, D. Schaller, L. N. Naden, J. D. Chodera, *Perses* (2023; https://zenodo.org/record/8092391).

42. I. Zhang, D. A. Rufa, I. Pulido, M. M. Henry, L. E. Rosen, K. Hauser, S. Singh, J. D. Chodera, Identifying and Overcoming the Sampling Challenges in Relative Binding Free Energy Calculations of a Model Protein:Protein Complex. *J. Chem. Theory Comput.* **19**, 4863–4882 (2023).

43. P. Eastman, J. Swails, J. D. Chodera, R. T. McGibbon, Y. Zhao, K. A. Beauchamp, L.-P. Wang, A. C. Simmonett, M. P. Harrigan, C. D. Stern, R. P. Wiewiora, B. R. Brooks, V. S. Pande, OpenMM 7: Rapid development of high performance algorithms for molecular dynamics. *PLoS Comput. Biol.* **13**, e1005659 (2017).

44. A. B. Santanilla, E. L. Regalado, T. Pereira, M. Shevlin, K. Bateman, L.-C. Campeau, J. Schneeweis, S. Berritt, Z.-C. Shi, P. Nantermet, Y. Liu, R. Helmy, C. J. Welch, P. Vachal, I. W. Davies, T. Cernak, S. D. Dreher, Nanomole-scale high-throughput chemistry for the synthesis of complex molecules. *Science*. **347**, 49–53 (2015).

45. D. Perera, J. W. Tucker, S. Brahmbhatt, C. J. Helal, A. Chong, W. Farrell, P. Richardson, N. W. Sach, A platform for automated nanomole-scale reaction screening and micromole-scale synthesis in flow. *Science*. **359**, 429–434 (2018).

46. N. J. Gesmundo, B. Sauvagnat, P. J. Curran, M. P. Richards, C. L. Andrews, P. J. Dandliker, T. Cernak, Nanoscale synthesis and affinity ranking. *Nature*. **557**, 228–232 (2018).

47. R. P. Thomas, R. E. Heap, F. Zappacosta, E. K. Grant, P. Pogány, S. Besley, D. J. Fallon, M. M. Hann, D. House, N. C. O. Tomkinson, J. T. Bush, A direct-to-biology high-throughput chemistry approach to reactive fragment screening. *Chem. Sci.* **12**, 12098–12106 (2021).

48. P. Gehrtz, S. Marom, M. Bührmann, J. Hardick, S. Kleinbölting, A. Shraga, C. Dubiella, R. Gabizon, J. N. Wiese, M. P. Müller, G. Cohen, I. Babaev, K. Shurrush, L. Avram, E. Resnick, H. Barr, D. Rauh, N. London, Optimization of Covalent MKK7 Inhibitors via Crude Nanomole-Scale Libraries. *J. Med. Chem.* **65**, 10341–10356 (2022).

49. J.-Q. Chen, J.-H. Li, Z.-B. Dong, A Review on the Latest Progress of Chan-Lam Coupling Reaction. *Adv. Synth. Catal.* **362**, 3311–3331 (2020).

50. E. Resnick, A. Bradley, J. Gan, A. Douangamath, T. Krojer, R. Sethi, P. P. Geurink, A. Aimon, G. Amitai, D. Bellini, J. Bennett, M. Fairhead, O. Fedorov, R. Gabizon, J. Gan, J. Guo, A. Plotnikov, N. Reznik, G. F. Ruda, L. Díaz-Sáez, V. M. Straub, T. Szommer, S. Velupillai, D. Zaidman, Y. Zhang, A. R. Coker, C. G. Dowson, H. M. Barr, C. Wang, K. V. M. Huber, P. E. Brennan, H. Ovaa, F. von Delft, N. London, Rapid Covalent-Probe Discovery by Electrophile-Fragment Screening. *J. Am. Chem. Soc.* **141**, 8951–8968 (2019).

51. M. N. L. Nalam, A. Ali, M. D. Altman, G. S. K. K. Reddy, S. Chellappan, V. Kairys, A. Ozen, H. Cao, M. K. Gilson, B. Tidor, T. M. Rana, C. A. Schiffer, 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).

52. J. Jansen, K. C. Reimer, J. S. Nagai, F. S. Varghese, G. J. Overheul, M. de Beer, R. Roverts, D. Daviran, L. A. S. Fermin, B. Willemsen, M. Beukenboom, S. Djudjaj, S. von Stillfried, L. E. van Eijk, M. Mastik, M. Bulthuis, W. den Dunnen, H. van Goor, J.-L. Hillebrands, S. H. Triana, T. Alexandrov, M.-C. Timm, B. T. van den Berge, M. van den Broek, Q. Nlandu, J. Heijnert, E. M. J. Bindels, R. M. Hoogenboezem, F. Mooren, C. Kuppe, P. Miesen, K. Grünberg, T. Ijzermans, E. J. Steenbergen, J. Czogalla, M. F. Schreuder, N. Sommerdijk, A. Akiva, P. Boor, V. G. Puelles, J. Floege, T. B. Huber, COVID Moonshot consortium, R. P. van Rij, I. G. Costa, R. K. Schneider, B. Smeets, R. Kramann, 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.

53. Y. Unoh, S. Uehara, K. Nakahara, H. Nobori, Y. Yamatsu, S. Yamamoto, Y. Maruyama, Y. Taoda, K. Kasamatsu, T. Suto, K. Kouki, A. Nakahashi, S. Kawashima, T. Sanaki, S. Toba, K. Uemura, T. Mizutare, S. Ando, M. Sasaki, Y. Orba, H. Sawa, A. Sato, T. Sato, T. Kato, Y. Tachibana, Discovery of S-217622, a Non-Covalent Oral SARS-CoV-2 3CL Protease Inhibitor Clinical Candidate for Treating COVID-19. *bioRxiv* (2022), p. 2022.01.26.477782.

54. 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)* (2021), (available at https://dndi.org/press-releases/2021/covid-moonshot-funded-by-wellcome-to-rapidly-develop-safe-globally-accessible-affordable-antiviral-pill/).

55. J. Strovel, S. Sittampalam, N. P. Coussens, M. Hughes, J. Inglese, A. Kurtz, A. Andalibi, L. Patton, C. Austin, M. Baltezor, M. Beckloff, M. Weingarten, S. Weir, *Early Drug Discovery and Development Guidelines: For Academic Researchers, Collaborators, and Start-up Companies* (Eli Lilly & Company and the National Center for Advancing Translational Sciences, 2016).

56. E. S. Lander, L. M. Linton, B. Birren, C. Nusbaum, M. C. Zody, J. Baldwin, K. Devon, K. Dewar, M. Doyle, W. FitzHugh, R. Funke, D. Gage, K. Harris, A. Heaford, J. Howland, L. Kann, J. Lehoczky, R. LeVine, P. McEwan, K. McKernan, J. Meldrim, J. P. Mesirov, C. Miranda, W. Morris, J. Naylor, C. Raymond, M. Rosetti, R. Santos, A. Sheridan, C. Sougnez, Y. Stange-Thomann, N. Stojanovic, A.

Subramanian, D. Wyman, J. Rogers, J. Sulston, R. Ainscough, S. Beck, D. Bentley, J. Burton, C. Clee, N. Carter, A. Coulson, R. Deadman, P. Deloukas, A. Dunham, I. Dunham, R. Durbin, L. French, D. Grafham, S. Gregory, T. Hubbard, S. Humphray, A. Hunt, M. Jones, C. Lloyd, A. McMurray, L. Matthews, S. Mercer, S. Milne, J. C. Mullikin, A. Mungall, R. Plumb, M. Ross, R. Shownkeen, S. Sims, R. H. Waterston, R. K. Wilson, L. W. Hillier, J. D. McPherson, M. A. Marra, E. R. Mardis, L. A. Fulton, A. T. Chinwalla, K. H. Pepin, W. R. Gish, S. L. Chissoe, M. C. Wendl, K. D. Delehaunty, T. L. Miner, A. Delehaunty, J. B. Kramer, L. L. Cook, R. S. Fulton, D. L. Johnson, P. J. Minx, S. W. Clifton, T. Hawkins, E. Branscomb, P. Predki, P. Richardson, S. Wenning, T. Slezak, N. Doggett, J. F. Cheng, A. Olsen, S. Lucas, C. Elkin, E. Uberbacher, M. Frazier, R. A. Gibbs, D. M. Muzny, S. E. Scherer, J. B. Bouck, E. J. Sodergren, K. C. Worley, C. M. Rives, J. H. Gorrell, M. L. Metzker, S. L. Naylor, R. S. Kucherlapati, D. L. Nelson, G. M. Weinstock, Y. Sakaki, A. Fujiyama, M. Hattori, T. Yada, A. Toyoda, T. Itoh, C. Kawagoe, H. Watanabe, Y. Totoki, T. Taylor, J. Weissenbach, R. Heilig, W. Saurin, F. Artiguenave, P. Brottier, T. Bruls, E. Pelletier, C. Robert, P. Wincker, D. R. Smith, L. Doucette-Stamm, M. Rubenfield, K. Weinstock, H. M. Lee, J. Dubois, A. Rosenthal, M. Platzer, G. Nyakatura, S. Taudien, A. Rump, H. Yang, J. Yu, J. Wang, G. Huang, J. Gu, L. Hood, L. Rowen, A. Madan, S. Qin, R. W. Davis, N. A. Federspiel, A. P. Abola, M. J. Proctor, R. M. Myers, J. Schmutz, M. Dickson, J. Grimwood, D. R. Cox, M. V. Olson, R. Kaul, C. Raymond, N. Shimizu, K. Kawasaki, S. Minoshima, G. A. Evans, M. Athanasiou, R. Schultz, B. A. Roe, F. Chen, H. Pan, J. Ramser, H. Lehrach, R. Reinhardt, W. R. McCombie, M. de la Bastide, N. Dedhia, H. Blöcker, K. Hornischer, G. Nordsiek, R. Agarwala, L. Aravind, J. A. Bailey, A. Bateman, S. Batzoglou, E. Birney, P. Bork, D. G. Brown, C. B. Burge, L. Cerutti, H. C. Chen, D. Church, M. Clamp, R. R. Copley, T. Doerks, S. R. Eddy, E. E. Eichler, T. S. Furey, J. Galagan, J. G. Gilbert, C. Harmon, Y. Hayashizaki, D. Haussler, H. Hermjakob, K. Hokamp, W. Jang, L. S. Johnson, T. A. Jones, S. Kasif, A. Kaspryzk, S. Kennedy, W. J. Kent, P. Kitts, E. V. Koonin, I. Korf, D. Kulp, D. Lancet, T. M. Lowe, A. McLysaght, T. Mikkelsen, J. V. Moran, N. Mulder, V. J. Pollara, C. P. Ponting, G. Schuler, J. Schultz, G. Slater, A. F. Smit, E. Stupka, J. Szustakowki, D. Thierry-Mieg, J. Thierry-Mieg, L. Wagner, J. Wallis, R. Wheeler, A. Williams, Y. I. Wolf, K. H. Wolfe, S. P. Yang, R. F. Yeh, F. Collins, M. S. Guyer, J. Peterson, A. Felsenfeld, K. A. Wetterstrand, A. Patrinos, M. J. Morgan, P. de Jong, J. J. Catanese, K. Osoegawa, H. Shizuya, S. Choi, Y. J. Chen, J. Szustakowki, International Human Genome Sequencing Consortium, Initial sequencing and analysis of the human genome. *Nature*. **409**, 860–921 (2001).

57. A. R. Williamson, Creating a structural genomics consortium. *Nat. Struct. Biol.* **7 Suppl**, 953 (2000).

58. U.S. Cancer Institute “Megaproject” to Target Common Cancer-Driving Protein. *Science*, (available at https://www.science.org/content/article/us-cancer-institute-megaproject-target-common-cancer-driving-protein).

59. A. Lee, J. Chodera, F. von Delft, Why we are developing a patent-free Covid antiviral therapy. *Knowable mag.* (2021), doi:10.1146/knowable-092721-1.

60. S. Legare, F. Heide, B. A. Bailey-Elkin, J. Stetefeld, Improved SARS-CoV-2 main protease high-throughput screening assay using a 5-carboxyfluorescein substrate. *J. Biol. Chem.* **298**, 101739 (2022).

61. M. G. Acker, D. S. Auld, Considerations for the design and reporting of enzyme assays in high-throughput screening applications. *Perspectives in Science*. **1**, 56–73 (2014).

62. T. R. Malla, A. Tumber, T. John, L. Brewitz, C. Strain-Damerell, C. D. Owen, P. Lukacik, H. T. H. Chan, P. Maheswaran, E. Salah, F. Duarte, H. Yang, Z. Rao, M. A. Walsh, C. J. Schofield, Mass spectrometry reveals potential of β-lactams as SARS-CoV-2 Mpro inhibitors. *Chem. Commun.* . **57**, 1430–1433 (2021).

63. G. D. Noske, A. M. Nakamura, V. O. Gawriljuk, R. S. Fernandes, G. M. A. Lima, H. V. D. Rosa, H. D. Pereira, A. C. M. Zeri, A. F. Z. Nascimento, M. C. L. C. Freire, D. Fearon, A. Douangamath, F. von Delft, G. Oliva, A. S. Godoy, A Crystallographic Snapshot of SARS-CoV-2 Main Protease Maturation Process. *J. Mol. Biol.* **433**, 167118 (2021).

64. G. Winter, D. G. Waterman, J. M. Parkhurst, A. S. Brewster, R. J. Gildea, M. Gerstel, L. Fuentes-Montero, M. Vollmar, T. Michels-Clark, I. D. Young, N. K. Sauter, G. Evans, DIALS: implementation and evaluation of a new integration package. *Acta Crystallogr D Struct Biol*. **74**, 85–97 (2018).

65. G. Winter, R. J. Gildea, N. G. Paterson, J. Beale, M. Gerstel, D. Axford, M. Vollmar, K. E. McAuley, R. L. Owen, R. Flaig, A. W. Ashton, D. R. Hall, How best to use photons. *Acta Crystallogr D Struct Biol*. **75**, 242–261 (2019).

66. G. Winter, K. E. McAuley, Automated data collection for macromolecular crystallography. *Methods*. **55**, 81–93 (2011).

67. W. Kabsch, Integration, scaling, space-group assignment and post-refinement. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 133–144 (2010).

68. G. Winter, C. M. C. Lobley, S. M. Prince, Decision making in xia2. *Acta Crystallogr. D Biol. Crystallogr.* **69**, 1260–1273 (2013).

69. C. Vonrhein, C. Flensburg, P. Keller, A. Sharff, O. Smart, W. Paciorek, T. Womack, G. Bricogne, Data processing and analysis with the autoPROC toolbox. *Acta Crystallogr. D Biol. Crystallogr.* **67**, 293–302 (2011).

70. T. Krojer, R. Talon, N. Pearce, P. Collins, A. Douangamath, J. Brandao-Neto, A. Dias, B. Marsden, F. von Delft, The XChemExplorer graphical workflow tool for routine or large-scale protein-ligand structure determination. *Acta Crystallogr D Struct Biol*. **73**, 267–278 (2017).

71. N. M. Pearce, T. Krojer, A. R. Bradley, P. Collins, R. P. Nowak, R. Talon, B. D. Marsden, S. Kelm, J. Shi, C. M. Deane, F. von Delft, A multi-crystal method for extracting obscured crystallographic states from conventionally uninterpretable electron density. *Nat. Commun.* **8**, 15123 (2017).

72. P. Emsley, B. Lohkamp, W. G. Scott, K. Cowtan, Features and development of Coot. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 486–501 (2010).

73. F. Long, R. A. Nicholls, P. Emsley, S. Graæulis, A. Merkys, A. Vaitkus, G. N. Murshudov, AceDRG: a stereochemical description generator for ligands. *Acta Crystallogr D Struct Biol*. **73**, 112–122 (2017).

74. T. Gorrie-Stone, *xchem/XChemReview: Jan_2021* (Zenodo, 2022; https://zenodo.org/record/5939112).

75. I. J. Bruno, J. C. Cole, M. Kessler, J. Luo, W. D. S. Motherwell, L. H. Purkis, B. R. Smith, R. Taylor, R. I. Cooper, S. E. Harris, A. G. Orpen, Retrieval of crystallographically-derived molecular geometry information. *J. Chem. Inf. Comput. Sci.* **44**, 2133–2144 (2004).

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

77. V. M. Corman, O. Landt, M. Kaiser, R. Molenkamp, A. Meijer, D. K. Chu, T. Bleicker, S. Brünink, J. Schneider, M. L. Schmidt, D. G. Mulders, B. L. Haagmans, B. van der Veer, S. van den Brink, L. Wijsman, G. Goderski, J.-L. Romette, J. Ellis, M. Zambon, M. Peiris, H. Goossens, C. Reusken, M. P. Koopmans, C. Drosten, Detection of 2019 novel coronavirus (2019-nCoV) by real-time RT-PCR. *Euro Surveill.* **25** (2020), doi:10.2807/1560-7917.ES.2020.25.3.2000045.

78. F. S. Varghese, E. van Woudenbergh, G. J. Overheul, M. J. Eleveld, L. Kurver, N. van Heerbeek, A. van Laarhoven, P. Miesen, G. den Hartog, M. I. de Jonge, R. P. van Rij, Berberine and Obatoclax Inhibit SARS-Cov-2 Replication in Primary Human Nasal Epithelial Cells In Vitro. *Viruses*. **13** (2021), doi:10.3390/v13020282.

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

80. R. Rosales, B. L. McGovern, M. Luis Rodriguez, D. K. Rai, R. D. Cardin, A. S. Anderson, PSP study group, E. M. Sordillo, H. van Bakel, V. Simon, A. García-Sastre, K. M. White, Nirmatrelvir, Molnupiravir, and Remdesivir maintain potent in vitro activity against the SARS-CoV-2 Omicron variant. *bioRxiv* (2022), p. 2022.01.17.476685.

81. V. Gapsys, L. Pérez-Benito, M. Aldeghi, D. Seeliger, H. van Vlijmen, G. Tresadern, B. L. de Groot,

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

82. J. A. Maier, C. Martinez, K. Kasavajhala, L. Wickstrom, K. E. Hauser, C. Simmerling, ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *J. Chem. Theory Comput.* **11**, 3696–3713 (2015).

83. I. S. Joung, T. E. Cheatham 3rd, Determination of alkali and halide monovalent ion parameters for use in explicitly solvated biomolecular simulations. *J. Phys. Chem. B*. **112**, 9020–9041 (2008).

84. I. S. Joung, T. E. Cheatham 3rd, Molecular dynamics simulations of the dynamic and energetic properties of alkali and halide ions using water-model-specific ion parameters. *J. Phys. Chem. B*. **113**, 13279–13290 (2009).

85. W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, M. L. Klein, Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **79**, 926–935 (1983).

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

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

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

89. M. Robinson, M. Henry, A. Morris, COVID_moonshot_submissions. *Zenodo* (2023), , doi:10.5281/zenodo.8196348.

90. J. Scheen, Covid Moonshot Free Energy Calculations Data. *zenodo* (2023), , doi:10.5281/zenodo.8215553.

91. D. Fearon, A. Aimon, J. C. Aschenbrenner, B. H. Balcomb, I. A. Barker, F. K. R. Bertram, J. Brandao-Neto, A. Dias, A. Douangamath, L. Dunnett, A. S. Godoy, T. J. Gorrie-Stone, L. Koekemoer, T. Krojer, R. M. Lithgo, P. Lukacik, P. G. Marples, H. Mikolajek, E. Nelson, K. H. V. Nidamarthi, C. D. Owen, A. J. Powell, V. L. Rangel, R. Skyner, C. M. Strain-Damerell, W. Thompson, C. W. E. Tomlinson, C. Wild, M. A. Walsh, F. von Delft, *Crystal structures of SARS-CoV-2 main protease screened against COVID Moonshot compounds by X-ray Crystallography at the XChem facility of Diamond Light Source* (2023; https://zenodo.org/record/8214528).

92. D. W. Kneller, G. Phillips, H. M. O'Neill, R. Jedrzejczak, L. Stols, P. Langan, A. Joachimiak, L. Coates, A. Kovalevsky, Structural plasticity of SARS-CoV-2 3CL Mpro active site cavity revealed by room temperature X-ray crystallography. *Nat. Commun.* **11**, 3202 (2020).

93. J. Jiménez-Luna, F. Grisoni, G. Schneider, Drug discovery with explainable artificial intelligence. Nature Machine Intelligence. 2, 573–584 (2020).

94. A. Özen, W. Sherman, C. A. Schiffer, Improving the Resistance Profile of Hepatitis C NS3/4A Inhibitors: Dynamic Substrate Envelope Guided Design. J. Chem. Theory Comput. 9, 5693–5705 (2013).

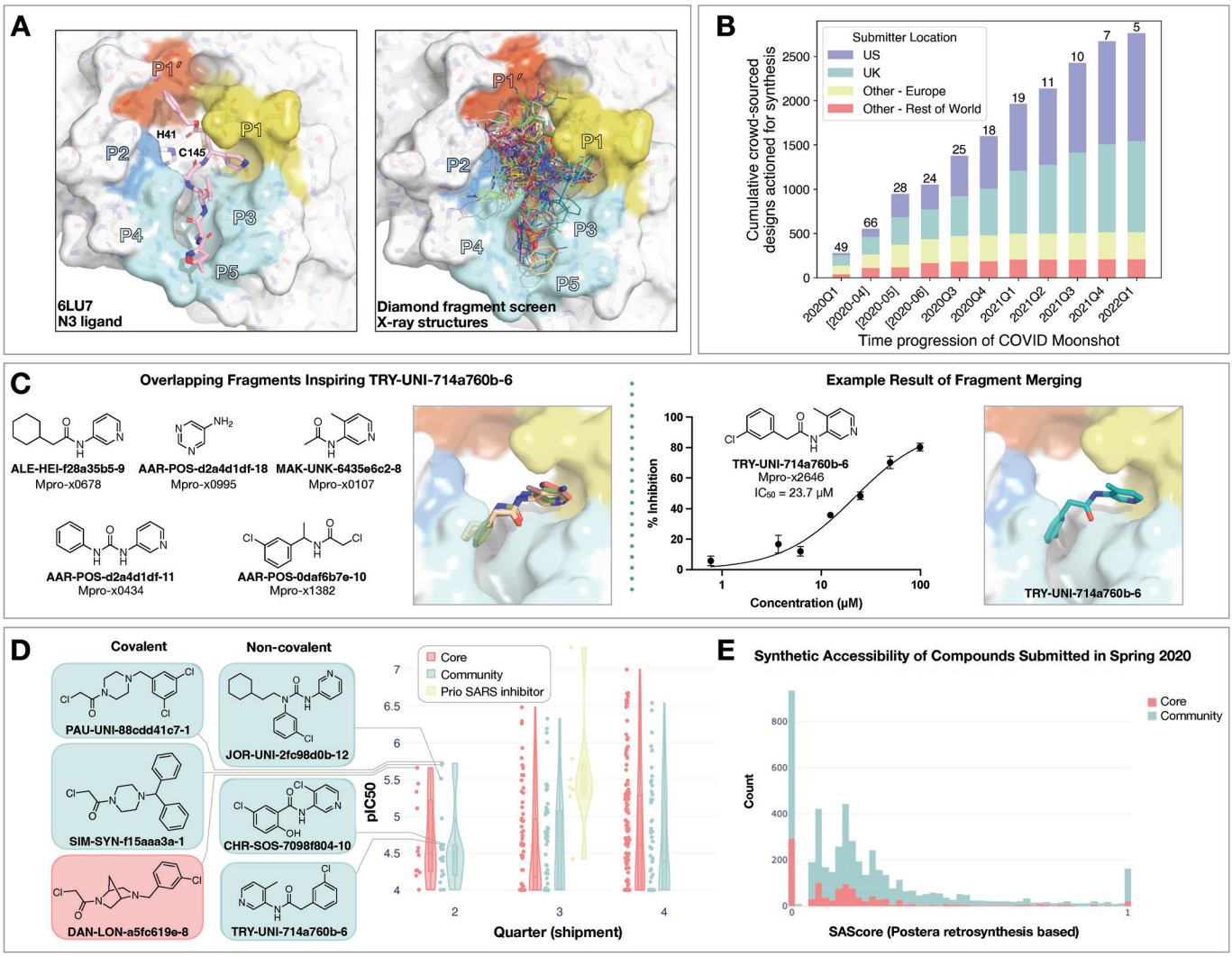


Figure 1: Crowdsourcing rapidly identified novel chemotype scaffolds by merging fragment hits.

A: A 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(*20*).

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 nucleophilic Cys145 reacts with the scissile peptide bond between P1 and P1'; His41-Cys145 form a catalytic dyad whose coupled charge states shuttle between zwitterionic and neutral states (*92*).

B: On March 18th 2020, the COVID Moonshot set up a crowdsourcing website to empower scientists across the globe to contribute molecule designs. The number of designs actioned for synthesis each quarter (except for the 2020Q2 which is shown per-month in brackets) is shown, subdivided by the region of the submitter of the design idea. The total number of unique submitters that contributed actioned designs for that quarter is shown on top of the bars.

C: Many submissions, such as TRY-UNI-714a760b-6, exploited spatially overlapping fragment hits to design potent leads that are synthetically facile.

D: Experimental biochemical potency of designs broken down by submission group. Multiple submissions in 2020 from the community were more potent than the best designs from the core team as seen for the top three chloroacetamide structures (left) and non-covalent structures (right).

E: Distribution of synthetic accessibility scores for designs contributed by the core team and the community. The concern that community submissions may be of poor quality is not supported by the fact that these were as synthetic accessible as those designed by the core team (median: community 0.17 core 0.13,). Half of the outliers (SAScore = 1) were primarily natural products, which are hard to achieve via organic chemistry.

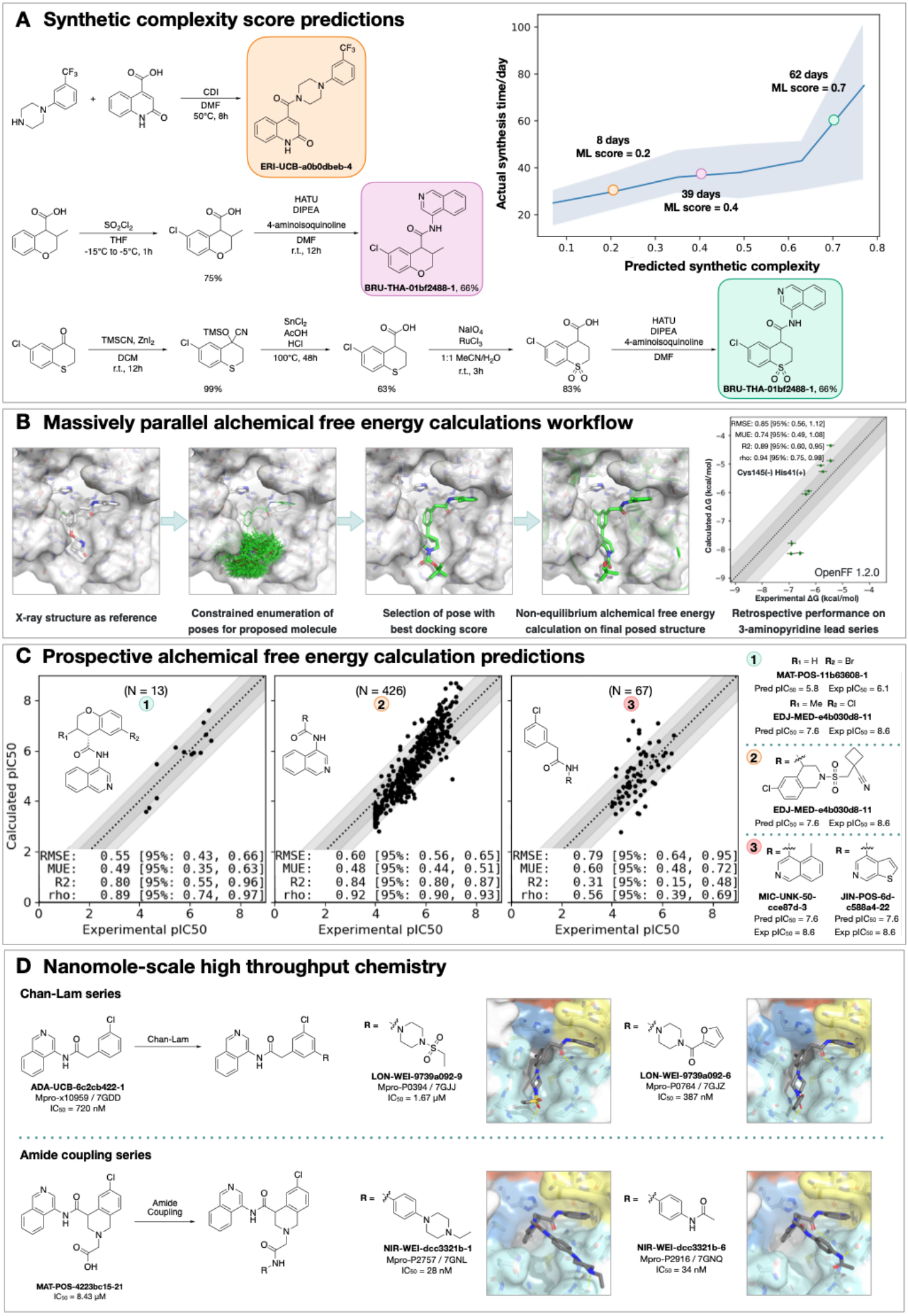

A Synthetic complexity score predictions
CDI
DMF
50°C, 8h
ERI-UCB-a0b0dbeb-4
SO2Cl2
THF
-15°C to -5°C, 1h
75%
HATU
DIPEA
4-aminoisoquinoline
DMF
r.t., 12h
BRU-THA-01bf2488-1, 66%
TMSCN, ZnI2
DCM
r.t., 12h
99%
SnCl2
AcOH
HCl
100°C, 48h
63%
NaIO4
RuCl3
1:1 MeCN/H2O
r.t., 3h
83%
HATU
DIPEA
4-aminoisoquinoline
DMF
BRU-THA-01bf2488-1, 66%
Actual synthesis time/day
Predicted synthetic complexity
8 days
ML score = 0.2
39 days
ML score = 0.4
62 days
ML score = 0.7
B Massively parallel alchemical free energy calculations workflow
X-ray structure as reference
Constrained enumeration of poses for proposed molecule
Selection of pose with best docking score
Non-equilibrium alchemical free energy calculation on final posed structure
Retrospective performance on 3-aminopyridine lead series
RMSE: 0.85 [95%: 0.56, 1.12]
MUE: 0.74 [95%: 0.49, 1.08]
R2: 0.89 [95%: 0.60, 0.95]
rho: 0.94 [95%: 0.75, 0.98]
Cys145(-) His41(+)
OpenFF 1.2.0
Calculated ΔG (kcal/mol)
Experimental ΔG (kcal/mol)
C Prospective alchemical free energy calculation predictions
(N = 13)
(N = 426)
(N = 67)
Calculated pIC50
Experimental pIC50
RMSE: 0.55 [95%: 0.43, 0.66]
MUE: 0.49 [95%: 0.35, 0.63]
R2: 0.80 [95%: 0.55, 0.96]
rho: 0.89 [95%: 0.74, 0.97]
RMSE: 0.60 [95%: 0.56, 0.65]
MUE: 0.48 [95%: 0.44, 0.51]
R2: 0.84 [95%: 0.80, 0.87]
rho: 0.92 [95%: 0.90, 0.93]
RMSE: 0.79 [95%: 0.64, 0.95]
MUE: 0.60 [95%: 0.48, 0.72]
R2: 0.31 [95%: 0.15, 0.48]
rho: 0.56 [95%: 0.39, 0.69]
R1 = H R2 = Br
MAT-POS-11b63608-1
Pred pIC50 = 5.8 Exp pIC50 = 6.1
R1 = Me R2 = Cl
EDJ-MED-e4b030d8-11
Pred pIC50 = 7.6 Exp pIC50 = 8.6
EDJ-MED-e4b030d8-11
Pred pIC50 = 7.6 Exp pIC50 = 8.6
MIC-UNK-50-cce87d-3
Pred pIC50 = 7.6
Exp pIC50 = 8.6
JIN-POS-6d-c588a4-22
Pred pIC50 = 7.6
Exp pIC50 = 8.6
D Nanomole-scale high throughput chemistry
Chan-Lam series
Chan-Lam
ADA-UCB-6c2cb422-1
Mpro-x10959 / 7GDD
IC50 = 720 nM
LON-WEI-9739a092-9
Mpro-P0394 / 7GJJ
IC50 = 1.67 μM
LON-WEI-9739a092-6
Mpro-P0764 / 7GJZ
IC50 = 387 nM
Amide coupling series
Amide Coupling
MAT-POS-4223bc15-21
IC50 = 8.43 μM
NIR-WEI-dcc3321b-1
Mpro-P2757 / 7GNL
IC50 = 28 nM
NIR-WEI-dcc3321b-6
Mpro-P2916 / 7GNQ
IC50 = 34 nM

Figure 2: Strategies to support rapid optimization cycles.

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 2020-05-01 to 2021-07-01 (n=898). 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. Retrospective assessment of our automated free energy calculation workflow on early compounds in the 3-aminopyridine series in the first month of the COVID Moonshot campaign suggested free energy calculations could provide good predictive utility, which inspired confidence for large scale deployment during this campaign. Here, the absolute free energy of binding ΔG is shown in the rightmost panel by adding a constant offset to the computed relative free energy differences.

C: Alchemical free energy predictions for all submissions elaborating on the depicted scaffold for three representative batches of prospective free energy calculations plotted as calculated (converted using Cheng-Prusoff equation) versus experimental pIC_{50}. Simulations were run using Mpro in dimer form, with neutral catalytic dyad and no restraints. Each batch (numbered 1-3 from left to right) is annotated with its scaffold and top-scoring candidates are shown on the right-hand side (numbered 1-3 from top to bottom)---for these the structure names are shown together with their predicted and experimental pIC_{50} ("Pred"/"P" and "Expt"/"E", resp.). Statistical performance with 95% confidence intervals for each batch is shown as a table in each scatterplot.

D: Two examples of nanomole scale high throughput chemistry (HTC) campaigns used to optimize the potency of intermediate binders, centering on the Chan Lam reaction (**Fig S7**) and amide couplings (**Fig S8**). Direct biochemical screening of crude reactions identified candidates that were resynthesized and in both cases were able to improve the potency of the parent compound. Soaking of crude reaction mixtures of the most potent biochemical hits into Mpro crystals provided complex structures with the identified hits (Chan-Lam - PDBs: 7GJJ/7GJZ resolution: 1.75Å/1.65Å; Amide coupling PDBs: 7GNL/7GNQ resolution: 1.68Å/1.53Å). In both cases new interactions were discovered, explaining the improved activity. It is interesting to note, that while for the Chan Lam reaction campaign, the extended compounds indeed occupied the intended P4, for the amide-coupling vector, all compounds extended into the P3/5 pockets.

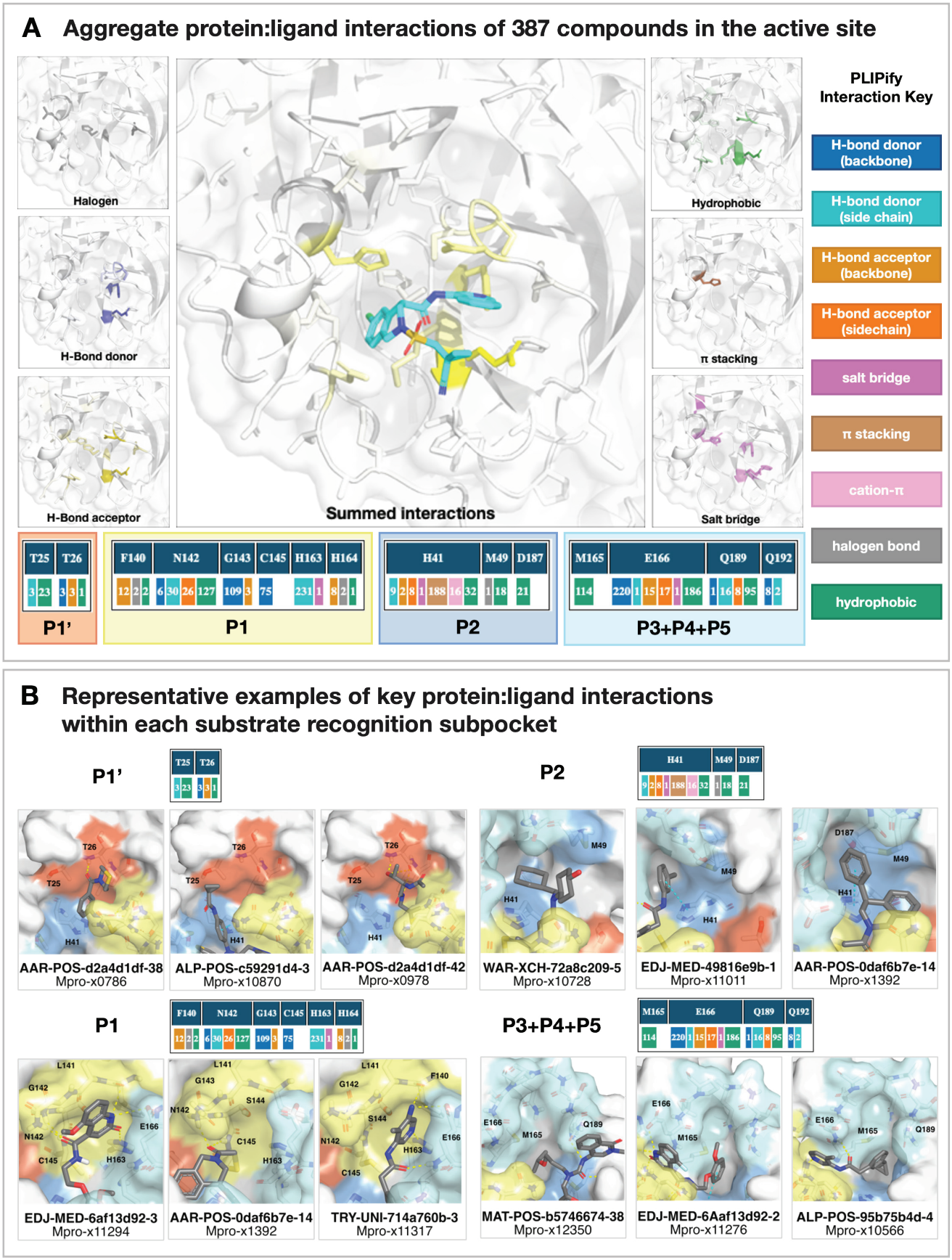


Figure 3: Analysis of 367 complex crystal structures reveals hotspots for ligand engagement, and a variety of ways to engage substrate recognition subpockets.

A: The five substrate recognition subpockets exhibit distinct preferences for intermolecular interactions. The figure highlights the locations of different types of interactions, with the shading indicating the frequency. The bottom row tallies the number of times each interaction was seen in our structures for different residues. The interaction map was generated using PLIPify (see Detailed Methods) and summarizes the interactions witnessed across 367 complexes from

the perspective of the protein, distinguishing between backbone (bb) and sidechain (sc) interactions (which might be more vulnerable to point mutations).

B: Representative examples of protein-ligand interactions engaging the P1', P1, P2, and P3-5 subpockets. Hydrogen bonds and π-stacking interactions are depicted as yellow and cyan dashed lines, respectively. The rows above each set of complexes tally the number of times each interaction was seen with the specific residues within the subpockets. See **Data S4** for PDB IDs and crystallography statistics.

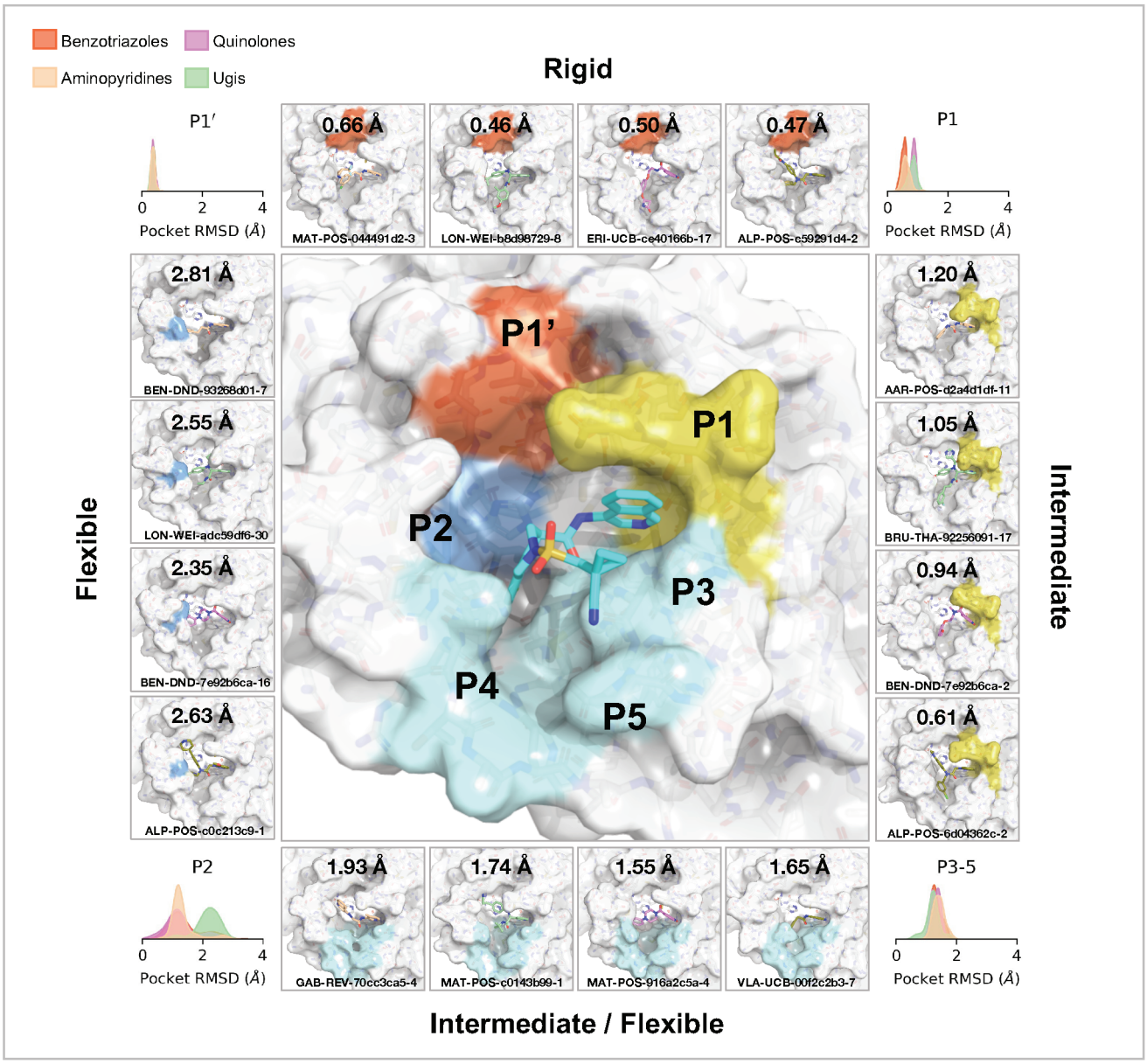
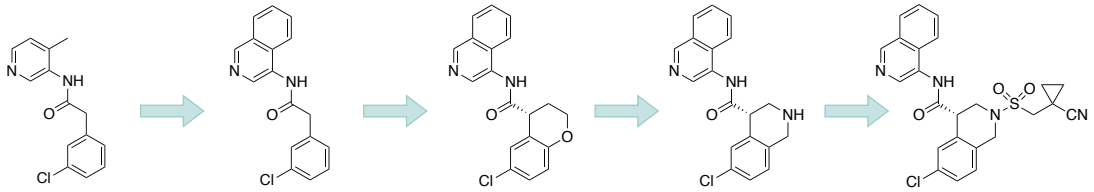


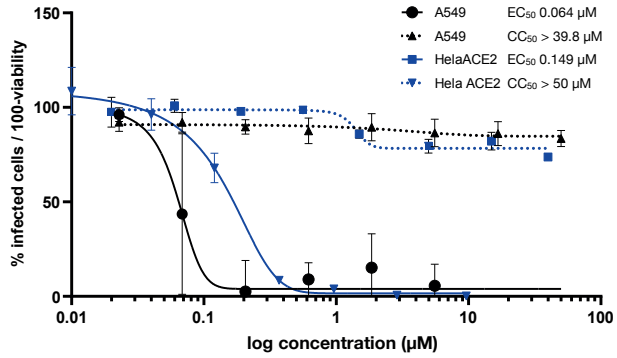
Figure 4: Structural plasticity of the binding subpockets. The subpockets have different degrees of plasticity, which is also dependent on the chemical series (**Fig S2**). The corners of the figure shows the distribution of side chain root-mean-square-deviation (RMSD) deviations from the structure of **MAT-POS-e194df51-1** (middle panel; PDB: 7GAW). The boxes exemplify ligands that significantly deform the pockets.

A

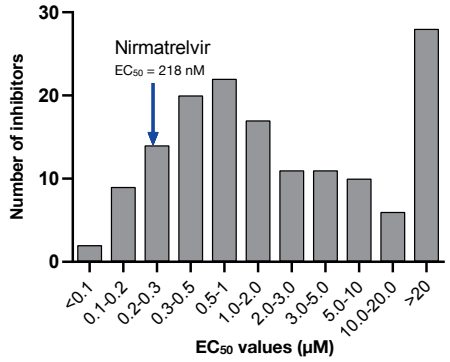


PostEra ID	TRY-UNI-714a760b-6	ADA-UCB-6c2cb422-1	MAT-POS-b3e365b9-1	MAT-POS-3ccb8ef6-1	MAT-POS-e194df51-1
Fragalysis/PDB ID	x2646 / 5RH2	x10959 / 7S3S	x11612 / 7GFB	P0744 / 7JGX	x0862 / 7GAW
Enamine Cat No.	Z1129289650	Z1530724813	Z4643752419	Z4943052515	Z5129808241
IC_{50} (Mpro)	23.7 [19.5, 28.9] μM	0.721 [0.647, 0.804] μM	0.255 [0.240, 0.270] μM	0.288 [0.273, 0.304] μM	0.0368 [0.0343, 0.0395] μM
EC_{50} (SARS-CoV-2, A549)	n.d.	4.5 μM	7.0 μM	1.9 μM	0.064 μM
$\Delta\Delta G_{exp}$		-2.07 [-1.89,-2.25] kcal/mol	-0.615 [-0.517, -0.715] kcal/mol	0.007 [0.006,0.139] kcal/mol	-1.22 [-1.14, -1.29] kcal/mol
$\Delta\Delta G_{FEP}$		-1.8±0.1 kcal/mol	-0.7±0.2 kcal/mol	0.0±0.1 kcal/mol	-2.4±0.1 kcal/mol

B



C



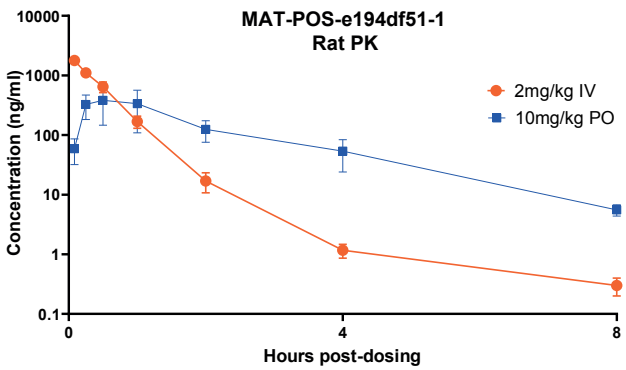
D

Compound	A549 (+ p-gp inhibitor) CPE			HelaAce2 (+ p-gp inhibitor) CPE			Calu-3 (no p-gp inhibitor) CPE			VeroE6 (no p-gp inhibitor) CPE	
	EC_{50} (μM)	EC_{90} (μM)	CC_{50} (μM)	EC_{50} (μM)	EC_{90} (μM)	CC_{50} (μM)	EC_{50} (μM)	EC_{90} (μM)	CC_{50} (μM)	EC_{50} (μM)	EC_{90} (μM)
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	n.d.	n.d.
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	n.d.	n.d.	n.d.	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	n.d.	n.d.	n.d.	> 10	n.d.	> 100	3.24	6.41
TRY-UNI-714a760b-6	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.

E

Virus	MAT-POS-e194df51-1		Nirmatrelvir	
	EC_{50} (μM)	CC_{50} (μM)	EC_{50} (μM)	CC_{50} (μM)
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 IC_{50} (μM)	0.037
A549 EC_{50} (μM)	0.064

In Vitro ADME				
LogD (measured)	2.5			
MDCK-LE $P_{app\ A\to B}$ (10^{-6} cm/s)	21.8			
	Rat	Dog	Mining	Human
Liver microsomes CL (μL/min/mg)	604	164	542	152
Liver microsomes $t_{1/2}$ (min)	2.4	8.5	2.6	9.1
Hepatocyte CL (μL/min/10^{-6} cells)	67.6	61.4	65.9	10.3
Hepatocyte $t_{1/2}$ (min)	10.3	11.3	10.5	67.5
PPB free fraction (%)	5.4	n.d.	n.d.	10.1

Safety / Drug-Drug Interactions	
Cyp450 (μIM) 2C9 /2D6 / 3A4	25 / 9.4 / 10.3
PXR risk	Low
Herg (μM)	> 30

In Vivo Pharmacokinetics	
Rat IV V_d (L/kg)	1.05
Rat IV CL (mL/min/kg)	34.8
Rat $t_{1/2}$ IV/PO (h)	0.448 / 1.4
Rat Bioavailability (%)	18

Figure 5: Iterative medicinal chemistry furnished an orally bioavailable inhibitor.

A: Summary of key medicinal chemistry milestones in developing the initial crowdsourced lead compound into a potent antiviral. X-ray structures for each milestone compound are available via Fragalysis, and each compound can be obtained from Enamine via the corresponding catalogue numbers. Retrospective alchemical free energy calculation

predictions for each transformation ($\Delta\Delta G_{FEP}$) are shown for each step between milestones, along with the corresponding experimental free energy difference ($\Delta\Delta G_{exp}$) derived from biochemical activities. As positive control, under our assay condition Nirmatrelvir has IC_{50} of 2.6nM.

B: Antiviral activity of **MAT-POS-e194df51-1** cellular antiviral assays, with an EC_{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 5A) 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 **MAT-POS-e194df51-1** in rats with a 2 mg/kg intravenous and 10 mg/kg oral dosing with good oral availability.

G: ADME characteristics of **MAT-POS-e194df51-1** demonstrate a favorable safety profile, indicating translational potential of the lead series.