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

The COVID Moonshot Consortium*

## Abstract

The COVID-19 pandemic was a stark reminder that a barren global antiviral pipeline has grave humanitarian consequences. Pandemics could be prevented in principle by accessible, easily deployable broad-spectrum oral antivirals. Here 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 novel chemical series that is differentiated from current Mpro inhibitors in that it maintains a new non-covalent, non-peptidic scaffold with nanomolar potency. Our approach leveraged crowdsourcing, high-throughput structural biology, machine learning, and exascale molecular simulations and high-throughput chemistry. In the process, 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. In a first for a structure-based drug discovery campaign, 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 COVID-19 culminating in more than 669 million confirmed cases worldwide by Jan 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*). Significant progress has been made over the last two years, with several directly acting oral antivirals now approved for emergency usage against SARS-CoV-2 infection, including nirmatrelvir(*4*), ensitrelvir (Japan)(*5*), and molnupiravir(*6*).

COVID-19 is not an isolated event, but the latest exemplar of a series of significant threats to human health caused by beta-coronaviruses also responsible for the SARS (2003) and MERS (2010) pandemics(*7*). Open knowledge bases and technology infrastructures for antiviral drug discovery will enable pandemic preparedness by kindling the currently barren global antivirals pipeline and providing multiple starting points for the development of therapeutics. Here, we report the open science discovery of a novel potent inhibitor of SARS-CoV-2 Mpro 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 is corroborated by recent in vitro and in vivo studies for SARS-CoV-2(*10*, *11*) and the recent 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 novel non-covalent non-peptidomimetic scaffolds. While first-generation oral Mpro inhibitors have now demonstrated clinical efficacy(*15*, *16*), the need for CYP3A4 inhibitor co-dosing (ritonavir, in the case of Paxlovid) to achieve sufficient human exposure may substantially

*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)

limit use in at-risk populations due to drug-drug interactions(*17*). There remains a significant 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 a unique opportunity for rapid lead generation by combining chemotypes from multiple fragments. Nonetheless, this requires heuristic chemical reasoning that accounts for the spatial orientation of fragments in the binding site -a feat that can challenge algorithms but is potentially also solvable by humans. Building on successes in crowdsourced protein(*25*) and 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 noncovalent 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 successfully soaked and were consequently picked-up for merging by the designers. Apart from the aminopyridine series, our campaign identified three major additional chemically distinct lead series with measurable potencies against SARS-CoV-2 Mpro inspired by reported SARS-CoV-1 inhibitors (**Fig S2**). Those compounds span the same binding pocket but feature different chemotypes, and the significant quantity of SAR subsequently generated for these series furnishes multiple backup series with different risk profiles. Other groups have subsequently further elaborated on the Ugi(*29*, *30*) and the benzotriazole series we generated(*31*).

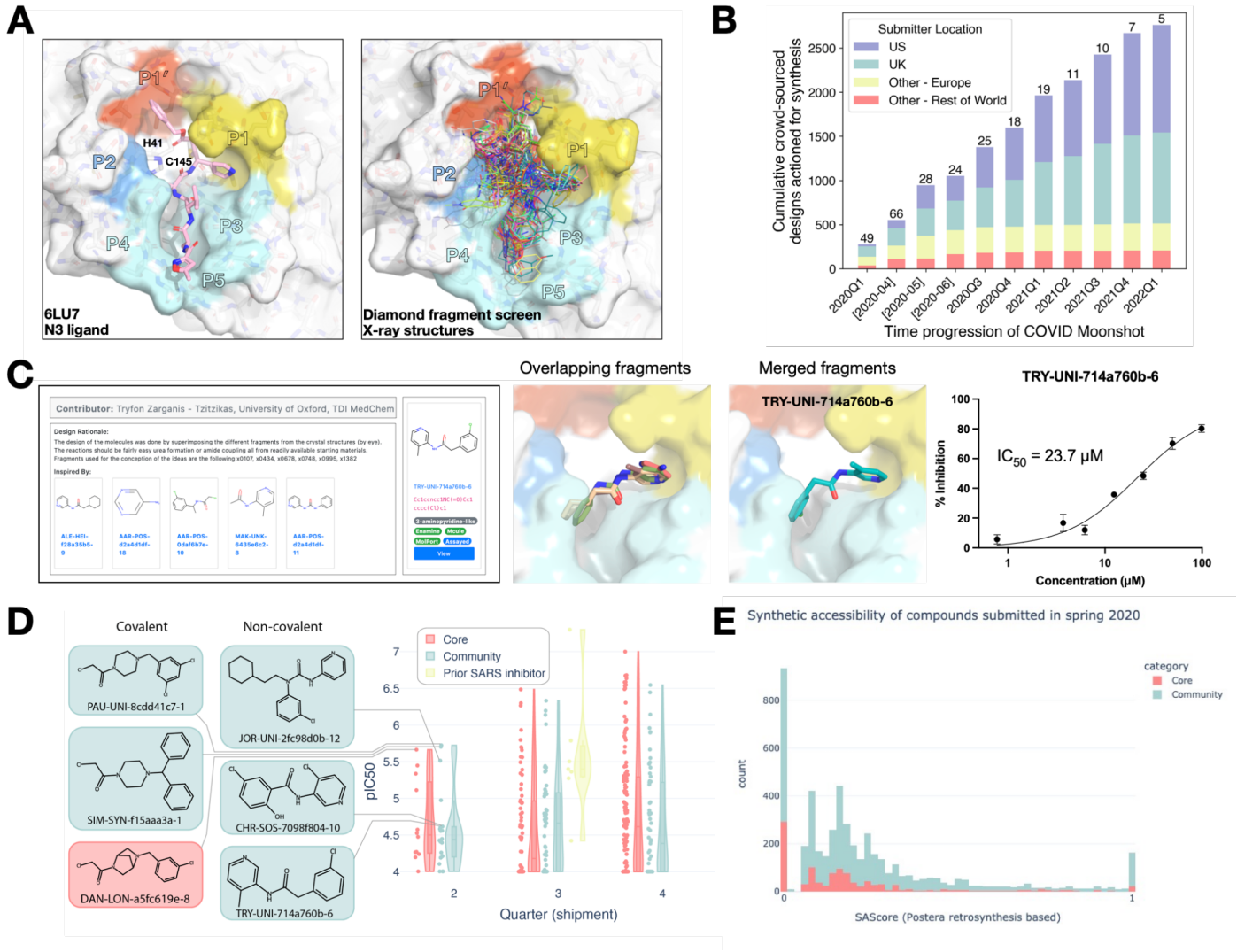



**Figure 1: Crowdsourcing rapidly identified novel chemotype scaffolds by merging fragment hits.**
**A:** A Diamond / XChem fragment screen that initiated this SARS-CoV-2 Mpro inhibitor discovery campaign generated 58 hits that completely cover the Mpro active site, with a variety of chemotypes engaging 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 (*32*).
**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 (see **Figure 3**).
**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-covalevent 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.

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 categorise 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 hypothesise 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 significant 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 the 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 4A**) 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**). While anecdotal, this clearly proves 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 aiming to progress multiple lead series, 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.

*Synthetic route predictions guided decision making to accelerate design-make-test-analyze (DTMA) 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 (over 10 million), we used a machine learning approach that plans efficient retrosynthetic routes, to predict synthetic tractability(*33*, *34*). 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 (**Figure 2A**). **Figure 2A** (right) demonstrates how the algorithm was able to pick out advanced intermediates as starting materials.

*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(*35*–*37*), 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(*38*) - 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(*39*) - 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(*40*) and nonequilibrium switching with the open source perses alchemical free energy toolkit(*41*) based on the GPU-accelerated OpenMM framework(*42*)(*39*) (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 4A**. 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 additional 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 significant synthetic effort would be required. Chemically-related groups of outliers frequently provided new chemical insight that informed modeling choices (**Fig S5**). The approach was not without drawbacks, including the ability 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**). Further, the method is not uniformly accurate across all chemical transformations, and its accuracy is challenging to *a priori* estimate. For example, isoquinoline replacements show lower correlation between calculated and predicted free energy (Panel C of **Figure 2**) compared to the benzopyran replacements (Panel B of **Figure 2**).

*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(*43*, *44*) (HTC), coupled with a 'direct to biology'(*45–47*) biochemical screening. Two examples include the optimization of the Chan-Lam reaction(*48*) to extend molecule **ADA-UCB-6c2cb422-1** and amide coupling to extend **MAT-POS-4223bc15-21** (**Figure 2D**). In both cases, the co-crystal structures of the parent compounds were available (**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 2E**). Upon resynthesis, one of the Chan-Lam compounds was able to improve the parent compound $IC_{50}$ by almost two-fold. 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.

*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(*49*) 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 choloacetamides (**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 away towards acrylamides by derivatizing potent reversible hits (*30*) (**Fig S11**). Ultimately, we focused on a non-covalent series altogether, but still, the chlorophenyl moiety that remained throughout the series (**Figure 4A**) was adopted from a chloroacetamide hit fragment (**AAR-POS-0daf6b7e-10; Figure 1C**).

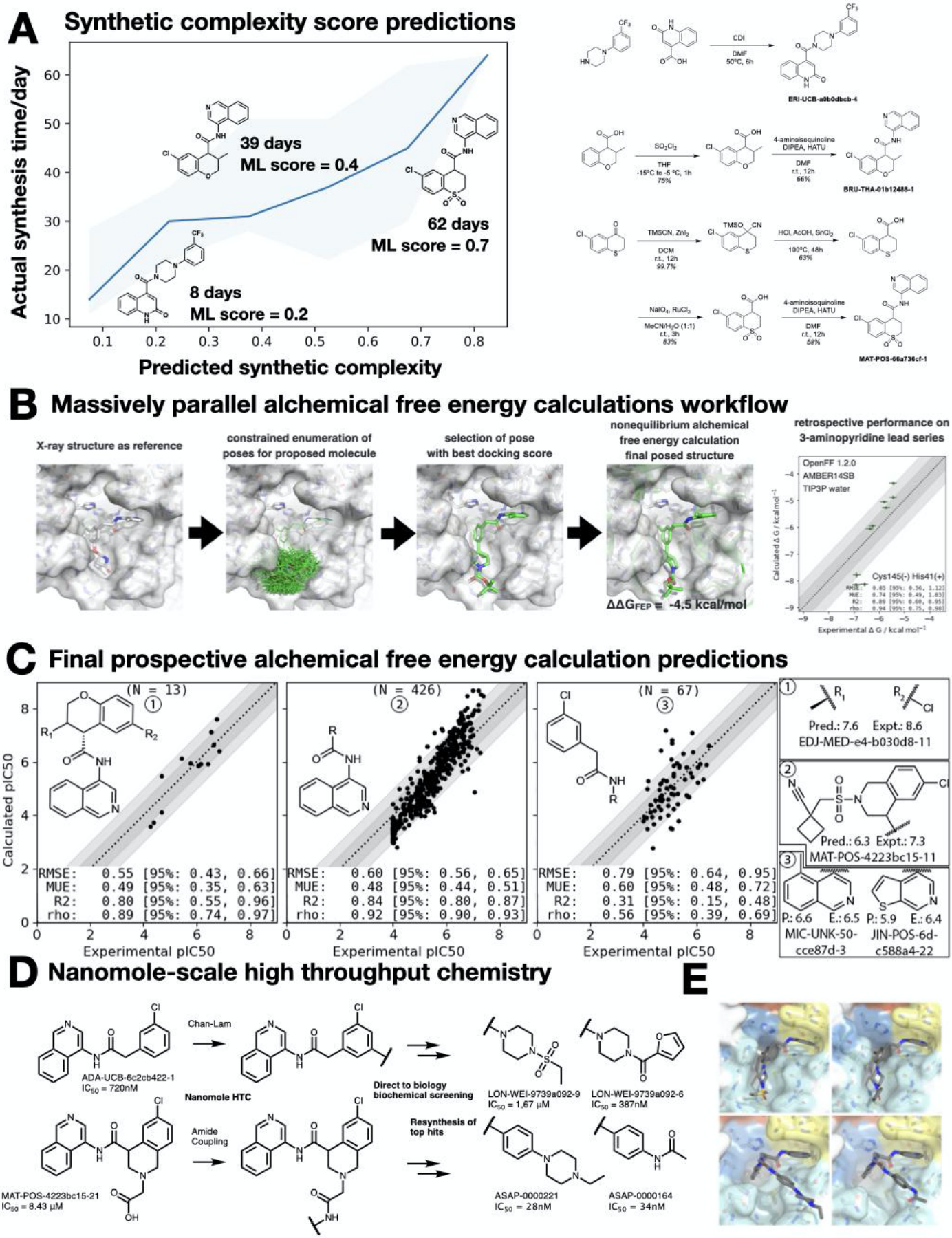

A
Synthetic complexity score predictions
Actual synthesis time/day
Predicted synthetic complexity
39 days
ML score = 0.4
62 days
ML score = 0.7
8 days
ML score = 0.2
ERI-UCB-a0b0dbcb-4
BRU-THA-01b12488-1
MAT-POS-66a736cf-1
B
Massively parallel alchemical free energy calculations workflow
X-ray structure as reference
constrained enumeration of poses for proposed molecule
selection of pose with best docking score
nonequilibrium alchemical free energy calculation final posed structure
ΔΔG_FEP = -4.5 kcal/mol
retrospective performance on 3-aminopyridine lead series
OpenFF 1.2.0
AMBER14SB
TIP3P water
Cys145(-) His41(+)
Calculated Δ G / kcal mol⁻¹
Experimental Δ G / kcal mol⁻¹
C
Final 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]
Pred.: 7.6 Expt.: 8.6
EDJ-MED-e4-b030d8-11
Pred.: 6.3 Expt.: 7.3
MAT-POS-4223bc15-11
P.: 6.6 E.: 6.5
MIC-UNK-50-cce87d-3
P.: 5.9 E.: 6.4
JIN-POS-6d-c588a4-22
D
Nanomole-scale high throughput chemistry
Chan-Lam
ADA-UCB-6c2cb422-1
IC50 = 720nM
Nanomole HTC
Direct to biology biochemical screening
LON-WEI-9739a092-9
IC50 = 1,67 μM
LON-WEI-9739a092-6
IC50 = 387nM
Amide Coupling
MAT-POS-4223bc15-21
IC50 = 8.43 μM
Resynthesis of top hits
ASAP-0000221
IC50 = 28nM
ASAP-0000164
IC50 = 34nM
E

**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:** Examples of two nanomole scale high throughput chemistry (HTC) campaigns to optimize the potency of intermediate binders. In one case we used the Chan-Lam reaction (**Fig S7**) and in another amide coupling (**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.
**E:** Soaking of crude reaction mixtures into Mpro crystals provided complex structures with the identified hits. In both cases new interactions were discovered, explaining the improved activity. It is interesting to note, that while for the Chan-Lam reaction, the extended compounds indeed occupied the intended P4, for the amide-coupling vector, all compounds extended into the P3/5 pockets.

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

Compounds selected based on synthetic tractability and alchemical free energy calculations were synthesized and tested. We profiled every compound through crystal soaking and X-ray diffraction totaling in 842 structures (see **Data Availability**). 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 3A** 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 significantly more rigid than P2 (**Figure 3B**). Interestingly, this degree of rigidity is also dependent on the chemical series (**Supp Figure 2**), with the Ugi and benzotriazole series being able to significantly deform the P2 pocket. Those series comprise more heavy atoms and span a larger region of the binding site, thus changes in P2 pocket interactions could be better tolerated.

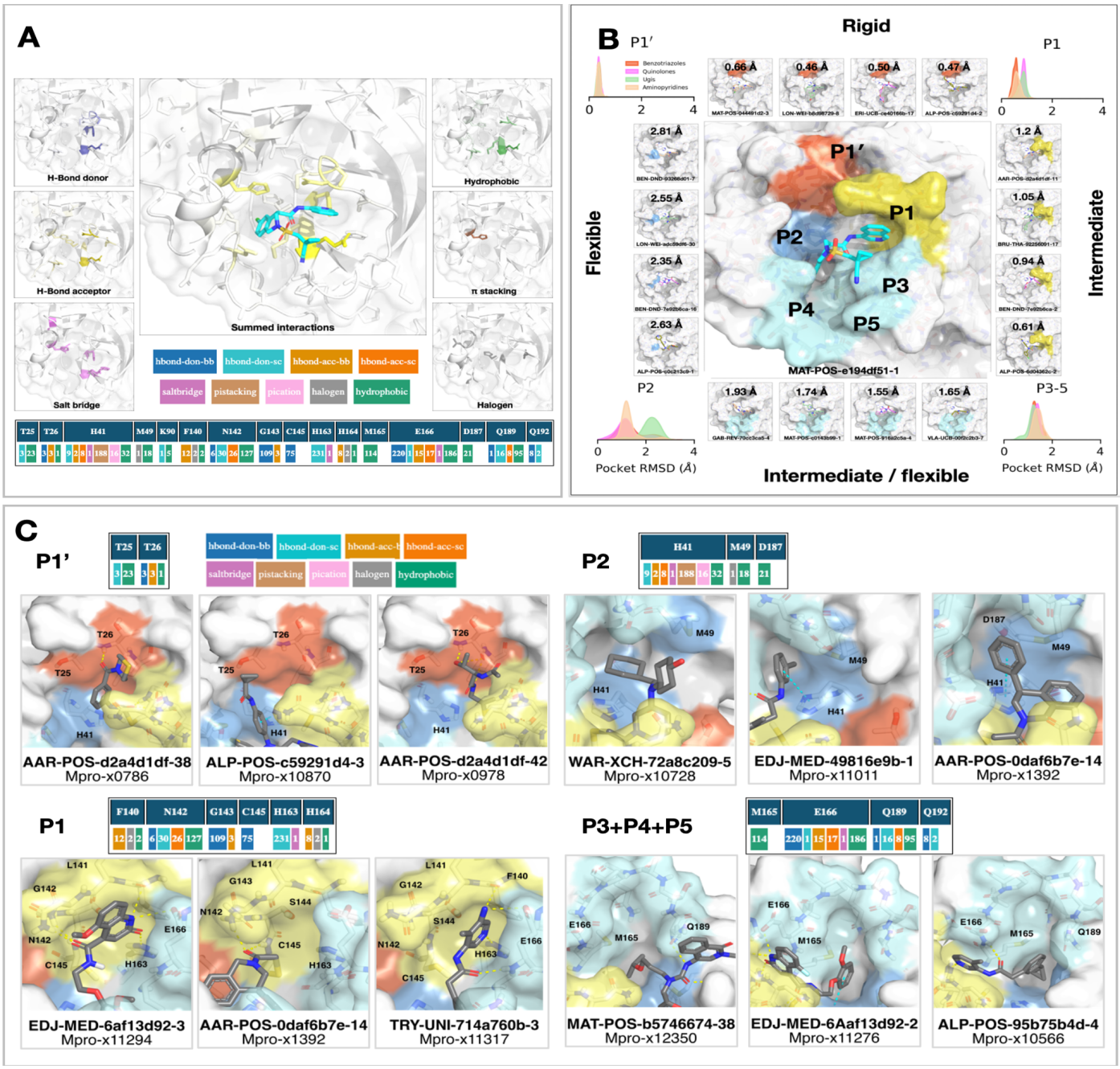



**Figure 3: Analysis of 367 complex crystal structures reveals hotspots for ligand engagement, structural plasticity of the binding sites, and a variety of ways to engage substrate recognition subpockets.**

**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:** 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 RMSD deviations from the structure of **PET-UNK-29afea89-2** (middle panel). The boxes exemplify ligands that significantly deform the pockets.

**C:** 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.

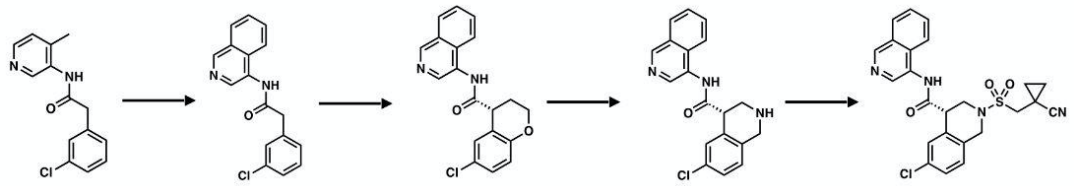
**Design of a novel SARS-CoV-2 Mpro inhibitor lead series with potent antiviral activity**

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(*50*). **Figure 4A** 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**). A significant increase in potency was unlocked by replacing pyridine with isoquinoline, which picks up additional hydrophobic interactions with N142. The SAR around the P2 pocket is considerably more tolerant to modifications, and broadly favours hydrophobic moieties. A step-change in potency was achieved by rigidifying the scaffold: We introduced 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 4A**), 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 functionalisable handle (**MAT-POS-3ccb8ef6-1, Figure 4A**), which maintained potency. Finally, we constructed a focused library realised via sulphonamide Schotten-Baumann coupling, furnishing a surprising increase in both enzymatic inhibition and cellular antiviral efficacy, leading to potent antiviral chemical series (**Figure 4A)** with a favourable safety profile, low brain penetrance (**Fig S12; Data S4**), improved oral bioavailability, but moderate in vitro-in vivo correlation in clearance (**Fig S13; Data S4**; all measured cellular antiviral data is available in **Data S5**).

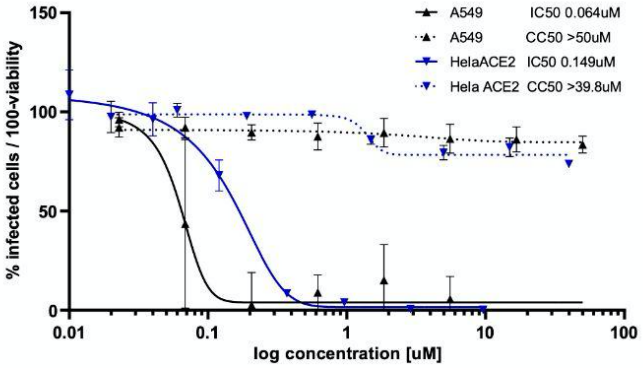
As an example for the aminopyridine lead series, we discuss antiviral efficacy, ADME and PK characteristics of compound **MAT-POS-e194df51-1:**. **MAT-POS-e194df51-1** was profiled in SARS-CoV-2 antiviral assays across multiple cell lines, exhibiting $EC_{50}$ of 64 nM in A549-ACE2-TMPRSS2 cells and 126 nM in HelaAce2, without measurable cytotoxicity (**Figure 4B**). This is in line with overall cellular efficacy for the chemical series; of 150 compounds with enzyme assay $IC_{50}$ <500nM assessed in A549-ACE2-TMPRSS2 cellular CPE assays, 15 compounds showed lower $IC_{50}$ values compared to the internal control Nirmatrelvir (Paxlovid) that was measured at an $IC_{50}$ of 218 nM in this assay (**Figure 4C**). 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 4D**). 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 variants (**Figure 4E**). Additionally, supporting the antiviral activity of our lead series in more complex organoid models, closely related molecules **PET-UNK-29afea89-2** and **MAT-POS-932d1078-3** with $IC_{50}$s in HelaAce2 CPE assays of 240 nM and 331 nM and 657 nM and 2.57 μM in A549-ACE2-TMPRSS2 CPE assays, respectively (**Fig S14A** and **B**) show an over 100-fold reduction of intracellular viral RNA and infectious virus secretion into the apical compartment of human induced pluripotent stem cell-derived kidney organoids (**Fig S14D** 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(*51*). **MAT-POS-e194df51-1** exhibits favourable properties required for an orally bioavailable inhibitor: using the rat pharmacokinetics (**Figure 4F**) and in vitro human metabolism data **(Figure 4G)**, early dose predictions based on one species only suggest an oral human dose (see **Detailed Methods**) in the range 450-2000 mg BID (two times a day) or 100-350 mg TID (three times a day). In addition, crystallographic studies reveal that the interaction pattern of **MAT-POS-e194df51-1** with the Mpro binding site is distinct to approved Mpro inhibitor nirmatrelvir and ensitrelvir (S-217622) (**Fig S15)**, thus potentially offering complementary resistance profiles and justifying further development.

**A**

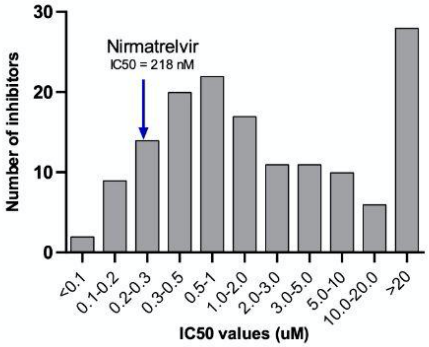


| | | | | | |
|---|---|---|---|---|---|
| **PostEra ID** | TRY-UNI-714a760b-6 | ADA-UCB-6c2cb422-1 | MAT-POS-b3e365b9-1 | MAT-POS-3ccb8ef6-1 | MAT-POS-e194df51-1 |
| **Fragalysis ID / PDB ID** | x2646 / 5RH2 | x10959 / 7S3S | x11612 / TBD | P0744 / TBD | P1788 / TBD |
| **Enamine Cat No** | Z1129289650 | 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**



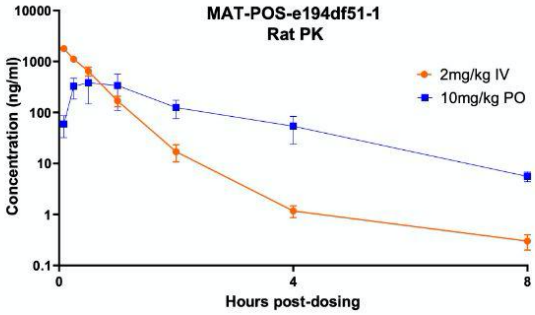

**D**

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

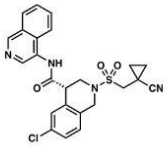
**E**

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

**F**




**G**



**MAT-POS-e194df51-1**

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

**Figure 4: 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 $IC_{50}$ of 64 nM in A549-ACE2-TMPRSS2 cells assessing cytopathic effect (CPE, black, plotted as 100 - % viability), and 126 nM in HelaAce2 assays (blue, plotted as % infected cells). Both assays were performed with p-gp inhibitors.
**C**: Histogram comparing antiviral efficacy of all COVID Moonshot compounds measured to date in an A549-ACE2-TMPRSS2 CPE cellular antiviral assay.
**D:** Detailed cellular antiviral assessment of key compounds comprising the synthetic strategy (Fig 4A) across different cell lines and assay techniques, with and without p-gp inhibitors, demonstrating efficacy of **MAT-POS-e194df51-1** in various set-ups and laboratories.
**E**: **MAT-POS-e194df51-1** shows good cross-reactivity against known circulating variants of SARS-CoV-2 in antiviral cellular assays in a CPE assay in HelaACE2 cells.
**F**: PK profile of **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 favourable safety profile, indicating translational potential of the lead series.

## Open Science presents a viable route to early drug discovery

The combined 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(*52*).

Here, we present the crowdsourced discovery and in-depth characterisation of a novel SARS-CoV-2 MPro lead inhibitor series. Despite our optimization and characterization efforts, considerable gaps from reporting a clinical candidate remain: The series requires further pharmacokinetic and pharmacodynamic optimisation; 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 more 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 enabling Phase I readiness. 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) (*53*). To reach Phase II readiness we expect further $7.5M will be required in addition to process route development costs(*54*). Downstream, we envisage a direct-to-generics production and distribution model where the eventual drug price will immediately reflect the cost of producing the compounds alongside a reasonable margin, rather than the monopoly rent associated with recovering substantial development costs and based on restrictive intellectual property positions on the compound. Working with

globally-focused development and manufacturing organisations and public health funding organisations united in our goal of global equitable access, we are building a streamlined, economically viable downstream development plan designed to deliver affordable treatment options from this chemical series, with a primary focus on patients in Low-and-middle-income countries (LMICs).

Open science efforts have transformed many areas of biosciences, with illustrious examples such as the Human Genome Project(*55*), the Structural Genomics Consortium(*56*), and the RAS Initiative(*57*). Yet its inroad in therapeutic discovery has been slow principally because of the perceived need for commercial return. We hope that The COVID Moonshot is an exemplar for future open science drug discovery(*58*) that paves the way towards a new paradigm for infectious diseases drug discovery - a disease area of grave public importance but one which is chronically underfunded by the private sector(*58*).

**Materials and Methods**

The Supplementary Material contains Detailed Methods and protocols for all the experiments reported in this manuscript.

**Data availability**

- **All compound submissions and experimental data are available via GitHub:** https://github.com/postera-ai/COVID_moonshot_submissions
- **All compound designs, datasets, and X-ray structures are indexed on the COVID Moonshot website:** https://postera.ai/covid
- **Bioactivity data can be interactively browsed at the COVID Moonshot website:** https://covid.postera.ai/covid/activity_data
- **Alchemical free energy calculations code and datasets are indexed on GitHub:** https://github.com/foldingathome/covid-moonshot
- **All X-ray structures are available for interactive viewing or bulk download on Fragalysis:** https://fragalysis.diamond.ac.uk/viewer/react/preview/target/Mpro
- **All chemical characterization data and PDBs of X-ray structures:** https://drive.google.com/drive/u/0/folders/1QANWRAjBG1xxsZ3iN-VkZJxmKpviv4NG

**Resource availability**

- **Synthesized compounds**: We have made all compounds assayed here are available from the current Enamine catalogue, and readily available for purchase from Enamine (and other suppliers) via the Manifold platform accessible for each compound page on the COVID Moonshot website: https://covid.postera.ai/covid/activity_data

**Acknowledgements**

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

**Funding Acknowledgments and Financial Disclosures**

Funding acknowledgements and Disclosures statements for each author are listed in the consortium spreadsheet (The COVID Moonshot Consortium). 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 licence to any Author Accepted Manuscript version arising from this submission.

**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, J. Prilusky, S. Duberstein, G. Cohen, 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* (2021), p. In Press, , doi:10.1101/2020.09.21.299776.

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. 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).

33. 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).

34. 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).

35. 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).

36. 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).

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

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

39. 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).

40. 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.

41. 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.

42. 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).

43. 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).

44. 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).

45. 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).

46. 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).

47. 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).

48. 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).

49. 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).

50. 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).

51. 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.

52. 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.

53. 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/).

54. 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).

55. 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).

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

57. 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).

58. 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.