# High-Throughput Virtual Screening and Validation of a SARS-CoV-2 Main Protease Noncovalent Inhibitor

Austin Clyde,* Stephanie Galanie, Daniel W. Kneller, Heng Ma, Yadu Babuji, Ben Blaiszik, Alexander Brace, Thomas Brettin, Kyle Chard, Ryan Chard, Leighton Coates, Ian Foster, Darin Hauner, Vilmos Kertesz, Neeraj Kumar, Hyungro Lee, Zhuozhao Li, Andre Merzky, Jurgen G. Schmidt, Li Tan, Mikhail Titov, Anda Trifan, Matteo Turilli, Hubertus Van Dam, Srinivas C. Chennubhotla, Shantenu Jha,* Andrey Kovalevsky,* Arvind Ramanathan,* Martha S. Head,* and Rick Stevens*

**ABSTRACT:** Despite the recent availability of vaccines against the acute respiratory syndrome coronavirus 2 (SARS-CoV-2), the search for inhibitory therapeutic agents has assumed importance especially in the context of emerging new viral variants. In this paper, we describe the discovery of a novel noncovalent small-molecule inhibitor, MCULE-5948770040, that binds to and inhibits the SARS-Cov-2 main protease ($M^{pro}$) by employing a scalable high-throughput virtual screening (HTVS) framework and a targeted compound library of over 6.5 million molecules that could be readily ordered and purchased. Our HTVS framework leverages the U.S. supercomputing infrastructure achieving nearly 91% resource utilization and nearly 126 million docking calculations per hour. Downstream biochemical assays validate this $M^{pro}$ inhibitor with an inhibition constant ($K_i$) of 2.9 $\mu$M (95% CI 2.2, 4.0). Furthermore, using room-temperature X-ray crystallography, we show that MCULE-5948770040 binds to a cleft in the primary binding site of $M^{pro}$ forming stable hydrogen bond and hydrophobic interactions. We then used multiple $\mu$s-time scale molecular dynamics (MD) simulations and machine learning (ML) techniques to elucidate how the bound ligand alters the conformational states accessed by $M^{pro}$, involving motions both proximal and distal to the binding site. Together, our results demonstrate how MCULE-5948770040 inhibits $M^{pro}$ and offers a springboard for further therapeutic design.
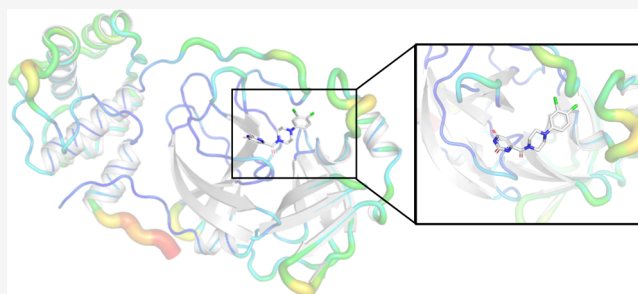
## INTRODUCTION

The ongoing novel coronavirus pandemic (COVID-19) has resulted in over 200 million infections and more than 4 million deaths worldwide[a]. Although vaccines against the COVID-19 causative agent, the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), are being deployed,[1,2] the discovery of drugs which can inhibit various SARS-CoV-2 proteins remains essential for treating patients.[3,4] Leveraging existing coronavirus treatments developed for SARS and middle eastern respiratory syndrome (MERS),[5] as well as broad international collaborations, researchers have quickly determined structures for over 15 viral proteins, including inhibitor/lead bound structures and fragment-based screening for several nonstructural proteins (NSPs) such as the main protease (3C-like protease/$M^{pro}$), adenosine diphosphate ribosyl phosphatase (ADPRP/NSP3/Mac1), endoribonuclease (NSP15), and helicase (NSP13),[6] all playing crucial roles in viral replication. Together, these collaborations have significantly accelerated the design and development of antiviral treatments targeting SARS-CoV-2.[7,8]

Of these proteins, $M^{pro}$ is an attractive drug target mainly because it plays a critical role in viral replication and does not have any closely related homologues within the human genome.[9,10] Drug discovery efforts have resulted in discovering/repurposing small molecules based on their ability to inhibit other coronavirus $M^{pro}$ from MERS and SARS; however, it has been a challenge to identify noncovalent inhibitors for SARS-CoV-2 $M^{pro}$ mainly due to the intrinsic flexibility of the primary binding site.[11]

High-throughput virtual screening (HTVS) is a common step of drug discovery, enabling rapid, low-cost screening of significantly larger compound libraries than feasible in experimental studies.[12] A number of efforts have focused on

**Table 1. Resource Utilization When Docking $126 \times 10^6$ and $205 \times 10^6$ Ligands with OpenEye Using RAPTOR on Frontera[a]**

| #nodes | #ligands [$\times 10^6$] | utilization [%] | docking time [s] | | | throughput [$\times 10^6$ docks/h] | |
|---|---|---|---|---|---|---|---|
| | | | min | max | mean | max | mean |
| 128 | 205 | 89.6 | 0.1 | 3582.6 | 28.8 | 17.4 | 5.0 |
| 3850 | 126 | 95.5 | 0.1 | 833.1 | 25.1 | 27.5 | 19.1 |
| 7000 | 126 | 90.0 | 1.0 | 180.0 | 10.1 | 144.0 | 126.0 |

[a]Docking time varies depending on the physical properties of each ligand, affecting the obtained docking throughput.

creating open HTVS infrastructure, taking advantage of cloud computing platforms or supercomputing resources to support large-scale ligand docking across various protein targets.[13] These platforms have leveraged open-source toolkits such as AutoDock/AutoDock-VINA (for molecular docking)[14] in conjunction with molecular modeling (MM) and molecular dynamics (MD) simulation engines to capture "modes" of interaction between a protein target and specific compounds from compound libraries (e.g., ZINC and[15] MCULE[16]). Of these approaches, the COVID-Moonshot project, using crowdsourced design strategies, high-throughput experimental screening, MD simulations, and machine learning (ML) were able to identify both covalent and noncovalent inhibitors against M[pro] which demonstrated viral inhibition in vitro.[17]

In this paper, we describe our discovery of a noncovalent small-molecule inhibitor for M[pro] using our HTVS platform that employs supercomputing resources, ensemble docking strategies, high-throughput experimental screening, X-ray crystallography, and MD. Complementary to efforts that scaled crowdsourcing approaches,[17] as well as HTVS across potentially O(billion) compounds,[13,18] we used a library consisting of 6.5 million in-stock compounds from the MCULE library.[16] Ensemble docking was carried out across available crystal structures for M[pro], from which O(1, 000) top consensus-scoring compounds across two popular docking programs (Autodock-VINA[14] and OpenEye FRED[19]) were experimentally characterized. From these compounds, we discovered one molecule that inhibits M[pro] with a $K_i$ of 2.9 $\mu$M and determined its room-temperature X-ray crystallographic structure to 1.8 Å resolution. Finally, we used $\mu$s-time scale atomistic MD simulations to characterize the binding mechanism to the M[pro] active site, while altering the enzyme's overall conformational dynamics. Our workflow provides a scalable framework for the rapid discovery of viable lead molecules against SARS-CoV-2. All of our generated data, including ensemble docking results and MD simulations, are freely available and can be used for future studies.

## ■ METHODS

**Molecular Library Generation.** We use a set of on-demand compounds from Mcule (ORD), which can be freely obtained on their website.[16] ORD consists of compounds from Mcule listed as available on their website, Mcule Purchasable (Known Stock Amounts).

**Molecular Docking Protocol with an OpenEye Toolkit.** The main protease structures screened included the following PDB structures: 7BQY, 6LU7, 6W63, 7C7P, and 7JU7. The receptors for OpenEye Chemgauss4 scoring were prepared using the known binding region of M[pro] with the OpenEye Docking Toolkit.[19]

For memory efficiency, conformer generation and tautomerization were performed on the fly. OMEGA[20] was used, sampling around 300−500 conformations for each ligand.

When ligand-binding information was available in the receptor, HYBRID was used due to its increased pose prediction accuracy over FRED.[19] HYBRID and FRED have the same scoring function; however, HYBRID uses a heuristic to reduce the search space for ligand positioning. The best score from the ensemble of tautomers and conformers is chosen as the representative "docking score" for the chemical species.

**Computational Workflow.** OpenEye toolkit's FRED docking program was deployed on Frontera at TACC. Docking scores for the M[pro] receptor were computed with individual runs per pocket. The docking protocol as described above requires the following steps for each compound: (1) load receptor into memory; (2) load compound data stored in MCULE[16] database from disk; (3) run the specific docking protocol over the receptor/compound pair; and (4) write the resulting docking score to persistent storage.

High-throughput docking was implemented using RADI-CAL-Pilot (RP) and RAPTOR.[21] RP is a pilot-enabled runtime system, while RAPTOR is a scalable master/worker overlay developed to improve the execution performance of many short-running tasks encoded as Python functions. The runs used between 128 and 7000 concurrent nodes. For each run, we measured throughput (the number of docking calls per hour) and resource utilization (the fraction of time that acquired nodes were kept busy). Resource utilization was dependent on the size of the run (number of compounds to dock and number of nodes to use) and was typically above 90%. In Table 1, we summarize the results from multiple runs, including one on 7000 nodes.

The runtimes for individual docking calculations have a long-tail distribution (see Table 1). The mean docking time determines the achievable maximal throughput, for example, the last row in Table 1 suggests an upper limit on average docking call throughput of 139M docking calls per hour (1 h * cores per node * number of nodes/mean docking time seconds = 3600 * 56 * 7000/10.1 = 139,722,772 ~ 139M docks/h). The achieved average throughput was ~126M docks/h.

**SARS-CoV-2 M[pro] Expression and Purification.** A gene construct encoding M[pro] (NSP5) from SARS-CoV-2 was cloned into plasmid pD451-SR (ATUM, Newark, CA), which was developed in ref 11, and expressed and purified consistent with the protocols detailed in ref 22. Protein purification supplies were purchased from Cytiva (Piscataway, New Jersey, USA). Briefly, an authentic N-terminus is achieved by a NSP4-NSP5 autoprocessing sequence (SAVLQ ↓ SGFRK where the arrow indicates the scissile bond) flanked by maltose binding protein and M[pro]. Following M[pro], a sequence encoding the human rhinovirus 3C (HRV-3C) cleavage site (SGVTFQ ↓ GP) is followed by a His6-tag. The N-terminal sequence is created by autocleavage during expression, while the C-terminus is generated by HRV-3C treatment following Ni-immobilized metal affinity chromatography.

**Primary M[pro] Inhibition Screen.** Compounds were purchased from Mcule, Inc. as 10 mM stock solutions in

dimethyl sulfoxide (DMSO) and stored at −20 °C. The assays were performed in 40 $\mu$L of total volume in black half area 96-well plates (Greiner PN 675076) at 25 °C. The assay buffer contained 20 mM Tris−HCl pH 7.3, 100 mM NaCl, 1 mM EDTA, and 2 mM reduced glutathione (6.15 mg added per 10 mL of buffer fresh for each experiment) with 5% v/v final DMSO concentration. $M^{pro}$ initial rates were measured using a previously established fluorescence resonance energy transfer (FRET) peptide substrate assay.[23] The FRET substrate DABCYL-KTSAVLQ↓SGFRKM-E(EDANS) trifluoroacetate salt was purchased from Bachem (PN 4045664), dissolved to 10 mM in DMSO, and stored in aliquots at −20 °C. 10 $\mu$L of enzyme solution was dispensed into wells (250 nM final concentration), followed by 10 $\mu$L of inhibitor solution (20 $\mu$M final concentration), centrifuged briefly, and incubated for 30 min. Reactions were initiated by adding 20 $\mu$L of the substrate at 40 $\mu$M final concentration. Fluorescence was detected every 24 s using a Biotek Synergy H1 plate reader with an excitation wavelength of 336 nm and an emission wavelength of 490 nm, 6.25 mm read height, low lamp energy, and 3 measurements per data point. After background subtraction of the average of no enzyme negative controls, product formation was quantified using a 0.05−22 $\mu$M calibration curve of the free EDANS acid (Sigma PN A6517). Product concentrations were adjusted for inner filter absorbance effects with correction factors generated by comparing the fluorescence of 2 $\mu$M EDANS in solution with each concentration of the substrate used to that with no substrate. Initial rates were determined for time points in the linear range by linear regression in Excel, residual activities were determined by normalizing candidate initial rates to the average of the positive controls, and Z-scores were determined by dividing the difference between the candidate initial rate and average positive control initial rate by the standard deviation of the positive control initial rates. The Z′-statistics for the plate was calculated using the published equation.

**Peptide Synthesis.** The unlabeled $M^{pro}$ substrate peptide AVLQ↓SGFRKK-amide and the isotopically labeled substrate and product peptide internal standards (A + 7)VLQSGFRKK-amide and (A + 7)VLQ-OH were synthesized by automated peptide synthesis using a Liberty PRIME peptide synthesizer (CEM). Reagents were peptide synthesis or biotechnology grade. Amino acids were purchased from P3Bio, and Fmoc-[$^{13}C_3$, $^{15}N$, $D_3$]-alanine (A + 7) was previously synthesized at Los Alamos National Laboratory following published protocols.[24] Other purchased reagents were dimethylformamide (DMF), pyrrole (prepared as 20% v/v in DMF), and high-performance liquid chromatography (HPLC)-grade acetonitrile (Alfa Aesar); diisopropylcarbodiimide and Oxyma Pure (AKScientific) N,N-diisopropyl ethyl amine (DIPEA), triisopropyl silane (TIPS), trifluoroacetic acid (TFA), thioanisole, Rink amide resin, and octaethylenglycol-dithiol (Sigma-Aldrich); and dichloromethane and Optima mass spectrometry (MS)-grade acetonitrile (Fisher Scientific).

Peptide syntheses were performed at a 0.1 mM scale under argon on a Rink amide resin with 0.1 M DIPEA added to the Oxyma solution to prevent hydrolysis of acid-labile side-chain protecting groups, obtaining average yields for double couple cycles of >99%. For stable isotope-labeled peptides, only two equivalents of the labeled amino acid were used and coupling time was extended to 20 min at 90 °C. Peptides were deprotected with the following mixture: 1.25 mL of TIPS, 0.625 mL of thioanisole, and 1.25 mL of octaethyleneglyco-

dithiol, and after 5 min, TFA was added to a total volume of 25 mL. Solutions were filtered and the filtrate was concentrated to 10 mL, followed by precipitation with ice-cold ether and collection by centrifugation.

Peptides were purified to >98% by Waters HPLC work-station (2545 pump with 2998 photodiode array detector) with a Waters BEH 130, 5 $\mu$m, 19 × 150 mm C18 column, and a linear gradient from 98:2 to 50:50 water/acetonitrile with 0.1% TFA at 20 mL/min. Absorbance at 215 nm was monitored, and peaks were collected and lyophilized to yield a white fluffy solid. Peptide purity was analyzed by analytical HPLC and Thermo LTQ MS with electrospray ionization in the positive mode with a Waters BEH 130, 5 $\mu$m, 4.6 × 150 mm C18 column, and a linear gradient from 96:2 to 60:40 water/acetonitrile with 0.1% TFA at 1.5 mL/min over 12 min (Figure S2).

**Quantitative MS $M^{pro}$ Inhibition Assay.** The quantitative MS inhibition assay was performed as described for the FRET-based primary screen with some modifications. Round-bottom polypropylene 96-well plates (Corning PN 3365) were used with 150 nM final $M^{pro}$ concentration and the unlabeled peptide substrate synthesized above. Five min after substrate addition, the assay was quenched 1:1 v/v with 2% formic acid in water with 2 $\mu$M each internal standard peptide from above and centrifuged for 10 min at 4 °C, and the supernatant was diluted 1:9 v/v into 1% formic acid. Substrate and product peptides and internal standards were quantified by high-throughput MS using a Sciex 5500 QTRAP with a custom open port sampling interface (OPSI).[25] Samples were introduced as 2 $\mu$L droplets, and the OPSI-MS analysis was performed using 10:90:0.1 v/v/v water/methanol/formic acid at 80 $\mu$L/min. Positive ion mode electrospray ionization parameters were CUR: 25, IS: 5000, TEM: 400, GS1: 90, GS2: 60, EP: 10, and CXP: 10. Optimized multiple reaction monitoring detection parameters were dwell: 50 msec, product DP: 100 and CE: 25, and substrate DP: 150 and CE: 34. The following mass-to-charge transitions were monitored: substrate AVLQSGFRKK, 566.9 → 722.3 and (A + 7)VLQSGFRKK, 570.4 → 722.3; product AVLQ, 430.3 → 260.3 and (A + 7)VLQ, 437.3 → 260.3. Product formation and the remaining substrate were quantified by dividing the peak area of the transitions by that of the corresponding internal standard transitions.

**$IC_{50}$ and $K_i$ Value Determination.** To determine the concentration at which a compound was able to achieve 50% inhibition of $M^{pro}$ activity in vitro ($IC_{50}$), the FRET and quantitative MS assays described above were performed at 10 concentrations of the inhibitor (0.56−100 $\mu$M) in triplicate with 150 nM enzyme. Initial rates, for FRET, or product formation in 5 min, for MS, were normalized to no inhibitor control (100% activity) and no enzyme control (0% activity), and nonlinear regression of the [inhibitor] versus normalized response $IC_{50}$ equation was performed to fit the data using GraphPad Prism 9.0.0, yielding $IC_{50}$ and its 95% confidence interval. To confirm the mechanism of inhibition and determine $K_i$, the FRET activity assay was performed at 8 concentrations of the substrate (20−500 $\mu$M) and 4 concentrations of the inhibitor (0−25 $\mu$M) in triplicate in two independent experiments. A global nonlinear regression was performed to fit the competitive inhibition equation to the entire data set using GraphPad Prism 9.0, yielding $K_M$, $K_i$, and $V_{max}$ and their associated 95% confidence intervals.
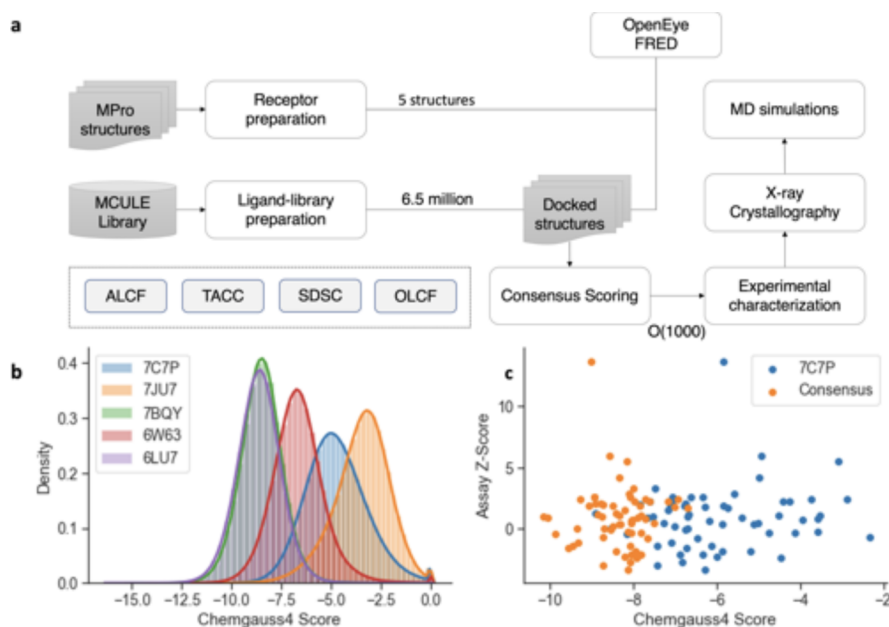
**Figure 1.** (A) Computational workflow used for screening on-demand chemical libraries against SARS-CoV2 M$^{pro}$ with computational docking techniques. Four major supercomputing centers were utilized, namely, Argonne Leadership Computing Facility (ALCF), Texas Advanced Computing Center (TACC), San Diego Supercomputing Center (SDSC), and Oak Ridge Leadership Computing Facility (OLCF). (B) Distribution of Chemgauss4 scores, from docking, from the docking a 6 million in-stock compound library. (C) Consensus scoring used shifted possible hits (higher Z-score is better) toward better scoring regions over just a single score from a single structure (7C7P is used for illustration). A lower consensus score implies a higher likelihood from the docking programs that the candidate compound will bind to the receptor.

**Crystallization.** Crystallization reagents were purchased from Hampton Research (Aliso Viejo, California, USA). Crystallographic tools were purchased from MiTeGen (Ithaca, New York, USA) and Vitrocom (Mountain Lakes, New Jersey, USA). M$^{pro}$ was concentrated to ~5.0 mg/mL in 20 mM Tris pH 8.0, 150 mM NaCl, and 1 mM TCEP, for crystallization. The presence of a reducing agent such as TCEP is essential for preventing oxidation of the catalytic cysteine side chain.[26] Conditions for growing crystalline aggregates of ligand-free (LF) M$^{pro}$ were identified by high-throughput screen at the Hauptman-Woodward Research Institute[27] and reproduced locally using 22% PEG3350 and 0.1 M Bis−Tris pH 6.5 in 20 μL drops with 1:1 ratio of the protein/well solution using sitting-drop vapor diffusion with microbridges. Crystal aggregates of the LF sample were converted to microseeds with Hampton Research Seed Beads and used for nucleating M$^{pro}$ crystals in subsequent co-crystallization experiments. Lyophilized MCULE-5948770040 for co-crystallization was dissolved in 100% DMSO as a 50 mM stock stored at −20 °C. MCULE-5948770040 was mixed with M$^{pro}$ at a 5:1 M ratio and allowed to incubate on ice for a minimum of 1 h. Crystals were grown in a 40 μL drop at a 1:1 mixture with 18% PEG3350 and 0.1 M Bis−Tris pH 7.0 with 0.2 μL of 1:200 dilution microseeds and incubated at 14 °C. A large crystal measuring ~1 × 0.5 × 0.3 mm suitable for room-temperature X-ray diffraction grew after 2 weeks (Figure S2).

**Room-Temperature X-ray Data Collection and Structure Refinement.** The protein crystal was mounted using a MiTeGen (Ithaca, NY) room-temperature capillary system (Figure S2). X-rays for crystallography were generated from a Rigaku HighFlux HomeLab employing a MicroMax-007 HF X-ray generator and Osmic VariMax optics allowing diffraction images to be collected using an Eiger R 4M hybrid photon-counting detector. Diffraction data were reduced and scaled using Rigaku CrysAlis Pro software package. Molecular replacement was performed using the LF room-temperature M$^{pro}$ structure (PDB code 6WQF)[11] using Molrep.[28] Structure refinement was performed with Phenix.refine from Phenix suite[29] and COOT[30] for manual refinement and Molprobity.[31] Data collection and refinement statistics are listed in Table S1. The structure and the corresponding structure factors of the room-temperature M$^{pro}$/MCULE-5948770040 complex have been deposited into the Protein Data Bank with the PDB accession code 7TLJ.

**MD Simulations of the M$^{pro}$ Complex with MCULE-5948770040.** The crystal structure of the protein dimer was modeled with the AMBER MM package[32] with the amber.ff14sb force field parameters (for the protein)[33] and with the GAFF parameters (for the ligand).[34] In order to better determine the partial charges for the ligand, quantum mechanical calculations were performed using NWChem[35] based on the RESP method at the B3LYP/6-31G* level of theory,[36] while all bonded parameters were taken from the GAFF force field.

The systems (both the ligand-bound/LB and LF) were solvated using the TIP3P water model and counter ions were added to neutralize the charge. After equilibrating the systems using previously published protocols,[37] we carried out production runs using the OpenMM[38] simulation package on Nvidia V100 GPUs using the Argonne Leadership Computing Facility's (ALCF) computing clusters. Each time step was integrated with a Langevin integrator at 310 K, 1 ps$^{-1}$ friction coefficient, and 2 fs interval with fixed lengths being maintained for atomic bonds involving hydrogen atoms. System pressure was maintained at 1 atm with the Monte Carlo Barostat. Nonbonded interactions were cut off at 1.0 nm, and particle mesh Ewald was implemented for long-range
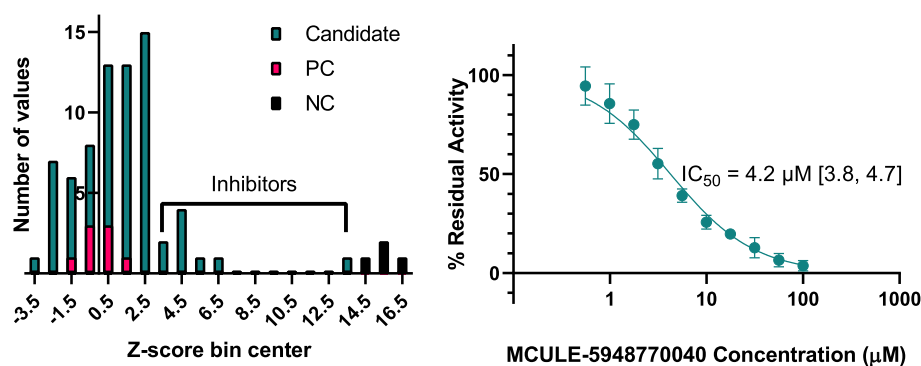
**Figure 2.** Plate-based M$^{pro}$ activity inhibition screening and hit confirmation. (A) Histogram of Z-scores of candidate inhibitors, no enzyme negative controls (NCs), and no inhibitor positive controls (PCs). (B) Inhibition of M$^{pro}$ activity in vitro with increasing concentration of MCULE-5948770040. Initial rates are normalized to no inhibitor control (100% activity) and no enzyme control (0% activity). Error bars are standard deviation of two independent experiments, each performed in triplicate. Lines indicate the nonlinear regression of the [inhibitor] vs normalized response IC$_{50}$ equation to the data with GraphPad Prism. Bracketed values indicate 95% confidence intervals from the regression.

interactions. The simulations were run for 1 $\mu$s and 50 ps reporting interval (four replicas).

### Quantifying Conformational Transitions in the Ligand-Bound and LF States of MPro with Anharmonic Conformational Analysis-Driven Autoencoders (ANCA-AE).

Conformational fluctuations within biomolecular simulations (and specifically proteins) show significant higher-order moments (see Figure S12); these fluctuations may relate to protein function.[37,39] To quantify such anharmonic fluctuations within our simulations, we used fourth-order statistics to describe atomistic fluctuations and to characterize the internal motions using a small number of anharmonic modes.[40] We projected the original data (306 C$^{\alpha}$ atoms per chain—$(x, y, z)$ coordinates) onto a 40 (or 50) dimensional space, depending on the set of simulations considered. Notably, for the ligand-bound states (in both protomers), 40 dimensions covered about 95% of overall variance, whereas we required 50 dimensions to cover 95% of the variance when we included the ligand-bound states from just one protomer.

Given the significant nonlinearity in the atomic fluctuations, we used an autoencoder to further delineate the intrinsic structure in the low-dimensional anharmonic space. Similar to approaches that use variational approximations to model molecular kinetics from MD simulations,[41] we used an autoencoder architecture consisting of a symmetric encoder and decoder network. The network is composed of a single dense layer with 32 dimensions and an 8-dimensional latent space. We trained the network for 50 epochs using the RMSprop optimizer to minimize the mean-squared error reconstruction loss with a learning rate of 0.001, a weight decay of 0.00001, and a batch size of 64. We used ReLU activation in all places except the final reconstruction layer, where we used Tanh activation. A mixture of Gaussian (MoG) model was used to cluster the conformations in the low-dimensional landscape, similar to the approach outlined in ref 42.

### ■ RESULTS

#### HTVS of M$^{pro}$ with On-Demand Molecular Libraries.

A docking screen against the main protease of SARS-CoV2, M$^{pro}$, was performed on an orderable on-demand compound library from Mcule.[16] Given the intrinsic flexibility of the M$^{pro}$'s primary binding pocket consisting of the four conserved binding sites (S1′, S1, S2, and S4),[11,43] we used five different crystal structures for an ensemble docking approach using PDB

identifiers 6LU7,[44] 6W63,[45] 7BQY,[44] 7C7P,[46] and 7JU7.[47] In addition to the structural ensemble, we used the docking protocols and scoring functions from the OpenEye Scientific FRED[19] toolkit. In total, over 63 million docking scores were computed over the five structures, two compound libraries, and two protocols. The overall workflow is summarized in Figure 1a. The workflows were deployed on HPC resources at the Argonne and Oak Ridge leadership computing facilities (ALCF/OLCF) using Theta and Summit supercomputers and using the Texas Advanced Computing Center (TACC; Frontera) and San Diego Supercomputing Center (SDSC; Comet). The resulting docking libraries (including scripts of preparation and docking) and the docking scores are available as a downloadable data set. The details of the computational performance and workflow optimization are described in the Supporting Information text (Sections S1 and S2) and Figure S1a−c.

The resulting compounds were ranked based on the docking scores in conjunction with visual inspection and availability at the time, and selected compounds were ordered for experimental validation studies. Interestingly, docking score distributions across each of the structures were slightly different (summarized in Figure 1b,c), and we therefore examined the top 0.1% of the overall distributions. Between receptors' respective docking, the highest correlation coefficient is 0.85 (7BQY and 6LU7) and the lowest was 0.001 (6W63 and 7JU7; see Table S1). In fact, 6W63's docking result is an outlier with respect to the other four receptors, with the highest correlation coefficient of only 0.003. Given the variation among docking results between receptors, a consensus score was deemed necessary. A consensus score was created by taking the minimum over the available series. A minimum was chosen rather than an average, or other aggregation techniques, due to the nature of our docking protocol. OpenEye FRED has a wide range of scores, unbounded above or below. A small steric difference between receptors can cause a wide numerical discrepancy or even lack of a result. We see in Figure 1d a significant difference between the correlation of consensus scores over using single samples (Table S1).

#### MCULE-5948770040 is a SARS-CoV-2 M$^{pro}$ Inhibitor.

Based on the consensus-scoring procedure mentioned above, 116 compounds from the Mcule database were selected for experimental screening using the top 20 from different M$^{pro}$
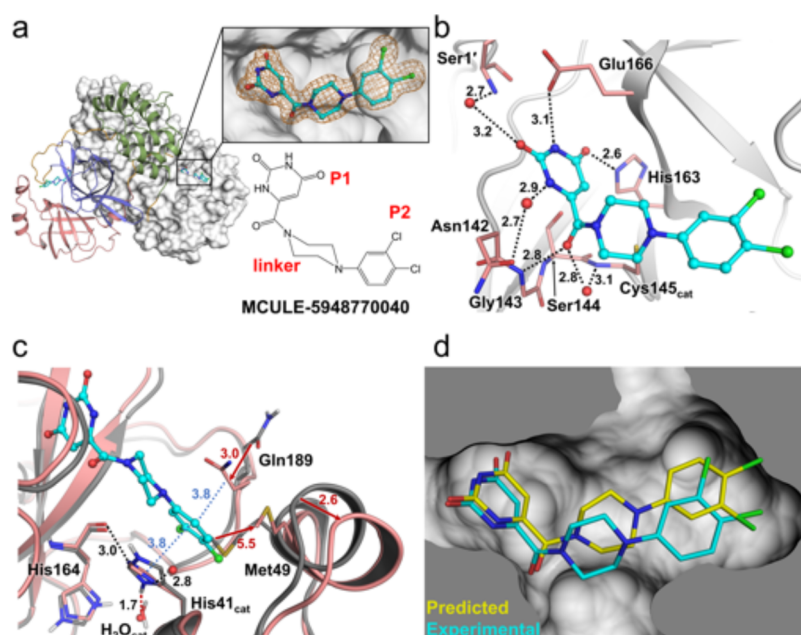
**Figure 3.** Room-temperature X-ray crystal structure of M[pro] in complex with MCULE-5948770040 and comparison with LF and docked structures. (A) Overall M[pro] homodimer in complex with MCULE-5948770040 (cyan carbon ball-and-stick representation). One protomer is shown as a cartoon representation with domains I, II, and III in pink, purple, and green, respectively, and orange interdomain loops. The other protomer is shown as a white surface. Insets show MCULE-5948770040 electron density (2Fo−Fc at 1.2σ as orange mesh) and 2D chemical diagram. (B) Intermolecular interactions between M[pro] (gray cartoon with salmon sticks) and the ligand. H bonds are shown as black dashes. Distances in Å. (C) Superposition of the M[pro]/9MCULE-5948770040 complex (salmon) with LF X-ray/neutron structure (gray, PDB code 7JUN). Red arrows indicate conformational shifts from the LF structure to complex structure. Blue dots show π−π interactions with the P2-dichlorobenzene group. Red dashes represent a lost H bond due to catalytic His41's imidazole side chain flip. (D) Comparison of computationally predicted (yellow carbons) and experimentally determined (cyan carbons) pose of MCULE-5948770040 bound to M[pro].

crystal structures. Of these 116 compounds, 5 were not available for ordering, 15 were excluded due to pan assay interference compounds (PAINS) violations based on the substructure filters of Baell and Holloway,[48] and 72 were ultimately delivered. These compounds were subjected to a primary SARS-CoV-2 M[pro] activity inhibition screen in which they were preincubated with the enzyme and the initial velocities of cleavage of a FRET peptide were determined.[23] The Z′-factor of the assay was 0.65, and the distribution of Z-scores of compounds and positive (no inhibitor) and negative (no enzyme) controls is shown in Figure 2A. At least, 25% inhibition was observed for 7 compounds, with MCULE-5948770040 resulting in the lowest residual activity at 20 μM (12%, Table S2).

**Mechanism of Inhibition.** The concentration dependence of MCULE-5948770040 in vitro M[pro] inhibition was measured at 40 μM substrate, giving an $IC_{50}$ of 4.2 μM (95% confidence interval 3.8, 4.7) (Figure 2B). An orthogonal quantitative high-throughput MS-based endpoint assay was also performed at 40 μM unlabeled peptide substrate, giving a similar $IC_{50}$ of 2.6 μM (95% CI 2.3, 2.9) (Figure S2). Initial rates measured at 20−500 μM substrate and 0−25 μM inhibitor were consistent with a competitive mechanism of inhibition with a $K_i$ of 2.9 μM (Figure S2).

**Room-Temperature X-ray Crystal Structure of M[pro] in Complex with MCULE-5948770040.** To elucidate the molecular basis of M[pro] inhibition by the MCULE-5948770040 compound, an X-ray crystal structure of M[pro] in complex with the compound was determined to 1.80 Å at near-physiological (room) temperature (Table S3 and Figure S3). The M[pro]/MCULE-5948770040 complex crystallized as the

biologically relevant homodimer with the protomers related by a twofold crystallographic axis (Figure 3). The tertiary fold is shown in Figure 3. Each protomer consists of three domains (I−III). The substrate-binding cleft is formed at the interface of catalytic domains I (residues 8−101) and II (residues 102−184), whereas the α-helical domain III (residues 201−303) creates a dimerization interface.[44,49] The substrate-binding cleft lies on the surface of the enzyme and accommodates amino acid residues or inhibitor groups at positions P1′−P5 in subsites S1′−S5, respectively.[50−52] Subsites S1, S2, and S4 have well-defined shapes, while S1′, S3, and S5 are surface-facing with poorly defined edges.[53] The noncanonical catalytic dyad composed of Cys145 and His41 lies deep within the substrate-binding cleft poised for peptide bond cleavage between the C-terminal P1′ and N-terminal P1 positions.

MCULE-5948770040 binds noncovalently to the active site of M[pro], occupying subsites S1 and S2. The electron density for the inhibitor is unambiguous (Figure 3a) enabling accurate determination of the protein−ligand interactions (Figure 3b). The uracil P1 group of the ligand is situated in the S1 subsite driven by polar contacts. Nε2 of the His163 imidazole side chain makes a close 2.6 Å H bond with the carbonyl at position 4 of the uracil substituent. Notably, His163 was previously determined to be singly protonated on the Nδ1 by neutron crystallography,[53] suggesting a possible rearrangement of the protonation state for His163 side chain upon ligand binding. The far end of the S1 subsite is formed such that the second protomer's N-terminal-protonated amine creates H bonds with the Glu166 side chain, Phe140 main chain carbonyl, and a water molecule. The amide NH at position 3 of the ligand's P1 heterocycle is situated within the hydrogen-bonding distance
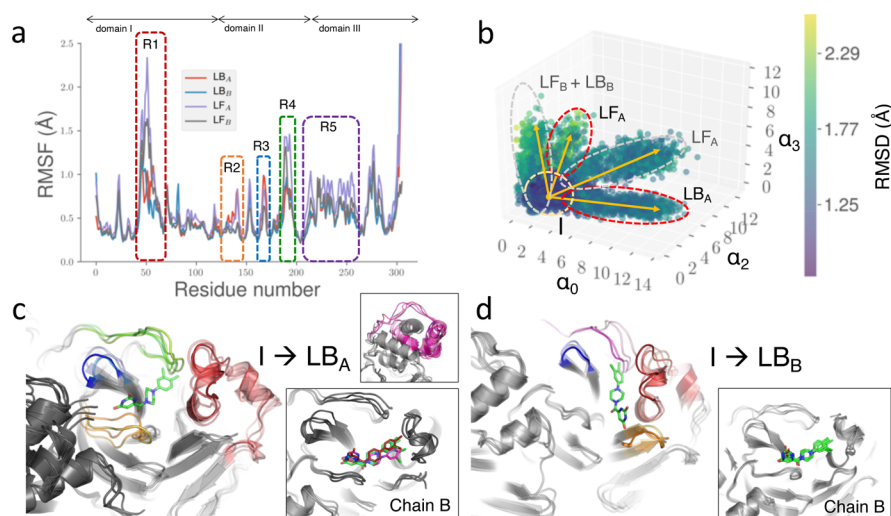
**Figure 4.** Conformational changes upon MCULE-5948770040 binding to M$^{pro}$ indicate changes within distinct regions, both close-to and farther-away from the primary binding site. (a) RMS fluctuations of the LF and LB state of M$^{pro}$ show several regions with decreased fluctuations that are highlighted within rounded rectangles. Although several regions within these regions are largely similar, amino-acid residues interacting with the ligand stabilize the binding site. (b) To further quantify the nature of these fluctuations, we characterized the collective motions which show distinct conformational states sampled by the LF and LB states. The yellow arrows indicate conformational transitions from the average structure toward the distinct conformational states (I, LF$_A$, LF$_B$, LB$_A$, and LB$_B$). These transitions are mapped in (c) I → LB$_A$ and (d) I → LB$_B$. (We show I → LF$_A$ and I → LF$_B$). In each case, we observed that M$^{pro}$ chain B of the dimer was more stable than the chain A (insets). Regions highlighted in (a) show the motions undergone by the different regions of M$^{pro}$.

with Glu166 and Phe140, although the geometry is unfavorable. The carbonyl and amide NH at positions 2 and 1 participate in water-mediated H bonds with Ser1′ and Asn142, respectively. M$^{pro}$ features an oxyanion hole created by the main chain amide NH groups of Gly143, Ser144, and Cys145 at the S1 subsite base. A carbonyl linking the P1 uracil to the central piperazine linker of MCULE-5948770040 is positioned on the perimeter of the oxyanion hole forming a direct 2.8 Å H bond with Gly143 and a water-mediated contact with Cys145. Piperazine is located above the catalytic Cys145 side chain that was determined to be a deprotonated, negatively charged thiolate in the neutron structure of the LF M$^{pro}$. The P2 dichlorobenzene substituent occupies the largely hydrophobic S2 subsite.

An overlay of the MCULE-5948770040 complex with the LF joint neutron/X-ray crystal structure of M$^{pro53}$ shows that the P2-dichlorobenzene group moves into the hydrophobic S2 pocket altering the position of the Met49 side chain and pushing out the P2-helix (residues 46–50) by as much as ∼2.6 Å (Figure 3c). The Met49 terminal methyl is shifted ∼5.5 Å away and its Cα atom moves by ∼1.1 Å. Furthermore, the position of the P2-dichlorobenzene group is stabilized by the π−π stacking interactions with Gln189 and the imidazole side chain of catalytic His41 with the interatomic distances of ∼3.8 Å. The Gln189 side chain amide is recruited from 3 Å away from its position in the LF structure and the Cα atom shifts by almost 1 Å. Thus, the P2-dichlorobenzene is sandwiched between the side chains of these two residues. Interestingly, the binding of MCULE-5948770040 to the M$^{pro}$ active site cleft causes the His41 side chain to flip and a χ2 angle rotation to create a favorable geometry for π−π stacking with P2-dichlorobenzene in the complex structure. Such a change in the His41 conformation severs a conserved H bond between the His41 Nδ1 and the conserved catalytic water molecule (H$_2$O$_{cat}$) normally seen in M$^{pro}$ structures[54,55] but replaces it with a direct H bond to the main chain carbonyl of His164 and

results in recruitment of an additional water molecule from the bulk solvent to make an H bond with the His imidazole ring.

The computationally predicted binding pose of MCULE-5948770040 to the M$^{pro}$ active site is in good agreement with the experimentally determined orientation (Figure 3d). Only minor discrepancies in the piperazine linker and P2-dichlorobenzyl are present. The ligand's uracil group forms the same polar interactions in the docked pose as observed in the crystal structure. The piperazine linker is best modeled as a chair conformation in the crystal structure, while the docked geometry scored the highest with it adopting a twisted boat. P2-dichlorobenzene fits well into the S2 pocket as observed in the experimental configuration, despite a 180° rotation of the aromatic ring.

**M$^{pro}$ Interacts with MCULE-5948770040 through Conformational Changes within the Binding Site.** In order to understand how the molecule interacts with the M$^{pro}$ binding site, we carried out μs-time scale atomistic MD simulations (see Methods). For the LB-state simulations, we modeled the ligand present in both M$^{pro}$ protomers. Within the time scales of our simulations (μs time scales), we observed that the ligand stays bound to the primary binding site (for both chains A and B in the dimer). The protein also does not undergo any large conformational changes (as seen from the root-mean squared deviations/rmsd from the starting structure Figures S4−S6).

We used the root mean-squared fluctuations (RMSFs) from LF and LB M$^{pro}$ simulations (Figure 4A) to understand how the ligand impacts the conformational dynamics. For convenience, we considered each protomer individually (although the simulations were run with the ligand bound to both monomers in the active dimer form) and observed that several distinct regions across M$^{pro}$ exhibit altered fluctuations. In all of our replicas, the RMSFs in chain A of the dimer were slightly higher than those in chain B. Distinct regions within M$^{pro}$ respond to the ligand (rounded rectangles in Figure 4A);

these mostly consist of flexible loops surrounding the immediate vicinity of the binding pocket, corresponding to the sites (S1—orange rounded rectangle and S2—red rounded rectangle). Other regions surrounding the binding site (S3—green and S4—blue rounded rectangles) also exhibit stabilization upon ligand binding. However, it is notable that not all regions exhibit stabilization within each protomer (e.g., region S4, the protomer chain A exhibits similar fluctuations to the LF state). Interestingly, regions farther away from the binding site, including domain III of each protomer [R5 in Figure 4 (purple rounded rectangle)], exhibit lower fluctuations in the LB simulations.

To elucidate the collective motions that are influenced by ligand binding, we used anharmonic conformational analysis-enabled autoencoder (ANCA-AE; see Methods) to embed the conformational landscape spanned by the LF and LB simulations in a low-dimensional manifold (summarized in Figure 4b). Notably, our simulations can be embedded within a 10-dimensional manifold (Figure S7) which best explains conformational fluctuations undergone by the protein. Of these embeddings, the LF and LB simulations occupied distinct projections, with the LF simulations sampling diverse conformational states (as quantified by the RMSD to the LB states). The predominant conformational changes in the LB simulations were confined to the binding pocket spanning S1—S4 and R5, shown in Figure 4c,d in protomer A, while we did not observe significant motions with respect to protomer B (shown in inset). The fluctuations observed were mostly a consequence of reorienting the ligand within protomer A from its primary interaction site (P1 and P2) to cover the complementary binding site of (P2 and P4; orange rounded rectangle in Figure 4c, labeled I → Holo$_B$). Notably, the P1-uracil forms new interactions with the S4 region (residues), while the P2-dichlorobenzene stays bound within the hydrophobic pocket. Although we do not observe such a ligand placement within the crystal structure, these ligand motions are prevalent across multiple replicas of our simulations and can form stable interactions between the uracil and protein side chains. Corresponding to these changes, motions in domain III of the protein (purple cartoon insets in Figure 4c,d) are also suppressed, showing that this region may be stabilized upon ligand binding. We also examined the hydrogen-bonding patterns between the ligand and protein from the LB simulations (Figure S8) and found that the hydrogen bonds between the ligand and the protein in protomer chain B are more stable than the hydrogen bonds in chain A.

## ■ DISCUSSION

Our study demonstrates a 2.9 $\mu$M potent M$^{pro}$ inhibitor discovered through HTVS. After computationally screening 6 million molecules, we located 100 promising compounds based on a consensus score across five different M$^{pro}$ crystal structures. Through X-ray crystallographic studies, we observed that the compound MCULE-5948770040 forms stable interactions within a hydrophobic pocket (S2) formed by the P2-dichlorobenzene group along with the P1-uracil group occupying the S1 site. Assay results also indicate that this molecule is a $\mu$M noncovalent inhibitor of this enzyme and can act as a competitive inhibitor. Our $\mu$s-time scale simulations indicate fairly stable interactions between the protein and the ligand, which suggest that several regions of M$^{pro}$—both in the vicinity of the binding site and distal from it—are impacted upon binding. This alters the conformational

states accessed by the protein as quantified by our ANCA-AE approach.

The combination of experimental validation with computational tools is essential to the rapid development of an inhibitor for M$^{pro}$. Without the ability to quickly obtain a compound and experimentally validate it, computational results alone will not be a solution. Virtual screens against M$^{pro}$ have ranged from small libraries of existing approved pharmaceuticals to natural product libraries and billion-scale combinatorial libraries.[18,56] Studies such as ref 57[57] use drug repurposing databases that are smaller (<50k) but have potential for faster lead to drug time. In this study, we aimed to balance library size and feasibility of validation; hence, we opted for a 6M in-stock chemical library from Mcule. While ref 58[58] used an approximately 120 million compound library to obtain hits for the D4 dopamine receptor, we were able to use our approximately 6 million compound library without any advanced filtering post consensus scoring to locate a $\mu$M hit for M$^{pro}$.

Although a number of sub-$\mu$M and some nanomolar inhibitors are now available,[59−63] our paper focused on the discovery of novel molecules from in-stock molecules that could potentially inhibit M$^{pro}$ activity. Furthermore, several molecules from the COVID-moonshot project[62] also possess similar scaffolds (like the piperazine linker or uracil) to MCULE-5948770040, which provides an indirect validation of how these fragments may be important for discovering additional inhibitors based on this molecule. Finally, we note that we have not tested this molecule for antiviral activity, which we plan to do as part of future work.

The collective conformational motions elucidated using ANCA-AE suggest an intrinsic asymmetry in how the ligand interacts with the two protomers. Our simulations point to a mechanism of complementary interactions and interdomain motions, whereby the ligand stabilizes the conformations of the loops around the binding site as well as a loop within domain III that is considerably far away from either binding sites. In order to elucidate if the ligand binding affects either protomer separately, we also carried out simulations where the ligand was bound to only one of the two protomer units. Our analysis (Supporting Information text and Figures S10 and S11) of these simulations further indicate that the fluctuations in domain III of the protein are only affected from the LB chain. Taken together, our simulations suggest that the primary mechanism by which MCULE-5948770040 binds to and interacts with M$^{pro}$ is by stabilizing the loops in and around the binding site. The binding of the ligand is asymmetric in the protomers; while it is stable in one of the protomers, it undergoes a slight conformational change (albeit stable) within the binding pocket while still maintaining the strong hydrophobic interactions within P2. Furthermore, our analysis indicates that the hydrogen-bonding patterns are different for the two chains, which also lends support to the idea that the collective motions as induced by the LB states may indeed be different.

Computational protocols for HTVS vary widely across existing SARS-CoV-2 M$^{pro}$ screens, from accurate but expensive simulations for MMGBSA/PBSA scoring to standard docking. Gorgulla et al.[18] screened Enamine Real, a billion-scale combinatorial product library, against various SARS-CoV-2 targets using QuickVina W—a slightly less accurate but computationally efficient flavor of AutoDock Vina.[64] Acharya et al.[13] also screened Enamine Real with Autodock-GPU.

Comparing one of the few other studies which involve assay and crystallographic experimental studies for $M^{pro}$ lead generation,[65] our work relies on a single computational workflow rather than community lead sourcing (where the methods of each contributor are not restricted). Rather than taking community input for prioritizing experimental leads, our work features an HTVS protocol based on ensemble docking with consensus scoring. Using the web portal to access hits from,[65] we compared the difference between leads from domain experts with the library we used for screening and found that both groups arrived at structurally similar hits independently, with the same P1-Liner-P2 topology and interaction with $M^{pro}$ S1 and S2 sites (Figure S12). Between the two groups, our best compounds share piperazine and the uracil groups.

The compound MCULE-5948770040 forms stable interactions with both the protomers as we observed from our X-ray crystallography and MD simulations. Our simulations provide insights into how the conformational fluctuations of the protein are altered in response to the ligand binding to the primary site. In fact, we observed that the fluctuations in region R5, which is over 20 Å away from the primary binding site in $M^{pro}$, are affected by the ligand binding. Furthermore, the ligand's interaction with $M^{pro}$ alters the conformational states accessed by the enzyme, notably along the substrate-binding loops. Compared to other ligands that have been structurally characterized (as well as the substrate peptide), MCULE-5948770040 is much smaller and interacts stably with both the S1 and S2 sites within $M^{pro}$. Thus, a design strategy that targets the S1 and S2 sites and mimics important features of the peptide side chains is sufficient for identifying inhibitors of $M^{pro}$. We have early indications that molecules structurally similar to MCULE-5948770040 also demonstrate inhibition, and efforts are currently underway to identify promising candidates, which we expect to publish shortly.

## ■ DATA AND SOFTWARE AVAILABILITY

The computational data produced in this study are freely available for academic use. The 2D version of the data set, containing the docking scores and 2D molecular structures, is hosted publicly by a third party provider, FigShare: https://doi.org/10.6084/m9.figshare.14745234. The rest of the data, including all the 3D resulting structures from docking, is hosted by Argonne's Leadership Computing Center and accessible via a Globus endpoint with documentation hosted by GitHub: https://doi.org/10.26311/BFKY-EX6P. The authors are confident that the data will be persistent across FigShare, GitHub, ALCF, and Globus. Experimental data discussed in this paper are shared in the Supporting Information. The room-temperature crystal structure determined is deposited in the Protein Data Bank (PDB ID: 7LTJ).

The OpenEye Scientific software used for docking (FRED) is available via an academic license for users. The workflow outlined in this paper, along with specific hyperparameters for the docking protocol, is available here: https://github.com/aclyde11/Model-generation. The MD simulations in this study were carried out using the OpenMM toolkit (https://openmm.org). Simulations were analyzed using the pyANCA toolkit (https://github.com/ChakraLabPitt/pyanca), and the autoencoder was implemented using PyTorch (http://pytorch.org); the hyperparameter settings as well as specifics of how the models were trained are described in the Methods section.

Python notebooks for running pyANCA are available here: https://csb.pitt.edu/anca/index.html.

## ■ ASSOCIATED CONTENT

### ⓢ Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jcim.1c00851.

> Detailed description of the experimental procedures for our HTVS, performance characterization of the computational models, MD simulations, and implementation of ANCA-AE; resource utilization for HTVS; quantitative high-throughput MS-based end point assay; protein crystal of $M^{pro}$; highlight root-mean-squared deviations of $M^{pro}$ simulations; characteristics of ANCA-AE training and hyperparameter settings; hydrogen-bond analysis across the two $M^{pro}$ protomers; ANCA-AE analysis of LB states of $M^{pro}$; long-tailed atomistic fluctuations within $M^{pro}$; similarity of molecule MCULE-5948770040 with other fragments from the $M^{pro}$ crystal structures; and other molecules tested in our assays and crystallographic parameters (PDF)

## ■ AUTHOR INFORMATION

### Corresponding Authors

**Austin Clyde** − *Data Science and Learning Division, Argonne National Laboratory, Lemont, Illinois 60439, United States; Department of Computer Science, University of Chicago, Chicago, Illinois 60615, United States; National Virtual Biotechnology Laboratory, Washington, District of Columbia 20585, United States*; Email: aclyde@anl.gov

**Shantenu Jha** − *Department of Electrical and Computer Engineering, Rutgers University, Piscataway, New Jersey 08854, United States; Computational Science Initiative, Brookhaven National Laboratory, Upton, New York 11973, United States; National Virtual Biotechnology Laboratory, Washington, District of Columbia 20585, United States*; Email: shantenu.jha@rutgers.edu

**Andrey Kovalevsky** − *Second Target Station, Oak Ridge National Laboratory, Oak Ridge, Tennessee 37831, United States; National Virtual Biotechnology Laboratory, Washington, District of Columbia 20585, United States;* ⓞ orcid.org/0000-0003-4459-9142; Email: kovalevskyay@ornl.gov

**Arvind Ramanathan** − *Data Science and Learning Division, Argonne National Laboratory, Lemont, Illinois 60439, United States; National Virtual Biotechnology Laboratory, Washington, District of Columbia 20585, United States;* ⓞ orcid.org/0000-0002-1622-5488; Email: ramanathana@anl.gov

**Martha S. Head** − *Joint Institute for Biological Sciences, Oak Ridge National Laboratory, Oak Ridge, Tennessee 37831, United States; National Virtual Biotechnology Laboratory, Washington, District of Columbia 20585, United States;* Email: headms@ornl.gov

**Rick Stevens** − *Department of Computer Science, University of Chicago, Chicago, Illinois 60615, United States; Computing Environment and Life Sciences Directorate, Argonne National Laboratory, Lemont, Illinois 60439, United States; National Virtual Biotechnology Laboratory, Washington, District of Columbia 20585, United States;* Email: stevens@anl.gov

## Authors

**Stephanie Galanie** − *Biosciences Division, Oak Ridge National Laboratory, Oak Ridge, Tennessee 37831, United States; National Virtual Biotechnology Laboratory, Washington, District of Columbia 20585, United States*

**Daniel W. Kneller** − *Neutron Scattering Division, Oak Ridge National Laboratory, Oak Ridge, Tennessee 37831, United States; National Virtual Biotechnology Laboratory, Washington, District of Columbia 20585, United States;* orcid.org/0000-0002-5416-5789

**Heng Ma** − *Data Science and Learning Division, Argonne National Laboratory, Lemont, Illinois 60439, United States; National Virtual Biotechnology Laboratory, Washington, District of Columbia 20585, United States*

**Yadu Babuji** − *Department of Computer Science, University of Chicago, Chicago, Illinois 60615, United States; National Virtual Biotechnology Laboratory, Washington, District of Columbia 20585, United States*

**Ben Blaiszik** − *Data Science and Learning Division, Argonne National Laboratory, Lemont, Illinois 60439, United States; National Virtual Biotechnology Laboratory, Washington, District of Columbia 20585, United States*

**Alexander Brace** − *Data Science and Learning Division, Argonne National Laboratory, Lemont, Illinois 60439, United States; Department of Computer Science, University of Chicago, Chicago, Illinois 60615, United States; National Virtual Biotechnology Laboratory, Washington, District of Columbia 20585, United States*

**Thomas Brettin** − *Computing Environment and Life Sciences Directorate, Argonne National Laboratory, Lemont, Illinois 60439, United States; National Virtual Biotechnology Laboratory, Washington, District of Columbia 20585, United States*

**Kyle Chard** − *Department of Computer Science, University of Chicago, Chicago, Illinois 60615, United States; National Virtual Biotechnology Laboratory, Washington, District of Columbia 20585, United States*

**Ryan Chard** − *Data Science and Learning Division, Argonne National Laboratory, Lemont, Illinois 60439, United States; Department of Computer Science, University of Chicago, Chicago, Illinois 60615, United States; National Virtual Biotechnology Laboratory, Washington, District of Columbia 20585, United States*

**Leighton Coates** − *Neutron Scattering Division, Oak Ridge National Laboratory, Oak Ridge, Tennessee 37831, United States; National Virtual Biotechnology Laboratory, Washington, District of Columbia 20585, United States;* orcid.org/0000-0003-2342-049X

**Ian Foster** − *Data Science and Learning Division, Argonne National Laboratory, Lemont, Illinois 60439, United States; Department of Computer Science, University of Chicago, Chicago, Illinois 60615, United States; National Virtual Biotechnology Laboratory, Washington, District of Columbia 20585, United States*

**Darin Hauner** − *Computational Biology Group, Biological Science Division, Pacific Northwest National Laboratory, Richland, Washington 99352, United States; National Virtual Biotechnology Laboratory, Washington, District of Columbia 20585, United States*

**Vilmos Kertesz** − *Biosciences Division, Oak Ridge National Laboratory, Oak Ridge, Tennessee 37831, United States; National Virtual Biotechnology Laboratory, Washington, District of Columbia 20585, United States;* orcid.org/0000-0003-0186-5797

**Neeraj Kumar** − *Computational Biology Group, Biological Science Division, Pacific Northwest National Laboratory, Richland, Washington 99352, United States; National Virtual Biotechnology Laboratory, Washington, District of Columbia 20585, United States;* orcid.org/0000-0001-6713-2129

**Hyungro Lee** − *Department of Electrical and Computer Engineering, Rutgers University, Piscataway, New Jersey 08854, United States; National Virtual Biotechnology Laboratory, Washington, District of Columbia 20585, United States*

**Zhuozhao Li** − *Data Science and Learning Division, Argonne National Laboratory, Lemont, Illinois 60439, United States; Department of Computer Science, University of Chicago, Chicago, Illinois 60615, United States; National Virtual Biotechnology Laboratory, Washington, District of Columbia 20585, United States*

**Andre Merzky** − *Department of Electrical and Computer Engineering, Rutgers University, Piscataway, New Jersey 08854, United States; National Virtual Biotechnology Laboratory, Washington, District of Columbia 20585, United States*

**Jurgen G. Schmidt** − *Bioscience Division, Los Alamos National Laboratory, Los Alamos, New Mexico 87545, United States; National Virtual Biotechnology Laboratory, Washington, District of Columbia 20585, United States*

**Li Tan** − *Department of Electrical and Computer Engineering, Rutgers University, Piscataway, New Jersey 08854, United States; National Virtual Biotechnology Laboratory, Washington, District of Columbia 20585, United States*

**Mikhail Titov** − *Department of Electrical and Computer Engineering, Rutgers University, Piscataway, New Jersey 08854, United States; National Virtual Biotechnology Laboratory, Washington, District of Columbia 20585, United States*

**Anda Trifan** − *University of Illinois at Urbana-Champaign, Champaign, Illinois 61820, United States; National Virtual Biotechnology Laboratory, Washington, District of Columbia 20585, United States*

**Matteo Turilli** − *Department of Electrical and Computer Engineering, Rutgers University, Piscataway, New Jersey 08854, United States; Computational Science Initiative, Brookhaven National Laboratory, Upton, New York 11973, United States; National Virtual Biotechnology Laboratory, Washington, District of Columbia 20585, United States*

**Hubertus Van Dam** − *Computational Science Initiative, Brookhaven National Laboratory, Upton, New York 11973, United States; National Virtual Biotechnology Laboratory, Washington, District of Columbia 20585, United States*

**Srinivas C. Chennubhotla** − *Department of Computational and Systems Biology, University of Pittsburgh, Pittsburgh, Pennsylvania 15260, United States; National Virtual Biotechnology Laboratory, Washington, District of Columbia 20585, United States*

Complete contact information is available at:
https://pubs.acs.org/10.1021/acs.jcim.1c00851

## Notes

The authors declare no competing financial interest.

## ■ ADDITIONAL NOTE

[a]https://covid19.who.int.

## ■ REFERENCES

(1) Knoll, M. D.; Wonodi, C. Oxford−AstraZeneca COVID-19 vaccine efficacy. *Lancet* **2021**, *397*, 72−74.

(2) Le, T. T.; Cramer, J. P.; Chen, R.; Mayhew, S. Evolution of the COVID-19 vaccine development landscape. *Nat. Rev. Drug Discovery* **2020**, *19*, 667−668.

(3) Cao, X. COVID-19: Immunopathology and its implications for therapy. *Nat. Rev. Immunol.* **2020**, *20*, 269−270.

(4) Wu, R.; Wang, L.; Kuo, H.-C. D.; Shannar, A.; Peter, R.; Chou, P. J.; Li, S.; Hudlikar, R.; Liu, X.; Liu, Z.; Poiani, G. J.; Amorosa, L.; Brunetti, L.; Kong, A.-N. An update on current therapeutic drugs treating COVID-19. *Curr. Pharmacol. Rep.* **2020**, *6*, 56−70.

(5) Yang, H.; Bartlam, M.; Rao, Z. Drug design targeting the main protease, the Achilles' heel of coronaviruses. *Curr. Pharm. Des.* **2006**, *12*, 4573−4590.

(6) Lubin, J. H.; et al. Evolution of the SARS-CoV-2 proteome in three dimensions (3D) during the first six months of the COVID-19 pandemic, **2020**. bioRxiv:10.1101/2020.12.01.406637.

(7) Rosas-Lemus, M.; Minasov, G.; Shuvalova, L.; Inniss, N. L.; Kiryukhina, O.; Brunzelle, J.; Satchell, K. J. F. High-resolution structures of the SARS-CoV-2 2'-O-methyltransferase reveal strategies for structure-based inhibitor design. *Sci. Signaling* **2020**, *13*, No. eabe1202.

(8) Shyr, Z. A.; Gorshkov, K.; Chen, C. Z.; Zheng, W. Drug discovery strategies for SARS-CoV-2. *J. Pharmacol. Exp. Ther.* **2020**, *375*, 127−138.

(9) Hegyi, A.; Ziebuhr, J. Conservation of substrate specificities among coronavirus main proteases. *J. Gen. Virol.* **2002**, *83*, 595−599.

(10) Pillaiyar, T.; Manickam, M.; Namasivayam, V.; Hayashi, Y.; Jung, S.-H. An overview of severe acute respiratory syndrome−coronavirus (SARS-CoV) 3CL protease inhibitors: peptidomimetics and small molecule chemotherapy. *J. Med. Chem.* **2016**, *59*, 6595−6628.

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

(12) Zhang, S.; Kumar, K.; Jiang, X.; Wallqvist, A.; Reifman, J. DOVIS: An implementation for high-throughput virtual screening using AutoDock. *BMC Bioinf.* **2008**, *9*, 126.

(13) Acharya, A.; Agarwal, R.; Baker, M. B.; Baudry, J.; Bhowmik, D.; Boehm, S.; Byler, K. G.; Chen, S. Y.; Coates, L.; Cooper, C. J.; Demerdash, O.; Daidone, I.; Eblen, J. D.; Ellingson, S.; Forli, S.; Glaser, J.; Gumbart, J. C.; Gunnels, J.; Hernandez, O.; Irle, S.; Kneller, D. W.; Kovalevsky, A.; Larkin, J.; Lawrence, T. J.; LeGrand, S.; Liu, S.-H.; Mitchell, J. C.; Park, G.; Parks, J. M.; Pavlova, A.; Petridis, L.; Poole, D.; Pouchard, L.; Ramanathan, A.; Rogers, D. M.; Santos-Martins, D.; Scheinberg, A.; Sedova, A.; Shen, Y.; Smith, J. C.; Smith, M. D.; Soto, C.; Tsaris, A.; Thavappiragasam, M.; Tillack, A. F.; Vermaas, J. V.; Vuong, V. Q.; Yin, J.; Yoo, S.; Zahran, M.; Zanetti-Polzi, L. Supercomputer-based ensemble docking drug discovery pipeline with application to COVID-19. *J. Chem. Inf. Model.* **2020**, *60*, 5832−5852.

(14) Trott, O.; Olson, A. J. AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J. Comput. Chem.* **2010**, *31*, 455−461.

(15) Irwin, J. J.; Shoichet, B. K. ZINC - A free database of commercially available compounds for virtual screening. *J. Chem. Inf. Model.* **2005**, *45*, 177−182.

(16) Kiss, R.; Sandor, M.; Szalai, F. A. http://Mcule.com: a public web service for drug discovery. *J. Cheminf.* **2012**, *4*, P17.

(17) Achdout, H.; Aimon, A.; Bar-David, E.; Barr, H.; Ben-Shmuel, A.; Bennett, J.; Bobby, M. L.; Brun, J.; Sarma, B.; Calmiano, M.; Carbery, A. COVID Moonshot: open science discovery of SARS-CoV-2 main protease inhibitors by combining crowdsourcing, high-throughput experiments, computational simulations, and machine learning, **2020**. bioRxiv:10.1101/2020.10.29.339317.

(18) Gorgulla, C.; Padmanabha Das, K. M.; Leigh, K. E.; Cespugli, M.; Fischer, P. D.; Wang, Z.-F.; Tesseyre, G.; Pandita, S.; Shnapir, A.; Calderaio, A.; Gechev, M.; Rose, A.; Lewis, N.; Hutcheson, C.; Yaffe, E.; Luxenburg, R.; Herce, H. D.; Durmaz, V.; Halazonetis, T. D.; Fackeldey, K.; Patten, J. J.; Chuprina, A.; Dziuba, I.; Plekhova, A.; Moroz, Y.; Radchenko, D.; Tarkhanova, O.; Yavnyuk, I.; Gruber, C.; Yust, R.; Payne, D.; Näär, A. M.; Namchuk, M. N.; Davey, R. A.; Wagner, G.; Kinney, J.; Arthanari, H. A multi-pronged approach targeting SARS-CoV-2 proteins using ultra-large virtual screening. *iScience* **2021**, *24*, 102021.

(19) McGann, M. FRED and HYBRID docking performance on standardized datasets. *J. Comput.-Aided Mol. Des.* **2012**, *26*, 897−906.

(20) Hawkins, P. C. D.; Skillman, A. G.; Warren, G. L.; Ellingson, B. A.; Stahl, M. T. Conformer generation with OMEGA: algorithm and validation using high quality structures from the protein databank and Cambridge structural database. *J. Chem. Inf. Model.* **2010**, *50*, 572−584.

(21) Merzky, A.; Turilli, M.; Titov, M.; Al-Saadi, A.; Jha, S. Design and performance characterization of RADICAL-Pilot on leadership-class platforms. *EEE Trans. Parallel Distrib. Syst.* **2022**, *33*, 818−829.

(22) Kneller, D. W.; Phillips, G.; Kovalevsky, A.; Coates, L. Room-temperature neutron and X-ray data collection of 3CL Mpro from SARS-CoV-2. *Acta Crystallogr., Sect. F: Struct. Biol. Commun.* **2020**, *76*, 483−487.

(23) Kuo, C.-J.; Chi, Y.-H.; Hsu, J. T.-A.; Liang, P.-H. Characterization of SARS main protease and inhibitor assay using a fluorogenic substrate. *Biochem. Biophys. Res. Commun.* **2004**, *318*, 862−867.

(24) Lodwig, S. N.; Unkefer, C. J. Stereoselective synthesis of stable isotope-labeled L-α-amino acids: Electrophilic amination of Oppolzer's acyl sultams in the synthesis of L-[$^{15}$N]alanine, L-[$^{15}$N]valine, L-[$^{15}$N]leucine, L-[$^{15}$N]phenylalanine and L-[1-$^{13}$C, $^{15}$N]valine. *J. Labelled Compd. Radiopharm.* **1996**, *38*, 239−248.

(25) Van Berkel, G. J.; Kertesz, V. An open port sampling interface for liquid introduction atmospheric pressure ionization mass spectrometry. *Rapid Commun. Mass Spectrom.* **2015**, *29*, 1749−1756.

(26) Kneller, D. W.; Phillips, G.; O'Neill, H. M.; Tan, K.; Joachimiak, A.; Coates, L.; Kovalevsky, A. Room-temperature X-ray crystallography reveals the oxidation and reactivity of cysteine residues in SARS-CoV-2 3CL Mpro: Insights into enzyme mechanism and drug design. *IUCrJ* 2020, 7, 1028−1035.

(27) Luft, J. R.; Collins, R. J.; Fehrman, N. A.; Lauricella, A. M.; Veatch, C. K.; DeTitta, G. T. A deliberate approach to screening for initial crystallization conditions of biological macromolecules. *J. Struct. Biol.* 2003, 142, 170−179.

(28) Winn, M. D.; Ballard, C. C.; Cowtan, K. D.; Dodson, E. J.; Emsley, P.; Evans, P. R.; Keegan, R. M.; Krissinel, E. B.; Leslie, A. G. W.; McCoy, A.; McNicholas, S. J.; Murshudov, G. N.; Pannu, N. S.; Potterton, E. A.; Powell, H. R.; Read, R. J.; Vagin, A.; Wilson, K. S. Overview of theCCP4 suite and current developments. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* 2011, 67, 235−242.

(29) Adams, P. D.; Afonine, P. V.; Bunkóczi, G.; Chen, V. B.; Davis, I. W.; Echols, N.; Headd, J. J.; Hung, L.-W.; Kapral, G. J.; Grosse-Kunstleve, R. W.; McCoy, A. J.; Moriarty, N. W.; Oeffner, R.; Read, R. J.; Richardson, D. C.; Richardson, J. S.; Terwilliger, T. C.; Zwart, P. H. PHENIX: A comprehensive Python-based system for macro-molecular structure solution. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* 2010, 66, 213−221.

(30) Emsley, P.; Cowtan, K. Coot: Model-building tools for molecular graphics. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* 2004, 60, 2126−2132.

(31) Chen, V. B.; Arendall, W. B.; Headd, J. J.; Keedy, D. A.; Immormino, R. M.; Kapral, G. J.; Murray, L. W.; Richardson, J. S.; Richardson, D. C. MolProbity: All-atom structure validation for macromolecular crystallography. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* 2010, 66, 12−21.

(32) Götz, A. W.; Williamson, M. J.; Xu, D.; Poole, D.; Le Grand, S.; Walker, R. C. Routine microsecond molecular dynamics simulations with AMBER on GPUs. 1. Generalized Born. *J. Chem. Theory Comput.* 2012, 8, 1542−1555.

(33) Maier, J. A.; Martinez, C.; Kasavajhala, K.; Wickstrom, L.; Hauser, K. E.; Simmerling, C. ff14SB: Improving the accuracy of protein side chain and backbone parameters from ff99SB. *J. Chem. Theory Comput.* 2015, 11, 3696−3713.

(34) Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. Development and testing of a general amber force field. *J. Comput. Chem.* 2004, 25, 1157−1174.

(35) Valiev, M.; Bylaska, E. J.; Govind, N.; Kowalski, K.; Straatsma, T. P.; Van Dam, H. J. J.; Wang, D.; Nieplocha, J.; Apra, E.; Windus, T. L.; de Jong, W. A. NWChem: A comprehensive and scalable open-source solution for large scale molecular simulations. *Comput. Phys. Commun.* 2010, 181, 1477−1489.

(36) Kozlowski, P. M. Quantum chemical modeling of CoC bond activation in B12-dependent enzymes. *Curr. Opin. Chem. Biol.* 2001, 5, 736−743.

(37) Ramanathan, A.; Parvatikar, A.; Chennubhotla, S. C.; Mei, Y.; Sinha, S. C. Transient unfolding and long-range interactions in viral BCL2 M11 enable binding to the BECN1 BH3 domain. *Biomolecules* 2020, 10, 1308.

(38) Eastman, P.; Swails, J.; Chodera, J. D.; McGibbon, R. T.; Zhao, Y.; Beauchamp, K. A.; Wang, L.-P.; Simmonett, A. C.; Harrigan, M. P.; Stern, C. D.; Wiewiora, R. P.; Brooks, B. R.; Pande, V. S. OpenMM 7: Rapid development of high performance algorithms for molecular dynamics. *PLoS Comput. Biol.* 2017, 13, No. e1005659.

(39) Ramanathan, A.; Ma, H.; Parvatikar, A.; Chennubhotla, S. C. Artificial intelligence techniques for integrative structural biology of intrinsically disordered proteins. *Curr. Opin. Struct. Biol.* 2021, 66, 216−224.

(40) Parvatikar, A.; Vacaliuc, G. S.; Ramanathan, A.; Chennubhotla, S. C. ANCA: anharmonic conformational analysis of biomolecular simulationsß. *Biophys. J.* 2018, 114, 2040−2043.

(41) Mardt, A.; Pasquali, L.; Wu, H.; Noé, F. VAMPnets for deep learning of molecular kinetics. *Nat. Commun.* 2018, 9, 4443.

(42) Ramanathan, A.; Savol, A. J.; Langmead, C. J.; Agarwal, P. K.; Chennubhotla, C. S. Discovering conformational sub-states relevant to protein function. *PLoS One* 2011, 6, No. e15827.

(43) Xue, X.; Yu, H.; Yang, H.; Xue, F.; Wu, Z.; Shen, W.; Li, J.; Zhou, Z.; Ding, Y.; Zhao, Q.; Zhang, X. C.; Liao, M.; Bartlam, M.; Rao, Z. Structures of two coronavirus main proteases: implications for substrate binding and antiviral drug design. *J. Virol.* 2008, 82, 2515−2527.

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

(45) Mesecar, A. *Structure of COVID-19 Main Protease Bound to Potent Broad-Spectrum Non-covalent Inhibitor X77*; RCSB Protein Data Bank, 2020.

(46) Qiao, J.; Zeng, R.; Wang, Y.; Li, Y.; Yao, R.; Zhou, Y.; Chen, P.; Lei, J.; Yang, S. *Crystal Structure of the SARS-CoV-2 Main Protease in Complex with Telaprevir*; RCSB Protein Data Bank, 2020.

(47) Drayman, N.; et al. Masitinib is a broad coronavirus 3CL inhibitor that blocks replication of SARS-CoV-2. *Science* 2021, 373, 931−936.

(48) Baell, J. B.; Holloway, G. A. New substructure filters for removal of pan assay interference compounds (PAINS) from screening libraries and for their exclusion in bioassays. *J. Med. Chem.* 2010, 53, 2719−2740.

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

(50) Sacco, M. D.; Ma, C.; Lagarias, P.; Gao, A.; Townsend, J. A.; Meng, X.; Dube, P.; Zhang, X.; Hu, Y.; Kitamura, N.; Hurst, B.; Tarbet, B.; Marty, M. T.; Kolocouris, A.; Xiang, Y.; Chen, Y.; Wang, J. Structure and inhibition of the SARS-CoV-2 main protease reveal strategy for developing dual inhibitors against Mpro and cathepsin L. *Sci. Adv.* 2020, 6, No. eabe0751.

(51) Hoffman, R. L.; et al. Discovery of ketone-based covalent inhibitors of coronavirus 3CL proteases for the potential therapeutic treatment of COVID-19. *J. Med. Chem.* 2020, 63, 12725−12747.

(52) Kneller, D. W.; Galanie, S.; Phillips, G.; O'Neill, H. M.; Coates, L.; Kovalevsky, A. Malleability of the SARS-CoV-2 3CL Mpro active-site cavity facilitates binding of clinical antivirals. *Structure* 2020, 28, 1313−1320.

(53) Kneller, D. W.; Phillips, G.; Weiss, K. L.; Pant, S.; Zhang, Q.; O'Neill, H. M.; Coates, L.; Kovalevsky, A. Unusual zwitterionic catalytic site of SARS-CoV-2 main protease revealed by neutron crystallography. *J. Biol. Chem.* 2020, 295, 17365−17373.

(54) Vuong, W.; Khan, M. B.; Fischer, C.; Arutyunova, E.; Lamer, T.; Shields, J.; Saffran, H. A.; McKay, R. T.; van Belkum, M. J.; Joyce, M. A.; Young, H. S.; Tyrrell, D. L.; Vederas, J. C.; Lemieux, M. J. Feline coronavirus drug inhibits the main protease of SARS-CoV-2 and blocks virus replication. *Nat. Commun.* 2020, 11, 4282.

(55) Rathnayake, A. D.; Zheng, J.; Kim, Y.; Perera, K. D.; Mackin, S.; Meyerholz, D. K.; Kashipathy, M. M.; Battaile, K. P.; Lovell, S.; Perlman, S.; Groutas, W. C.; Chang, K. O. 3C-like protease inhibitors block coronavirus replication in vitro and improve survival in MERS-CoV-infected mice. *Sci. Transl. Med.* 2020, 12, No. eabc5332.

(56) Abo-Zeid, Y.; Ismail, N. S. M.; McLean, G. R.; Hamdy, N. M. A molecular docking study repurposes FDA approved iron oxide nanoparticles to treat and control COVID-19 infection. *Eur. J. Pharm. Sci.* 2020, 153, 105465.

(57) Marinho, E. M.; Batista de Andrade Neto, J.; Silva, J.; Rocha da Silva, C.; Cavalcanti, B. C.; Marinho, E. S.; Nobre Júnior, H. V. Virtual screening based on molecular docking of possible inhibitors of COVID-19 main protease. *Microb. Pathog.* 2020, 148, 104365.

(58) Lyu, J.; Wang, S.; Balius, T. E.; Singh, I.; Levit, A.; Moroz, Y. S.; O'Meara, M. J.; Che, T.; Algaa, E.; Tolmachova, K.; Tolmachev, A. A.;

Shoichet, B. K.; Roth, B. L.; Irwin, J. J. Ultra-large library docking for discovering new chemotypes. *Nature* **2019**, *566*, 224−229.

(59) Zhang, C.-H.; et al. Potent noncovalent inhibitors of the main protease of SARS-CoV-2 from molecular sculpting of the drug perampanel guided by free energy perturbation calculations. *ACS Cent. Sci.* **2021**, *7*, 467−475.

(60) Zhang, C.-H.; et al. Optimization of Triarylpyridinone Inhibitors of the Main Protease of SARS-CoV-2 to Low-Nanomolar Antiviral Potency. *ACS Med. Chem. Lett.* **2021**, *12*, 1325−1332.

(61) Deshmukh, M. G.; Ippolito, J. A.; Zhang, C.-H.; Stone, E. A.; Reilly, R. A.; Miller, S. J.; Jorgensen, W. L.; Anderson, K. S. Structure-guided design of a perampanel-derived pharmacophore targeting the SARS-CoV-2 main protease. *Structure* **2021**, *29*, 823−833.

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

(63) Rizzuti, B.; Ceballos-Laita, L.; Ortega-Alarcon, D.; Jimenez-Alesanco, A.; Vega, S.; Grande, F.; Conforti, F.; Abian, O.; Velazquez-Campoy, A. Sub-Micromolar Inhibition of SARS-CoV-2 3CLpro by Natural Compounds. *Pharmaceuticals* **2021**, *14*, 892.

(64) Hassan, N. M.; Alhossary, A. A.; Mu, Y.; Kwoh, C.-K. Protein-ligand blind docking using QuickVina-W with inter-process spatio-temporal integration. *Sci. Rep.* **2017**, *7*, 15451.

(65) Chodera, J.; Lee, A.; London, N.; Delft, F. v. Open Science Discovery of Oral Non-Covalent SARS-CoV-2 Main Protease Inhibitors. *ChemRxiv*, 2021. https://chemrxiv.org/engage/chemrxiv/article-details/61684497be107474529cb137.