Showcasing research from Professor Piquemal's laboratory, Department of Chemistry, Sorbonne Université, Paris, France and from Qubit Pharmaceuticals, Paris, France.

Computationally driven discovery of SARS-CoV-2 M^{pro} inhibitors: from design to experimental validation

We report a fast-track computationally driven discovery of new SARS-CoV-2 main protease (Mpro) inhibitors whose potency ranges from mM for the initial non-covalent ligands to sub-μM for the final covalent compound (IC50 = 830 ± 50 nM). The project extensively relied on high-resolution all-atom molecular dynamics simulations and absolute binding free energy calculations performed using the polarizable AMOEBA force field. While simulations extensively use high performance computing to strongly reduce the time-to-solution, they were systematically coupled to nuclear magnetic resonance experiments to drive synthesis and for in vitro characterization of compounds.

As featured in:

See Jean-Philip Piquemal, Davide Sabbadin *et al.*, *Chem. Sci.*, 2022, **13**, 3674.

ROYAL SOCIETY
OF **CHEMISTRY**

rsc.li/chemical-science

Chemical Science

EDGE ARTICLE

Computationally driven discovery of SARS-CoV-2 Mpro inhibitors: from design to experimental validation†‡

Léa El Khoury, [ID] §[a] Zhifeng Jing, [ID] §[a] Alberto Cuzzolin,[b] Alessandro Deplano,[c] Daniele Loco,[a] Boris Sattarov,[a] Florent Hédin, [ID] [a] Sebastian Wendeborn,[d] Chris Ho,[a] Dina El Ahdab, [ID] [n] Theo Jaffrelot Inizan,[n] Mattia Sturlese,[g] Alice Sosic,[e] Martina Volpiana,[e] Angela Lugato,[e] Marco Barone,[e] Barbara Gatto,[e] Maria Ludovica Macchia,[e] Massimo Bellanda,[f] Roberto Battistutta,[f] Cristiano Salata, [ID] [h] Ivan Kondratov,[i] Rustam Iminov,[i] Andrii Khairulin,[i] Yaroslav Mykhalonok,[i] Anton Pochepko,[i] Volodymyr Chashka-Ratushnyi,[i] Iaroslava Kos,[i] Stefano Moro, [ID] [g] Matthieu Montes,[j] Pengyu Ren,[k] Jay W. Ponder, [ID] [lm] Louis Lagardère,[n] Jean-Philip Piquemal [ID] *[no] and Davide Sabbadin*[a]

We report a fast-track computationally driven discovery of new SARS-CoV-2 main protease (Mpro) inhibitors whose potency ranges from mM for the initial non-covalent ligands to sub-μM for the final covalent compound ($IC_{50} = 830 \pm 50$ nM). The project extensively relied on high-resolution all-atom molecular dynamics simulations and absolute binding free energy calculations performed using the polarizable AMOEBA force field. The study is complemented by extensive adaptive sampling simulations that are used to rationalize the different ligand binding poses through the explicit reconstruction of the ligand–protein conformation space. Machine learning predictions are also performed to predict selected compound properties. While simulations extensively use high performance computing to strongly reduce the time-to-solution, they were systematically coupled to nuclear magnetic resonance experiments to drive synthesis and for *in vitro* characterization of compounds. Such a study highlights the power of *in silico* strategies that rely on structure-based approaches for drug design and allows the protein conformational multiplicity problem to be addressed. The proposed fluorinated tetrahydroquinolines open routes for further optimization of Mpro inhibitors towards low nM affinities.

1. Introduction

Since December 2019, the COVID-19 global pandemic has put the entire world on edge.[1,2] The disease is due to a coronavirus (CoV) called SARS-CoV-2 (severe acute respiratory syndrome, SARS) that has triggered the start of an unprecedented research effort.[3–5] While the vaccination strategy[6] has been particularly successful with the rise of mRNA techniques, additional programs have

[a]Qubit Pharmaceuticals, Incubateur Paris Biotech Santé, 24 Rue du Faubourg Saint Jacques, 75014 Paris, France. E-mail: davide@qubit-pharmaceuticals.com

[b]Chiesi Farmaceutici S.p.A, Nuovo Centro Ricerche, Largo Belloli 11a, 43122, Parma, Italy

[c]Pharmacelera, Torre R, 4a planta, Despatx A05, Parc Cientific de Barcelona, Baldiri Reixac 8, 08028 Barcelona, Spain

[d]University of Applied Sciences and Arts Northwestern Switzerland, School of LifeSciences, Hofackerstrasse 30, CH-4132 Muttenz, Switzerland

[e]Department of Pharmaceutical and Pharmacological Sciences, University of Padova, via Marzolo 5, 35131, Padova, Italy

[f]Department of Chemistry, University of Padova, via Marzolo 1, 35131, Padova, Italy

[g]Molecular Modeling Section, Department of Pharmaceutical and Pharmacological Sciences, University of Padua, via F. Marzolo 5, 35131, Padova, Italy

[h]Department of Molecular Medicine, University of Padua, via Gabelli 63, 35121, Padova, Italy

[i]Enamine Ltd, 78 Chervonotkats'ka Str., Kyiv 02094, Ukraine

[j]Laboratoire GBCM, EA7528, Conservatoire National des Arts et Métiers, Hesam Université, 2 Rue Conte, 75003 Paris, France

[k]University of Texas at Austin, Department of Biomedical Engineering, TX 78712, USA

[l]Department of Chemistry, Washington University in Saint Louis, MO 63130, USA

[m]Department of Biochemistry and Molecular Biophysics, Washington University School of Medicine, MO 63110, USA

[n]Sorbonne Université, Laboratoire de Chimie Théorique, UMR 7616 CNRS, 75005, Paris, France. E-mail: jean-philip.piquemal@sorbonne-universite.fr

[o]Institut Universitaire de France, 75005, Paris, France

† Qubit Pharmaceuticals and Sorbonne Université have submitted a preliminary patent application on the compounds reported in this study.

‡ Electronic supplementary information (ESI) available. See DOI: 10.1039/d1sc05892d

§ These authors contributed equally to this work.

been launched to obtain antivirals able to reduce the impact of COVID-19 on ill patients. Despite these efforts, few potential treatments are presently available with the exception of Paxlovid, a nirmatrelvir/ritonavir combo proposed by Pfizer.[7] Due to the persistence of the pandemic, it remains essential to propose new antiviral drugs. A possible strategy consists in designing small molecules to interact with one of the main proteins of the SARS-CoV-2 virus, thus blocking its activity. Among the potential targets, the main protease protein, denoted as M^{pro} or $3CL^{pro}$, is a primary choice[8] as it has no human homolog and it is well conserved among coronaviruses,[9] especially in terms of the structure of its active site, catalytic dyad, and dimer interface. Furthermore, M^{pro} is required to release viral proteins for particle assembly, and is thus essential to the virus replication cycle.

Developing a new drug targeting the viral M^{pro} is challenging as it requires extensive resources and the success rate is notoriously low.[10] Relying on *in silico* driven rational design could accelerate the process. In fact, it diminishes the cost by reducing the need for synthetic iterations while also providing an interpretation of the interactions occurring between the target protein and potential inhibitors.

It is important to note that theoretical modeling of M^{pro} is challenging as the protein exhibits high structural flexibility[11–13] leading to high conformational complexity. M^{pro} is also involved in a variety of complex protein–ligand–solvent interaction networks.[12,13] These challenges can be tackled using a high-resolution modeling approach[12,13] going beyond rigid docking procedures (see ref. 14 for a detailed discussion of the difficulties of docking approaches in predicting the native binding modes of small molecules within M^{pro}).

Many studies have been devoted to the design of new M^{pro} inhibitors[3,5,15–25] through joint computational and experimental approaches. In particular, a recent study by the Jorgensen group highlighted the usefulness of relative binding free energy (RBFE) computations as part of the drug design process.[26]

In this paper, we present a computationally driven discovery and binding mode rationalization of new SARS-CoV-2 M^{pro} inhibitors. In doing so, we build on our previous high-resolution M^{pro} molecular dynamics studies.[12,13] Here, we explore more deeply some specific subpockets of the substrate binding site of the protease using absolute binding free energy (ABFE) calculations and adaptive sampling grounded on extensive molecular dynamics simulations with high-resolution polarizable force fields (PFFs). Using the GPU-accelerated module[27] (GPU = Graphics Processing Unit) of the Tinker-HP molecular dynamics package[28] coupled to the AMOEBA PFF,[29–32] it has been shown that simulations can reach the required level of accuracy and μs timescales needed to explore the structural rearrangement and interactions profile of this flexible protein.[12,13] More precisely, the modeling of M^{pro} necessitates the ability to evaluate at high resolution various types of key interactions including hydrogen bonds, salt bridges, π–π stacking, and specific solvation effects. Long timescales are required to achieve sufficient sampling. This is now possible by using the large number of graphics processing units (GPUs) that are presently available on supercomputers and high-performance cloud computing platforms. In this study, we combine our computationally driven strategy, using absolute

binding free energy computations[33–37] and unsupervised adaptive sampling,[12,13] with machine learning-assisted property predictions, while conducting extensive characterization experiments including nuclear magnetic resonance (NMR), mass spectrometry (MS), and FRET-based assays to evaluate the activity of the newly designed compounds.

In the following, we introduce our design strategy, which led to non-covalent and covalent inhibitors of M^{pro} (ESI-Fig. 1‡). Then, we describe how an interplay between experiments and molecular simulations allowed the discovery of a final compound (QUB-00006-Int-07) with a high affinity to the protease ($IC_{50} = 830 \pm 50$ nM).

2. Computational details

2.1. Systems preparation

The protease dimer structure (PDB code: 7L11) was used for all the MD simulations and it was prepared at physiological pH (pH = 7). This structure has a higher resolution (1.80 Å) than the PDB structure (PDB code: 6LU7) used in our previous work[12] (resolution of 2.16 Å). Both structures are of the holo state in complex with covalent inhibitors, and the rotamers of the key residues at the catalytic site (Cys145, His41, His162, His163, and His172) are virtually identical. The protonation states of His residues were assigned based on previous work,[38] where His41 and His80 are protonated at the delta carbon atom and all other His residues are epsilon-protonated, which is favorable for substrate binding.[38] This is different from our previous work where His64 and His80 are protonated at the delta carbon atom and all other histidines are epsilon-protonated.[12] All water molecules were retained except for those that might collide with the ligands.

2.2. Simulation protocols

All-atom simulations were performed using Qubit Pharmaceuticals' Atlas platform which enables the use of any type of High-Performance Computing (HPC) system including cloud supercomputing infrastructures. Among its possibilities, Atlas has the ability to efficiently handle polarizable force field molecular dynamics simulations using a custom version of the multi-GPU module[27] of the Tinker-HP molecular dynamics package,[28,39] to perform docking runs using either Autodock-Vina[40] or Autodock-GPU,[41] and to enable machine learning predictions of molecular properties.

2.2.1. Molecular dynamics simulations. All Tinker-HP MD simulations (for a total of several μs) were performed in mixed precision to benefit from a strong acceleration of simulations using GPUs.[27] The AMOEBA polarizable force field[29–32] was used to describe the full systems including the protein, ions and water. Several utilities (TinkerTools) from Tinker 8 (ref. 42) were used. Periodic boundary conditions were applied within the framework of smooth particle mesh Ewald summation[43,44] with a grid of dimensions 120 × 120 × 120 using a cubic box with side lengths of 97 Å. The Ewald cutoff was set to 7 Å, and the van der Waals cutoff was 12 Å. Langevin molecular dynamics simulations were performed using the recently introduced BAOAB-RESPA1 integrator (10 fs outer timestep),[45]

a preconditioned conjugate gradient polarization solver (with a 10^{-5} convergence threshold) to solve polarization at each time step,[46] hydrogen-mass repartitioning (HMR) and random initial velocities. Absolute binding free energy simulations following a protocol described in the next section were performed as well as adaptive sampling runs that are also described further in the text. Absolute binding free energy computations were both performed on the HPE Jean Zay Supercomputer (IDRIS, GENCI, France) and on Amazon Web Services (AWS). All adaptive sampling computations were performed using AWS. Simulations on AWS used both p3.2x (NVIDIA V100 GPU cards) and p4d.24xlarge (NVIDIA A100 GPU cards) instances whereas computations on the Jean Zay supercomputer were powered by V100 cards.

2.2.2. Molecular docking protocol. The protonation states of the ligands were calculated at a neutral pH and the hydrogen atoms were added using Chimera. Next, we docked the ligands QUB-00006-Int-01(R) and QUB-00006-Int-01(S) into the M^{pro} dimer structure using Autodock Vina 1.1.2.[40] AutoDock Vina requires the pdbqt format for the input files of the receptor and the ligand. Therefore, using the scripts 'prepare_receptor4.py' (v 1.13) and 'prepare_ligand4.py' (v 1.10) provided by Autodock Tools,[47] we generated pdbqt files corresponding to the receptor and the ligands, respectively. We set the exhaustiveness search to 100 and the num_mode option to 50.

Since molecular docking could suggest reasonable potential binding modes, but does not always rank the most likely binding mode as the best docked pose,[14,48] we visually inspected the generated docked poses and chose an ensemble of binding poses with different binding orientations that we used to run MD and ABFE calculations in order to explore the binding mode of QUB-00006-Int-01, as described in the Results and discussion section.

2.2.3. Equilibration. A detailed description of the equilibration protocol used for MD simulations can be found in the ESI.‡

2.2.4. High-resolution adaptive sampling simulations. Starting from several binding poses as described above we ran adaptive sampling simulations using the AMOEBA force field[29–32] in order to explore their stability and more generally to explore the conformational space of the ligands in the pocket of the M^{pro}. Because of the flexibility of the pocket and the role it may play in the exploration of the potential binding modes of the ligand, we chose to keep the whole system (ligand + protein) flexible during this sampling phase. The restart strategy (similar to the one introduced in ref. 12) was the following: first, all the previously generated conformations of the protein were loaded and aligned with MDTraj,[49] then PCAs of the conformations of the ligand were computed using Scikitlearn[50] and these frames were projected on the first four PCAs. Finally, the same scheme as the one described in ref. 12 was used to generate new starting points, favoring points that were less explored during the previous phases. In practice, a first set of 5 simulations of 10 nanoseconds were performed using different random seeds, and then 4 iterations of 10 times 10 nanoseconds were generated using the adaptive sampling protocol described above, for a total of 450 nanoseconds.

2.2.5. Absolute binding free energy calculations. In order to benefit from the high-accuracy evaluation of free energies using the AMOEBA force field,[33–37] we used the same clustering algorithms as described above to analyze the adaptive molecular dynamics simulations. The largest clusters were used for absolute free energy calculations. The double-decoupling protocol and the Bennett acceptance ratio (BAR)[51] method were used to calculate the standard binding free energy for each binding pose.[33,37] There were 27 or 26 thermodynamic states for the decoupling in the complex phase or the aqueous phase. A distance restraint between two groups of atoms in the ligand and in the protein binding pocket was applied when decoupling the ligand in the complex to accelerate the convergence when the ligand is fully decoupled, and the restraint was removed at an additional step at the full interaction state. A harmonic restraint with a force constant of 15.0 kcal mol^{-1} Å^{-2} and radius of 2.0 Å was used. An analytical correction was added to the binding free energy to account for the standard state at 1.0 mol L^{-1} in the fully decoupled state. 10 ns simulations were performed for each thermodynamic state for the simulations of M^{pro} in complex with x0195, QUB-00006(S), QUB-00006(R), and QUB-00006-Int-07. For the simulations of M^{pro} in complex with QUB-00006-Int-01(R) and QUB-00006-Int-01(S), we ran each thermodynamic state for 20 ns. We used the BAOAB-RESPA1 integrator with a 10 fs time step and we calculated the electrostatic interactions using Ewald summation with a real space cutoff of 7 Å. van der Waals interactions were calculated using a cutoff of 12 Å with long-range correction.

2.3. Quantitative structure–property relationship (QSPR) modeling: predicting solubility using machine learning

Qubit Pharmaceuticals' Atlas internal machine learning-based QSPR module was used to predict the water solubility (log S, S measured in mol L^{-1}) and octanol/water partition coefficient (log P). To build a water solubility QSPR predictor, the AqSolDB dataset[52] was used as a training set. To predict octanol/water partition coefficients (log P), the dataset from EPA's OPERA[53] was used as a training set.

Selected datasets were preprocessed and standardized to some extent by authors of the corresponding publications. However, the need for additional processing was identified when doing exploratory data analysis. We discarded compounds with less than two carbon atoms and kept molecules with molecular weight between 50 and 750 daltons. Additional rules of fragment standardization developed at Qubit Pharmaceuticals were applied.

2.3.1. Similarity analysis. Tanimoto similarity[54] to the x0195 compound was calculated for each molecule using the MAACS fingerprint from the RDKit Open-Source Cheminformatics Software (https://www.rdkit.org). The Morgan circular fingerprint[55] with radius = 2 and nBits = 2048 from RDKit was also tested and the results (not shown) exhibit a similar ranking of the compounds.

3. Experimental protocol

3.1. Recombinant expression of SARS-CoV-2 M^{pro} in *E. coli*

The plasmid pGEX-6P-1 encoding SARS-CoV-2 M^{pro}[56] was a generous gift from Prof. Rolf Hilgenfeld, University of Lübeck,

Lübeck, Germany. Protein expression and purification were adapted from Zhang et al.[56] The expression plasmid was transformed into E. coli strain BL21 (DE3) and then pre-cultured in YT medium at 37 °C (100 μg mL^{-1} ampicillin) overnight. The pre-culture was used to inoculate fresh YT medium supplied with an antibiotic and the cells were grown at 37 °C to an OD600 of 0.6–0.8 before induction of overexpression with 0.5 mM iso-propyl-D-thiogalactoside (IPTG). After 5 h at 37 °C, cells were harvested by centrifugation (5000g, 4 °C, 15 min) and frozen. The pellets were resuspended in buffer A (20 mM Tris, 150 mM NaCl, pH 7.8) supplemented with lysozyme, DNase I and PMSF for the lysis. The lysate was clarified by centrifugation at 12 000g at 4 °C for 1 h and loaded onto a HisTrap HP column (GE Healthcare) equilibrated with 98% buffer A/2% buffer B (20 mM Tris, 150 mM NaCl, 500 mM imidazole, pH 7.8). The column was washed with 95% buffer A/5% buffer B and then His-tagged M^{pro} was eluted with a linear gradient of imidazole ranging from 25 mM to 500 mM. Pooled fractions containing the target protein were subjected to buffer exchange with buffer A using a HiPrep 26/10 desalting column (GE Healthcare). Next, PreS-cission protease was added to remove the C-terminal His tag (20 μg of PreScission protease per mg of target protein) at 12 °C overnight. Protein solution was loaded onto a HisTrap HP column connected to a GSTrap FF column (GE Healthcare) equilibrated in buffer A to remove the GST-tagged PreScission protease, the His-tag, and the uncleaved protein. M^{pro} was finally purified with a Superdex 75 prep-grade 16/60 (GE Healthcare) SEC column equilibrated with buffer C (20 mM Tris, 150 mM NaCl, 1 mM EDTA, 1 mM DTT, pH 7.8). Fractions containing the target protein at high purity were pooled, concentrated at 25 mg mL^{-1} and flash-frozen in liquid nitrogen for storage in small aliquots at −80 °C.

3.2. Protein characterization and enzymatic activity

The molecular mass of the recombinant SARS-CoV-2 M^{pro} was determined by direct infusion electrospray ionization mass spectrometry (ESI-MS) on a Xevo G2-XS QTOF mass spectrometer (Waters). Samples were diluted in 50% acetonitrile with 0.1% formic acid to achieve a final 1 μM concentration of protein. The detected species displayed a mass of 33 796.64 Da, which matches very closely the value of 33 796.81 Da calculated from the theoretical full-length protein sequence (residues 1–306). To characterize the enzymatic activity of our recombinant M^{pro}, we adopted a FRET-based assay using the fluorogenic substrate 5-FAM-AVLQ′SGFRK(DABCYL)K (ProteoGenix) harbouring the cleavage site of SARS-CoV-2 M^{pro} (′ indicates the cleavage site). The fluorescence of the intact peptide is very low since the fluorophore 5-FAM and the quencher Dabcyl are in close proximity. When the substrate is cleaved by the protease, the fluorophore and the quencher are separated, increasing the fluorescence signal. Freshly unfrozen recombinant SARS-CoV-2 M^{pro} was used in our assays. The assay was performed by mixing 0.05 μM M^{pro} with different concentrations of substrate (1–128 μM) in the reaction buffer (20 mM Tris–HCl, 100 mM NaCl, 1 mM EDTA and 1 mM DTT, pH 7.3) in the final volume of 100 μL. Fluorescence intensity (Ex = 485 nm/Em = 535 nm) was

monitored at 37 °C with a Victor3 microplate reader (Perki-nElmer) for 50 min. A calibration curve was created by measuring multiple concentrations (from 0.001 to 5 μM) of free fluorescein in a final volume of 100 μL reaction buffer. Initial velocities were determined from the linear section of the curve, and the corresponding relative fluorescence units per unit of time (ΔRFU/s) were converted to the amount of the cleaved substrate per unit of time (μM s^{-1}) by fitting to the calibration curve of free fluorescein. Inner-filter effect corrections were applied for the kinetic measurements according to ref. 57. The catalytic efficiency k_{cat}/k_m resulted in 4819 ± 399 s^{-1} M^{-1}, in line with literature data.[56,58]

3.3. Nuclear magnetic resonance

All the NMR screening experiments were performed with a Bruker Neo 600 MHz spectrometer, equipped with a nitrogen cooled 5 mm Prodigy CryoProbe at 298 K. The ligand binding was monitored by WaterLOGSY (wLogsy)[59] and Saturation Transfer Difference (STD)[60] experiments in the presence and in the absence of the protein. Samples contained 10 μM M^{pro} and 100 μM to 2 mM ligand dissolved in 150 mM NaCl, 20 mM phosphate, 5% D$_2$O, and 4% DMSO-d6 (pH = 7.3). WaterLOGSY experiments were performed with a 180° inversion pulse applied to the water signal at 4.7 ppm using a Gaussian-shaped selective pulse of 5 ms. Each WaterLOGSY spectrum was acquired with 320 scans, a mixing time of 1.5 s and a relaxation delay of 4.5 s. STD experiments were performed with 256 scans. Selective saturation of the protein at 0.4 ppm frequency was carried out by a 2 s pulse train (60 Gaussian pulses of 50 ms separated by 1 ms intervals) included in the relaxation delay and a 30 ms spin-lock was used to reduce the broad background protein signal. The estimation of the KD was achieved by a STD titration according to a previously reported procedure and fitting the curves using OriginPro 2018 (OriginPro version 2018 developed by OriginLab Corporation, Northampton, MA, USA). The water suppression was achieved by the excitation sculpting pulse scheme.

3.4. Screening of potential M^{pro} inhibitors and hits validation

A FRET-based assay employed to test the enzymatic activity of the recombinant SARS-CoV-2 M^{pro} was used to evaluate the ability of the compounds to inhibit its activity in vitro. In fact, inhibition of M^{pro} by the tested compounds results in a decrease of the fluorescence signal compared to the M^{pro} activity in the absence of an inhibitor. A preliminary screening was first performed at a single compound concentration to rapidly identify the ability of the compounds to inhibit M^{pro} activity and to rank them according to their inhibitory activity. The protein was diluted in the reaction buffer (20 mM Tris–HCl, 100 mM NaCl, 1 mM EDTA and 1 mM DTT, pH 7.3) and pipetted into a 96-well plate to a final protein concentration of 0.02 μM in a final volume of 100 μL. Each compound at the final concentration of 100 μM was incubated with M^{pro} for 20 minutes at room temperature. After incubation, the peptide substrate (5 μM final) was added to initiate the reaction which

was monitored for 50 min at 37 °C. The final DMSO amount was 3.75%. Two controls were prepared for each experiment: the peptide substrate in the absence of M^{pro} (0% M^{pro} activity, hence minimal fluorescence intensity detected) and the reaction mixture in the absence of the compounds (100% M^{pro} activity, therefore maximal fluorescence intensity detected). Following the preliminary screening, the most active compounds (hits) were tested at increasing concentrations (0.25, 0.5, 1, 5, 25, 50, 100, and 150 μM) to determine the dose–response curves and calculate IC_{50} values fitted using Graph-Pad Prism 5 software. Each experiment was performed in triplicate and the results were used to calculate an average and a standard deviation.

3.5. Binding studies by mass spectrometry

Samples were prepared by mixing appropriate volumes of M^{pro} (10 μM final) with each compound in the reaction buffer (20 mM Tris–HCl, 100 mM NaCl, 1 mM EDTA and 1 mM DTT, pH 7.3). The final mixtures had a 1 : 1 or 10 : 1 compound : protein molar ratio. Samples were incubated at room temperature for 20 min before analysis. Control experiments were performed on 10 μM solutions of M^{pro} in the absence of the compounds. Mass spectrometric analyses were carried out in positive ion mode by ESI-MS under denaturing conditions, *i.e.* water/acetonitrile 50 : 50 with 0.1% formic acid on a Q-Tof Xevo G2S (Waters, Manchester, UK). Data were processed using MassLynx V4.1 software.
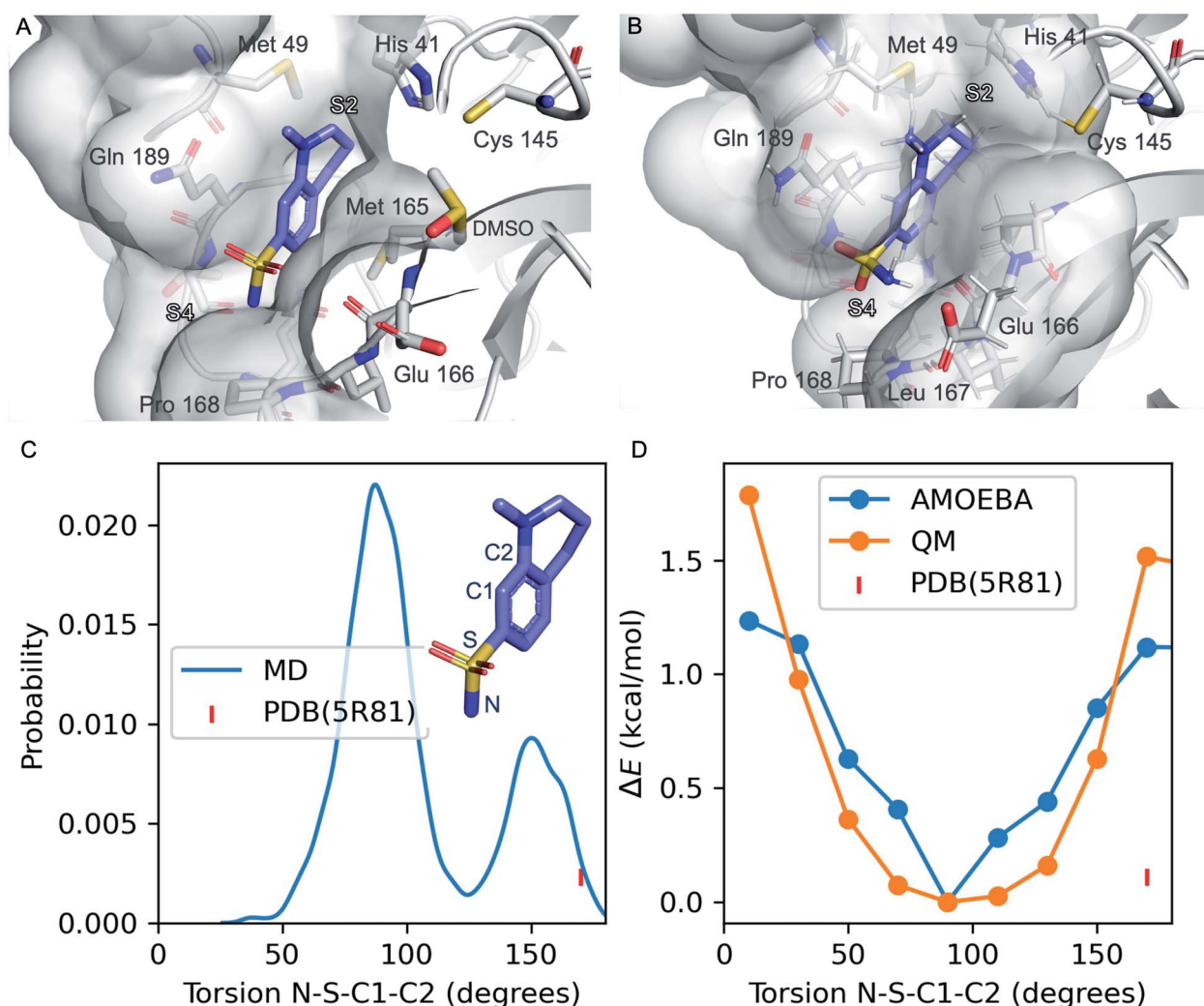

Fig. 1   Refinement of the co-crystal structure of x0195 and M^{pro} using MD simulations. (A) An unusual conformation of x0195 (in purple) located in the binding pocket formed by His41, Met49, Glu166, Gln189, and Pro168 and their surroundings (PDB code: 5R81), and (B) the relaxed structure of x0195 (in purple), obtained after the equilibration step, interacting with the amino acid residues of the substrate binding site. M^{pro} is shown in light grey. (C) Torsion angle distribution for the sulfonamide group during 20 ns of MD simulations (in blue) performed on the M^{pro} dimer in complex with x0195; the torsion angle of the sulfonamide group in the co-crystal structure is shown in pink. (D) Torsion energy scan calculated by AMOEBA (in blue) and QM (in orange); the torsion angle of the sulfonamide group in the co-crystal structure is shown in pink. QM level = ωB97x-D/6-31g*.[65–67]

3.6. Synthesis

The detailed synthetic protocol used to prepare all molecules can be found in the ESI.‡

4. Results and discussion

Several diverse fragments binding the viral M^{pro} have been identified by high-throughput crystallographic screening of this protease. Among the screened fragments, x0195 (PDB ID: 5R81 (ref. 61) – Fig. 1A) shows one of the highest binding affinities[62] and therefore provides a reasonable starting point for fragment-based design of novel M^{pro} inhibitors.

The crystal structure shows that x0195 is located within the M^{pro} substrate binding pocket, at the interface of the two sub-pockets S2 and S4 as described by Cannalire et al.[63] S4 is a solvent exposed subpocket that is partially composed of a flexible loop delimited by Gln189 and Gln192, while S2 is defined by the side chain residues of Phe140, Asn142, His163, Glu166, and His172, and the backbone atoms of Phe140 and Leu141.

In the co-crystal structure corresponding to M^{pro} in complex with x0195 (see Fig. 1A), the aromatic portion of the molecule is located between the side chains of Gln189 and Met 165, while the unsaturated region of the tetrahydroquinoline scaffold establishes a hydrophobic interaction with the side chains of His41 and Met49. The N-methyl group attached to the tetrahydroquinoline core is solvent exposed, while the sulfonamide moiety is in contact with Pro168 and Glu166. In particular, the aromatic ring of the small molecule is bisecting the SO$_2$ unit and the polar sulfonamide nitrogen (–NH$_2$) is reaching the boundaries of the hydrophobic part of the binding pocket composed of the alkyl chain of Pro168.

After comparing the available X-ray structural information with previously conducted studies on small molecule conformational preferences derived from crystal structure data,[64] we noticed that x0195 was modeled in a high energy conformation and that an unusual high-energy (i.e. repulsive) contact occurs between the sulfonamide oxygen and the carbonyl oxygen of the Glu166 backbone. Additionally, the tetrahydroquinoline scaffold was not fully exploring S2 subpocket boundaries. As reported by Cannalire et al.[63] and Zhang et al.,[8] the volume of the S2 subpocket in SARS-CoV M^{pro} is very similar to that of the MERS-CoV homologue. However, the volume of S2 in SARS-CoV M^{pro} (252 Å^3) is significantly larger than in other CoV homologues of the α-genus, such as the HCoV-NL63 M^{pro} (45 Å^3).[8,63]

Therefore, exploiting this knowledge might be key to designing specific inhibitors of CoV M^{pro}.

In order to refine the available X-ray structural model and to gather more structural insights (e.g. protein flexibility and binding pocket rearrangements[12,13]) to guide the design of better binders of the subpocket S2, we ran all-atom molecular dynamics simulations using the AMOEBA polarizable force field[29–32] on M^{pro} (PDB code: 7L11) in complex with x0195 (PDB code: 5R81).

Our simulations show that the unusual high-energy contacts between the sulfonamide oxygen and the carbonyl oxygen of the Glu166 backbone no longer occurred. Also, regarding the electronic structure, we noticed that the p orbitals of the aromatic carbon C1 bisect (e.g. are parallel to) the SO$_2$ angle, compared with a 90° value for the same angle as reported in the crystal structure (see Fig. 1). Moreover, the NH$_2$ of the sulfonamide group is engaged in favorable polar interactions with the Gln189 side chain and the solvent.

Then, we performed absolute binding free energy calculations on the refined protein–ligand structure. Our results show that x0195 binds to the protein with a binding free energy of −2.83 kcal mol^{-1} at 283 K, which is comparable to the experimental binding energy (−3.59 ± 0.1 kcal mol^{-1}, see Table 1).

We obtained the experimental binding free energy by converting the experimental K_d (1.7 mM ± 0.2) provided in the literature[62] using the Gibbs free energy equation and the experimental temperature used in the binding assays (283 K). The agreement of the computed free energy prediction with the experimental results is reasonable. Further analysis of MD simulations suggests that the tetrahydroquinoline scaffold of x0195 is sub-optimally occupying the binding pocket.

We put in place design strategies to modify the chemical moieties of x0195 and potentially increase its binding affinity. Here, we introduce the design of a new molecule, namely QUB-00006 (Fig. 2), where we added two fluorines and a methyl group on the tetrahydroquinoline core of x0195. Also, we substituted the sulfonamide group on the aromatic ring of the molecule by a methanethiol. Fluorination at position 3 of the tetrahydroquinoline core could increase ligand occupancy with no disruption of the water network surrounding the binding pocket,[12,13] while methylation at position 4 seemed an interesting modification to increase the potential interactions of the ligand with binding pocket residues. QUB-00006 was generated based on the structure and position of x0195 in the co-crystal (5R81), and then placed in the receptor structure (M^{pro} dimer
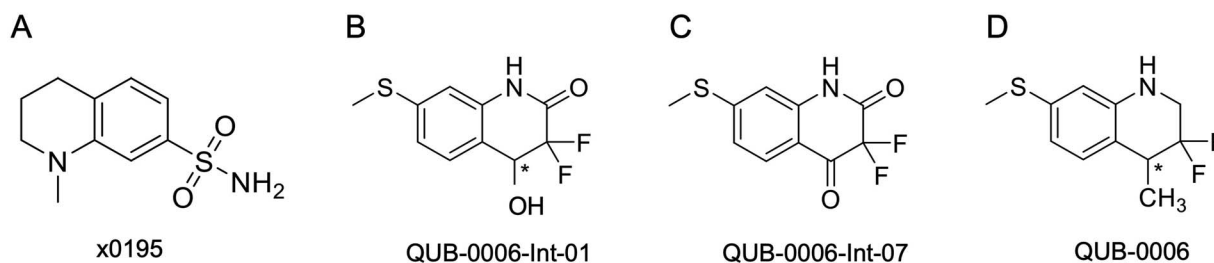

Fig. 2 2D structures of (A) x0195, (B) QUB-00006-Int-01, (C) QUB-00006-Int-07, and (D) QUB-00006. The asterisk represents a chiral center.

Table 1 Experimental and computed binding free energies (kcal mol^{-1}) for the non-covalent compounds. N.A. = not available (see text for details)

Compound	Computed ΔG	Experimental ΔG
QUB-00 006(R)	−2.73 ± 0.34	N.A.
QUB-00 006(S)	−2.72 ± 0.22	
QUB-00 006-Int-01(R)	−4.30 ± 0.35	−3.71 ± 0.2
QUB-00 006-Int-01(S)	−4.45 ± 0.29	
x0195	−2.83 ± 0.66	−3.59 ± 0.1
QUB-00 006-Int-07	−5.37 ± 0.23	Covalent binder

with the PDB code: 7ILL); next, the M^{pro}–QUB-00006 complex was equilibrated using MD simulations (see ESI Section 1‡ for the detailed protocol), followed by free energy calculations. To explore the potential of our computational platform in designing new binders with no or few experimental data such as ligand–M^{pro} co-crystal structures, we leveraged all-atom molecular dynamics simulations on QUB-00006 complexed with M^{pro}. The aim of this approach is to gather insights on the binding conformation of the newly *in silico* designed ligand, assess pocket fitness, and evaluate its binding affinity using ABFE calculations.

The initial molecular conformation is mostly anchored at the binding pocket, with the α,α-difluoro-methyl group attached to the tetrahydroquinoline core fully occupying the buried part of the S2 subpocket, which is composed of the side chains of Met49 and His41, while the sulfonamide moiety extends to S4 (Leu167 and Pro168). We note that methylation at position 4 of the tetrahydroquinoline core introduces a chiral center, however no significant differences in terms of pocket occupancy between the R and S enantiomers were observed.

The computed absolute binding free energies for QUB-00006(R) and QUB-00006(S) are −2.73 ± 0.34 kcal mol^{-1} and

−2.72 ± 0.22 kcal mol^{-1}, respectively (Table 1). These results suggest that the designed fluorinated fragment is a binder at the M^{pro} S2 subpocket and could represent a starting point for structure-based design of novel M^{pro} inhibitors.

The identified binding mode is defined by several favorable intermolecular interactions occurring between the newly designed ligand and the M^{pro} binding pocket: (i) the sulfur group of QUB-00006(R) interacts with the oxygen of the carbonyl belonging to the backbone of Glu166 with a distance of 3.3 Å, (ii) the α,α-difluoro moiety points towards His41, and (iii) the sulfur of Met49 establishes a favorable interaction with one of the two fluorines of the substrate (distance 3.3 Å). In fact, the sulfur–oxygen contact observed in our simulations is in agreement with the findings of a study conducted by Iwaoka *et al.*,[68] where they found that a total of 1200 and 626 fragments from the Cambridge Structural Database (CSD) and Protein Data Bank (PDB), respectively, have close intermolecular S–O contacts (with a distance of 3.52 Å or less). Another study analyzing the protein structures deposited in the Protein Data Bank reports 1133 interactions between His and halogen atoms found in 3833 PDB entries with one or more halogenated ligands co-crystallized with a protein.[69] Moreover, the strong S–F interaction identified during the simulations is in good agreement with experimentally observed distances for fluorine–sulfur contacts in crystal structures (2.8–3.4 Å).[70] It is worth noting that such interactions involving sulfur and halogen atoms are usually better captured with polarizable models than with their classical counterparts.[71–73]

QUB-00006 was then synthesized following the path in Fig. 4 in order to validate *in vitro* the simulation outcomes.

The ligand orientation in the MD simulations and the computed hydration ratio of the different atoms of QUB-00006 during ABFE simulations suggest that proton C is solvent exposed, while the protons of the methyl thioether group (group
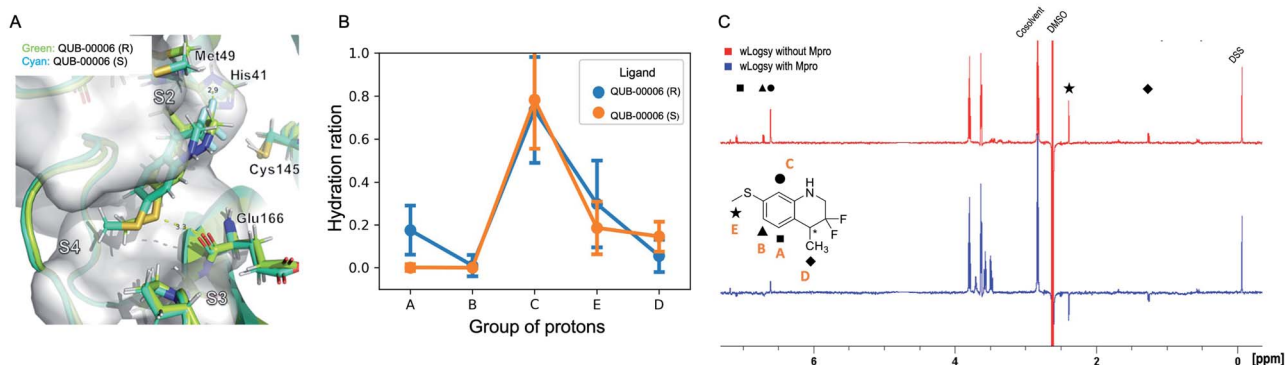

Fig. 3 Computational and experimental characterization of QUB-00006 binding within the M^{pro} binding pocket. (A) QUB-00006(R) (in light green) and QUB-00006(S) (in cyan) binding in a similar fashion at the interface of subpockets S2 and S4; the binding poses shown here were clustered and extracted from the trajectories of the binding free energy calculations performed on QUB-00006(R) and QUB-00006(S). (B) The analysis of our binding free energy trajectories showing that protons in groups A, B, E, and D have a low hydration ratio (less than 0.5), while the proton of group C has a high hydration ratio of 0.8. Hydration ratios calculated for the different proton groups of QUB-00006(R) correlate with those calculated for QUB-00006(S). (C) The WaterLOGSY spectra of QUB-00006 in the presence and absence of the M^{pro}. The assignment scheme is reported along with the 2D structure of the ligand. The strong negative intensity of the signals of the hydrogens of groups A, D, and E suggests that they are orientated towards the protein, while the hydrogen atom in C is solvent exposed. These experimental findings confirm the hydration ratio calculated during our binding free energy simulations and described in panel B.

E) and the methyl group at position 4 of the tetrahydroquinoline core (group D) are buried (Fig. 3B).

Those findings strongly correlate with the NMR characterization of QUB-00006 obtained *via* WaterLOGSY experiments. In fact, WaterLOGSY epitope mapping confirms that QUB-00006 binds to the protein binding pocket. We leveraged the experimental approach to better identify the region of the ligand in contact with the protein. In Fig. 3C, the proton signals arising from the two methyl groups (D and E) in the presence of M^{pro} show a change in the sign suggesting that these protons are in close contact with the protein. Similarly, the aromatic protons A and B undergo a sign inversion. In contrast, the aromatic proton C is not significantly perturbed, which suggests that this position is solvent exposed. The binding mode suggested by NMR is in agreement with the MD-derived hydration ratios confirming the predictive power of our MD-based approach to characterize the binding mode of novel ligands at the experimental level of accuracy (Fig. 3B and C).

Although we were able to gather structural information about the binding mode of QUB-00006 using a WaterLOGSY assay, we could not measure its experimental binding affinity *via* STD NMR due to solubility challenges.

Several synthetic steps were performed in order to obtain QUB-00006, as detailed in Fig. 4. Through this synthetic scheme, we obtained different intermediates characterized by a better solubility profile (Table 2). Interestingly, the hydroxyquinolinone QUB-00006-Int-01 displayed the best solubility profile of all the synthetic intermediates, making it a strong candidate for *in vitro* evaluation.

Before conducting NMR STD experiments to determine the dissociation constant (K_d) of the more polar QUB-00006-Int-01

Table 2 Prediction of the properties of compounds using our machine learning workflow. MW represents the molecular weight of the compounds in daltons, log S is the predicted solubility of the different compounds, log P represents the differential solubility, and the Tanimoto coefficient reflects the similarity of the selected compounds relative to x0195

	MW (Da)	log S	log P	Tanimoto (MACCS)
QUB-00 006	229.07	−3.99	3.56	0.391
QUB-00 006-Int-07	243.02	−3.73	1.96	0.371
QUB-00 006-Int-01	245.03	−2.73	1.66	0.338
x0195	226.08	−1.94	0.56	1

compound, we decided to predict its binding conformation at the binding pocket and compute the respective absolute binding free energy. Modification of the molecular scaffolds, especially in fragment-like molecules, might affect the binding mode[74] compared to a reference structure (*e.g.* x0195 as per PDB ID:5R81).

We used a combination of docking, MD and ABFE calculations to explore the putative binding mode of QUB-00006-Int-01. Those calculations identified two dominant binding modes for QUB-00006-Int-01(R) and QUB-00006-Int-01(S) (Fig. 5A) with computed binding free energies of −4.4 and −4.3 kcal mol^{-1}, respectively. Then, we estimated the binding affinity of QUB-00006-Int-01 towards M^{pro} by a STD NMR titration and we found a dissociation constant in the low millimolar range, with an estimated K_d of 1.9 ± 0.6 mM (−3.71 ± 0.2 kcal mol^{-1}), which agrees reasonably well with our binding free energy calculations (Table 1). As shown in Fig. 5A, both enantiomers bind to the S2 and S4 subpockets with the thioether group being fully buried
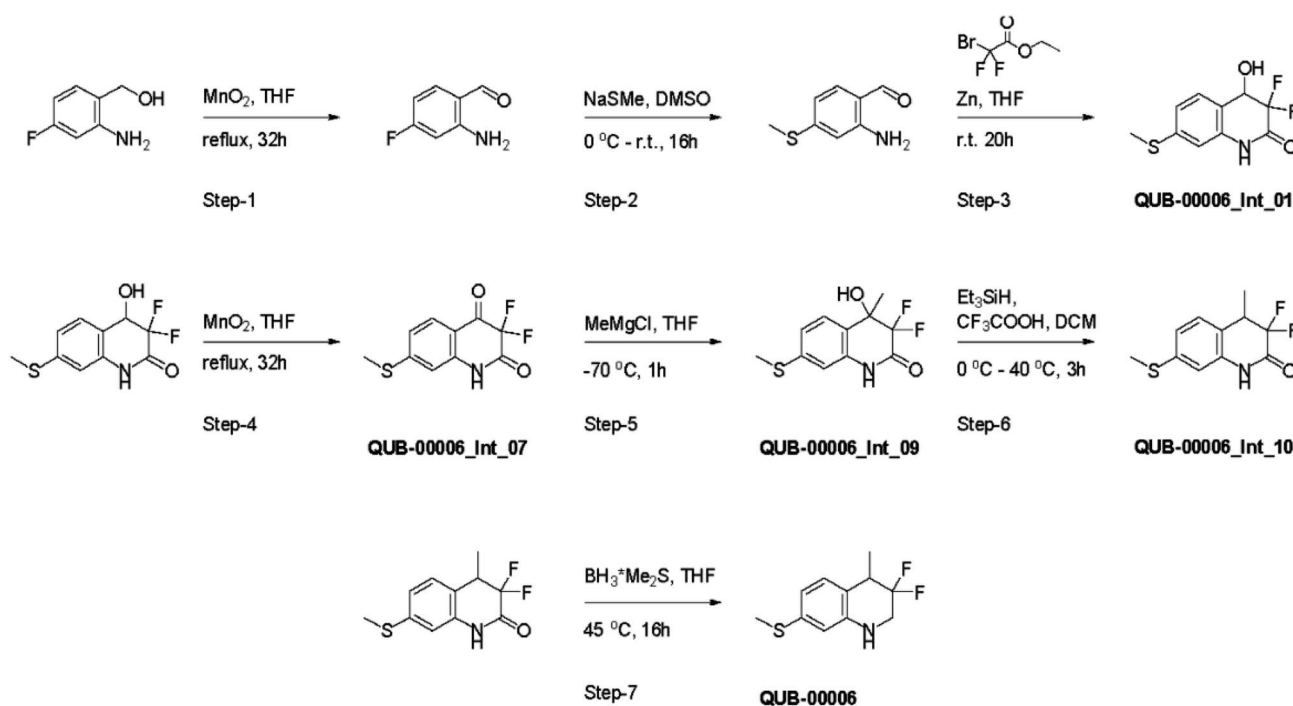

Fig. 4 Synthesis path of 3,3-difluoro-4-methyl-7-(methylsulfanyl)-1,2,3,4-tetrahydroquinoline named QUB-00006.
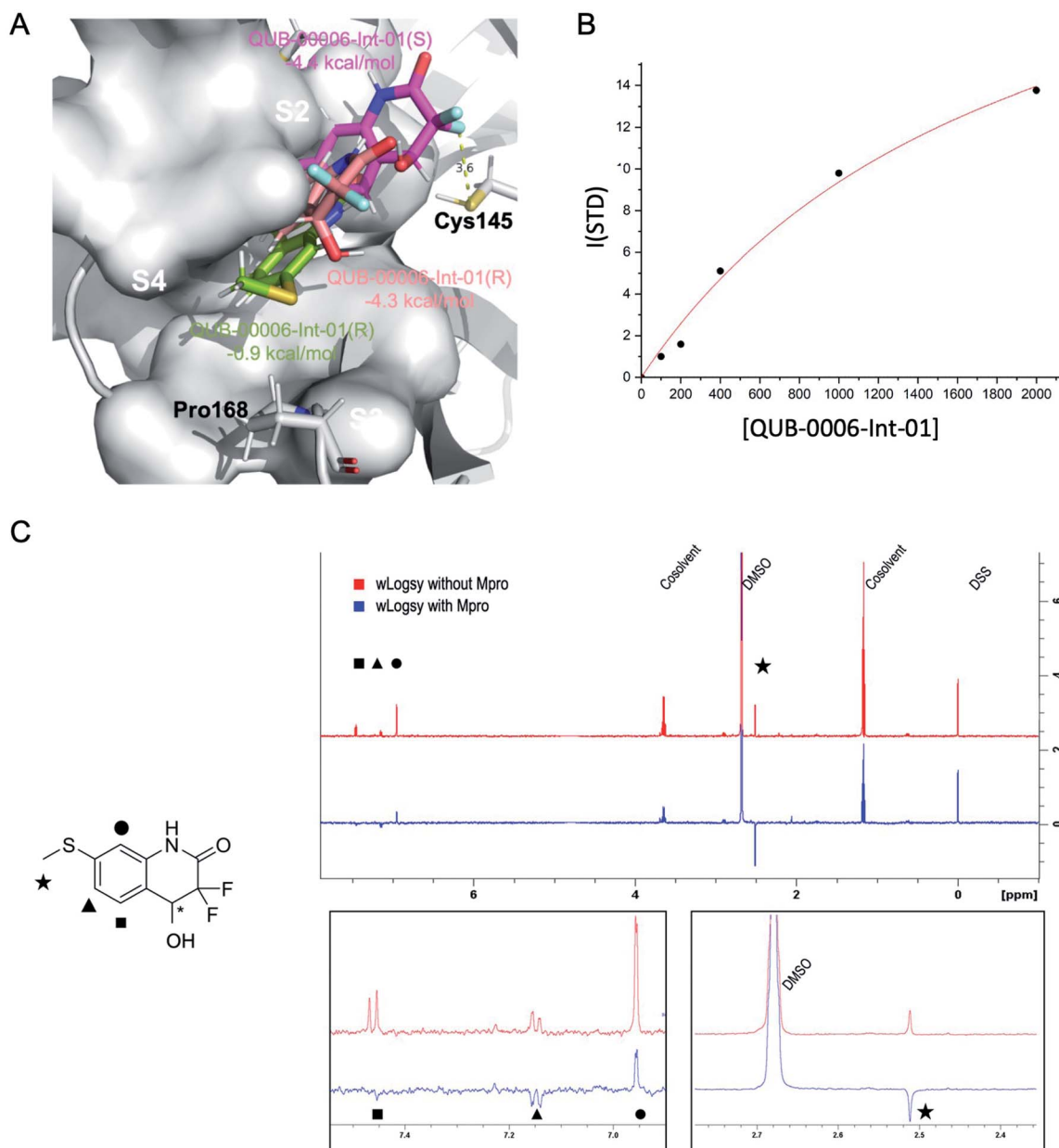
Fig. 5   Computational and experimental characterization of QUB-00006-Int-01 in the M^{pro} binding pocket. (A) The dominant binding modes of QUB-00006-Int-01(R) (in pink) and QUB-00006-Int-01(S) (in magenta), identified during ABFE simulations. They have computed binding free energies of −4.4 and −4.3 kcal mol^{-1}, respectively; also, they bind to the S2 and S4 subpockets in a similar fashion with the thioether group being fully buried in S2. On the other hand, starting with a QUB-00006 like binding mode, we ran an additional absolute binding free energy calculation on M^{pro} in complex with QUB-00006-Int-01(R) and obtained a second binding mode for QUB-00006-Int-01(R) (in green) with a binding free energy of −0.9 kcal mol^{-1}. (B) STD titration profile of QUB-00006-Int-01. The ligand concentration ranges from 100 μM to 2 mM against 10 μM of M^{pro}. (C) The WaterLOGSY spectra of QUB-00006-Int-01 with M^{pro} (in blue) and without M^{pro} (in red). The assignment of the signals is reported on the 2D structure of the fragment. The methyl and the aromatic signals of the two protons adjacent to the hydroxyl group undergo a significant change, which suggests that these groups are in close contact with the protein's cavity. In contrast, the aromatic proton adjacent to the lactamic nitrogen undergoes a reduction of its intensity, suggesting that this proton is partially exposed to the solvent. These STD results confirm our computational characterization of the binding mode of QUB-00006-Int-01 (panel A).

in subpocket S2, which correlates with WaterLOGSY experiments (Fig. 5C). Additionally, QUB-00006-Int-01(R) and QUB-00006-Int-01(S) fill up a binding pocket space that is different from the one occupied by QUB-00006. On the other hand, starting with a QUB-00006-like binding mode, we ran an additional absolute binding free energy calculation on an M^{pro}–

QUB-00006-Int-01(R) complex and obtained a binding free energy of −0.9 kcal mol^{-1}. These results suggest that QUB-00006-Int-01 and QUB-00006 might have different dominant binding conformations (see Fig. 3A and 5A).

Since a fragment-like molecule could have multiple binding modes and the ligand conformation is unlikely to be fully

Table 3 Population of the clusters generated by adaptive sampling performed on M^{pro} in complex with QUB-00 006-Int-01(R) and (S). $\Delta\Delta G$ (kcal mol^{-1}) is the relative free energy at 298 K. The relative binding free energies reported for QUB-00 006-Int-01(R) and (S) are calculated using the respective cluster 1 as a reference ligand

QUB-00 006-Int-01(R)			QUB-00 006-Int-01(S)		
Cluster	Fraction	$\Delta\Delta G$	Cluster	Fraction	$\Delta\Delta G$
1	0.101	0	1	0.103	0
2	0.083	0.05	2	0.093	0.03
3	0.067	0.11	3	0.088	0.04
4	0.053	0.17	4	0.065	0.12
5	0.042	0.23	5	0.059	0.14
6	0.035	0.27	6	0.054	0.17
7	0.034	0.28	7	0.049	0.19
8	0.033	0.29	8	0.039	0.25
9	0.032	0.30	9	0.033	0.29
10	0.032	0.30	10	0.031	0.31

Fig. 6 Conformations of QUB-00006-Int-01 sampled during 20 ns of ABFE calculations and 450 ns of adaptive sampling simulations. The conformation was plotted as a function of two distances: (i) the distance between C2 (carbon of QUB-00006-Int-01 connected to the hydroxyl group) and the sulfur of Cys145, and (ii) the distance between the methyl thioether group in QUB-00006-Int-01 and the beta carbon of Gln189. "0" indicates the starting structure, "1" indicates the largest cluster, and "i" indicates the ith largest cluster. The frames were taken at 10 ps time intervals.

sampled during 20 ns of binding free energy simulations, we used unsupervised adaptive sampling (AS) to further explore the conformational space of QUB-00006-Int-01. AS can be used here as an interpretative tool able to gather structural insights on the various potential M^{pro}–ligand interactions (see the ESI‡ for details). The AS trajectories were clustered using average-linkage hierarchical clustering algorithms and the top ten largest clusters were chosen for analysis. These clusters have comparable populations (the smallest clusters have 3–4 times smaller populations or 0.3 kcal mol^{-1} higher free energy than the largest clusters, see Table 3), indicating the coexistence of multiple binding modes.

More precisely, starting from these clusters, absolute binding free energies would yield results within 0.3 kcal mol^{-1} of what was previously obtained. The simulations of QUB-00006-Int-01(R) and QUB-00006-Int-01(S) converged to similar

ensembles containing several possible binding modes. Clusters 3, 5, and 6 of QUB-00006-Int-01(R) and cluster 4 of QUB-00006-Int-01(S) (ESI-Fig. 2‡) correspond to the respective dominant binding modes predicted by ABFE simulations (Fig. 5A). For both enantiomers, the most conserved interactions are the hydrophobic contacts between C9 (methyl thioether) and Gln189, and between C5 (proton B) and His41, Arg188, and Gln189.

Overall, our computational findings on QUB-00006-Int-01 confirm that the structural approach we introduce in this work using a sequence of MD-based techniques (classical MD simulations, adaptive sampling, and absolute binding free energy calculations) is able to capture potential binding orientations of fragment-like compounds in the binding pocket of a protein, and to accurately predict their binding free energies.

Then, we analyzed the clustered QUB-00006-Int-01 binding conformations from the adaptive sampling simulations plotted as a function of the distance between the methyl thioether group in QUB-00006-Int-01 and the beta carbon of Gln189, and the distance between C2 (carbon connected to the hydroxyl group) and the sulfur atom (SG) of the catalytic side chain of the Cys145 residue (Fig. 6). We noticed that the distance of C2–SG in the most populated cluster generated by the AS simulations is around 4 Å. To reinforce our analysis, we leveraged another unsupervised reduction of dimension technique: TICA (time-lagged independent component analysis),[75] which aims at finding the slow collective variables of the data, and applied it to QUB-00006-Int-01(R). We then used the k-means clustering method on the data projected on this space and built a Hidden Markov State Model (HMSM).[76] Three clusters emerged, whose characteristics also show the coexistence of several binding modes of QUB-00006-Int-01(R), one of which corresponds to a distance between C2 and SG below 4 Å. Detailed results can be found in the ESI.‡

Targeting Cys145 with covalent warheads has been used by several researchers to discover novel potent inhibitors of M^{pro}.[38,63,77] As a matter of fact, a simple chemical modification to QUB-00006-Int-01 would lead to QUB-00006-Int-07 bearing an α,α-difluoro-keto moiety, which is prone to a nucleophilic attack by the vicinal R-SH of Cys145. In order to enable the latter, QUB-00006-Int-07 would need to access the M^{pro} substrate pocket and adopt a stable binding conformation prior to the covalent binding. Thus, we conducted absolute binding free energy simulations on the M^{pro}–QUB-00006-Int-07 complex, which confirmed a favorable binding energy of QUB-00006-Int-07 to the M^{pro} substrate pocket (-5.37 ± 0.23 kcal mol^{-1}). As reported in Fig. 7, compound QUB-00006-Int-07 is bound to the S2 and S4 subpockets with the thioether group being fully buried in sub-pocket S2 and the α,α-difluoro-keto moiety facing Cys145. More precisely, the average distance between SG and the C is 3.65 angstroms (± 0.33) and the average distance between the SG and C2 is 3.61 angstroms (± 0.43) as can be seen in Fig. 7.

Our computational findings motivated us to test the compound with a FRET-based proteolytic assay. This assay should detect potent functional binders to the viral M^{pro}. Being a fluorogenic assay, compounds with fluorescence quenching properties can suppress the fluorescence signal generated by

Fig. 7 The dominant binding mode of QUB-00006-Int-07 during ABFE simulations. (A) Time evolution of key distances in the simulation. "SG" stands for the sulfur atom in Cys145, "C" is the amide carbon in QUB-00006-Int-07, and "C2" is the carbonyl carbon in QUB-00006-Int-07. The average distances for SG−C and SG−C2 are 3.61 Å and 3.65 Å, respectively. (B) The dominant binding mode of QUB-00006-Int-07 within the M^{pro} binding pocket. QUB-00006-Int-07 is shown in pink and the protein is shown in silver sticks and surfaces. The binding mode is very stable during the simulation, where the hydroxyl group is close to Cys145 and forms a hydrogen-bond with the Glu166 backbone, and the difluoro group interacts with the carbonyl group of Asn142. These binding modes are also comparable to the dominant binding mode of QUB-00006-Int-01(S) identified during ABFE calculations.

Fig. 8 Dose−response curves obtained by plotting the percentage of SARS-CoV-2 M^{pro} residual activity as a function of increasing concentrations of QUB-00006-Int-07 (0–150 μM). [M^{pro}] = 20 nM, [PS1] = 5 μM, %DMSO = 3.75%. Experiments were performed in triplicate. A counter screening control experiment was performed by testing increasing concentrations of QUB-00006-Int-07 in the presence of 0.5 μM free fluorescein.

(Fig. 9A). The sample obtained after incubation of QUB-00006-Int-07 with M^{pro} (compound : protein ratio = 10 : 1) was analyzed by ESI-MS under denaturing conditions, and a representative spectrum is provided in Fig. 9B. In addition to the signals corresponding to multiple charge states of the initial protein (red dots), we identified the distribution of signals corresponding to the M^{pro} modified by the presence of the compound (green asterisks) which is therefore covalently linked to the protein given the non-native conditions of the experiment. The nature of the adduct and the molecular mechanism of binding are under investigation and will be the subject of further studies.

Finally, in this work, the introduction of multiple modifications (*e.g. gem*-difluoro, thioether, hydroxyl and methyl groups) to the tetrahydroquinoline scaffold of x0195, and the design and synthesis of novel molecular scaffolds, enabled the exploration of binding pocket boundaries and provided additional information related to druggability of the S2 subpocket. Other
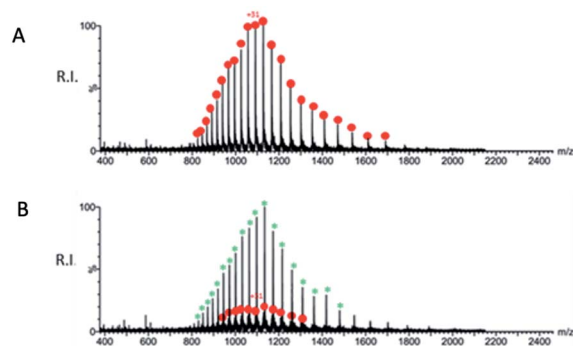
the protease activity. To eliminate false positive results, we conducted a preliminary counter screen and verified that the tested compound possesses negligible fluorescence quenching effects. Subsequently, to assess the potential inhibitory activity of the compound against SARS-CoV-2 M^{pro}, increasing concentrations of QUB-00006-Int-07 (0.25–150 μM) were incubated with 20 nM M^{pro} before the addition of 5 μM FRET substrate. As shown in Fig. 8, QUB-00006-Int-07 inhibited M^{pro} with 50% inhibitory concentration (IC_{50} value of 830 ± 50 nM), thus resulting in a fairly potent inhibitor of the M^{pro} enzymatic activity. The binding of QUB-00006-Int-07 to M^{pro} was confirmed by electrospray ionization (ESI) mass spectrometry.

A preliminary determination of the initial protein showed an experimental mass of 33 796.40 Da, which matches very closely the expected value of 33 796.64 Da calculated from the sequence

Fig. 9 (A) Representative ESI-MS spectrum of a solution containing 10 μM SARS-CoV-2 M^{pro} in water/acetonitrile (50 : 50) added with 0.1% formic acid. The spectrum was acquired in positive ion mode. (B) Representative ESI-MS spectrum of a mixture containing 10 μM SARS-CoV-2 M^{pro} after incubation with QUB-00006-Int-07 (compound : protein ratio = 10 : 1) in water/acetonitrile (50 : 50) added with 0.1% formic acid. The spectrum was acquired in positive ion mode. The red dots correspond to the unmodified protein, and the green asterisks correspond to the modified protein.

molecules were produced over the course of this research but, due to their weaker activity, their detailed analysis is not provided here. Their list can be found in the ESI.‡ These compounds were either designed computationally without leading to improved affinities or were synthesis intermediates. All resulting molecules were submitted to biological testing, but none of them were found to be as potent as QUB-00006-Int-07 nor presented a strongly druggable profile, compared to the previously discussed compounds.

5. Conclusion and perspectives

We presented a computationally driven discovery of a new set of non-covalent and covalent inhibitors of M^{pro} that have been further characterized experimentally. The best compound, QUB-00006-Int-07, has been found to be a covalent binder that resulted in a potent inhibition of the M^{pro} enzymatic activity ($IC_{50} = 830 \pm 50$ nM). The results of the innovative scaffold design described here were obtained within three months *via* a fast-track project that took place in the summer of 2021. It involved a small consortium of theoreticians, organic chemists and drug designers, and demonstrated the effectiveness of a computation-guided synthetic strategy. Indeed, GPU-accelerated high-performance computing platforms can now provide access to high-resolution molecular dynamics simulations, which are able to predict detailed protein conformational maps and provide accurate absolute binding free energy results. Such computations can be further rationalized by means of adaptive sampling simulations, an approach which is able to decipher multiple binding modes. Coupled to NMR, *in vitro* experiments and machine learning, such high-resolution predictions yield structural insights regarding the design of new active compounds, while offering an atomic level understanding of binding affinities.

Beyond this preliminary proof of concept study, the next research steps will be devoted to the QM/MM modeling[78,79] of the warhead reaction mechanism[38,77,80] leading to the covalent binding of QUB-00006-Int-07, and to optimization of active compounds with the goal of reaching low nanomolar activity.

Data availability

All data have been provided in the main text and ESI.†

Author contributions

L. E. K., Z. J., D. L., T. J. I., D. E. A., B. S., and L. L. performed the simulations. L. L., T. J. I., D. S., and J.-P. P. contributed new code. M. S., M. B., A. S., and M. V. performed NMR and mass spectroscopy experiments. I. K., R. I., A. K., Y. M., A. P., and V. C.-R. performed the synthesis. L. L., B. S., P. R., J. W. P., D. S., and J.-P. P. contributed new methodology. L. L., B. S., F. H., P. R., J. W. P., and J.-P. P. contributed analytical tools. All authors analyzed the data. L. E. K., Z. J., L. L., M. S., A. S., D. S. and J.-P. P. wrote the paper. D. S. and J.-P. P. designed the research.

Conflicts of interest

P. R., M. M., L. L., J. W. P. and J.-P. P. are co-founders and shareholders of Qubit Pharmaceuticals.

Acknowledgements

References

1 P. Zhou, X.-L. Yang, X.-G. Wang, B. Hu, L. Zhang, W. Zhang, H.-R. Si, Y. Zhu, B. Li, C.-L. Huang, *et al.*, *nature*, 2020, **579**, 270–273.

2 F. Wu, S. Zhao, B. Yu, Y.-M. Chen, W. Wang, Z.-G. Song, Y. Hu, Z.-W. Tao, J.-H. Tian, Y.-Y. Pei, *et al.*, *Nature*, 2020, **579**, 265–269.

3 E. N. Muratov, R. Amaro, C. H. Andrade, N. Brown, S. Ekins, D. Fourches, O. Isayev, D. Kozakov, J. L. Medina-Franco, K. M. Merz, T. I. Oprea, V. Poroikov, G. Schneider, M. H. Todd, A. Varnek, D. A. Winkler, A. V. Zakharov, A. Cherkasov and A. Tropsha, *Chem. Soc. Rev.*, 2021, **50**, 9121–9151.

4 F. von Delft, M. Calmiano, J. Chodera, E. Griffen, A. Lee, N. London, T. Matviuk, B. Perry, M. Robinson and A. von Delft, *A white-knuckle ride of open COVID drug discovery*, 2021.

5 J. Breidenbach, C. Lemke, T. Pillaiyar, L. Schäkel, G. Al Hamwi, M. Diett, R. Gedschold, N. Geiger, V. Lopez, S. Mirza, V. Namasivayam, A. C. Schiedel, K. Sylvester, D. Thimm, C. Vielmuth, L. Phuong Vu, M. Zyulina, J. Bodem, M. Gütschow and C. E. Müller, *Angew. Chem., Int. Ed.*, 2021, **60**, 10423–10429.

6 F. Krammer, *Nature*, 2020, **586**, 516–527.

7 H. Ledford, *et al.*, *Nature*, 2021, **599**, 358–359.

8 L. Zhang, D. Lin, Y. Kusov, Y. Nian, Q. Ma, J. Wang, A. von Brunn, P. Leyssen, K. Lanko, J. Neyts, A. de Wilde, E. J. Snijder, H. Liu and R. Hilgenfeld, *J. Med. Chem.*, 2020, **63**, 4562–4578.

9 H. Yang, W. Xie, X. Xue, K. Yang, J. Ma, W. Liang, Q. Zhao, Z. Zhou, D. Pei, J. Ziebuhr, *et al.*, *PLoS Biol.*, 2005, **3**, e324.

10 Z. Cournia, B. Allen and W. Sherman, *J. Chem. Inf. Model.*, 2017, **57**, 2911–2937.

11 D. W. Kneller, G. Phillips, H. M. O'Neill, R. Jedrzejczak, L. Stols, P. Langan, A. Joachimiak, L. Coates and A. Kovalevsky, *Nat. Commun.*, 2020, **11**, 1–6.

12 T. Jaffrelot Inizan, F. Célerse, O. Adjoua, D. El Ahdab, L.-H. Jolly, C. Liu, P. Ren, M. Montes, N. Lagarde, L. Lagardère, P. Monmarché and J.-P. Piquemal, *Chem. Sci.*, 2021, **12**, 4889–4907.

13 D. El Ahdab, L. Lagardère, T. J. Inizan, F. Célerse, C. Liu, O. Adjoua, L.-H. Jolly, N. Gresh, Z. Hobaika, P. Ren, R. G. Maroun and J.-P. Piquemal, *J. Phys. Chem. Lett.*, 2021, **12**, 6218–6226.

14 S. Zev, K. Raz, R. Schwartz, R. Tarabeh, P. K. Gupta and D. T. Major, *J. Chem. Inf. Model.*, 2021, **61**, 2957–2966.

15 Z. Li, X. Li, Y.-Y. Huang, Y. Wu, R. Liu, L. Zhou, Y. Lin, D. Wu, L. Zhang, H. Liu, X. Xu, K. Yu, Y. Zhang, J. Cui, C.-G. Zhan, X. Wang and H.-B. Luo, *Proc. Natl. Acad. Sci. U. S. A.*, 2020, **117**, 27381–27387.

16 H. T. H. Chan, M. A. Moesser, R. K. Walters, T. R. Malla, R. M. Twidale, T. John, H. M. Deeks, T. Johnston-Wood, V. Mikhailov, R. B. Sessions, W. Dawson, E. Salah, P. Lukacik, C. Strain-Damerell, C. D. Owen, T. Nakajima, K. Świderek, A. Lodola, V. Moliner, D. R. Glowacki, J. Spencer, M. A. Walsh, C. J. Schofield, L. Genovese, D. K. Shoemark, A. J. Mulholland, F. Duarte and G. M. Morris, *Chem. Sci.*, 2021, **12**, 13686–13703.

17 A. Morris, W. McCorkindale, T. C. M. Consortium, N. Drayman, J. D. Chodera, S. Tay, N. London and A. A. Lee, *Chem. Commun.*, 2021, **57**, 5909–5912.

18 D. Shcherbakov, D. Baev, M. Kalinin, A. Dalinger, V. Chirkova, S. Belenkaya, A. Khvostov, D. Krut'ko, A. Medved'ko, E. Volosnikova, E. Sharlaeva, D. Shanshin, T. Tolstikova, O. Yarovaya, R. Maksyutov, N. Salakhutdinov and S. Vatsadze, *ACS Med. Chem. Lett.*, 2021, **13**(1), 140–147.

19 E. Glaab, G. B. Manoharan and D. Abankwa, *J. Chem. Inf. Model.*, 2021, **61**, 4082–4096.

20 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 and H. Yang, *Nature*, 2020, **582**, 289–293.

21 J. Gossen, S. Albani, A. Hanke, B. P. Joseph, C. Bergh, M. Kuzikov, E. Costanzi, C. Manelfi, P. Storici, P. Gribbon, A. R. Beccari, C. Talarico, F. Spyrakis, E. Lindahl, A. Zaliani, P. Carloni, R. C. Wade, F. Musiani, D. B. Kokh and G. Rossetti, *ACS Pharmacol. Transl. Sci.*, 2021, **4**, 1079–1095.

22 A. Manandhar, V. Srinivasulu, M. Hamad, H. Tarazi, H. Omar, D. J. Colussi, J. Gordon, W. Childers, M. L. Klein, T. H. Al-Tel, M. Abou-Gharbia and K. M. Elokely, *J. Chem. Inf. Model.*, 2021, **61**, 4745–4757.

23 G. Amendola, R. Ettari, S. Previti, C. Di Chio, A. Messere, S. Di Maro, S. J. Hammerschmidt, C. Zimmer, R. A. Zimmermann, T. Schirmeister, M. Zappalà and S. Cosconati, *J. Chem. Inf. Model.*, 2021, **61**, 2062–2073.

24 M. M. Ghahremanpour, J. Tirado-Rives, M. Deshmukh, J. A. Ippolito, C.-H. Zhang, I. Cabeza de Vaca, M.-E. Liosi, K. S. Anderson and W. L. Jorgensen, *ACS Med. Chem. Lett.*, 2020, **11**, 2526–2533.

25 C.-H. Zhang, K. A. Spasov, R. A. Reilly, K. Hollander, E. A. Stone, J. A. Ippolito, M.-E. Liosi, M. G. Deshmukh, J. Tirado-Rives, S. Zhang, Z. Liang, S. J. Miller, F. Isaacs, B. D. Lindenbach, K. S. Anderson and W. L. Jorgensen, *ACS Med. Chem. Lett.*, 2021, **12**, 1325–1332.

26 C.-H. Zhang, E. A. Stone, M. Deshmukh, J. A. Ippolito, M. M. Ghahremanpour, J. Tirado-Rives, K. A. Spasov, S. Zhang, Y. Takeo, S. N. Kudalkar, Z. Liang, F. Isaacs, B. Lindenbach, S. J. Miller, K. S. Anderson and W. L. Jorgensen, *ACS Cent. Sci.*, 2021, **7**, 467–475.

27 O. Adjoua, L. Lagardère, L.-H. Jolly, A. Durocher, T. Very, I. Dupays, Z. Wang, T. J. Inizan, F. Célerse, P. Ren, J. W. Ponder, J.-P. Piquemal, *et al.*, *J. Chem. Theory Comput.*, 2021, **17**, 2034–2053.

28 L. Lagardère, L.-H. Jolly, F. Lipparini, F. Aviat, B. Stamm, Z. F. Jing, M. Harger, H. Torabifard, G. A. Cisneros, M. J. Schnieders, N. Gresh, Y. Maday, P. Y. Ren, J. W. Ponder and J.-P. Piquemal, *Chem. Sci.*, 2018, **9**, 956–972.

29 P. Ren and J. W. Ponder, *J. Phys. Chem. B*, 2003, **107**, 5933–5947.

30 J. W. Ponder, C. Wu, P. Ren, V. S. Pande, J. D. Chodera, M. J. Schnieders, I. Haque, D. L. Mobley, D. S. Lambrecht, R. A. DiStasio Jr, *et al.*, *J. Phys. Chem. B*, 2010, **114**, 2549–2564.

31 Y. Shi, Z. Xia, J. Zhang, R. Best, C. Wu, J. W. Ponder and P. Ren, *J. Chem. Theory Comput.*, 2013, **9**, 4046–4063.

32 C. Zhang, C. Lu, Z. Jing, C. Wu, J.-P. Piquemal, J. W. Ponder and P. Ren, *J. Chem. Theory Comput.*, 2018, **14**, 2084–2108.

33 D. Jiao, P. A. Golubkov, T. A. Darden and P. Ren, *Proc. Natl. Acad. Sci. U. S. A.*, 2008, **105**, 6290–6295.

34 Y. Shi, C. Z. Zhu, S. F. Martin and P. Ren, *J. Phys. Chem. B*, 2012, **116**, 1716–1727.

35 R. Qi, Z. Jing, C. Liu, J.-P. Piquemal, K. N. Dalby and P. Ren, *J. Phys. Chem. B*, 2018, **122**, 6371–6376.

36 Z. Jing, J. A. Rackers, L. R. Pratt, C. Liu, S. B. Rempe and P. Ren, *Chem. Sci.*, 2021, **12**, 8920–8930.

37 Y. Shi, M. L. Laury, Z. Wang and J. W. Ponder, *J. Comput.-Aided Mol. Des.*, 2021, **35**, 79–93.

38 K. Arafet, N. Serrano-Aparicio, A. Lodola, A. J. Mulholland, F. V. González, K. Świderek and V. Moliner, *Chem. Sci.*, 2021, **12**, 1433–1444.

39 L.-H. Jolly, A. Duran, L. Lagardère, J. W. Ponder, P. Ren and J.-P. Piquemal, *Living Journal of Computational Molecular Science*, 2019, **1**, 10409.

40 O. Trott and A. J. Olson, *J. Comput. Chem.*, 2010, **31**, 455–461.

41 D. Santos-Martins, L. Solis-Vasquez, A. F. Tillack, M. F. Sanner, A. Koch and S. Forli, *J. Chem. Theory Comput.*, 2021, **17**, 1060–1073.

42 J. A. Rackers, Z. Wang, C. Lu, M. L. Laury, L. Lagardère, M. J. Schnieders, J.-P. Piquemal, P. Ren and J. W. Ponder, *J. Chem. Theory Comput.*, 2018, **14**, 5273–5289.

43 U. Essmann, L. Perera, M. L. Berkowitz, T. Darden, H. Lee and L. G. Pedersen, *J. Chem. Phys.*, 1995, **103**, 8577–8593.

44 L. Lagardère, F. Lipparini, E. Polack, B. Stamm, E. Cancès, M. Schnieders, P. Ren, Y. Maday and J.-P. Piquemal, *J. Chem. Theory Comput.*, 2015, **11**, 2589–2599.

45 L. Lagardère, F. Aviat and J.-P. Piquemal, *J. Phys. Chem. Lett.*, 2019, **10**, 2593–2599.

46 L. Lagardère, F. Lipparini, E. Polack, B. Stamm, E. Cancès, M. Schnieders, P. Ren, Y. Maday and J.-P. Piquemal, *J. Chem. Theory Comput.*, 2015, **11**, 2589–2599.

47 G. M. Morris, R. Huey, W. Lindstrom, M. F. Sanner, R. K. Belew, D. S. Goodsell and A. J. Olson, *J. Comput. Chem.*, 2009, **30**, 2785–2791.

48 G. L. Warren, C. W. Andrews, A.-M. Capelli, B. Clarke, J. LaLonde, M. H. Lambert, M. Lindvall, N. Nevins, S. F. Semus, S. Senger, G. Tedesco, I. D. Wall, J. M. Woolven, C. E. Peishoff and M. S. Head, *J. Med. Chem.*, 2006, **49**, 5912–5931.

49 R. T. McGibbon, K. A. Beauchamp, M. P. Harrigan, C. Klein, J. M. Swails, C. X. Hernández, C. R. Schwantes, L.-P. Wang, T. J. Lane and V. S. Pande, *Biophys. J.*, 2015, **109**, 1528–1532.

50 F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, *et al.*, *Journal of Machine Learning Research*, 2011, **12**, 2825–2830.

51 C. H. Bennett, *J. Comput. Phys.*, 1976, **22**, 245–268.

52 M. C. Sorkun, A. Khetan and S. Er, *Sci. Data*, 2019, **6**, 143.

53 K. Mansouri, C. M. Grulke, R. S. Judson and A. J. Williams, *J. Cheminf.*, 2018, **10**, 10.

54 D. Bajusz, A. Rácz and K. Héberger, *J. Cheminf.*, 2015, **7**, 1–13.

55 H. L. Morgan, *J. Chem. Doc.*, 1965, **5**, 107–113.

56 L. Zhang, D. Lin, X. Sun, U. Curth, C. Drosten, L. Sauerhering, S. Becker, K. Rox and R. Hilgenfeld, *Science*, 2020, **368**, 409–412.

57 Y. Liu, W. Kati, C.-M. Chen, R. Tripathi, A. Molla and W. Kohlbrenner, *Anal. Biochem.*, 1999, **267**, 331–335.

58 C. Ma, M. D. Sacco, B. Hurst, J. A. Townsend, Y. Hu, T. Szeto, X. Zhang, B. Tarbet, M. T. Marty, Y. Chen and J. Wang, *Cell Res.*, 2020, **30**, 678–692.

59 C. Dalvit, G. Fogliatto, A. Stewart, M. Veronesi and B. Stockman, *J. Biomol. NMR*, 2001, **21**, 349–359.

60 A. Mier, I. Maffucci, F. Merlier, E. Prost, V. Montagna, G. U. Ruiz-Esparza, J. V. Bonventre, P. K. Dhal, B. Tse Sum Bui, P. Sakhaii and K. Haupt, *Angew. Chem., Int. Ed.*, 2021, **60**, 20849–20857.

61 A. Douangamath, D. Fearon, P. Gehrtz, T. Krojer, P. Lukacik, C. D. Owen, E. Resnick, C. Strain-Damerell, A. Aimon, P. Ábrányi Balogh and others, *Nat. Commun.*, 2020, **11**, 1–11.

62 A. L. Kantsadi, E. Cattermole, M.-T. Matsoukas, G. A. Spyroulias and I. Vakonakis, *J. Biomol. NMR*, 2021, **75**, 167–178.

63 R. Cannalire, C. Cerchia, A. R. Beccari, F. S. Di Leva and V. Summa, *J. Med. Chem.*, 2020, DOI: 10.1021/acs.jmedchem.0c01140.

64 K. A. Brameld, B. Kuhn, D. C. Reuter and M. Stahl, *J. Chem. Inf. Model.*, 2008, **48**, 1–24.

65 J.-D. Chai and M. Head-Gordon, *Phys. Chem. Chem. Phys.*, 2008, **10**, 6615–6620.

66 V. A. Rassolov, J. A. Pople, M. A. Ratner and T. L. Windus, *J. Chem. Phys.*, 1998, **109**, 1223–1229.

67 M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, G. A. Petersson, H. Nakatsuji, X. Li, M. Caricato, A. V. Marenich, J. Bloino, B. G. Janesko, R. Gomperts, B. Mennucci, H. P. Hratchian, J. V. Ortiz, A. F. Izmaylov, J. L. Sonnenberg, D. Williams-Young, F. Ding, F. Lipparini, F. Egidi, J. Goings, B. Peng, A. Petrone, T. Henderson, D. Ranasinghe, V. G. Zakrzewski, J. Gao, N. Rega, G. Zheng, W. Liang, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, K. Throssell, J. A. Montgomery Jr, J. E. Peralta, F. Ogliaro, M. J. Bearpark, J. J. Heyd, E. N. Brothers, K. N. Kudin, V. N. Staroverov, T. A. Keith, R. Kobayashi, J. Normand, K. Raghavachari, A. P. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, J. M. Millam, M. Klene, C. Adamo, R. Cammi, J. W. Ochterski, R. L. Martin, K. Morokuma, O. Farkas, J. B. Foresman and D. J. Fox, *Gaussian ~16 Revision C.01*, Gaussian Inc, Wallingford CT, 2016.

68 M. Iwaoka, S. Takemoto and S. Tomoda, *J. Am. Chem. Soc.*, 2002, **124**, 10613–10620.

69 S. Kortagere, S. Ekins and W. J. Welsh, *J. Mol. Graphics Modell.*, 2008, **27**, 170–177.

70 M. R. Bauer, R. N. Jones, M. G. J. Baud, R. Wilcken, F. M. Boeckler, A. R. Fersht, A. C. Joerger and J. Spencer, *ACS Chem. Biol.*, 2016, **11**, 2265–2274.

71 K. El Hage, J.-P. Piquemal, Z. Hobaika, R. G. Maroun and N. Gresh, *J. Comput. Chem.*, 2013, **34**, 1125–1135.

72 X. Mu, Q. Wang, L.-P. Wang, S. D. Fried, J.-P. Piquemal, K. N. Dalby and P. Ren, *J. Phys. Chem. B*, 2014, **118**, 6456–6465.

73 J. Melcr and J.-P. Piquemal, *Front. Mol. Biosci.*, 2019, **6**, 143.

74 C. P. Mpamhanga, D. Spinks, L. B. Tulloch, E. J. Shanks, D. A. Robinson, I. T. Collie, A. H. Fairlamb, P. G. Wyatt, J. A. Frearson, W. N. Hunter, I. H. Gilbert and R. Brenk, *J. Med. Chem.*, 2009, **52**, 4454–4465.

75 G. Pérez-Hernández, F. Paul, T. Giorgino, G. De Fabritiis and F. Noé, *J. Chem. Phys.*, 2013, **139**, 07B604_1.

76 M. K. Scherer, B. Trendelkamp-Schroer, F. Paul, G. Pérez-Hernández, M. Hoffmann, N. Plattner, C. Wehmeyer, J.-H. Prinz and F. Noé, *J. Chem. Theory Comput.*, 2015, **11**, 5525–5542.

77 S. Martí, K. Arafet, A. Lodola, A. J. Mulholland, K. Świderek and V. Moliner, *ACS Catal.*, 2022, **12**, 698–708.

78 D. Loco, L. Lagardère, S. Caprasecca, F. Lipparini, B. Mennucci and J.-P. Piquemal, *J. Chem. Theory Comput.*, 2017, **13**, 4025–4033.

79 D. Loco, L. Lagardère, G. A. Cisneros, G. Scalmani, M. Frisch, F. Lipparini, B. Mennucci and J.-P. Piquemal, *Chem. Sci.*, 2019, **10**, 7200–7211.

80 K. Świderek and V. Moliner, *Chem. Sci.*, 2020, **11**, 10626–10630.