

Team name	SARStroopers
Team member(s)	Goutam Mukherjee ^{1,2} , Giulia D'Arrigo ³ , Eleonora Gianquinto ³ , Sandra Kovachka ³ , S. Kashif Sadiq ¹ , Daria B. Kokh ¹ , Athanasios-Alexandros Tsengenes ^{1,4,5} , Christina Athanasiou ^{1,4,5} , Abraham Muniz Chicharro ^{1,4,6} , Ariane Nunes-Alves ^{1,2} , Anton Hanke ^{1,4} , Giulia Paiardi ¹ , Jonas Gosse ^{7,8} , Simone Albani ^{7,8} , Benjamin Philipp Joseph ^{7,8} , Francesco Musiani ⁹ , Candida Manelfi ¹⁰ , Carmine Talarico ¹⁰ , Andrea Beccari ¹⁰ , Paolo Carloni ^{7,8,11} , Giulia Rossetti ^{7,12,13} , Francesca Spyrokis ³ and Rebecca C. Wade ^{1,2,4,14}
Affiliation	<p>1. Heidelberg Institute for Theoretical Studies (HITS), Schloss-Wolfsbrunnenweg 35, 69118 Heidelberg, Germany.</p> <p>2. Center for Molecular Biology (ZMBH), DKFZ-ZMBH Alliance, Heidelberg University, Im Neuenheimer Feld 282, 69120 Heidelberg, Germany.</p> <p>3. Department of Drug Science and Technology, University of Turin, Italy, via Giuria 9, 10125, Turin.</p> <p>4. Faculty of Biosciences, Heidelberg University, 69120 Heidelberg, Germany</p> <p>5 Heidelberg Biosciences International Graduate School (HBIGS), Heidelberg University, 69120 Heidelberg Germany</p> <p>6. Heidelberg Graduate School Mathematical and Computational Methods for the Sciences (HGSMathComp), Heidelberg University, 69120 Heidelberg, Germany</p> <p>7. Institute for Neuroscience and Medicine (INM-9) and Institute for Advanced Simulations (IAS-5) "Computational biomedicine", Forschungszentrum Jülich, 52425 Jülich, Germany.</p> <p>8. Faculty of Mathematics, Computer Science and Natural Sciences, RWTH Aachen, 52425 Aachen, Germany</p> <p>9. Department of Pharmacy and Biotechnology, University of Bologna, Viale Giuseppe Fanin 44, 40127, Italy</p> <p>10. R&D Platform Services, Dompè farmaceutici SpA Via Santa Lucia 6, 20122 Milano, Italy</p> <p>11. Institute for Neuroscience and Medicine (INM-11)"Molecular Neuroscience and Neuroimaging", Forschungszentrum Jülich, 52425 Jülich</p> <p>12. Jülich Supercomputing Center (JSC), Forschungszentrum Jülich, 52425 Jülich, Germany</p> <p>13. Department of Oncology, Hematology, Oncology, Hemostaseology, and Stem Cell Transplantation University Hospital Aachen, RWTH Aachen University, Pauwelsstraße 30, 52074 Aachen, Germany.</p> <p>14. Interdisciplinary Center for Scientific Computing (IWR), Heidelberg University, Im Neuenheimer Feld 368, 69120 Heidelberg, Germany.</p>
Contact email	Goutam.mukherjee@h-its.org
Contact phone number (optional)	+496221 533 247
Protein targets 3 required	3CIPro, PLpro, RdRp, Spike RBD

Section 1: methods & metrics

Describe what methods you have used, how they are independent from one another, what your workflow was, how you performed the cross-correlation between your methods. If applicable, please report estimated performance metrics of your methods, such as accuracy, sensitivity, false-discovery rate, etc., and how those metrics were obtained (e.g. cross-validation). Please provide key references if available.

Methods:

The computational pipeline is shown in **Figure 1**. The pipeline was carried out by three research groups, who independently conducted different parts of the pipeline, which included rapid virtual screening using several methods (RASPD+ (a machine learning method developed at HITS), ligand similarity searches and ROCS ligand pharmacophore screens), molecular docking with three independent methods (GOLD, Glide, FRED), molecular dynamics simulation of docked complexes (with analysis with our MD-IFP workflow), and ADMET screening (Qikprop). The results were then compared to identify consensus compounds that could be taken through the further steps of the pipeline (e.g. MD simulation and ADMET assessment). The pipeline allowed us to assign compounds to classes 1-7 according to their ranking as potential binders to the four targets studied. The top five compounds in class 1A were ranked. The compounds were initially screened for membership of one of the classes 3-7 and could then move up to classes 1-6, respectively, according to the results of the subsequent screening steps. The classes were defined as shown in Table 1.

Table 1. Definitions of the classes used for the selection and ranking of compounds

Class	Requirements
1A	Best compound selected after docking, MD simulation and ADMET
1B	Compounds identified by a ligand similarity search on class 1A compounds
1C	Known inhibitors that the pipeline allocated ‘blindly’ to class 1A
2A	Compounds from consensus of GOLD and Glide docking
2B	Compounds from consensus of Glide and FRED docking
3A	Compounds with good GOLD scores and poses but not taken further
3B	Compounds with good FRED scores and poses but not taken further
3C	Compounds with good Glide scores and poses but not taken further
5	Compounds with good ROCS scores from ligand-based screening
6A	Compounds selected using RASPD+ but not passing GOLD docking threshold
6B	Compounds docked with FRED and Glide but not passing the threshold scores
7	Compounds selected using RASPD+ but not taken further

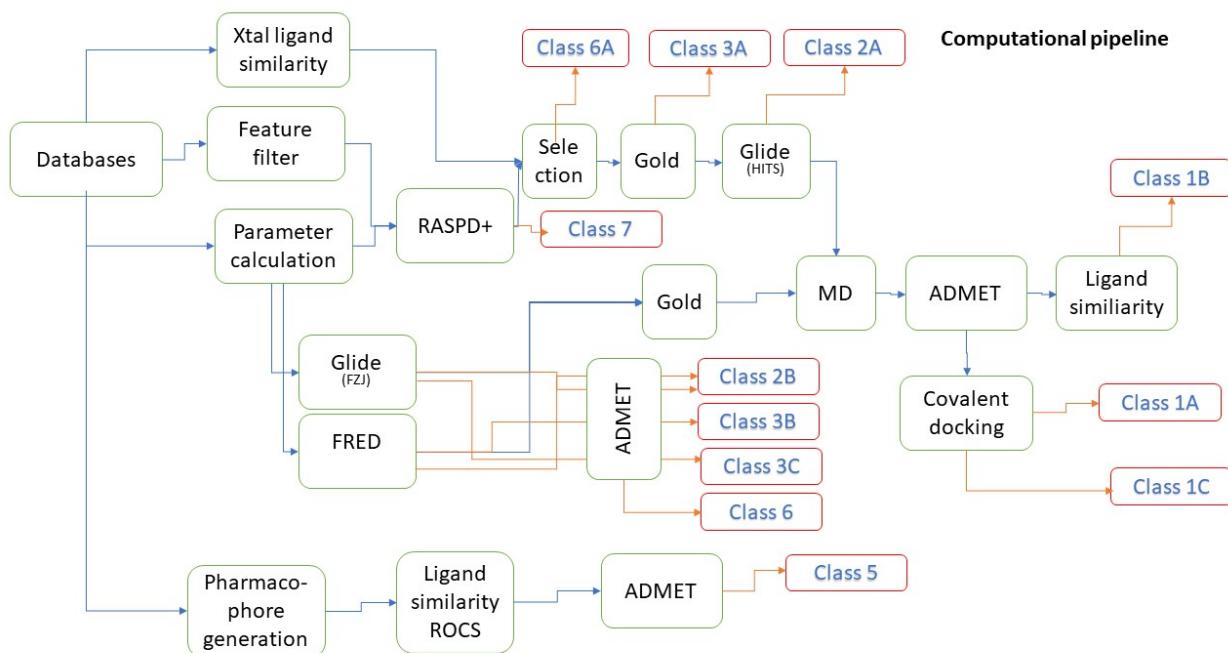


Figure 1. The computational pipeline employed in this study. Methodological steps are shown in the green boxes and the compound classifications as a result of these steps are indicated in red boxes.

The methodological details including validation steps and definition of target-specific criteria for compound selection are given in **Appendix 1**.

Section 2: targets

Describe for each protein target: why you chose it, from which source you obtained it (e.g., insidecorona.net / covid.molssi.org / rcsb.org) and why this is the best quality structure, if any pre-processing (e.g., energy minimization, residue correction, alternative folding, ...) was performed.

We focused on the virus proteins that were best structurally characterized as of May 2020: 3CL main protease (MP), the papain-like protease (PLpro), the RNA-dependent RNA polymerase (RdRp), and the spike receptor binding domain (RBD). We targeted the active sites of the three enzymes and the binding site on the RBD of the host cell receptor, ACE2, whose inhibition may lead to anti-viral activity. Binding sites were defined for the molecular docking analyses using Flapsite, paying attention to the distribution and nature of the corresponding Molecular Interaction fields³⁴.

The variation of the binding site shape and druggability due to conformational flexibility was analysed for MP and the RBD using the TRAPP workflow⁴⁵

Target 1: 3CL main protease (MP).

The MP structure of SARS-CoV-2 complexed with an α -ketoamide inhibitor was used as template for virtual screening, molecular docking and MD simulations (PDB ID 6Y2G; 2.2 Å resolution¹²). The missing C-terminal residues were reconstructed with Sybyl, extending both monomers up to Gln306, according to the identified site of auto-cleavage³⁷. Each monomer had a different conformation of the C-terminus: in one case, the C-terminus pointed away from the binding site and in another it lined the binding site. Both monomers A and B were used for molecular docking analyses. For MD

simulations, the catalytic histidine 41 was modelled in the HID tautomeric state to simulate the condition in which Nε can sequester the catalytic Cys145 proton and allow the cysteine to perform a nucleophilic attack. All other histidines were assumed to be in the HIE tautomeric state.

In addition, the structure PDB ID 6LU7³⁸ was used for docking with FRED and Glide . The TRAPP analysis for MP (Hanke et al, in preparation) supported the choice of the two crystal structures used for docking. The crystal structures 6Y2G and 6LU7 had high druggability scores but it was found that variations in the conformations of the N- and C-terminii and the loops lining the pocket could further affect druggability.

Target 2: papain-like protease (Plpro)

The structure of SARS-CoV-1 Plpro co-crystallized with a naphthalene-based inhibitor³⁹ (PDB ID 3MJ5; 2.63 Å resolution), and that of SARS-CoV-2 Plpro co-crystallized with the peptide inhibitor VIR250 (PDB ID 6WUU; 2.79 Å resolution) were used as templates for VS, molecular docking and MD simulations. The structure of the SARS-CoV-1 Plpro was used due to the absence of SARS-CoV-2 Plpro structures co-crystallized with non-peptidomimetic inhibitors, and to the significant induced fit effect exerted by the ligand binding on the Thr265-His273 loop. In particular, residues Tyr268 and Gln269 assume quite different positions that can stabilize the binding of the ligand in the orthosteric site. We carefully checked the sequence similarity of Plpro from SARS-CoV-1 and SARS-CoV-2 and found 100% identity of the binding site residues. On the other hand, the SARS-CoV-2 Plpro co-crystallized with the peptide inhibitor (PDB ID 6WUU) was used (with the inhibitor removed) to screen and dock peptidomimetic ligands that would be difficult to accommodate in the structure with PDB ID 3MJ5 because of the different binding site conformation.

Even though the protein has been reported as a trimer in some crystal structures, only one monomer (chain A) was used. The enzyme is functional as a monomer and responsible for the cleavage of three recognized junctions in the replicase polyprotein. Furthermore, it recognizes two ubiquitin binding domains (UBDs) that would be completely excluded sterically from the active site in any oligomeric structure⁴⁰. The dimeric and trimERIC oligomeric states of Plpro observed in crystal structures are therefore not relevant to the physiological mechanism of the virus but rather arise during the crystallization process⁴¹.

Plpro has two structural zinc ions positioned remote from the binding site. The coordinating cysteines in MD simulations were deprotonated and named CYM. All other ionizable residues were modelled at physiological pH.

Target 3: RNA-dependent RNA polymerase (RdRp)

We retrieved a model of the SARS-CoV-2 apo-RdRp from the SWISS-MODEL website. This model was built using the structure with PDB ID 6M71 (determined by cryo-EM with a resolution of 2.9 Å)⁴² as a template, and lacks the nsp7 and nsp8 cofactors. At the time of our modelling, no SARS-CoV-2 RdRp had been co-crystallized with either RNA or Mn²⁺ ions. As RdRp divalent metals and their chelating residues are conserved among retroviral species⁴³, we took as a reference the metal coordination reported in RdRp structures co-crystallized with nucleotide analogues from Norovirus and HCV (PDB IDs: 3H5Y, 4WTG and 1GX6, respectively). Moreover, we inserted the RNA primer and template duplex from the structure with PDB ID 3H5Y in our model by structural alignment. Thus, the final RdRp model comprised the Nsp12 polypeptide chain, two Mn²⁺ ions and an RNA double strand.

Target 4: spike receptor binding domain (RBD)

The RBD coordinates obtained from the structure of its complex with human ACE2 (PDB ID 6M0J; 2.45 Å resolution)⁴⁴ was used for virtual screening, molecular docking and MD simulations. The structure of the RBD-ACE2 complex was inspected to identify the interface region, defined as

stabilized by thirteen hydrogen bonds and the key viral residues involved in the interactions. This analysis led to the definition of a putative binding crevice lined by residues Tyr449, Gln493, Gln496, Gln498, Asn501 and Tyr505 of the RBD. The initial screen with RASPD+ was performed for this region for the complete RBD-ACE2 complex and for the RBD alone with the ACE2 protein removed. For the subsequent docking and simulations, only the RBD was retained as we aimed to identify small molecules capable of interfering with formation of a complex with the host ACE2. Four disulfide bridges were modelled for the MD simulations with the cysteines named CYM. The TRAPP analysis of the RBD using the structure of the RBD from the RBD-ACE2 complex PDB ID 6M0J showed that the binding site had a low druggability score in the crystal structure. The shallow crevices in the binding site region were highly flexible due to loop mobility. Higher druggability could be obtained by exposing hydrophobic regions and stabilizing deeper subpockets.

Section 3: libraries

Describe which libraries you have used, how they were combined, if any compounds were removed / added, why additions are relevant, any unique features of your library, etc. Please provide the sources you obtained the libraries from (if publicly available). Describe the procedure of data preparation (removal of duplicates, standardization, etc). Indicate if different libraries were used for different targets, and why. If possible, provide a download link to your version of the library.

Library 1:

We screened compounds from the following sources: (1) ZINC15 database^{1,2} (<https://zinc.docking.org/>). (2) Drugbank³; (3) Enamine (<https://enamine.net/>); (4) Vitas-M STK - stock of HTS compounds for pharmaceutical and agrochemical research collection (<https://vitasmab.biz/compound-libraries/stock-libraries>); (5) Specs (<https://www.specs.net/>); (6) SWEETLEAD⁴; (7) Merck (<https://www.aldrichmarketselect.com> on 28.04.2020); (8) Pubchem⁵; (9) CAS Covid-19 antiviral candidate compounds dataset (<https://www.cas.org/covid-19-antiviral-compounds-dataset>); (10) Fraunhofer Broad dataset (<https://www.ime.fraunhofer.de>); (11) GHDDI (<https://ghddi-ailab.github.io/Targeting2019-nCoV/preclinical/>); (12) Literature reporting compounds binding to MP.⁶⁻³² In total, these libraries contained about 800 million compounds.

The **table in accompanying Excel file (Appendix 2)** shows the numbers of compounds screened from these libraries (after prefiltering) for the four targets. For the Zinc 15 database, the 3D-coordinates in mol2 format of 11760023 protomers in the “in-stock” tranches were downloaded. For the DrugBank dataset, the 3D-coordinates in mol2 format of 8687 molecules were retrieved from the ZINC database DrugBank catalog vendors collection (DrugBank-approved (3722), DrugBank-experimental (4963) and DrugBank-investigational (3839)). A total of 653691 molecules were obtained from the Enamine data set, which were filtered according to logD values using MoKa³³ (logD<0, for the purpose of identifying protein-protein interface inhibitors), and processed with FLAP³⁴ to generate the three-dimensional coordinates, resulting in 109728 compounds. The same treatment was applied to the SPECS dataset, resulting in 8316 compounds for screening. 1413073 molecules were retrieved from the Vitas M data set of in-stock compounds and processed with FLAP to generate the three-dimensional coordinates, resulting in 1215468 compounds.

All 9127 compounds from the SWEETLEAD dataset were retrieved. 8498849 compounds were provided in the MERCK dataset by Jedi (<https://www.aldrichmarketselect.com> on 28.04.2020). 102761607 compounds from PubChem were considered. The CAS Covid-19 antiviral candidate compounds dataset contained 46541 compounds at the time of retrieval. The Frauenhofer Broad dataset of 9329 compounds is a set of compounds available to the Frauenhofer IME Screeningport for screening, particularly with the aim of drug repurposing. GHDDI provides a listing of in-vitro

preclinical data on assays of compounds against the SARS-CoV-2 virus and its targets. At the time of this work, data on 986 compounds targeting MP and 94 compounds targeting PLpro were retrieved. In addition, we created a dataset of 130 compounds from the literature with reported activity against MP. All compounds in the Merck, SWEETLEAD, CAS, Fraunhofer and GHDDI and literature datasets were converted to 3D coordinates and prepared with Schrödinger's LigPrep tool³⁵. This process generated multiple states for stereoisomers, tautomers, ring conformations (1 stable ring conformer by default) and protonation states. The Schrödinger package Epik was used to assign tautomers and protonation states that would be dominant at a selected pH range (pH=7±1). Ambiguous chiral centers were enumerated, allowing a maximum of 32 isomers to be produced from each input structure. Then, an energy minimization was performed with the OPLS3 force field³⁶.

Section 4: results

Briefly describe your key findings, any interesting trends in your data, a description of your top 5 compounds for each target. If possible, provide a link to a code and/or data repository. Please do not submit randomly selected compounds!

Results:

For each of the four targets, a final list of 10000 putative ligands is given in the accompanying **Excel sheet** and **csv** files. The ligands are assigned to the classes described above with the most promising compounds being in class 1 and the top five compounds being listed first in class 1A. We started considering datasets containing a total of about 800 million compounds, but after a simple filtering were left with about 100 million compounds for target-specific screening. The pipeline incorporated methods of different complexity and similar methods that were applied independently to identify consensus compounds, which were subjected to a hierarchy of tests. We performed molecular docking with three different methods for thousands of molecules and MD simulations for nearly 100 protein-ligand complexes. At the end, we identified the following numbers of compounds in class 1: Mpro: 1A: 10, 1B: 54, 1C: 3; PLpro: 1A: 7, 1B: 87; RdRp: 1A: 11, 1B: 30; RBD: 1A: 7, 1B: 115.

Target 1: 3CL main protease (MP).

We identified the main methodological challenges for inhibitor design against MP, homodimeric cysteine protease, as: (I) accounting for covalent binding, (II) accounting for the effects of motion of the C-terminal residues and the flexible loops lining the binding site. A large number of putative inhibitors were identified from the virtual screening procedure. As proteases are established drug targets, the libraries contained known protease inhibitors and many peptidomimetics. A number of these were identified as potential covalent and non-covalent inhibitors. Covalent inhibitors were required to have appropriate reactive groups and to be positioned by docking in the vicinity of Cys145. Considering the known inhibitor, N3, we also required a subset of the compounds to contain a reactive double bond and a lactam ring that fits in a subpocket of the binding site.

The 5 top-ranked compounds are shown in **Figure MP-1**. The first three are peptide-like and are expected to form a covalent bond to Cys145. The first two contain a lactam ring, and a common scaffold. The compounds have substructure similarities to N3 in the crystal structure 6LU7 but have a shorter peptidic backbone. The second compound shares a common phenyl ring substituent with N3. Their superposition with the covalently bound N3 ligand is shown in **Figure MP-2**. All three compounds showed stable binding modes with good interactions in the MD simulations. The 4th and 5th ranked compounds are expected to be non-covalent inhibitors. The 5th ranked compound is smaller but the 4th ranked compound displays a similar number of stable hydrogen-bonding

simulations (e) are reported. The RMSD and RMSF values were computed for all non-hydrogen atoms and all atoms, respectively.

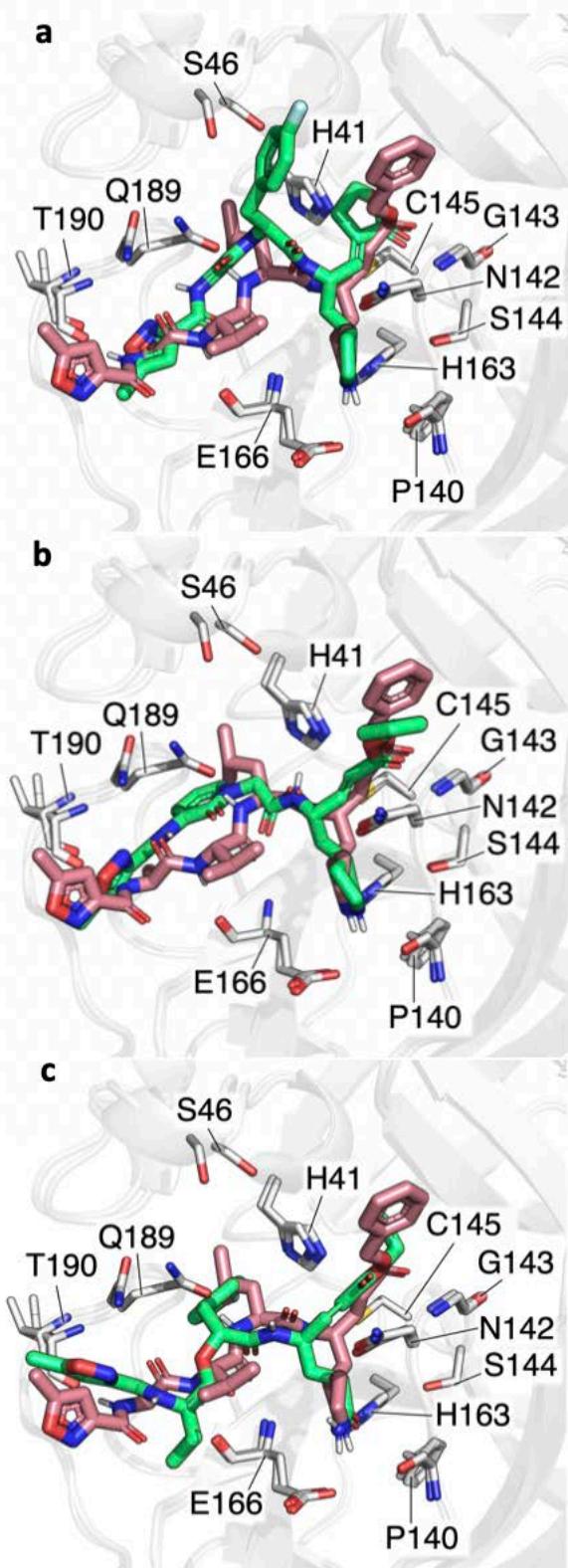


Figure MP-2. Superposition of first 3 ranked compounds for MP (which could make a covalent bond with Cys145) with the covalent N3 ligand (PDBID 6LU7). **a.** First top ranked compound (P70). **b.** Second top-ranked compound (C36). **c.** Third top ranked compound (P64). The compounds docked non-covalently are colored salmon, the covalently bound N3 is colored green.

Target 2: papain-like protease (PLpro)

We found that the main methodological challenges for inhibitor design against PLpro, a monomeric cysteine protease were: (I) accounting for covalent binding, (II) accounting for the effects of motion of the flexible loops and side-chain (especially aromatic) rotamers around the binding site. A large number of compounds were identified as putative binders. The active site Cys111 is present at one end of a long crevice and few of the ligands came in close proximity with this Cys. As inhibitors of the SARS-CoV1 PLPro containing a naphthalene moiety had been described, we screened for compounds containing this moiety and identified a large number of compounds with this moiety that could be well accommodated in the binding site between Pro249 and Tyr269. Considerable adjustment in the flexible loop over the binding site and the tyrosine residues lining the binding site (Tyr265, Tyr269, Tyr274) was observed during simulations, suggesting possibilities for induced fit.

The 5 top-ranked compounds are shown in **Figure PLpro**. The top ranked compound is peptide-like and the second and third ranked compounds each contain a peptide bond. The third ranked compound contains a naphthalene double fused ring. The second and fourth ranked compounds contain naphthalene with two substituents. The first and fifth ranked compounds also contain fused ring systems that are predicted to occupy a similar part of the binding site. The number of contacts was high and included aromatic interactions maintained during MD simulations of the first four compounds. The fifth compound made fewer contacts that decreased during MD, suggesting less stable binding. All five compounds contain a rather centrally positioned positively charged quaternary nitrogen that made a salt-link with Asp165 carboxylate group in docking poses for all compounds except for the 4th ranked one.

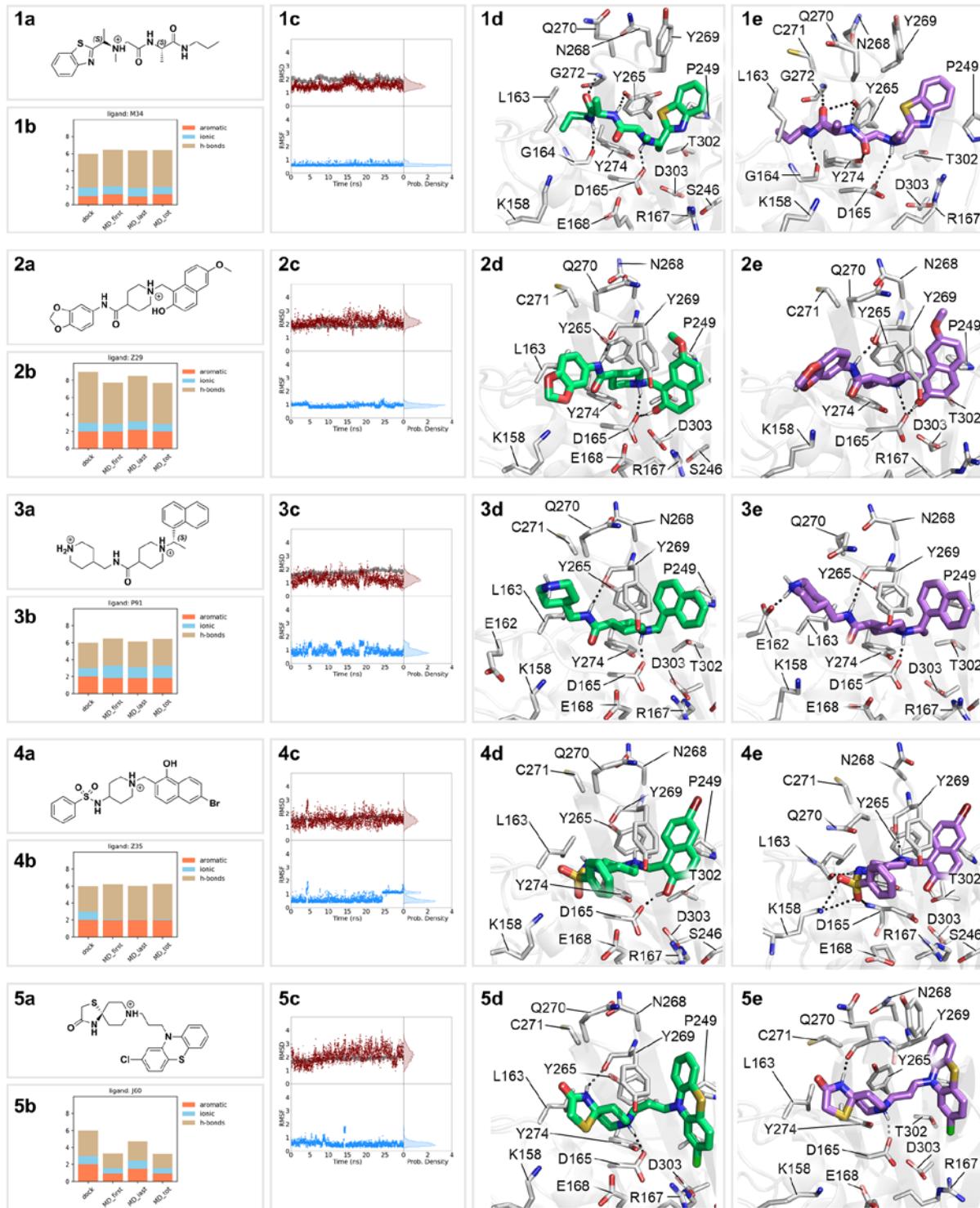


Figure PLpro. Panel of the 5 top-ranked compounds selected for Plpro, reported from 1 to 5 according to the ranking. For each line the compound's 2D chemical structure (a), the fingerprint indicating the protein-ligand interactions (b), the variation of the RMSD of the protein (gray) and the ligand (crimson) and the RMSF of the ligand (blue), during the production simulation, and the corresponding probability density for the ligand (c), the docking pose (d) and the complex at the end of the MD simulations (e) are reported. The RMSD and RMSF values were computed for all non-hydrogen atoms and all atoms, respectively.

Target 3: RNA-dependent RNA polymerase (RdRp)

We found that the main methodological challenges for inhibitor design against RdRp were: (I) accounting for covalent binding or base-stacking interactions with the RNA, (II) correctly describing metal ion coordination, and (III) the relatively poor resolution of the target model. Notably, Structures have been determined by cryo-EM and the model used is based on structures from other viruses. For the polymerase (RdRp: PDBID 6M71), no molecules with experimental IC₅₀ were found in GHDDI. However, Remdesivir is a drug approved for use in the clinic. The structure of its complex with RdRp (PDBID 7BV2) has been solved but was not suited as a target for virtual screen as Remdesivir was incorporated in the RNA and the metal ions were not well resolved (see **Appendix 1**).

Due to the presence of the two manganese ions in the binding site, we required selected compounds to contain negatively charged moieties, such as carboxylate or phosphate groups. We also favored compounds that could pack well against the RNA. These requirements were essentially independent of the rest of the protein structure which was of moderate resolution. The compounds showed quite a large range in size as the pocket was rather large with the RNA and metal ions adjacent to one "corner" of the pocket.

The 5 top-ranked compounds are shown in **Figure RdRp**. The third ranked compound contained three negatively charged functional groups while all others compounds in the top 5 contained two negatively charged functional groups: two carboxylate groups in the compounds ranked first and second; one carboxylate in the third ranked compound, one sulfate in the fourth ranked compound and two sulfates in the fifth ranked compound. The metal ions were coordinated by at least one of these negatively charged moieties and both ions were coordinated by oxygen atoms from the ligands. No direct interactions were made with the structural water molecule. All compounds made van der Waals or aromatic ring interactions with the RNA but H-bonds were only observed for the first compound. The binding site is surrounded by positively charged residues and these were able to interact with the negatively charged regions of the ligands as well. All compounds maintained or increased contacts during MD simulations and approximately retained their binding pose. The compounds had much worse scores for drug-likeness in the QikProp analysis than for the other targets. This was primarily due to the polar nature of the binding site, resulting in charged ligands being most complementary. However, the 4th compound is an approved drug (ZINC3830428, Cefonicid, a second-generation cephalosporin antibiotic that binds bacterial penicillin-binding proteins) and the first two compounds are investigational compounds (the moxalactam antibiotic is ranked 1st and pelitrexol is ranked 2nd).

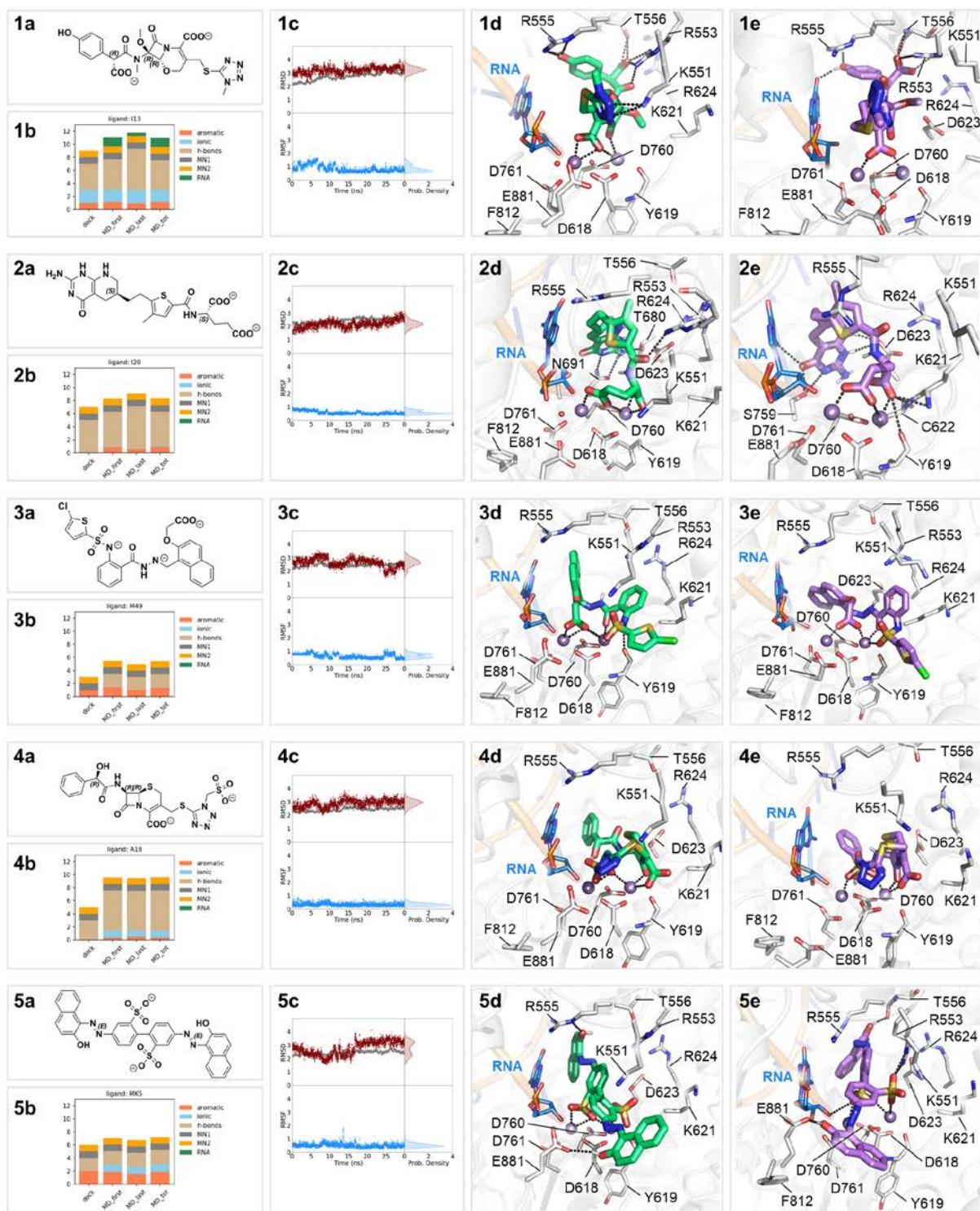


Figure RdRp. Panel of the 5 top-ranked compounds selected for RdRp, reported from 1 to 5 according to the ranking. For each line the compound's 2D chemical structure (a), the fingerprint indicating the protein-ligand interactions (b), the variation of the RMSD of the protein (gray) and the ligand (crimson) and the RMSF of the ligand (blue), during the production simulation, and the corresponding probability density for the ligand (c), the docking pose (d) and the complex at the end of the MD simulations (e) are reported. The RMSD and RMSF values were computed for all non-hydrogen atoms and all atoms, respectively.

Target 4: spike receptor binding domain (RBD)

The RBD is a domain of the spike protein S1 subunit. We targeted the interface with ACE2 observed in the crystal structure of their complex determined at 2.45 Angstroms resolution (PDB ID 6M0J), shown in **Figure RBD-1**. The main methodological challenges for inhibitor design against the spike RBD were: (I) the relative flatness of the binding site with only shallow flexible crevices, (II) the flexibility of the binding site containing protein loops and sidechains that can switch rotamers, (III) the relatively polar nature of the site, (IV) the lack of known small-molecule inhibitors that bind to the target site.

We screened libraries of relatively polar compounds ($\log D < 0$) with RASPD+ against the RBD-ACE2 complex and the RBD only, and then docked the compounds. We identified relatively small compounds that could fit in the small crevices in the binding site and make hydrogen-bonds. MD simulations showed that the docked positions were rather unstable. Small hotspot pocket(s) were identified where binding was usually retained despite ligand mobility during the MD simulations. The existence of these hotspots was supported by the TRAPP analysis. However, in replica MD simulations the ligands often explored different binding positions. The compounds identified could be considered as fragments, probably with weak affinity, that could provide the basis for fragment-based compound design.

The 5 top-ranked compounds are shown in **Figure RBD-2**. The top 5 compounds had quite linear structures with single ring systems that docked into shallow crevices on the surface. The MD simulations indicated, e.g. for the first ranked compound, that a 5-membered ring could fit stably into a small subpocket, while the rest of the structure was highly mobile. The fourth and fifth ranked compounds contained a positively charged moiety. This interacted with Glu406 in the pocket but, after MD simulation of the 4th ranked compound, moved to the solvent. Putting all docking poses obtained for the different ligands together indicated two crevices joined at the small pocket next to Tyr505 and Phe497, suggesting a possibility for connecting the compounds in a fragment-based approach.

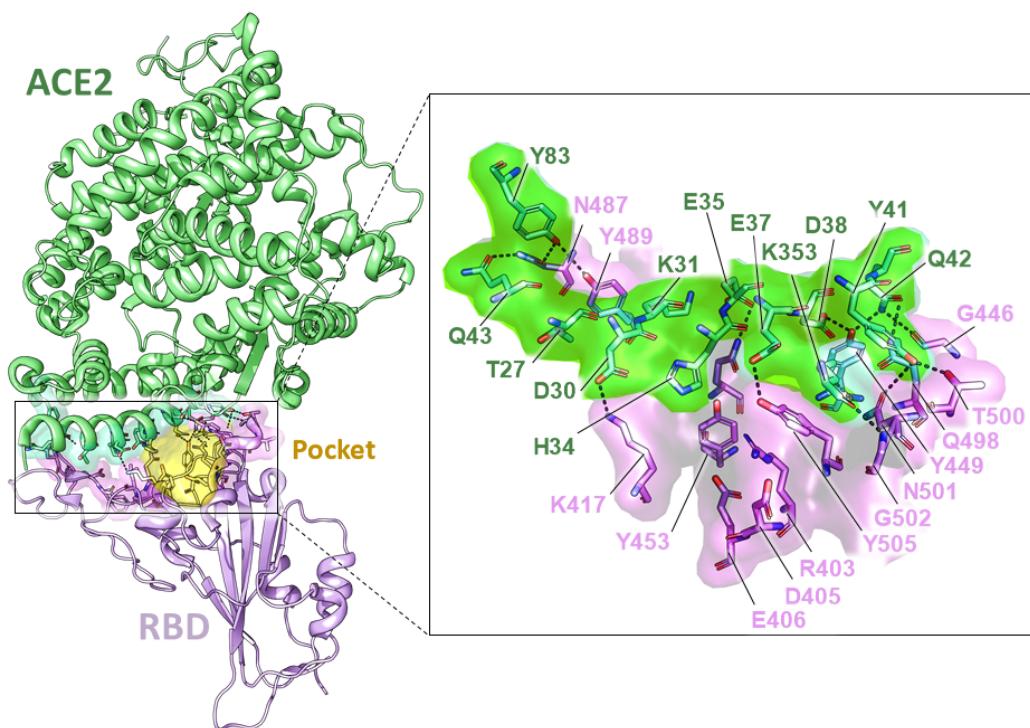


Figure RBD-1. RBD-ACE2 complex. The interface region targeted by our screening/docking experiments is shown in the inset. RBD and ACE2 are colored lilac and green, respectively.

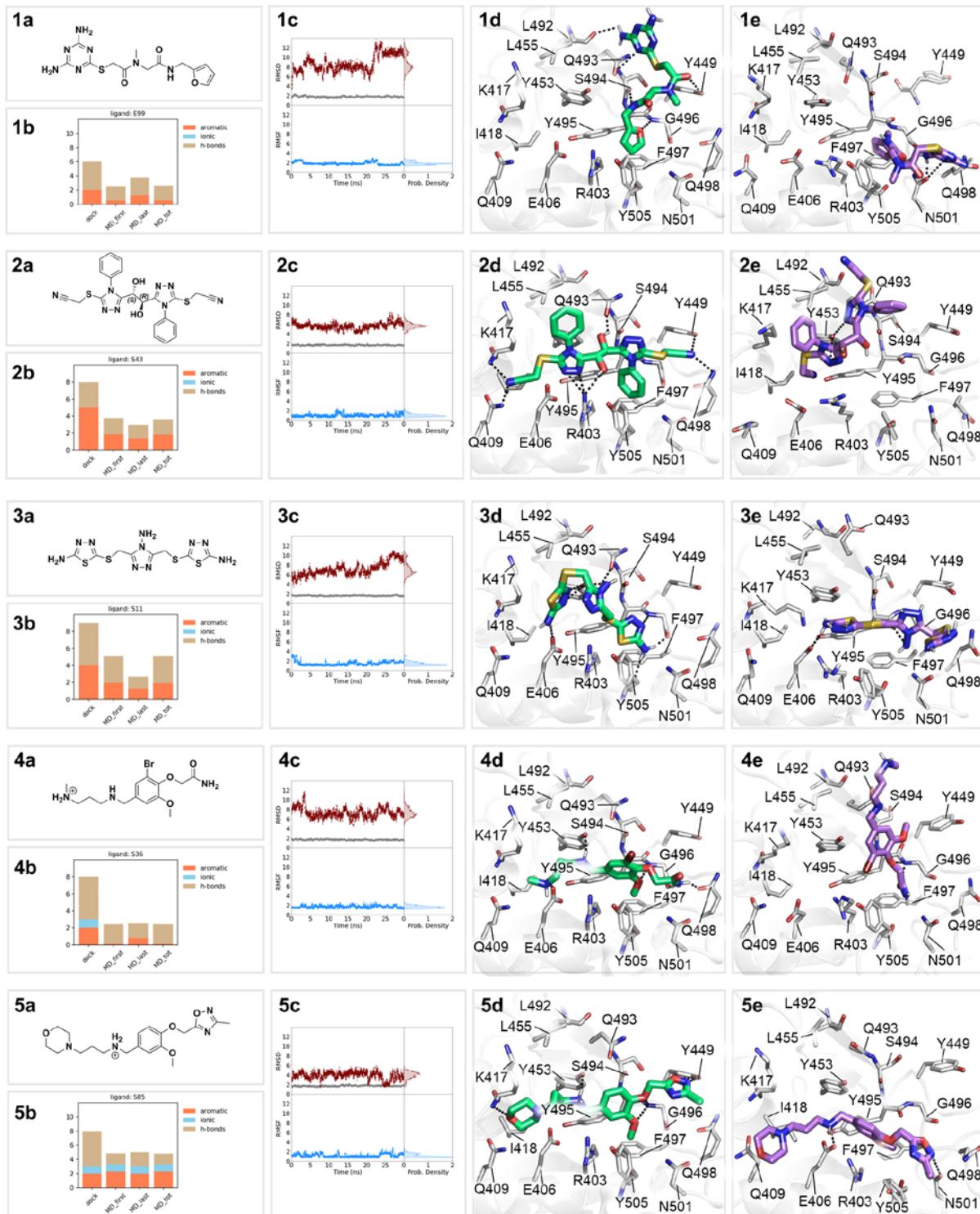


Figure RBD-2. Panel of the 5 top-ranked compounds selected for RDB, reported from 1 to 5 according to the ranking. For each line the compound's 2D chemical structure (a), the fingerprint indicating the protein-ligand interactions (b), the variation of the RMSD of the protein (gray) and the ligand (crimson) and the RMSF of the ligand (blue), during the production simulation, and the corresponding probability density for the ligand (c), the docking pose (d) and the complex at the end of the MD simulations (e) are reported. The RMSD and RMSF values were computed for all non-hydrogen atoms and all atoms, respectively.

Other comments:

Summary:

We applied a computational pipeline to identify small molecule hits against three virus enzymes - the 3CL main protease, the papain-like protease, and the RNA-dependent RNA polymerase - and the spike receptor binding domain of the virus. We employed a multifaceted, hierarchical pipeline to screen several libraries containing about 800 million compounds from different sources. The different steps of the pipeline, including rapid physiochemical property-based screening, docking and molecular dynamics simulation, were validated for the given targets using known inhibitors, when available. For each target, we identified 7-11 particularly promising compounds that satisfied all our computational tests for target binding. For each target, we also made a ranked list of 10000 compounds that are putative binders, that satisfy the early stages of the pipeline or parts of the pipeline.

Data availability:

All relevant data for this work have been deposited on Sciebo:

<https://fz-juelich.sciebo.de/s/j3S598yadmudMow>

Acknowledgements

The work carried out at HITS was supported by the Klaus Tschira Foundation. The authors Giulia Rossetti, Paolo Carloni, Carmine Talarico, Andrea Beccari, Candida Manelfi and Jonas Gossen acknowledge the project “EXaSCale smArt pLatform Against paThogEns for Corona Virus – Exscalate4CoV” funded by the EU’s H2020-SC1-PHE-CORONAVIRUS-2020 call, grant N. 101003551. Giulia Rossetti gratefully acknowledges OpenEye for providing an academic license. Francesca Spyros kindly acknowledges the Centro di Competenza sul Calcolo Scientifico (C3S) at the University of Turin (c3s.unito.it) for providing the computational time and resources, BiKi Technologies for providing the BiKi LiFe Sciences suite and Molecular Discovery Ltd for providing the FLAP software and for supporting GD.

Appendix 1. Methodological details, including validation.

Initial filtering and screening of the compound libraries

Three initial screening methods have been applied:

- (i) Filtering and screening by RASPD+
 - (ii) Screening by ligand similarity
 - (iii) Filtering followed by docking
-
- (i) Depending on the properties of the target and the library, different filters, e.g. to require the presence of a specific chemical fragment or a net charge, were first defined. These filters are listed in Table 1. For example, for MP, based on the inhibitor in the structure PDB ID 6LU7,

some searches were done that required γ -lactam (C1CC(=O)NC1) and vinyl-methyl acetate (VMA; CCCC=CC(=O)OC) moieties to be present, and required that the Wiener index ranged between 4000 and 4500. Then, the libraries were screened using the RASPD+ method⁴⁸ to identify compounds with properties complementary to that of the protein binding sites. This method is fast because it does not involve doing a docking of the ligands to the protein binding sites. RASPD+ screening was performed using 7 machine learning methods. Compounds were selected for docking with GOLD, and therefore assigned to class 6A, based on threshold binding free energy scores that were chosen by validation against known binders for the studied targets. For each of the screens of the individual datasets with specific filters, the resultant selected compounds were taken forward to docking with GOLD if the number of selected compounds was less than 1000. If the number exceeded 1000, then these compounds were retained in class 7, along with the compounds that did not satisfy the threshold binding free energy.

- (ii) A ligand similarity search was carried out for screening MP against PubChem based on similarity to the N3 ligand in the structure with PDB ID 6LU7. The search was carried out using the PubChem website by entering the SMILES string of the N3 inhibitor. Similar compounds from this search were required to have a molecular weight of 400 to 900 Da for docking with GOLD.
- (iii) The best conformers of compounds were selected and these were filtered for satisfying the Lipinski rule of 5 before docking with FRED and Glide.

Non-covalent Docking

Docking was performed with three programs (i) GOLD, (ii) Glide and (iii) FRED.

- (i) **GOLD:** Molecular docking was performed using the GOLD suite, version 5.5⁴⁹. The default GOLD parameters were used for all targets, and the compounds were subjected to 40 genetic algorithm runs using the CHEMPLP fitness function, retaining the 5 top ranked poses. Docking was performed without any water molecules in the binding site, apart from RdRp, for which one water molecule coordinating one of the manganese ions was retained. Binding pockets were defined as spheres centered on a centroid atom. The sphere radius was defined according to the volume and spatial properties of the binding site.

The following centroids and pocket radii were used for the different targets: MP: His164 O, radius 12 Å; Plpro: Tyr265 CB, radius 9 Å; RdRp: Cys592 H, radius 12 Å; Spike RBD: Tyr495 CB, radius 13 Å.

The GOLD scores were combined in a target-specific weighted function (GoldGLOB). All molecules were first visualized in order to assign a heuristic cutoff value of the GoldGLOB score so that unsuitable entities for the targets were discarded (and retained in class 6A) whereas candidates with a GoldGLOB score above the fixed GoldGLOB threshold value were classified as class 3A. The GoldGOLB score was computed from the following equation.

$$\text{GoldGLOB} = ((x-a) + ((y-b)*k) + ((z-c)*w))$$

x = Gold Score

a = Gold Score threshold

$y = S(Hbond)$
 $b = S(Hbond) \text{ threshold}$
 $k = H\text{-bonds weight}$
 $z = S(\text{metal})$
 $c = S(\text{metal}) \text{ threshold}$
 $w = \text{metal coordination weight}$

The last term was only non-zero for RdRp and allowed the contribution of the metal coordination to be considered. The parameters are given in **Table GoldGLOB**.

Table GoldGLOB: Parameters for computing the GoldGLOB score

	MP	Plpro	RdRp	RBD
GoldGLOB cutoff	-33.51	74.26	8.61	8.21
a	75	70	90	55
b	2	-	2	3.5
k	40	40	10	20
c	NA	NA	8	NA
w	NA	NA	10	NA

The molecules passing the GoldGLOB threshold were then subjected to a careful visual inspection of the docking poses. Molecules that appeared too large or small, or had an unfavourable core, or were outside the defined site were not selected for consensus docking with Glide. Furthermore, in libraries containing several variations of the same scaffold, only the top-ranked and chemically diverse derivatives were forwarded to the next docking phase with Glide and, therefore, for consideration for class 2A.

GOLD was also used to check the binding pose of compounds allocated to class 2B that had a consensus pose from Glide and Fred (see below). If the pose was reproduced in GOLD, these compounds were selected for MD simulation.

The GOLD docking procedure was validated by performing redocking. For two structures of MP co-crystallized with the non-covalent ligand X77 (PDB ID 6W63) and with a flavonoid-like inhibitor (PDB ID 6M2N), the RMSDs of the best ranked poses with respect to the crystallographic ligand were 1.34 Å and 1.25 Å (**Figure GOLD a,b**), respectively. For other MP structures co-crystallized with mechanism-based covalent ligands (PDB ID 7BQY, having the same N3 ligand as co-crystallized in 6LU7³⁸, PDB ID 6Y2F and 6Y2G, respectively¹²), the redocking was performed for the ligands in the configuration prior to covalent bond formation and comparable poses to the crystallographic poses were obtained (**Figure GOLD c,d**). For Plpro, an RMSD of 0.83 Å was obtained for redocking of the naphthylmethyl derivative 15h (PDB ID 3MJ5³⁹). The covalently bound peptide inhibitor VIR250 was also successfully redocked non-covalently in the corresponding crystallographic structure (PDB ID 6WUU) using the configuration prior to covalent bond formation (**Figure GOLD e,f**). No such

validation was possible for RdRp, for which the only complex available with an inhibitor was the structure determined by cryo-EM with remdesivir covalently incorporated in the RNA.⁵⁰ However, docking calculations were done with remdesivir to assess the compatibility of the non-covalent docking modes with the experimentally determined structure. In particular, the quality of the coordination of the metal ions and the interaction with the RNA were used as criteria for evaluating docking poses by visual inspection (**Figure GOLD g,h**). For the spike RBD, no experimentally determined structures of complexes with small molecules were available for target-specific validation.

A further validation of the GOLD docking procedure was carried out by docking compounds to MP that have a measured inhibitory activity towards SARS-CoV MP. Given the high similarity of the MP binding site in SARS-CoV and SARS-CoV-2, it is reasonable to expect that these compounds are active towards SARS-CoV-2 MP as well, even if to a different extent. 52 compounds with measured activities corresponding to a binding free energy better than -9 kcal/mol were extracted from the GHDDI dataset (https://ghddi-ailab.github.io/Targeting2019-nCoV/CoV_Experiment_Data/) and docked in the MP orthosteric site. Using the GoldGLOB score as described above, 83% of the compounds would have been classified in class 3A, that is as good binders in the GOLD procedure and suitable for subsequent analysis.

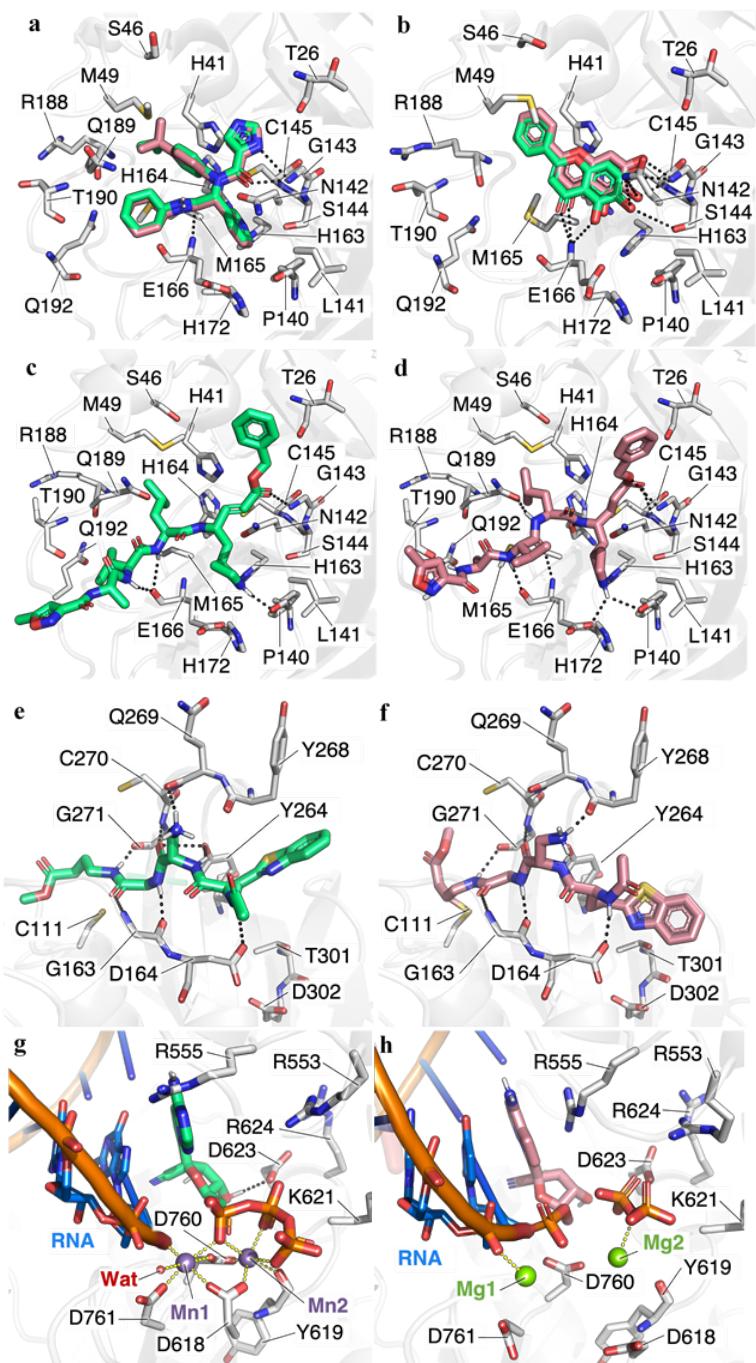


Figure GOLD: Redocking results with the GOLD non-covalent docking procedure. **a-d** MP: Redocking of the non-covalent ligand co-crystallized in PDB **a.** 6W63. **b.** 6M2N, and comparison of the docking of the covalent ligand N3 in MP (**c**), prior to the formation of the covalent bond with the crystallographic covalent complex (**d**; PDB ID 7BQY). **e-f** Plpro: Comparison of the docking of the covalent peptidomimetic inhibitor VIR250 prior to the formation of the covalent bond (**e**) with the crystallographic covalent complex (**f**; PDB ID 6WUU). **g-h** RdRp: Comparison of the **g** docking of remdesivir prior to its inclusion in the RNA in the triphosphate form (**g**) with the crystallographic covalent complex (**h**; PDB ID 7BV2). The protein is shown in gray cartoon representation with important residues shown in stick representation and colored by atom type. The experimentally observed co-crystallized ligands are shown in salmon colored stick representation and the redocked ligands are shown in green stick representation. In RdRp, the RNA, metal ions and structural water molecule are labelled.

- (ii) **Glide:** Docking to confirm or discard the binding modes obtained by GOLD was performed using Glide (Schrodinger Suite versions 2019-1, 2019-4 and 2020-2). The targets were processed with the Protein Preparation Wizard with the default parameters. Energy minimization of the targets was performed using the OPLS3e force field. All water molecules were removed except for RdRp, where one water molecule in the binding site next to one of the metal ions was kept. For the RdRp, it was necessary to change the manganese ions into magnesium ions to account for metal coordination during docking. Small molecules were prepared using LigPrep with the default parameters and the OPLS3e force field. The protonation states for each ligand and each side-chain in the protein were generated using Epik for pH = 7 ± 2. The grid for docking was prepared using the default parameters, with the grid box centered on the centroid of the co-crystallized ligand, when available. Both extra precision (XP) and standard precision (SP) Glide docking was carried out (see details of the parameters and validation in **Table Glide_details**). Glide score was used to rank the binding poses. Compounds for which one of the top four Glide poses reproduced the Gold pose, as judged by manual inspection, were assigned to class 2A and selected for performing molecular dynamics simulations to assess their stability. The other compounds were retained in class 3A.

Table Glide_details: Glide non-covalent docking parameters for docking of compounds that were selected from docking with GOLD.

Target	PDB ID	Glide grid generation	Glide docking	Comments
MP	6Y2G	Docking was performed to both active sites of the homodimer. The grid box was centered on the centroid of the co-crystallized ligand. The hydroxyl and thiol groups of Ser1, Ser46, Tyr54, Ser144 and Cys145, which are located in the binding pockets of both monomers, were allowed to rotate to produce the most favorable interactions during docking.	For all datasets, except for the Merck set, four protocols were used. A standard precision (SP) docking, an SP docking with the amide bonds frozen in the input conformation, an extra precision (XP) docking, and an SP-peptide docking for enhanced retention of poses. For the Merck library, only the SP and XP protocols were used.	The protocols, SP with the amide bonds frozen in the input conformation and SP-peptide, were not used for the Merck library because it does not contain peptidomimetic compounds. The best compounds were chosen based on the consensus of the top 20 molecules of the four protocols. For the Merck library, a docking score threshold of -6 kcal/mol was also set. In a validation redocking of the non-covalent form of the co-crystallized inhibitor with the SP protocol, the RMSD between the co-crystallized ligand and the docking pose was 1.88 Å.
Plpro	3MJ5	The grid box was centered on the	SP docking	Validation of the docking procedure for PLpro was performed by cross-
Plpro	6WUU		XP docking	

		centroid of the co-crystallized ligand.		docking ligands in PDB files 4OW0, 4OVZ, 3E9S and 6WUU to protein structures 3MJ5 (non-peptide inhibitors) or 6WUU (peptide-like inhibitors) and comparing the top poses with the crystallographic ones for PLprotease. RMSD values of 0.45, 0.47, 0.93 and 2.80 Å were obtained, respectively.
RdRp	6M71	The grid box was centered on the centroid of one of the Gold results (APVD 2074 compound).	SP docking with a VdW factor of 1.0. Conformational sampling was enhanced by 2 times.	To get good metal coordination, it was necessary to change the manganese ions to magnesium ions.
RBD	6M0J	The grid box was centered at (-36, 25, 4) in the Cartesian system, with the inner box for ligand center placement assigned a size of 15^3 \AA^3 . All hydroxyl and thiol groups inside the grid box were allowed to rotate.	The standard precision (SP) protocol was used with the default sampling parameters, while 50 poses per ligand were sent for post-docking minimization, and the output was set at a maximum of 5 poses per ligand.	The grid was centered at a point where docking would be able to produce poses that would potentially occupy the RBD/ACE2 interface and thus block the interaction.

(iii) An independent **Glide** docking procedure was also used to screen compounds from the Fraunhofer, GHDDI and SWEETLEAD datasets. For this purpose, the MP, PLpro and RdRp proteins were preprocessed with the Protein Preparation Wizard from the Schrodinger Suite version 2019-4^{51–54} with the default parameters. The protonation states of each side-chain were generated using EPIK^{51,55,56} for pH = 7 ± 2. Protein minimization was performed using the OPLS3e force field^{36,57}. All water molecules were removed for the MP and PLpro. For RdRp, the manganese ions were changed into magnesium ions to account for metal coordination during docking and one water molecule was kept because it resides close to the coordinating manganese ion. Glide Version 85012 was used for all docking calculations.^{58,59} A standard precision Glide docking was carried out, generating 20 poses per docked molecule. For the Mpro, we also applied H-Bond constraints with D166, H163 and H164. To extract the best binding pose for each ligand, Glide score version 5.0 was used⁶⁰. The resulting ligands were next filtered based on ADMET properties to exclude the ones with poor drug-like features. Specifically, all ligands with more than one violation of the Lipinski's rule of 5 were excluded using Qikprop⁵³. Within the ADMET-selected compounds, poses satisfying a target-specific Glide score threshold were assigned to class 3C⁶⁰.

(iv) **Fred:** Fred was used to dock the same set of compounds as docked by Glide above. The MP, PIpro and RdRp were pre-processed using OpenEye Spruce4Docking tool. Ligand conformers were generated using OpenEye OMEGA⁶¹. Conformers with internal clashes and duplicates were discarded by the software and the remaining ones were clustered on the basis of the RMSD. For this virtual screening, a maximum of 200 conformers per compound, clustered with a RMSD of 0.5 Å, was used. If the number of conformers generated exceeds the specified maximum, only those with the lowest energies were retained. Rigid docking was then performed using OpenEye FRED⁶², which is included in the OEDocking 3.4.0.2 suite (OpenEye Scientific Software, Santa Fe, NM. <http://www.eyesopen.com>). Each conformer was docked by FRED in the negative image of the active site of the target protein, which consists of a shape potential field in the binding site volume. The highest values in this field represent the points where molecules can have a high number of contacts, without clashing into the protein structure. In its exhaustive search, FRED translates and rotates the structure of each conformer within the negative image of the active site, scoring each pose. FRED's first step has a default translational and rotational resolution of 1.0 and 1.5 Å, respectively. The 100 best scoring poses were then optimized with translational and rotational single steps of 0.5 and 0.75 Å, respectively, exploring all the 729 (six degrees of freedom with three positions = 36) nearby poses. The best scoring pose was retained and assigned to the compound. The binding poses were evaluated by using the Chemgauss4 scoring function implemented in OpenEye FRED⁶². ADMET properties were next calculated with Qikprop⁶³ and used to filter out compounds with more than one violation of Lipinski's rule of 5. Among the remaining compounds, poses satisfying a target-specific score threshold was assigned to class 3B. The poses were also compared with those obtained with Glide (class 3C). If there was a consensus pose that was top-ranked by both the Glide and the Fred score, the compound was assigned to class 2B.

The FRED and Glide non-covalent docking procedures applied to identify class 3B and 3C compounds and the consensus class 2B, were evaluated for their ability to identify known binders against SARS-1 or SARS-2 from the literature and GHDDI.^{6,7,16–25,8,26–32,9–15}

For MP, we identified 478 inhibitors from both the literature and the GHDDI dataset. For PIpro, we used 85 inhibitors from the GHDDI dataset. For RdRp, such a validation was not possible due to the lack of known inhibitors of the holoprotein. For MP and PIPro, two training sets of 'active' molecules were established, one containing only molecules with pIC50 > 5, another with pIC50 > 6. The quality of the docking was assessed by the enrichment factor (EF), and the area under the ROC curve (AUC), see **Table EF_Validation**. For MP, an EF of 9-to 19 in FRED and Glide, respectively, for compound with pIC50 > 6 in the top ranking 1% was found. The Fred algorithm was further validated for MP by redocking the N3 ligand co-crystallized in 6LU7. The redocking was performed for the ligands in the configuration prior to covalent bond formation and comparable poses to the crystallographic poses were obtained with a RMSD of 1.85 Å. For PIpro, an RMSD of 1.89 Å was obtained for redocking of the naphthylmethyl derivative 15h (PDB ID 3MJ5). No such validation was possible for RdRp, since the only complex available with an inhibitor was the structure determined by cryo-EM with remdesivir covalently incorporated in the RNA.

Table EF_validation: Fred and Glide enrichment factors (EF) on virtual screening on the 3CL main protease using the crystal structure with PDBID 6LU7.

Method	pIC50	EF 0.5%	EF 1%	EF 2%	EF 5%	EF 10%	AUC
Glide	> 6	30.98	19.04	13.45	8.29	4.94	0.71
Glide	> 5	17.1	12.86	9.06	5.92	3.61	0.65
Fred	> 6	19.05	9.45	6.90	3.92	2.46	0.61
Fred	> 5	8.86	5.12	4.27	2.87	2.14	0.56

Covalent Docking

Three potential covalent inhibitors of MP were investigated via a combined approach consisting of induced fit docking followed by covalent docking. As starting structures, the GOLD docked pose of each molecule to the catalytic dyad of MP was taken. Next, the Schrödinger induced fit docking (IFD) protocol⁶⁴ was used. The receptor grid center was specified from the bound ligand of the first docking. In the first stage of the IFD protocol, softened-potential docking was performed to generate 20 initial poses. The scaling factors to soften the potentials of the receptors and ligands were set to 0.5 in both cases. A 2.5 kcal/mol energy window was used for ligand conformational sampling. For each of the top 20 poses (with respect to GlideScore) from the initial softened-potential docking step, all residues within 5.0 Å of ligand poses were refined using Prime molecular dynamics. Refinement was performed with the Prime package to accommodate the ligand by reorienting nearby side chains. Prime uses the OPLS parameter set^{36,57} and a surface Generalized Born implicit solvent model^{65,66}. The complexes were ranked by Prime energy (molecular mechanics plus solvation) and those within 30 kcal/mol of the minimum energy structure were passed through for a final round of Glide docking and scoring. The ligands were then redocked into their corresponding receptor structures using Extra Precision (XP) scoring in Glide. Each docking result was analyzed by comparing by Glide Docking scores⁵⁸. Induced fit docking was followed by covalent docking.

The CovDock program^{67,68} of the Schrödinger suite was used for covalent docking. In this step, only cysteine residue 145 was defined as a ‘reactive’ residue. CovDock uses both Glide and Prime^{58,59}. It is developed to mimic covalent ligand binding by first positioning the pre-reaction form of the ligand in the binding site close to the receptor reactive residue using Glide docking with positional constraints and only then generating the covalent attachment. In the pre-reaction docking step, the reactive residue is mutated to alanine to enable a closer approach by the reactive group of the ligand. In the following steps after the reaction, the reactive residue is mutated back, and its rotamers are sampled using the Prime VSGB2.0 energy model⁶⁹ with the OPLS3 force field⁵⁷. Ligand poses that place the

reacting pair of atoms within 5 Å are kept, and a covalent bond is formed according to the reaction specified accompanied by any bond order changes and hydrogen abstractions/additions. The reaction studied in this work was a Michael addition to residue C145 of the catalytic dyad.

The procedure of non-covalent followed by covalent docking was able to reproduce the crystal structure in a redocking test. For the complex in PDB 6Y2G, the cocrystallized ligand was first noncovalently docked by SP Glide docking (RMSD 4.95 Å), then by Induced-Fit docking (RMSD 1.85 Å), and then covalent docking was performed, giving a final RMSD of the docked ligand, from the experimental binding pose of 1.49 Å.

Pharmacophore based screening

ROCS (Rapid overlay of Chemical Structures) (OpenEye scientific Inc.)⁷⁰ was used to perform a ligand-based virtual screening of MP, PIPro and RdRp. For MP and PIPro, the ligands present in the crystal structures with PBD IDs 6Y2G and 6WUU, respectively, were used as a scaffold for the creation of a query. These structures were selected because of the large volume of the ligand. On the other hand, for the RdRp, the query was created using the top 10 structures resulting from the structure-based virtual screening in class 2B because of the absence of a crystal structure of this protein crystallized in the presence of a ligand. The queries are reported in **Figure ROCS**. ROCS superimpose the three-dimensional structures of the molecules in a database onto the shape of a given query and scores them according to the overlapping volume using a Shape Tanimoto score⁷⁰. The program uses a smooth Gaussian function to approximate the molecular volume. A Tanimoto Color score is calculated similarly for the overlap of the functional groups, which are categorized in 6 different “colors” (Donor, Acceptor, Ring, Anion, Cation, Hydrophobic) and are modelled with hard Gaussian curves. The sum of the Tanimoto Shape Score and the Tanimoto Color Score gives the Tanimoto Combo score, which was used to rank the results. Compounds with a score above a target specific threshold that also had suitable ADMET properties (with at most one violation of Lipinski's Rule of Five) were assigned to class 5.

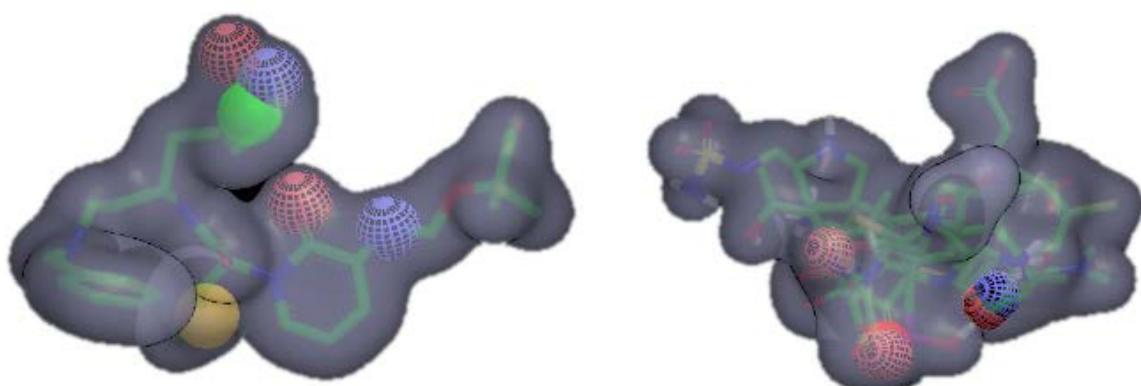


Figure ROCS: Volumes for the pharmacophore queries used for the ROCS ligand-based screening on MP (on the left) and RdRp (on the right). The spheres highlight different chemical features of the

ligand: red mesh spheres indicate an acceptor, blue mesh spheres a donor, green spheres a ring and yellow spheres a hydrophobic feature.

Molecular dynamics simulation protocol and analysis

Molecular dynamics (MD) simulations were performed to assess the structural stability of the shortlisted set of predicted complexes from classes 2A and 2B. Complexes that were stable in the MD simulations and had suitable ADMET properties were assigned to class 1A. The complexes docked with GOLD were processed using the following automated pipeline:

- 1) The coordinates in pdb format of each of the chosen protein-ligand complexes was extracted from the results of the docking. For RdRp, the structure of both the template and primer RNA strands, the positions of the two manganese ions and an additional co-localized water molecule were also retained. For Plpro, the zinc atom in the zinc binding domain was also retained.
- 2) The standard AMBER ff14sb force field⁷¹ was used to assign all protein parameters. RNA parameters were assigned using the RNA_Cl.lib library in AMBER 18. Ionic parameters for zinc (ZN) were based on Allner et al⁷² and those for manganese (MN) were obtained from the AMBER parameter database (<http://research.bmh.manchester.ac.uk/bryce/amber/>).⁷³ Ligand parameters were assigned using the GAFF forcefield⁷⁴. Partial charges were assigned to the ligands using the restrained electrostatic surface potential (RESP) approach⁷⁵, as implemented in the BiKi Life Science software suite⁷⁶ (<http://www.bikitech.com>).
- 3) Each docked complex was solvated using atomistic TIP3P water⁷⁷ with a minimum of 12 Å of complex-padding to form a rectangular periodic box and then electrically neutralized with an ionic concentration of 0.15 M NaCl using ionic parameters based on Joung and Cheatham⁷⁸. This resulted in atomic-detail explicit solvent systems with sizes of approximately 90000, 62000, 125000 and 40000 atoms for the MP, Plpro, RdRp and RDB systems, respectively.
- 4) A standardized energy minimization, equilibration and simulation protocol consisting of 11 stages was employed for all systems. Two sets of restraints were applied to each system at various stages of equilibration and/or production MD: Restraint set 1 (RS1) consisted of restraining all heavy (non-hydrogen) target and ligand atoms, including those of the RNA, the MN ions, the oxygen of the single MN-coordinated water molecule in RdRp and the ZN ion in Plpro. Restraint set 2 (RS2) consisted of restraining the backbone atoms of a small set of residues within specifically chosen secondary structures in each protein, to prevent global reorientation of the complex during the corresponding simulations.
- 5) Each system was first energy minimized in four stages, each of 1500 steps (500 steepest-descent + 1000 conjugate-gradient) of minimization, applying restraints RS1 with different force constants in each sequential state: Stage 1: 10 kcal/mol/ Å², Stage 2: 5 kcal/mol/ Å², Stage 3: 1 kcal/mol/ Å², Stage 4: unrestrained.
- 6) MD simulations were performed in all subsequent stages. The SHAKE algorithm was employed on all atoms covalently bonded to a hydrogen atom. A time-step of 2 fs was used. The long-range Coulomb interaction was treated using a GPU implementation of the particle

mesh Ewald summation method (PME). A nonbonded cutoff distance of 10 Å was used. In Stage 5, each system was heated from 10 K to 300 K in 1 ns with application of RS1 ($k=10$ kcal/mol/ Å²). The temperature was subsequently maintained at 300 K using a Langevin thermostat with a damping constant of $\gamma=5.0$ ps⁻¹ and in Stage 6, the systems were equilibrated for 1 ns at constant volume (i.e. in the NVT ensemble). Subsequently, the pressure was maintained at 1 atm using a Berendsen barostat with a pressure relaxation time of $\tau_p = 1.0$ ps and the systems simulated in the NPT ensemble for 100 ps for each of the subsequent stages with RS1: Stage 7: $k=10$ kcal/mol/ Å², Stage 8: $k=5$ kcal/mol/ Å², Stage 9: $k=1$ kcal/mol/ Å², Stage 10: $k=0.5$ kcal/mol/ Å². Finally, in Stage 11, the RS1 restraints were removed and replaced with RS2 restraints with $k=0.5$ kcal/mol/ Å², and the systems were simulated in the NPT ensemble for a further 5 ns.

- 7) Following the equilibration, a production simulation of 30 ns was performed for each system in the NPT ensemble with the same conditions as Stage 11. Coordinate snapshots from production simulations were generated every 10 ps, resulting in a trajectory of 3000 snapshots per system. One production simulation was run for each complex except in the case of the RBD-ligand complexes for which three independent simulations were run to sample these systems better due to their greater mobility.
- 8) Structural stability was analysed by computing the root-mean squared deviation (RMSD) of the protein and the ligand non-hydrogen atoms with respect to the initial docked structure (after aligning the protein C-alpha atoms) as well as the ligand all-atom root-mean-squared fluctuations (RMSF) with respect to the average structure in the production MD. We also carried out an interaction fingerprint (IFP) analysis using the MD-IFP procedure (Kokh et al, submitted, <http://arxiv.org/abs/2006.11066>) to compute the number of interactions between each ligand and the individual residues of the protein. The interactions were categorized as aromatic, ionic or hydrogen-bonding. For the RdRp system, we considered direct interactions with the RNA and each of the MN ions as separate categories. The IFP analysis was performed for the initial docked structure as well as the averages of the first 2 ns, the last 2 ns and the entire 30 ns MD simulation. Furthermore, the number of interactions lost on average during the entire MD simulation with respect to the original docked structure was computed for each interaction category.

ADMET property evaluation

The properties of compounds selected for class 1A after MD simulation were computed using Qikprop⁶³. The following properties were computed: number of rotational bonds, number of reactive groups, molecular weight, dipole moment, solvent-accessible surface area, volume, number of hydrogen-bond donors, number of hydrogen-bond acceptors, globularity, polarizability, octanol/water partition coefficient, aqueous solubility, IC₅₀ value for hERG K⁺ channel blockage, Caco-2 cell permeability (model for the gut-blood barrier), skin permeability, number of likely metabolic reactions, binding to HAS, and human oral absorption. Compounds were then assigned to class 1A or returned to their original classes 2A or 2B after manually assessing – on a per-target basis - the number of out-of-range violations and the agreement QikProp's property ranges for existing drugs.

Ligand similarity search

After the final assignment of compounds to class 1A, the compound libraries were searched for ligands with a Morgan fingerprint and Dice similarity score ≥ 0.8 to compounds assigned to class 1A using the software RdKit. These compounds were assigned to class 1B.

This search revealed that three compounds assigned to class 1A for MP were in fact known inhibitors of the SARS-CoV2 MP. These three compounds were therefore considered to be blind validation compounds and they were assigned to class 1C. The screened datasets were randomly, and by chance, populated with compounds with known activity towards SARS-CoV-2 or SARS-CoV. These could be molecules co-crystallized with the targets studied in this work, compounds from the GHDDI set or inhibitors already in use to treat other viral infections such as, for instance, rupintrivir. All the compounds were submitted to the same pipeline and were finally classified in the different categories.

One of the three compounds was the ligand N3 that was co-crystallized with MP in PDB 6LU7. It is one of the best molecules in terms of Gold and Glide scores, number of interactions and stability along the MD trajectory. This result is particularly interesting since: i) the compound was simulated in the pre-covalent configuration, ii) the superposition with the co-crystallized covalently bound form highlighted how the main contacts with His163 and Glu166 are also maintained in the non-covalent state, iii) the protease structure used for the docking simulation (PDB code 6Y2G) is somewhat different from that of 6LU7, since due to the intrinsic flexibility of the protein, the positions of residues, Met49 and Gln189, in particular, were rearranged, iv) the reactive double bond is positioned exactly in front of the catalytic Cys145, also in the non-covalent docked conformation. Overall, the entire procedure led to the identification of an active compound, without taking into account the formation of the covalent bond.

The other two compounds moved to class 1C were non-covalent inhibitors of MP that were in the GHDDI dataset, having IC₅₀ values of 40 and 180 nM. In addition, it is noteworthy that the antiviral compound Rupintrivir was classified in class 2A but was not submitted to MD simulations since we realized at this point that it has activity against SARS-CoV-2 MP. Altogether, these results support the ability of the computational pipeline to identify compounds active against MP.

Appendix 2:

Table of libraries screened showing number of compounds screened per target and the corresponding numbers of compounds selected from each initial screen together with the specific filters and criteria used. (see also corresponding Excel file).

卷之三

Open
Title
2010

Figure 11 shows the mean time per 20 m of swimming measured at 115 locations across all 100 m lanes.

No cycle (original library size) also performed satisfactorily. The original library can reduce the number of compounds after subtracting the compounds from the source and applying the solvent filter. For example, from the ca. 689 million compounds in ZINC15, we reduced the size to 13 million compounds.

Enrichment and Enrichment Factor: The compounds were analyzed and classified with a ZINC10 requirement to give the "original library".

After parameterization, the analysis of compounds in the same or different batches of source compounds did not show a significant increase of additional initial filters, e.g., for ZINC10 requirement to give the "original library".

Figure 8 shows the original library size, the number of unique molecules, and the number of unique molecules per batch.

References

- (1) Irwin, J. J.; Sterling, T.; Mysinger, M. M.; Bolstad, E. S.; Coleman, R. G. ZINC: A Free Tool to Discover Chemistry for Biology. *J. Chem. Inf. Model.* **2012**, *52*, 1757–1768. <https://doi.org/10.1021/ci3001277>.

(2) Sterling, T.; Irwin, J. J. ZINC 15 – Ligand Discovery for Everyone. *J. Chem. Inf. Model.* **2015**, *55*, 2324–2337. <https://doi.org/10.1021/acs.jcim.5b00559>.

(3) Wishart, D. S.; Feunang, Y. D.; Guo, A. C.; Lo, E. J.; Marcu, A.; Grant, J. R.; Sajed, T.; Johnson, D.; Li, C.; Sayeeda, Z.; Assempour, N.; lynkkaran, I.; Liu, Y.; Maciejewski, A.; Gale, N.; Wilson, A.; Chin, L.; Cummings, R.; Le, D.; Pon, A.; Knox, C.; Wilson, M. DrugBank 5.0: A Major Update

- to the DrugBank Database for 2018. *Nucleic Acids Res.* **2018**, *46*, D1074–D1082. <https://doi.org/10.1093/nar/gkx1037>.
- (4) Novick, P. A.; Ortiz, O. F.; Poelman, J.; Abdulhay, A. Y.; Pande, V. S. SWEETLEAD: An In Silico Database of Approved Drugs, Regulated Chemicals, and Herbal Isolates for Computer-Aided Drug Discovery. *PLoS One* **2013**, *8*, e79568.
- (5) Kim, S.; Chen, J.; Cheng, T.; Gindulyte, A.; He, J.; He, S.; Li, Q.; Shoemaker, B. A.; Thiessen, P. A.; Yu, B.; Zaslavsky, L.; Zhang, J.; Bolton, E. E. PubChem 2019 Update: Improved Access to Chemical Data. *Nucleic Acids Res.* **2019**, *47*, D1102–D1109. <https://doi.org/10.1093/nar/gky1033>.
- (6) Thanigaimalai, P.; Konno, S.; Yamamoto, T.; Koiwai, Y.; Taguchi, A.; Takayama, K.; Yakushiji, F.; Akaji, K.; Kiso, Y.; Kawasaki, Y.; Chen, S.-E.; Naser-Tavakolian, A.; Schön, A.; Freire, E.; Hayashi, Y. Design, Synthesis, and Biological Evaluation of Novel Dipeptide-Type SARS-CoV 3CL Protease Inhibitors: Structure–Activity Relationship Study. *Eur. J. Med. Chem.* **2013**, *65*, 436–447. [https://doi.org/https://doi.org/10.1016/j.ejmech.2013.05.005](https://doi.org/10.1016/j.ejmech.2013.05.005).
- (7) Wu, C.-Y.; King, K.-Y.; Kuo, C.-J.; Fang, J.-M.; Wu, Y.-T.; Ho, M.-Y.; Liao, C.-L.; Shie, J.-J.; Liang, P.-H.; Wong, C.-H. Stable Benzotriazole Esters as Mechanism-Based Inactivators of the Severe Acute Respiratory Syndrome 3CL Protease. *Chem. Biol.* **2006**, *13*, 261–268. <https://doi.org/https://doi.org/10.1016/j.chembiol.2005.12.008>.
- (8) Shie, J.-J.; Fang, J.-M.; Kuo, T.-H.; Kuo, C.-J.; Liang, P.-H.; Huang, H.-J.; Wu, Y.-T.; Jan, J.-T.; Cheng, Y.-S. E.; Wong, C.-H. Inhibition of the Severe Acute Respiratory Syndrome 3CL Protease by Peptidomimetic α,β -Unsaturated Esters. *Bioorg. Med. Chem.* **2005**, *13*, 5240–5252. <https://doi.org/https://doi.org/10.1016/j.bmc.2005.05.065>.
- (9) Jacobs, J.; Grum-Tokars, V.; Zhou, Y.; Turlington, M.; Saldanha, S. A.; Chase, P.; Eggler, A.; Dawson, E. S.; Baez-Santos, Y. M.; Tomar, S.; Mielech, A. M.; Baker, S. C.; Lindsley, C. W.; Hodder, P.; Mesecar, A.; Stauffer, S. R. Discovery, Synthesis, And Structure-Based Optimization of a Series of N-(Tert-Butyl)-2-(N-Arylamido)-2-(Pyridin-3-Yl) Acetamides (ML188) as Potent Noncovalent Small Molecule Inhibitors of the Severe Acute Respiratory Syndrome Coronavirus (SARS-CoV) 3CL Pr. *J. Med. Chem.* **2013**, *56*, 534–546. <https://doi.org/10.1021/jm301580n>.
- (10) Jain, R. P.; Pettersson, H. I.; Zhang, J.; Aull, K. D.; Fortin, P. D.; Huitema, C.; Eltis, L. D.; Parrish, J. C.; James, M. N. G.; Wishart, D. S.; Vederas, J. C. Synthesis and Evaluation of Keto-Glutamine Analogues as Potent Inhibitors of Severe Acute Respiratory Syndrome 3CLpro. *J. Med. Chem.* **2004**, *47*, 6113–6116. <https://doi.org/10.1021/jm0494873>.
- (11) Kumar, V.; Shin, J. S.; Shie, J.-J.; Ku, K. B.; Kim, C.; Go, Y. Y.; Huang, K.-F.; Kim, M.; Liang, P.-H. Identification and Evaluation of Potent Middle East Respiratory Syndrome Coronavirus (MERS-CoV) 3CLPro Inhibitors. *Antiviral Res.* **2017**, *141*, 101–106. <https://doi.org/https://doi.org/10.1016/j.antiviral.2017.02.007>.
- (12) 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 (80-.).* **2020**, *368*, 409–412. <https://doi.org/10.1126/science.abb3405>.
- (13) Liu, W.; Zhu, H.-M.; Niu, G.-J.; Shi, E.-Z.; Chen, J.; Sun, B.; Chen, W.-Q.; Zhou, H.-G.; Yang, C. Synthesis, Modification and Docking Studies of 5-Sulfonyl Isatin Derivatives as SARS-CoV 3C-like Protease Inhibitors. *Bioorg. Med. Chem.* **2014**, *22*, 292–302. <https://doi.org/https://doi.org/10.1016/j.bmc.2013.11.028>.
- (14) Wang, L.; Bao, B.-B.; Song, G.-Q.; Chen, C.; Zhang, X.-M.; Lu, W.; Wang, Z.; Cai, Y.; Li, S.; Fu, S.; Song, F.; Yang, H.; Wang, J.-G. Discovery of Unsymmetrical Aromatic Disulfides as Novel Inhibitors of SARS-CoV Main Protease: Chemical Synthesis, Biological Evaluation, Molecular

- Docking and 3D-QSAR Study. *Eur. J. Med. Chem.* **2017**, *137*.
<https://doi.org/10.1016/j.ejmech.2017.05.045>.
- (15) Chuck, C. P.; Ke, Z. H.; Chen, C.; Wan, D. C. C.; Chow, H. F.; Wong, K. B. Profiling of Substrate-Specificity and Rational Design of Broad-Spectrum Peptidomimetic Inhibitors for Main Proteases of Coronaviruses. *Hong Kong Med. J.* **2014**, *20*, 22–25.
- (16) 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.
<https://doi.org/10.1021/acs.jmedchem.5b01461>.
- (17) Regnier, T.; Sarma, D.; Hidaka, K.; Bacha, U.; Freire, E.; Hayashi, Y.; Kiso, Y. New Developments for the Design, Synthesis and Biological Evaluation of Potent SARS-CoV 3CL(pro) Inhibitors. *Bioorg. Med. Chem. Lett.* **2009**, *19*, 2722–2727.
<https://doi.org/10.1016/j.bmcl.2009.03.118>.
- (18) Turlington, M.; Chun, A.; Tomar, S.; Eggler, A.; Grum-Tokars, V.; Jacobs, J.; Daniels, J. S.; Dawson, E.; Saldanha, A.; Chase, P.; Baez-Santos, Y. M.; Lindsley, C. W.; Hodder, P.; Mesecar, A. D.; Stauffer, S. R. Discovery of N-(Benzo[1,2,3]Triazol-1-Yl)-N-(Benzyl)Acetamido)Phenyl Carboxamides as Severe Acute Respiratory Syndrome Coronavirus (SARS-CoV) 3CLpro Inhibitors: Identification of ML300 and Noncovalent Nanomolar Inhibitors with an Induced-Fit Binding. *Bioorg. Med. Chem. Lett.* **2013**, *23*, 6172–6177.
<https://doi.org/https://doi.org/10.1016/j.bmcl.2013.08.112>.
- (19) Kumar, V.; Tan, K.-P.; Wang, Y.-M.; Lin, S.-W.; Liang, P.-H. Identification, Synthesis and Evaluation of SARS-CoV and MERS-CoV 3C-like Protease Inhibitors. *Bioorg. Med. Chem.* **2016**, *24*, 3035–3042. <https://doi.org/10.1016/j.bmc.2016.05.013>.
- (20) Ramajayam, R.; Tan, K.-P.; Liu, H.-G.; Liang, P.-H. Synthesis, Docking Studies, and Evaluation of Pyrimidines as Inhibitors of SARS-CoV 3CL Protease. *Bioorg. Med. Chem. Lett.* **2010**, *20*, 3569–3572. <https://doi.org/10.1016/j.bmcl.2010.04.118>.
- (21) Wen, C. C.; Kuo, Y. H.; Jan, J. T.; Liang, P. H.; Wang, S. Y.; Liu, H. G.; Lee, C. K.; Chang, S. T.; Kuo, C. J.; Lee, S. S.; Hou, C. C.; Hsiao, P. W.; Chien, S. C.; Shyur, L. F.; Yang, N. S. Specific Plant Terpenoids and Lignoids Possess Potent Antiviral Activities against Severe Acute Respiratory Syndrome Coronavirus. *Journal of Medicinal Chemistry*. August 23, 2007, pp 4087–4095.
<https://doi.org/10.1021/jm070295s>.
- (22) Nguyen, T. T. H.; Ryu, H.-J.; Lee, S.-H.; Hwang, S.; Cha, J.; Breton, V.; Kim, D. Discovery of Novel Inhibitors for Human Intestinal Maltase: Virtual Screening in a WISDOM Environment and in Vitro Evaluation. *Biotechnol. Lett.* **2011**, *33*, 2185. <https://doi.org/10.1007/s10529-011-0675-8>.
- (23) Mandadapu, S. R.; Weerawarna, P. M.; Prior, A. M.; Uy, R. A. Z.; Aravapalli, S.; Alliston, K. R.; Lushington, G. H.; Kim, Y.; Hua, D. H.; Chang, K.-O.; Groutas, W. C. Macrocyclic Inhibitors of 3C and 3C-like Proteases of Picornavirus, Norovirus, and Coronavirus. *Bioorg. Med. Chem. Lett.* **2013**, *23*, 3709–3712. <https://doi.org/https://doi.org/10.1016/j.bmcl.2013.05.021>.
- (24) Pakravan, P.; Kashanian, S.; Khodaei, M. M.; Harding, F. J. Biochemical and Pharmacological Characterization of Isatin and Its Derivatives: From Structure to Activity. *Pharmacol. Reports* **2013**, *65*, 313–335. [https://doi.org/10.1016/S1734-1140\(13\)71007-7](https://doi.org/10.1016/S1734-1140(13)71007-7).
- (25) Shimamoto, Y.; Hattori, Y.; Kobayashi, K.; Teruya, K.; Sanjoh, A.; Nakagawa, A.; Yamashita, E.; Akaji, K. Fused-Ring Structure of Decahydroisoquinolin as a Novel Scaffold for SARS 3CL Protease Inhibitors. *Bioorg. Med. Chem.* **2015**, *23*, 876–890.
<https://doi.org/https://doi.org/10.1016/j.bmc.2014.12.028>.
- (26) Yang, S.; Chen, S.-J.; Hsu, M.-F.; Wu, J.-D.; Tseng, C.-T. K.; Liu, Y.-F.; Chen, H.-C.; Kuo, C.-W.; Wu, C.-S.; Chang, L.-W.; Chen, W.-C.; Liao, S.-Y.; Chang, T.-Y.; Hung, H.-H.; Shr, H.-L.; Liu, C.-Y.;

- Huang, Y.-A.; Chang, L.-Y.; Hsu, J.-C.; Peters, C. J.; Wang, A. H.-J.; Hsu, M.-C. Synthesis, Crystal Structure, Structure–Activity Relationships, and Antiviral Activity of a Potent SARS Coronavirus 3CL Protease Inhibitor. *J. Med. Chem.* **2006**, *49*, 4971–4980. <https://doi.org/10.1021/jm0603926>.
- (27) Zhang, J.; Pettersson, H. I.; Huitema, C.; Niu, C.; Yin, J.; James, M. N. G.; Eltis, L. D.; Vedera, J. C. Design, Synthesis, and Evaluation of Inhibitors for Severe Acute Respiratory Syndrome 3C-Like Protease Based on Phthalhydrazide Ketones or Heteroaromatic Esters. *J. Med. Chem.* **2007**, *50*, 1850–1864. <https://doi.org/10.1021/jm061425k>.
- (28) Galasiti Kankanamalage, A. C.; Kim, Y.; Damalanka, V. C.; Rathnayake, A. D.; Fehr, A. R.; Mehzabeen, N.; Battaile, K. P.; Lovell, S.; Lushington, G. H.; Perlman, S.; Chang, K.-O.; Groutas, W. C. Structure-Guided Design of Potent and Permeable Inhibitors of MERS Coronavirus 3CL Protease That Utilize a Piperidine Moiety as a Novel Design Element. *Eur. J. Med. Chem.* **2018**, *150*, 334–346. <https://doi.org/https://doi.org/10.1016/j.ejmech.2018.03.004>.
- (29) Prior, A. M.; Kim, Y.; Weerasekara, S.; Moroze, M.; Alliston, K. R.; Uy, R. A. Z.; Groutas, W. C.; Chang, K.-O.; Hua, D. H. Design, Synthesis, and Bioevaluation of Viral 3C and 3C-like Protease Inhibitors. *Bioorg. Med. Chem. Lett.* **2013**, *23*, 6317–6320. <https://doi.org/https://doi.org/10.1016/j.bmcl.2013.09.070>.
- (30) Zhang, L.; Lin, D.; Kusov, Y.; Nian, Y.; Ma, Q.; Wang, J.; von Brunn, A.; Leyssen, P.; Lanko, K.; Neyts, J.; de Wilde, A.; Snijder, E. J.; Liu, H.; Hilgenfeld, R. α -Ketoamides as Broad-Spectrum Inhibitors of Coronavirus and Enterovirus Replication: Structure-Based Design, Synthesis, and Activity Assessment. *J. Med. Chem.* **2020**, *63*, 4562–4578. <https://doi.org/10.1021/acs.jmedchem.9b01828>.
- (31) Nascimento Junior, J. A. C.; Santos, A. M.; Quintans-Júnior, L. J.; Walker, C. I. B.; Borges, L. P.; Serafini, M. R. SARS, MERS and SARS-CoV-2 (COVID-19) Treatment: A Patent Review. *Expert Opin. Ther. Pat.* **2020**, *1*–13. <https://doi.org/10.1080/13543776.2020.1772231>.
- (32) Shao, Y.-M.; Yang, W.-B.; Kuo, T.-H.; Tsai, K.-C.; Lin, C.-H.; Yang, A.-S.; Liang, P.-H.; Wong, C.-H. Design, Synthesis, and Evaluation of Trifluoromethyl Ketones as Inhibitors of SARS-CoV 3CL Protease. *Bioorg. Med. Chem.* **2008**, *16*, 4652–4660. <https://doi.org/https://doi.org/10.1016/j.bmcl.2008.02.040>.
- (33) Milletti, F.; Storchi, L.; Sforza, G.; Cross, S.; Cruciani, G. Tautomer Enumeration and Stability Prediction for Virtual Screening on Large Chemical Databases. *J. Chem. Inf. Model.* **2009**, *49*, 68–75. <https://doi.org/10.1021/ci800340j>.
- (34) Baroni, M.; Cruciani, G.; Sciabola, S.; Perruccio, F.; Mason, J. S. A Common Reference Framework for Analyzing/Comparing Proteins and Ligands. Fingerprints for Ligands and Proteins (FLAP): Theory and Application. *J. Chem. Inf. Model.* **2007**, *47*, 279–294. <https://doi.org/10.1021/ci600253e>.
- (35) Schrödinger Release 2019-4: LigPrep. Schrödinger, LLC: New York, NY, 2019.
- (36) Harder, E.; Damm, W.; Maple, J.; Wu, C.; Reboul, M.; Xiang, J. Y.; Wang, L.; Lupyan, D.; Dahlgren, M. K.; Knight, J. L.; Kaus, J. W.; Cerutti, D. S.; Krilov, G.; Jorgensen, W. L.; Abel, R.; Friesner, R. A. OPLS3: A Force Field Providing Broad Coverage of Drug-like Small Molecules and Proteins. *J. Chem. Theory Comput.* **2016**, *12*, 281–296. <https://doi.org/10.1021/acs.jctc.5b00864>.
- (37) Muramatsu, T.; Takemoto, C.; Kim, Y. T.; Wang, H.; Nishii, W.; Terada, T.; Shirouzu, M.; Yokoyama, S. SARS-CoV 3CL Protease Cleaves Its C-Terminal Autoprocessing Site by Novel Subsite Cooperativity. *Proc. Natl. Acad. Sci. U. S. A.* **2016**, *113*, 12997–13002. <https://doi.org/10.1073/pnas.1601327113>.
- (38) 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. <https://doi.org/10.1038/s41586-020-2223-y>.
- (39) Ghosh, A. K.; Takayama, J.; Rao, K. V.; Ratia, K.; Chaudhuri, R.; Mulhearn, D. C.; Lee, H.; Nichols, D. B.; Baliji, S.; Baker, S. C.; Johnson, M. E.; Mesecar, A. D. Severe Acute Respiratory Syndrome Coronavirus Papain-like Novel Protease Inhibitors: Design, Synthesis, Protein-Ligand X-Ray Structure and Biological Evaluation. *J. Med. Chem.* **2010**, *53*, 4968–4979. <https://doi.org/10.1021/jm1004489>.
- (40) Baez-Santos, Y. M.; Mielech, A. M.; Deng, X.; Baker, S.; Mesecar, A. D. Catalytic Function and Substrate Specificity of the Papain-Like Protease Domain of Nsp3 from the Middle East Respiratory Syndrome Coronavirus. *J. Virol.* **2014**, *88*, 12511–12527. <https://doi.org/10.1128/jvi.01294-14>.
- (41) Báez-Santos, Y. M.; St. John, S. E.; Mesecar, A. D. The SARS-Coronavirus Papain-like Protease: Structure, Function and Inhibition by Designed Antiviral Compounds. *Antiviral Research*. 2015, pp 21–38. <https://doi.org/10.1016/j.antiviral.2014.12.015>.
- (42) Gao, Y.; Yan, L.; Huang, Y.; Liu, F.; Zhao, Y.; Cao, L.; Wang, T.; Sun, Q.; Ming, Z.; Zhang, L.; Ge, J.; Zheng, L.; Zhang, Y.; Wang, H.; Zhu, Y.; Zhu, C.; Hu, T.; Hua, T.; Zhang, B.; Yang, X.; Li, J.; Yang, H.; Liu, Z.; Xu, W.; Guddat, L. W.; Wang, Q.; Lou, Z.; Rao, Z. Structure of the RNA-Dependent RNA Polymerase from COVID-19 Virus. *Science* **2020**, *368*, 779–782. <https://doi.org/10.1126/science.abb7498>.
- (43) Venkataraman, S.; Prasad, B. V. L. S.; Selvarajan, R. RNA Dependent RNA Polymerases: Insights from Structure, Function and Evolution. *Viruses*. 2018, p 76. <https://doi.org/10.3390/v10020076>.
- (44) Lan, J.; Ge, J.; Yu, J.; Shan, S.; Zhou, H.; Fan, S.; Zhang, Q.; Shi, X.; Wang, Q.; Zhang, L.; Wang, X. Structure of the SARS-CoV-2 Spike Receptor-Binding Domain Bound to the ACE2 Receptor. *Nature* **2020**, *581*, 215–220. <https://doi.org/10.1038/s41586-020-2180-5>.
- (45) Kokh, D. B.; Richter, S.; Henrich, S.; Czodrowski, P.; Rippmann, F.; Wade, R. C. TRAPP: A Tool for Analysis of Transient Binding Pockets in Proteins. *J. Chem. Inf. Model.* **2013**, *53*, 1235–1252. <https://doi.org/10.1021/ci4000294>.
- (46) Kokh, D. B.; Czodrowski, P.; Rippmann, F.; Wade, R. C. Perturbation Approaches for Exploring Protein Binding Site Flexibility to Predict Transient Binding Pockets. *J. Chem. Theory Comput.* **2016**, *12*, 4100–4113. <https://doi.org/10.1021/acs.jctc.6b00101>.
- (47) Yuan, J.-H.; Han, S. B.; Richter, S.; Wade, R. C.; Kokh, D. B. Druggability Assessment in TRAPP Using Machine Learning Approaches. *J. Chem. Inf. Model.* **2020**, *60*, 1685–1699. <https://doi.org/10.1021/acs.jcim.9b01185>.
- (48) Holderbach, S.; Adam, L.; Jayaram, B.; Wade, R.; Mukherjee, G. RASPD+: Fast Protein-Ligand Binding Free Energy Prediction Using Simplified Physicochemical Features. 2020. <https://doi.org/10.26434/chemrxiv.12636704>.
- (49) Jones, G.; Willett, P.; Glen, R. C.; Leach, A. R.; Taylor, R. Development and Validation of a Genetic Algorithm for Flexible Docking. *J. Mol. Biol.* **1997**, *267*, 727–748. <https://doi.org/10.1006/jmbi.1996.0897>.
- (50) Yin, W.; Mao, C.; Luan, X.; Shen, D.-D.; Shen, Q.; Su, H.; Wang, X.; Zhou, F.; Zhao, W.; Gao, M.; Chang, S.; Xie, Y.-C.; Tian, G.; Jiang, H.-W.; Tao, S.-C.; Shen, J.; Jiang, Y.; Jiang, H.; Xu, Y.; Zhang, S.; Zhang, Y.; Xu, H. E. Structural Basis for Inhibition of the RNA-Dependent RNA Polymerase from SARS-CoV-2 by Remdesivir. *Science (80-.).* **2020**, *368*, 1499–1504. <https://doi.org/10.1126/science.abc1560>.
- (51) Schrödinger Release 2019-4: Epik. Schrödinger,LLC: New York, NY, 2016.

- (52) Schrödinger Release 2019-4: Impact. Schrödinger, LLC: New York, NY, 2016.
- (53) Schrödinger Release 2019-4: Prime. Schrödinger, LLC: New York, NY, 2019.
- (54) Madhavi Sastry, G.; Adzhigirey, M.; Day, T.; Annabhimoju, R.; Sherman, W. Protein and Ligand Preparation: Parameters, Protocols, and Influence on Virtual Screening Enrichments. *J. Comput. Aided. Mol. Des.* **2013**, *27*, 221–234. <https://doi.org/10.1007/s10822-013-9644-8>.
- (55) Greenwood, J. R.; Calkins, D.; Sullivan, A. P.; Shelley, J. C. Towards the Comprehensive, Rapid, and Accurate Prediction of the Favorable Tautomeric States of Drug-like Molecules in Aqueous Solution. *J. Comput. Aided. Mol. Des.* **2010**, *24*, 591–604. <https://doi.org/10.1007/s10822-010-9349-1>.
- (56) Shelley, J. C.; Cholleti, A.; Frye, L. L.; Greenwood, J. R.; Timlin, M. R.; Uchimaya, M. Epik: A Software Program for PKaprediction and Protonation State Generation for Drug-like Molecules. *J. Comput. Aided. Mol. Des.* **2007**, *21*, 681–691. <https://doi.org/10.1007/s10822-007-9133-z>.
- (57) Jorgensen, W. L.; Maxwell, D. S.; Tirado-Rives, J. Development and Testing of the OPLS All-Atom Force Field on Conformational Energetics and Properties of Organic Liquids. *J. Am. Chem. Soc.* **1996**, *118*, 11225–11236. <https://doi.org/10.1021/ja9621760>.
- (58) Friesner, R. A.; Banks, J. L.; Murphy, R. B.; Halgren, T. A.; Klicic, J. J.; Mainz, D. T.; Repasky, M. P.; Knoll, E. H.; Shelley, M.; Perry, J. K.; Shaw, D. E.; Francis, P.; Shenkin, P. S. Glide: A New Approach for Rapid, Accurate Docking and Scoring. 1. Method and Assessment of Docking Accuracy. *J. Med. Chem.* **2004**, *47*, 1739–1749. <https://doi.org/10.1021/jm0306430>.
- (59) Halgren, T.A.; Murphy, R.B.; Friesner, R.A.; Beard, H.S.; Frye, L.L.; Pollard, W.T.; Banks, J. L.; Glide. Glide: A New Approach Fo Rrapid, Accurate Docking and Scoring. II. Enrichment Factors in Database Screening. *J. Med. Chem.* **2004**, *2*, 1750–1759.
- (60) Friesner, R. A.; Murphy, R. B.; Repasky, M. P.; Frye, L. L.; Greenwood, J. R.; Halgren, T. A.; Sanschagrin, P. C.; Mainz, D. T. Extra Precision Glide: Docking and Scoring Incorporating a Model of Hydrophobic Enclosure for Protein-Ligand Complexes. *J. Med. Chem.* **2006**, *49*, 6177–6196. <https://doi.org/10.1021/jm051256o>.
- (61) 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.
- (62) McGann, M. FRED Pose Prediction and Virtual Screening Accuracy. *J. Chem. Inf. Model.* **2011**, *51*, 578–596.
- (63) Schrödinger Release 2019-4: QikProp. Schrödinger, LLC: New York, NY, 2019.
- (64) Sherman, W.; Day, T.; Jacobson, M. P.; Friesner, R. A.; Farid, R. Novel Procedure for Modeling Ligand/Receptor Induced Fit Effects. *J. Med. Chem.* **2006**, *49*, 534–553. <https://doi.org/10.1021/jm050540c>.
- (65) Gallicchio, E.; Zhang, L. Y.; Levy, R. M. The SGB/NP Hydration Free Energy Model Based on the Surface Generalized Born Solvent Reaction Field and Novel Nonpolar Hydration Free Energy Estimators. *J. Comput. Chem.* **2002**, *23*, 517–529. <https://doi.org/10.1002/jcc.10045>.
- (66) Ghosh, A.; Rapp, C. S.; Friesner, R. A. Generalized Born Model Based on a Surface Integral Formulation. *J. Phys. Chem. B* **1998**, *102*, 10983–10990. <https://doi.org/10.1021/jp982533o>.
- (67) Toledo Warshaviak, D.; Golan, G.; Borrelli, K. W.; Zhu, K.; Kalid, O. Structure-Based Virtual Screening Approach for Discovery of Covalently Bound Ligands. *J. Chem. Inf. Model.* **2014**, *54*, 1941–1950. <https://doi.org/10.1021/ci500175r>.
- (68) Zhu, K.; Borrelli, K. W.; Greenwood, J. R.; Day, T.; Abel, R.; Farid, R. S.; Harder, E. Docking Covalent Inhibitors: A Parameter Free Approach to Pose Prediction and Scoring. *J. Chem. Inf. Model.* **2014**, *54*, 1932–1940. <https://doi.org/10.1021/ci500118s>.

- (69) Li, J.; Abel, R.; Zhu, K.; Cao, Y.; Zhao, S.; Friesner, R. A. The VSGB 2.0 Model: A next Generation Energy Model for High Resolution Protein Structure Modeling. *Proteins* **2011**, *79*, 2794–2812. <https://doi.org/10.1002/prot.23106>.
- (70) Hawkins, P. C. D.; Skillman, A. G.; Nicholls, A. Comparison of Shape-Matching and Docking as Virtual Screening Tools. *J. Med. Chem.* **2007**, *50*, 74–82.
- (71) 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. <https://doi.org/10.1021/acs.jctc.5b00255>.
- (72) Allnér, O.; Nilsson, L.; Villa, A. Magnesium Ion–Water Coordination and Exchange in Biomolecular Simulations. *J. Chem. Theory Comput.* **2012**, *8*, 1493–1502. <https://doi.org/10.1021/ct3000734>.
- (73) M. Bradbrook, G.; Gleichmann, T.; J. Harrop, S.; Habash, J.; Raftery, J.; Kalb (Gilboa), J.; Yariv, J.; H. Hillier, I.; R. Helliwell, J. X-Ray and Molecular Dynamics Studies of Concanavalin-A Glucoside and Mannoside Complexes Relating Structure to Thermodynamics of Binding. *J. Chem. Soc., Faraday Trans.* **1998**, *94*, 1603–1611. <https://doi.org/10.1039/A800429C>.
- (74) 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. <https://doi.org/10.1002/jcc.20035>.
- (75) Bayly, C. I.; Cieplak, P.; Cornell, W.; Kollman, P. A. A Well-Behaved Electrostatic Potential Based Method Using Charge Restraints for Deriving Atomic Charges: The RESP Model. *J. Phys. Chem.* **1993**, *97*, 10269–10280. <https://doi.org/10.1021/j100142a004>.
- (76) Sciabola, S.; Benedetti, P.; D'Arrigo, G.; Torella, R.; Baroni, M.; Cruciani, G.; Spyrosakis, F. Discovering New Casein Kinase 1d Inhibitors with an Innovative Molecular Dynamics Enabled Virtual Screening Workflow. *ACS Med. Chem. Lett.* **2019**, *10*, 487–492. <https://doi.org/10.1021/acsmmedchemlett.8b00523>.
- (77) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of Simple Potential Functions for Simulating Liquid Water. *J. Chem. Phys.* **1983**, *79*, 926–935. <https://doi.org/10.1063/1.445869>.
- (78) Joung, I. S.; Cheatham, T. E. Determination of Alkali and Halide Monovalent Ion Parameters for Use in Explicitly Solvated Biomolecular Simulations. *J. Phys. Chem. B* **2008**, *112*, 9020–9041. <https://doi.org/10.1021/jp8001614>.

