



Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.



## Investigating novel thiazolyl-indazole derivatives as scaffolds for SARS-CoV-2 M<sup>Pro</sup> inhibitors



Justin Airas<sup>a</sup>, Catherine A. Bayas<sup>a</sup>, Abdellah N'Ait Ousidi<sup>b</sup>, Moulay Youssef Ait Itto<sup>b</sup>, Aziz Auhmani<sup>b</sup>, Mohamed Loubidi<sup>b</sup>, M'hamed Esseffar<sup>b</sup>, Julie A. Pollock<sup>a</sup>, Carol A. Parish<sup>a,\*</sup>

<sup>a</sup> Department of Chemistry, University of Richmond, Gottwald Center for the Sciences, Richmond, VA, 23173, USA

<sup>b</sup> Département de Chimie, Faculté des Sciences Semlalia, Cadi Ayyad University, BP, 2390, Marrakech, Morocco

### ARTICLE INFO

#### Keywords:

SARS-CoV-2  
Coronavirus  
COVID-19  
M<sup>Pro</sup>  
Protease  
Substrate  
Indazole  
Thiazolyl  
Thiazolyl-indazole  
Inhibitors  
Scaffold  
Molecular dynamics  
Binding assay  
Drug discovery

### ABSTRACT

COVID-19 is a global pandemic caused by infection with the SARS-CoV-2 virus. Remdesivir, a SARS-CoV-2 RNA polymerase inhibitor, is the only drug to have received widespread approval for treatment of COVID-19. The SARS-CoV-2 main protease enzyme (M<sup>Pro</sup>), essential for viral replication and transcription, remains an active target in the search for new treatments. In this study, the ability of novel thiazolyl-indazole derivatives to inhibit M<sup>Pro</sup> is evaluated. These compounds were synthesized via the heterocyclization of phenacyl bromide with (R)-carvone, (R)-pulegone and (R)-menthone thiosemicarbazones. The binding affinity and binding interactions of each compound were evaluated through Schrödinger Glide docking, AMBER molecular dynamics simulations, and MM-GBSA free energy estimation, and these results were compared with similar calculations of M<sup>Pro</sup> binding various 5-mer substrates (VKLQA, VKLQS, VKLQG) and a previously identified M<sup>Pro</sup> tight-binder X77. From these simulations, we can see that binding is driven by residue specific interactions such as  $\pi$ -stacking with His41, and S/ $\pi$  interactions with Met49 and Met165. The compounds were also experimentally evaluated in a M<sup>Pro</sup> biochemical assay and the most potent compound containing a phenylthiazole moiety inhibited protease activity with an IC<sub>50</sub> of 92.9  $\mu$ M. This suggests that the phenylthiazole scaffold is a promising candidate for the development of future M<sup>Pro</sup> inhibitors.

### 1. Introduction

Coronavirus disease-2019 (COVID-19) is a global pandemic caused by infection with the severe acute respiratory syndrome-coronavirus 2 (SARS-CoV-2) [1–4]. While closely related to the severe acute respiratory syndrome-coronavirus (SARS-CoV), SARS-CoV-2 demonstrates faster human transmission than both SARS-CoV and the Middle Eastern respiratory syndrome-coronavirus (MERS-CoV) [1–3]. Symptoms of COVID-19 include cough, difficulty breathing, fever, fatigue, and loss of taste, smell and hearing that can progress to viral pneumonia in severe cases [5–8]. A significantly higher fatality rate is noted in those above the age of 60 and those with pre-existing conditions such as cardiovascular disease, chronic lung disease, hypertension, diabetes, congestive heart failure, chronic kidney disease, and cancer [7,9,10]. Currently, medicinal therapeutic options remain limited.

Owing to their significance in the viral life cycle and lack of related human homologues, the SARS-CoV-2 RNA-dependent RNA polymerase (RdRp) and 3C-like main protease (3CL<sup>Pro</sup> or M<sup>Pro</sup>) have been identified

as potentially promising drug targets [11,12]. The RdRp catalyzes the synthesis of viral RNA and is the target of the nucleotide analog prodrug remdesivir [11,13]. With the Food and Drug Administration's Emergency Use Authorization of remdesivir, the RdRp is currently the only SARS-CoV-2 drug target with an approved medicinal therapy [13]. M<sup>Pro</sup>, a homodimeric enzyme characterized by a catalytic His-Cys dyad, functions to cleave polyproteins pp1a and pp1ab at the recognition sequence [P4:Small] [P3:X] [P2:Leu/Phe/Val/Met] [P1:Gln] ↓ [P1':Gly/Ala/Ser/Asn] (↓ indicates the cleaved bond) into several nonstructural proteins essential for viral replication and transcription [12,14,15]. This function makes M<sup>Pro</sup> a potential target of therapeutic drugs [14,16]. While the Pfizer PF-07321332 covalent M<sup>Pro</sup> inhibitor has recently shown remarkably effective clinic results, there are currently no FDA-approved non-covalent inhibitors.

The M<sup>Pro</sup> catalytic site consists of four binding pockets or subsites. These are the S1', S1, S2, and S4 subsites, occupied by the substrate P1', P1, P2, and P4 residues, respectively [12,17]. The His41, Val42, Asn119, Thr25, Cys145, and Gly143 sidechains and the Thr26 backbone form the

\* Corresponding author. Department of Chemistry, Gottwald Center for the Sciences, University of Richmond, Richmond, VA, USA.

E-mail address: [cparish@richmond.edu](mailto:cparish@richmond.edu) (C.A. Parish).

<https://doi.org/10.1016/j.ejmcr.2022.100034>

Received 25 August 2021; Received in revised form 20 November 2021; Accepted 7 February 2022

Available online 10 February 2022

2772-4174/© 2022 The Authors. Published by Elsevier Masson SAS. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

S1' subsite. His41 and Cys145 form the catalytic dyad. The S1 subsite, which accommodates the P1-Gln, is formed by the Phe140, Asn142, Ser144, Cys145, His163, His172, and Glu166 sidechains and the Leu141, Gly142, His164, and Met165 backbones. The P2-Leu accommodating S2 subsite is formed by the His41, Met49, Tyr54, Met165 and Asp187 sidechains and the Arg188 and Gln189 backbone. Lastly, the S4 subsite is formed by the Met165, Leu167, Pro168, Ala191, and Gln192 sidechains and the Glu166, Arg188, and Thr190 backbones. These subsites are displayed in Fig. 1. Various studies have shown that the residues forming these subsites are essential targets of M<sup>Pro</sup> inhibitors [17–24]. Currently no drug has been approved for M<sup>Pro</sup> inhibition despite the screening of many structurally diverse compounds [25–36]. Notable candidates have included the peptidyl Michael acceptor N3 ( $k_{\text{obs}}/[I] = 11,300 \pm 880 \text{ M}^{-1} \text{ s}^{-1}$ ) and the broad-spectrum picornavirus-like M<sup>Pro</sup> inhibitor GC376 ( $\text{IC}_{50} = 26.4 \pm 1.1 \text{ nM}$ ,  $K_I = 12 \pm 1.4 \text{ nM}$ ) [12,34–36]. Both inhibitors function by covalent bond formation with the catalytic Cys145 residue and form binding interactions with additional S1' – S4 subsite residues [12,37–40].

Thiazole and indazole derivatives have displayed a wide array of biological function and there is significant interest in their pharmacologic applications. Many bioactive natural products, such as Vitamin B1, bacitracin, and penicillin contain thiazole ring structures [31]. Additionally, synthetic thiazole-based compounds have been shown to function as antineoplastic agents, anti-HIV drugs, antifungal agents, antiparasitic agents, anti-inflammatory agents, and antiulcer agents [41, 42]. Most recently, thiazole-based inhibitors have been reported for SARS-CoV-2 [43,44]. Synthetic indazole-based compounds have exhibited function as anti-inflammatory, antiarrhythmic, antitumor, antifungal, antibacterial, and anti-HIV drugs [45]. Indazole-based drugs are also being considered as SARS-CoV-2 M<sup>Pro</sup> inhibitors [46].

Previously, we reported the synthesis and characterization of two novel thiazolyl-indazole derivatives [47]. These compounds were synthesized through the heterocyclization of phenacyl bromide with (R)-carvone and (R)-pulegone thiosemicarbazones producing (R)-2-(2-(5-isopropyl-2-methylcyclohex-2-enylidene)hydrazinyl)-4-phenylthiazole **I** and (3*aR*,6*R*)-3,3,6-trimethyl-2-(4-phenylthiazol-2-yl)-3,3a,4,5,6,7-hexahydro-2H-indazol-3a-ol **II**, respectively. While it was initially expected that these reactions would result in compounds **I** and (R)-2-(2-(5-methyl-2-(propan-2-ylidene) cyclohexylidene)hydrazinyl)-4-phenylthiazole **III**, an unexpected rearrangement of the (R)-pulegone thiosemicarbazone occurred in which the

thioureido group underwent an N–H addition to the C=C double bond. Subsequent condensation of this intermediate with phenacyl bromide followed by an unexpected oxidation reaction resulted in compound **II** with two potential diastereomers (3*aR*,6*R*) and (3*aS*,6*R*). X-Ray and computational analysis indicated that the (3*aR*,6*R*) diastereomer was synthesized and is energetically favored [47]. In addition to exploring the inhibitory properties of **I** – **III**, we are also reporting the synthesis of a version of **II** lacking the hydroxy group (**IV**), and a version of **III** with a single bond connecting the isopropyl group to the cyclohexane ring (**V**). These compounds will provide further insight into the structural features underlying the binding of the thiazolyl-indazole derivatives. All structures are shown in Scheme 1.

Thiazole-based compounds have shown promise as SARS-CoV-2 M<sup>Pro</sup> inhibitors with best-in-class compounds demonstrating  $K_I$  values of 2.2  $\mu\text{M}$  and  $\text{IC}_{50}$  as low as 3  $\mu\text{M}$  [43,44]. A machine learning approach to identifying investigational or off-market drug targets for SARS-CoV-2 identified an indazole containing compound as one of the most promising with an estimated binding affinity of  $-9 \text{ kcal/mol}$  [46]. Given the wide-ranging and beneficial pharmacological impacts of thiazole- and indazole-containing compounds, and the community's interest in these drugs as potential SARS-CoV-2 M<sup>Pro</sup> inhibitors, we seek to elucidate the usefulness of thiazolyl-indazole derivatives as potential scaffolds for further pharmaceutical development and exploration [41–46,48].

In this work, we explore the potential for compounds **I–V** to function as reversible, noncovalent inhibitors of M<sup>Pro</sup> through chemical property prediction, biochemical assays, docking analysis and molecular dynamics simulations. For benchmarking purposes, we also include a detailed binding comparison with the viral substrate as well as a recently reported tight-binding non-covalent inhibitor **X77** (PDB code 6w63: estimated  $K_D = 0.057 \mu\text{M}$ ) [26]. We have identified M<sup>Pro</sup> amino acids that consistently interact with these compounds and our simulations reveal the importance of  $\pi$  interactions in the mechanism of binding. We demonstrate experimentally that compound **I** is capable of inhibiting the activity of M<sup>Pro</sup>; and this, taken together with our binding analyses, suggests that the phenylthiazole scaffold is a good candidate for future M<sup>Pro</sup> inhibitor drug development.

## 2. Methods

### 2.1. Compound property analysis

Two *in-silico* Pan-Assay Interference Compounds (PAINS) identification screens were conducted (<http://www.cbligand.org/PAINS/> [49], and <http://zinc15.docking.org/patterns/home/>) for **I** – **V**. Additionally, molecular properties and predicted absorption, distribution, metabolism, and excretion (ADME) values were calculated using Schrödinger's Qik-Prop program [50]. All default settings were used. In the Results section, we note any properties that violate the 95% range of known drugs.

### 2.2. Synthesis

Compounds **I** and **II** have been previously reported and their spectral details are included in Supplemental Information (Figures S1–S6) [47]. The synthesis of the thiazolyl-indazole heterocycles **IV** and **V** is being reported for the first time in this report. Each compound was prepared individually in a one pot reaction of the corresponding natural monoterpene ketone, thiosemicarbazide, and the phenacyl bromide in refluxing ethanol conditions (Scheme 2).

The general synthetic procedure was as follows. To an ethanolic solution (50 mL of absolute ethanol), 1 equivalent each of phenacyl bromide **1** and thiosemicarbazide **2** was added to 1 equivalent of (R)-Pulegone to produce **IV**, or to (R)-Menthone to produce **V**. The reaction mixture was heated under reflux for 1 h. After evaporating the solvent, the residue was diluted with water (10 mL) and extracted with dichloromethane. The organic layer was separated, dried on anhydrous  $\text{MgSO}_4$ , evaporated to dryness, and then purified by silica gel column

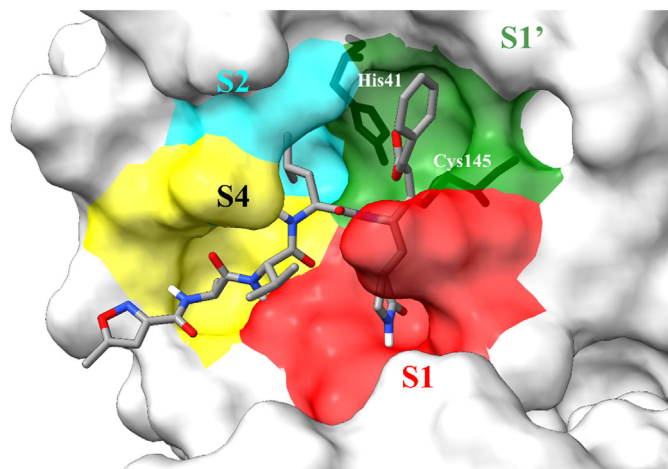
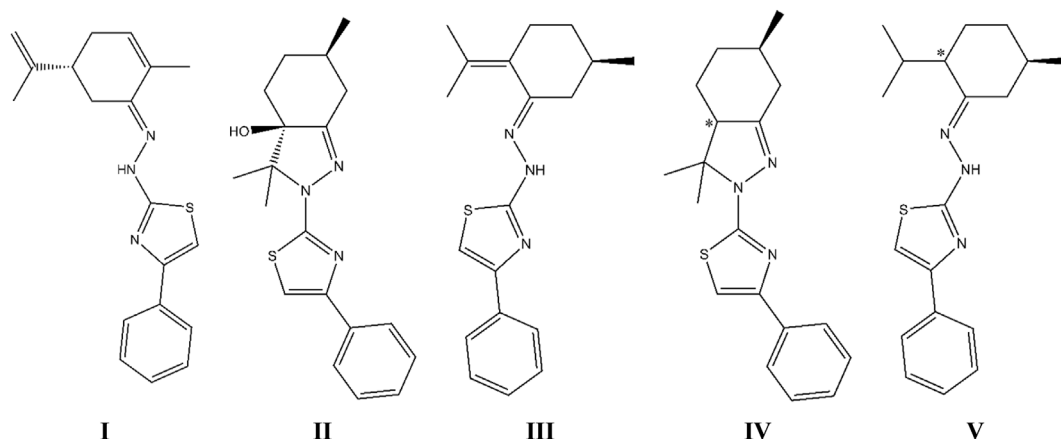
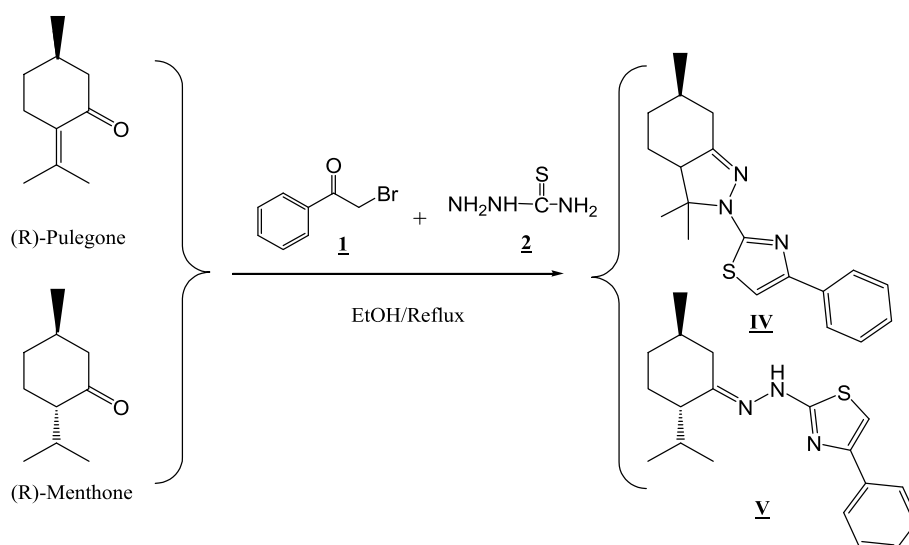


Fig. 1. Subsites of the SARS-CoV-2 M<sup>Pro</sup> with the N3 inhibitor covalently bound (PDB Code 6lu7). Subsites S1', S1, S2 and S4 are displayed in green, red, cyan, and yellow, respectively. Catalytic residues His41 and Cys145 are shown in black below the surface of the subsites. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)



**Scheme 1.** Structures of thiazolyl-indazole derivatives under consideration for SARS-CoV-2 M<sup>Pro</sup> inhibition.



**Scheme 2.** Synthesis of the thiazolyl-indazole heterocycles **IV** and **V**

chromatography, using hexane/ethyl acetate as eluent, to obtain the corresponding thiazolyl-indazole compounds in good yield (**IV**: 82%; **V**: 70%). The structures of the two synthesized products were confirmed based on their spectral data (Figures S7 – S12). The <sup>1</sup>H NMR spectra showed the appearance of the thiazolic proton H<sub>5'</sub> as a singlet at 6.82 ppm and at 6.86 ppm for **IV** and **V**, respectively. In addition, each compound displayed peaks in the range of  $\delta_{\text{H}} = 7.29\text{--}7.92$  for **IV** and  $\delta_{\text{H}} = 7.29\text{--}7.82$  for **V** corresponding to a phenyl group resonance. The <sup>13</sup>C NMR spectra reveal the thiazolic carbon C<sub>5'</sub> at 102.40 ppm and 102.79 ppm for **IV** and **V**, respectively. Further NMR spectral details can be found in the Supplemental Information.

### 2.3. Biochemical assay

All compounds were tested using the 3CL Protease, Untagged (SARS-CoV-2) Assay Kit (BPS Bioscience, San Diego, CA). The manufacturer's protocol was followed with slight modifications. Briefly, 5X serial dilutions of test compounds or covalent inhibitor control GC376 were prepared in water. M<sup>Pro</sup> was diluted in the Assay Buffer containing 1 mM DTT at 0.5 ng/μL. The diluted enzyme solution was pipetted in duplicate into a 96-well half area opaque plate (30 μL/well) while buffer alone was pipetted into “blank wells.” Dilutions of compounds **I**, **II**, and **V**, GC376, or 5% DMSO in water was added to wells containing the enzyme to 1X concentrations indicated in Fig. 2, S13, and S14 and incubated at room

temperature for 30 min with slow shaking. Protease Substrate was added to all wells and incubated with slow shaking for an additional 4 h. Fluorescence measurements were taken using a SpectraMax i3 plate reader (Molecular Devices) with fixed excitation (360 nm) and emission (460 nm). The endpoint values were recorded using the Softmax Pro software and all raw data was normalized with the “blank” solutions.

### 2.4. Structure retrieval and preparation

The structure of the SARS-CoV-2 M<sup>Pro</sup> was obtained from the Protein Data Bank (PDB code 6lu7) [51]. All waters and N3 were removed. A dimerized model of M<sup>Pro</sup> was created by aligning 6lu7 polypeptides to chains A and B of Human Coronavirus NL63 M<sup>Pro</sup> (PDB code 5gwy). Previous studies have shown that the monomer is functionally inactive due to the role N-terminal residues of the second chain play in forming the active site [52]. Schrödinger's Protein Preparation Wizard was used together with Prime, Epik, and PROPKA to prepare the protein [53–55]. Further protein preparation details can be found in the Supplemental Information [53–56].

### 2.5. Receptor Grid Generation, ligand preparation, and Glide Docking

Schrödinger's Receptor Grid Generation program was used to generate a 40 Å by 40 Å by 40 Å receptor grid with a ligand size cutoff of

20 Å. This grid was centered on the previously identified catalytic site of M<sup>Pro</sup> and used for all subsequent ligand docking using default parameters [12]. Structures of compound *I* – *V* were manually built and optimized according to the GAFF force field using Avogadro 1.2 [57,58]. Geometry optimizations in PCM water at the B3LYP/6-31G\* level of theory were also performed using Gaussian 16 [59]. Both diastereomers (*3aR,6R*) and (*3aS,6R*) of *IV* were constructed (*IV-E* and *IV-Z* respectively). Compound *V*, with the exocyclic double bond removed, resulted in an altered cyclohexane structure that placed the methyl and isopropyl groups cis (*V-E*) or trans (*V-Z*) to each other (Table 1). As such, *V-E* and *V-Z* were also built. Schrödinger's Glide Docking program was used to dock these molecules into the catalytic site of M<sup>Pro</sup> [60,61]. Two poses, the primary pose (1°) that reports the most favorable GScore and an additional top scoring secondary pose (2°) with different binding interactions, were selected for further MD evaluation. Docking tools such as Glide are best used as idea generators and selecting poses with different binding interactions helps speed surface coverage in subsequent MD simulation. These poses and their GScores are viewable in Table S1. Additional detail on Glide docking is provided in the Supplemental Information.

## 2.6. Benchmarking: substrate preparation and docking

Three 5-mer P4-P3-P2-P1-|P1' substrates, based on the best-recognized proteogenic amino acid M<sup>Pro</sup> substrate reported by Rut et al., were built using Schrödinger's Maestro program [62,63]. The C-terminal ACC dye was replaced with common P1'-site residues Ser, Ala, and Gly [15]. To protect against the influence of terminal charges, all substrates were ACE and NME capped. The sequences of these substrates are ACE-Val-Lys-Leu-Gln-[Ser, Ala, Gly]-NME (Figure S15A). These substrates will be referred to as VKLQS, VKLQA, and VKLQG.

As with our treatment of *I* – *V*, binding site docking using the Schrödinger Glide Docking program was conducted. However, as this program is not intended for peptide docking, this expectedly failed to generate poses aligning with previously described substrate recognition sequence – catalytic site interactions [15,62,64]. Therefore, using UCSF Chimera, each substrate was manually positioned within the M<sup>Pro</sup> catalytic site to maximize the interactions between the substrate recognition sequence residues P1-Gln and P2-Leu and the M<sup>Pro</sup> S1 and S2 pockets [65]. The S1 and S2 pockets have been previously noted to be invariably occupied by the P1-Gln and P2-Leu residues, respectively, while low specificity is noted for the P3 and P4 residues [62,64].

Our docking results were used merely as a starting structure for subsequent molecular dynamics simulations. To speed surface coverage for each system studied, we utilized two distinct conformations as simulation starting structures. Ideally, experimental structures (or modifications thereof) are best for initiating MD simulation; however, in cases such as the M<sup>Pro</sup>:substrate complex, where there is no experimental structure available, subsequent surface sampling imparted by the use of molecular dynamics mitigates concerns associated with our use of manually docked initial structures.

**Table 1**

Molecular properties and ADME values predicted by QikProp. Starred values fall outside of the 95% range of known drugs. The potential for QT-prolongation issues is noted, with QikProp flagging predicted logIC<sub>50</sub> values below –5 (IC<sub>50</sub> of 10000 nM). Approximate IC<sub>50</sub> values for each compound were calculated. H-bond donor and acceptor values estimate the number of such bonds between solute and water, averaged over a number of configurations.

| Cmpd.       | MW (Da) | H-Bond Donor | H-Bond Acceptor | logP Octanol/Water | logS Aqueous Solubility | logBB Brain/Blood | logIC <sub>50</sub> HERG K <sup>+</sup> | IC <sub>50</sub> (nM) HERG K <sup>+</sup> |
|-------------|---------|--------------|-----------------|--------------------|-------------------------|-------------------|---|---|
| <i>I</i>    | 323.455 | 1            | 3.50            | 4.749              | –5.203                  | –0.001            | –5.275                                  | ~5308.8                                   |
| <i>II</i>   | 341.470 | 1            | 3.25            | 4.943              | –6.329                  | 0.165             | –5.363                                  | ~4335.1                                   |
| <i>III</i>  | 325.471 | 1            | 3.50            | 5.228              | –6.182                  | 0.308             | –5.666                                  | ~2157.7                                   |
| <i>IV-E</i> | 325.471 | 0            | 2.50            | 5.471              | –6.739*                 | 0.449             | –5.423                                  | ~3775.7                                   |
| <i>IV-Z</i> | 325.471 | 0            | 2.50            | 5.496              | –6.773*                 | 0.453             | –5.447                                  | ~3572.7                                   |
| <i>V-E</i>  | 327.487 | 1            | 3.50            | 5.269              | –6.241                  | 0.234             | –5.547                                  | ~2837.9                                   |
| <i>V-Z</i>  | 327.487 | 1            | 3.50            | 5.188              | –6.510*                 | 0.147             | –5.924                                  | ~1191.2                                   |

After manual positioning, a minimization calculation was then performed under default settings using UCSF Chimera's Minimize Structure tool. The positions of the P1-Gln and P2-Leu sidechains were fine-tuned using Schrödinger's Maestro program [63]. Atoms/sidechains were manually moved, with localized minimizations performed with each movement. The P1-Gln sidechain was positioned to allow for contacts with Phe140, His163, and Gln166 in the S1 pocket, while the P2-Leu sidechain was moved into the S2 pocket. Additionally, the P1-Gln – P1' amide bond was positioned between catalytic residues His41 and Cys145. Following these movements, optimization and minimization was performed using Schrödinger's Protein Preparation Wizard according to the same steps previously detailed. This resulted in a common binding pose for all three substrates. This pose is detailed in Figure S15B. Additionally, ligand interaction diagrams for each substrate are viewable in Table S2.

## 2.7. Benchmarking: X77 inhibitor preparation and docking

The structure of *X77* was extracted from PDB code 6w63. Schrödinger's Glide Docking program was used to dock *X77* into the catalytic site of M<sup>Pro</sup>. This was conducted using the same protocol as previously detailed for *I* – *V*. Two poses were selected for further MD evaluation. The first pose reported the most favorable GScore and the second pose was obtained from 6w63 by aligning the structure onto chain A of our dimerized 6lu7 model and subsequently removing all 6w63 waters and peptide. These poses are viewable in Table S1.

## 2.8. Molecular dynamics simulations

Unrestrained molecular dynamics (MD) simulations were conducted on M<sup>Pro</sup> bound to compounds *I* – *V*, *X77*, and three 5-mer substrates, using the AMBER18 suite [66]. The 1° and 2° poses for *I* – *V*, and *X77*, were selected for MD analysis. The manually oriented binding poses detailed above were used to initiate MD for the 5-mer substrate. The *ff14SB* force field was applied to M<sup>Pro</sup> and the 5-mer substrates [67]. The program *antechamber* was used to apply the GAFF force field and AM1-BCC charges to *I* – *V* and *X77* [57,68–70]. All models were neutralized with Na<sup>+</sup> ions and explicitly solvated in a TIP3P unit cell using the program *tleap* [71]. All simulations were performed using the GPU-accelerated *pmemd* code of AMBER18 [72,73]. Further details describing the MD protocol can be found in the Supplemental Information. In total, nineteen 1000 ns ensembles were generated (two for each compound *I* – *III*, *IV-E,Z*, and *V-E,Z*), one for each of the three possible 5-mer substrates, and two for *X77*). The total number of simulations performed is detailed in Table S3.

## 2.9. Molecular dynamics analyses

From each 1000 ns ensemble, all frames in which the ligand either sampled binding positions outside of the catalytic site or dissociated from

the protein entirely where removed. All data analysis was conducted on the resultant truncated ensembles including only frames in which the ligand interacts with the catalytic site. As such, percent occurrence and thermodynamic data reported below are relative to each truncated ensemble. This approach allows us to evaluate the specific interaction energies between the active site of the M<sup>Pro</sup> and the various ligands without the variability associated with normal thermal motion and the random components of the MD algorithm. To avoid any binding overestimation associated with this approach, we also track percent dissociation times for each compound.

Trajectory visualization was conducted using UCSF Chimera and UCSF ChimeraX [65,74]. MM-GBSA binding free energy, per-residue decomposition, and normal mode entropy analyses were conducted using the AmberTools *MMPBSA.py* package [75]. Entropic analysis has been shown to scale binding free energies closer to experimental values while also providing improved comparison of binding affinities across models [76]. This is especially important for comparison between diverse structures like peptide substrates and small molecule inhibitors. Entropy calculations were performed on each truncated ensemble of *I* and *III* (with *II*, *IV* and *V* excluded due to experimental and computational shortcomings reported in the Results section) with a 12.5 ns interval. Entropy calculations were also performed on the substrates and the X77 inhibitor. Hydrogen bonding, center-of-mass distance (COM), and root-mean-squared deviation (RMSD) analyses were conducted on *I* - *III* using the AmberTools *cpptraj* module. Potential aromatic -  $\pi$  interactions were screened using literature-based COM distance cutoffs as detailed in Table S4 [77–82].

Ensemble averaged binding structures were created for the *I* - *III* truncated ensembles using *cpptraj* RMSD-based clustering. Compounds *IV* and *V* were excluded from this analysis due to shortcomings relative to *I* - *III*. The RMSD of each frame within each ensemble relative to its respective average binding structure was then calculated. Frames reporting an RMSD value below 1.75 Å were considered to sample the average structure. This cutoff value, intended to account for the dynamic behavior of each structure, was determined from trajectory visualization. Note that the percent occurrence of a structure that undergoes more fluctuation will be underreported. When RMSD analysis alone is deemed insufficient, center-of-mass (COM) analysis is additionally used with cutoff distances selected on a case-by-case basis.

Our simulation protocol is optimized for non-covalent inhibitors and cannot be applied to the covalent inhibitors that have been shown to bind to the SARS-CoV-2 M<sup>Pro</sup>. As such, a computational comparison to the biochemical assay control compound GC376 could not be conducted. In lieu of this, comparison to X77 and the substrates is provided.

### 3. Results and discussion

#### 3.1. PAINS and QikProp analysis

All compounds passed both PAINS screens with no points of concern. There were no 95% range violations or reactive functional groups reported for *I* - *III*, however *IV* and *V* generated range violations, and ADME screening indicated a reactive functional group in *V* (Table 1). Aqueous solubilities for *II*, *III* and *V-E* lie notably close to the logS -6.5 lower limit, while both isomers of *IV* and *V-Z* exceed it. Predicted brain/blood partition coefficients show that *I* displays the lowest potential for CNS activity while *III* and *IV* display the most. All compounds have predicted HERG K<sup>+</sup> channel logIC<sub>50</sub> values below -5 and as such are flagged as potential concerns. Compound *III* and *V-Z* report the largest potential for HERG K<sup>+</sup> channel blockage. It should be noted that this does not necessarily suggest that these or other thiazolyl-indazole compounds are not viable drug candidates. As noted in the QikProp version 6.1 User Manual, failed drug candidates displaying QT-prolongation issues exhibit a large range of IC<sub>50</sub> values, with cisapride (IC<sub>50</sub> 6.5 nM, logIC<sub>50</sub> -8.187) and grepafloxacin (IC<sub>50</sub> 50000 nM, logIC<sub>50</sub> -4.301) noted as examples. Ultimately, drug compound class, required dosage, and bioavailability

determine the allowable IC<sub>50</sub> limit [50]. Taken together, these results suggest that *I* displays the most favorable drug-like properties.

Property analysis largely eliminates *IV*-(*E,Z*) and *V*-(*E,Z*) as potential drug candidates, as both display shortcomings relative to *I* - *III*. ADME screening indicates that the removal of the hydroxy group from *II*, to form *IV*, results in *IV* having an aqueous solubility outside the 95% range of known drug-like molecules, along with an increased potential for CNS activity and HERG K<sup>+</sup> channel blockage. This suggests that the hydroxy group of *II* may be necessary for maintaining drug-like properties. Likewise, ADME screening indicates that removal of the isopropyl - cyclohexane double bond in *III* to produce *V* results in the cyclohexane - linking imine becoming a potentially reactive carbonyl center (Table 1).

#### 3.2. Biochemical evaluation

In order to evaluate the ability of the compounds to disrupt the activity of M<sup>Pro</sup>, we employed a commercially available kit (see methods). We confirmed the validity of the assay using the known covalent inhibitor GC376 obtaining an IC<sub>50</sub> of 32.5 nM, similar to literature values (Figure S13) [37]. Since the assay relies on the cleavage of an internally quenched EDANS fluorophore ( $\lambda_{ex}$  360 nm,  $\lambda_{em}$  460 nm), we first examined the background fluorescence of the synthesized compounds *I*, *II*, *IV*, and *V*. As seen in Figure S16, compound *IV* exhibited high fluorescence when excited at 360 nm so it was excluded from biological testing. Due to solubility issues, we were only able to test compounds *II* and *V* up to 100  $\mu$ M and did not observe inhibition of M<sup>Pro</sup> activity at these concentrations (Figure S14). There were no solubility issues with compound *I*, in agreement with our ADME prediction that *I* is the most water soluble, while *II* - *V* lie close to or even surpass the 95% solubility range of known drug-like molecules, and therefore we were able to characterize the inhibitory activity of *I* across the full range of solubilities (Fig. 2, IC<sub>50</sub> 92.9  $\mu$ M).

#### 3.3. Molecular dynamics analysis

To better understand the atomistic nature of the experimental results, MD simulations initiated from the structurally distinct 1° and 2° Glide poses (Table S1) of each compound bound to M<sup>Pro</sup> were conducted, resulting in the production of a 1000 ns ensemble for all of the 1° and 2° Glide poses. Initial data analysis began with the removal of frames where the compound either dissociated from M<sup>Pro</sup> or sampled a binding position outside of the S1, S1', S2, and/or S4 subsites. A simple analysis based

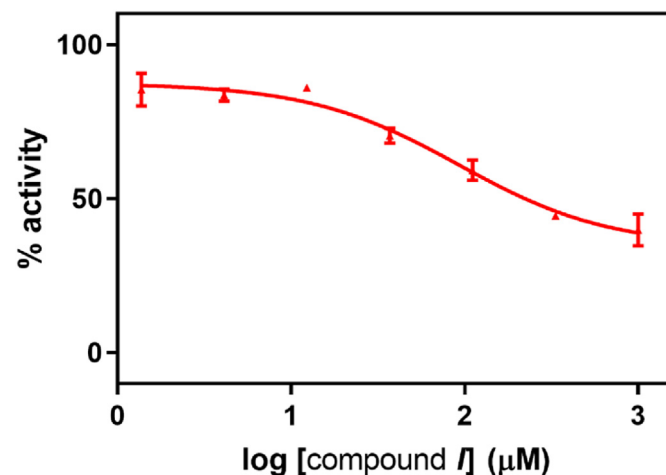


Fig. 2. M<sup>Pro</sup> activity at varying concentrations of compound *I*. Fluorescence intensity ( $\lambda_{ex}$  360 nm,  $\lambda_{em}$  460 nm) was measured after 4 h of reaction time as described in the methods and normalized by subtracting the blank and dividing by the 0  $\mu$ M control. The curve was fit in GraphPad Prism to give an IC<sub>50</sub> of 92.9  $\mu$ M (95% confidence interval 55.2–158.6  $\mu$ M).

**Table 2**

Full ensemble percent dissociations and truncated ensemble binding free energies for *I* – *V*. Frames encompassed by the full ensemble % dissociation [A] were removed, and the remaining frames [B] were used to create truncated ensembles. MM-GBSA binding free energies for the frames of these truncated ensembles and their standard errors are reported.

| Compound    | Ensemble | % Dissociation <sup>[A]</sup> /<br>% Truncated<br>Ensemble <sup>[B]</sup> | MM-GBSA<br>Binding Energies<br>(kcal/mol) |              |
|-------------|----------|---|---|--------------|
|             |          |   | Avg.                                      | Std.<br>Err. |
| <i>I</i>    | 1°       | 19.0/81.0   | -25.56                                    | 0.06         |
|             | 2°       | 6.0/94.0  | -26.83                                    | 0.05         |
| <i>II</i>   | 1°       | 29.0/71.0   | -26.21                                    | 0.07         |
|             | 2°       | 12.9/87.1   | -24.66                                    | 0.05         |
| <i>III</i>  | 1°       | 10.5/89.5   | -29.51                                    | 0.07         |
|             | 2°       | 9.4/90.6  | -26.09                                    | 0.07         |
| <i>IV-E</i> | 1°       | 28.2/71.8   | -24.23                                    | 0.08         |
|             | 2°       | 50.8/49.2   | -26.16                                    | 0.10         |
| <i>IV-Z</i> | 1°       | 39.3/60.7   | -23.04                                    | 0.07         |
|             | 2°       | 20.0/80.0   | -24.37                                    | 0.05         |
| <i>V-E</i>  | 1°       | 38.7/61.3   | -24.99                                    | 0.07         |
|             | 2°       | 38.3/61.7   | -22.73                                    | 0.07         |
| <i>V-Z</i>  | 1°       | 11.8/88.2   | -27.13                                    | 0.07         |
|             | 2°       | 23.9/76.1   | -23.37                                    | 0.07         |

solely on the percentage of frames of *I* – *V* that dissociated from M<sup>Pro</sup> suggest that *I*, *III* and *V-Z* show promise as good binders (Table 2). Increased dissociation of *II* and *IV* is likely the result of increased structural rigidity caused by the presence of the indazole group. Trajectory visualization suggests that this increased rigidity results in a diminished ability of *II* and *IV* to conform to catalytic site dynamics. To further evaluate binding, we calculated MM-GBSA binding free energies and average structures for *I* – *V* using truncated ensembles that included only frames where ligand dissociation did not occur (Table 2). Taken together, the estimated binding free energies, percent dissociation values, and properties prediction suggest that *I*, *III* and *V-Z* show the most promise as potential inhibitors. This agrees with our experimental results for *I*; unfortunately, we were unable to synthesize *III* for experimental testing, and *V* did not show inhibition at concentrations lower than 100  $\mu$ M. Limited solubility of *V* prevented us from testing at higher concentrations. It is noteworthy that our synthesis of *V* produced both the E and Z isomers and so it is possible that our experimental binding analysis at these concentrations was diminished by the presence of the E isomer. Data for each individual seed and the full non-truncated 1000 ns ensembles is detailed in Tables S5 – S11.

Compound *V* is not likely to be a viable candidate due to its physical properties; however, a comparison to the results obtained with this molecule provide structural insight into the role of the exocyclic isopropyl moiety on the binding of *I* and *III*. While our trajectory analysis of *I* and *III* does not provide evidence of any specific interactions between the isopropyl – cyclohexane double bond and the M<sup>Pro</sup> catalytic site (see below), the slightly reduced MM-GBSA binding estimations for *V* suggest that the positioning of the isopropyl group and overall conformation of the cyclohexane ring in *I* and *III* affect binding affinity. It is notable that *I*, with its non-planar isopropyl group, displays a somewhat less favorable binding affinity and higher percent dissociation than *III* which has a planar isopropyl. Likewise, *V-E* with a non-planar isopropyl oriented similarly to *I*, displays a less favorable binding affinity and higher percent dissociation than the *V-Z* diastereomer with the isopropyl group oriented to the opposite side of the molecule. These results suggest that the orientation of the isopropyl is important and that a *cis*-oriented isopropyl may be less suitable than a planar or *trans*-oriented isopropyl.

It is similarly possible that the presence of the double bond in or adjacent to the cyclohexane group affects binding affinity. Geometry optimizations performed on *I* at the B3LYP/6-31G\* level of theory in solvent (PCM water) indicate that this ring is largely planar with a pucker at the isopropyl group. The cyclohexane of *III* is also largely planar,

however the pucker is more pronounced and present at the two cyclohexane carbons between the methyl and isopropyl groups. Keeping in mind that *II* and *V* did not produce inhibition in our experimental studies, it is notable that the cyclohexane ring in *II* is a chair-like whereas in *V-E* and *V-Z* the cyclohexane adopts twist boat- and chair-like conformations, respectively (Table 2). Taken together, our experimental binding results and molecular modeling suggests that some planarity in the cyclohexane ring may be important to binding.

Per-residue energy decomposition analysis was conducted on each truncated ensemble of *I* – *III* (Table 3). Compound *II* was included for comparison to *I* and *III*, even though *II* did not show good binding at lower concentrations. The catalytic His41, and Met49, Met165, and Gln189 report highly favorable decomposition energies across all compounds, suggesting their importance for binding interactions. Apart from Gln189, these interactions are largely driven by van der Waals effects.

Visualization of the truncated ensembles suggests close interactions between many of the above residues and the phenyl and thiazole groups of *I* – *III*. To further elucidate the importance of this interaction, COM distance analysis was conducted between these aromatic rings and the functional groups of the noted sidechains of *I* – *III* (Table S12). This analysis suggests the occurrence of various interactions, most notably  $\pi$ -stacking interactions with His41, and S/ $\pi$  interactions with Met49 and Met165. These compounds display similar per-residue decomposition profiles, and thus the occurrence of similar  $\pi$  interactions is not surprising. Additionally, the potential for OH/ $\pi$  interactions with Ser144, SH/ $\pi$  interactions with Cys145, and anion/ $\pi$  interactions with Glu166 are noted with *II* and *III*. These Glu166 interactions resemble Asp anion/ $\pi$  interactions reported in the work of Ellenbarger et al. (face-on packing to the aromatic  $\pi$ -cloud) [83]. The Cys SH/ $\pi$  interaction occurs only with *II*, while potential Ser OH/ $\pi$  interactions occur only with the phenyl group of *II* and with the thiazole of *III*. While the interactions in *II* suggest promise for the indazole scaffold as the basis of a potential M<sup>Pro</sup> inhibitor,

**Table 3**

MM-GBSA per-residue total energy decomposition across all compounds/ensembles. All residues with an average total energy contribution of < -0.5 kcal/mol are shown. Similarly, if a residue contributed an average total energy of > -0.5 kcal/mol for a particular ensemble, that energy is not reported. The four bolded residues, His41, Met49, Met165, and Gln189 are shown to contribute energy decompositions important for binding across all compounds and truncated ensembles.

| Residue       | Average Total Energy Decomposition (kcal/mol) |             |              |              |               |               |
|---------------|---|-------------|--------------|--------------|---------------|---------------|
|               | <i>I</i> 1°                                   | <i>I</i> 2° | <i>II</i> 1° | <i>II</i> 2° | <i>III</i> 1° | <i>III</i> 2° |
| Leu27         | -   | -           | -            | -0.75        | -             | -0.85         |
| <b>His41</b>  | -1.14   | -1.11       | -            | -2.13        | -1.05         | -2.24         |
|               | $\pm 0.74$                                    | $\pm 0.54$  | -            | $\pm 0.75$   | $\pm 0.61$    | $\pm 1.41$    |
| <b>Met49</b>  | -1.47   | -0.89       | -0.52        | -1.51        | -1.04         | -1.73         |
|               | $\pm 0.95$                                    | $\pm 0.68$  | $\pm 0.53$   | $\pm 0.96$   | $\pm 0.62$    | $\pm 0.78$    |
| Leu141        | -   | -           | -            | -            | -1.27         | -             |
| Asn142        | -   | -           | -1.46        | -            | -0.92         | -             |
|               | -   | -           | $\pm 1.95$   | -            | $\pm 0.71$    | -             |
| Gly143        | -   | -           | -0.69        | -            | -             | -             |
| Ser144        | -   | -           | $\pm 1.12$   | -            | -             | -             |
|               | -   | -           | -0.70        | -            | -0.88         | -             |
| Cys145        | -   | -           | $\pm 0.71$   | -            | $\pm 0.57$    | -             |
|               | -   | -           | -0.52        | -0.63        | -0.64         | -0.60         |
| His164        | -   | -           | $\pm 0.59$   | $\pm 0.40$   | $\pm 0.49$    | $\pm 0.38$    |
|               | -   | -           | -            | -            | -1.66         | -             |
| <b>Met165</b> | -1.31   | -1.49       | -1.71        | -1.33        | -1.82         | -1.36         |
|               | $\pm 0.64$                                    | $\pm 0.90$  | $\pm 1.37$   | $\pm 0.73$   | $\pm 0.80$    | $\pm 0.87$    |
| Glu166        | -   | -           | -0.68        | -            | -0.83         | -             |
| Asp187        | -1.07   | -0.84       | -            | -0.83        | -             | -             |
|               | $\pm 0.56$                                    | $\pm 0.63$  | -            | $\pm 0.70$   | -             | -             |
| <b>Gln189</b> | -1.63   | -2.26       | -1.15        | -0.81        | -0.60         | -0.98         |
|               | $\pm 1.24$                                    | $\pm 0.99$  | $\pm 1.32$   | $\pm 1.11$   | $\pm 0.82$    | $\pm 1.16$    |

we do not see experimental inhibition at lower concentrations. In addition to the  $\pi/\pi$ ,  $S/\pi$ ,  $OH/\pi$  and anion/ $\pi$  interactions, we also looked at hydrogen bonding patterns for the binding of *I*. Compound *I* donates a hydrogen bond to Gln189 for 28.56% of the truncated ensembles, in agreement with the decomposition energies reported in Table 3. Solvent-bridged hydrogen bonds are additionally noted with Glu166 (11.17%) and Gln189 (9.33%). Hydrogen bonding patterns for *I*–*III* can be found in the Supplemental Information (Tables S13 – S14).

### 3.4. Average binding structures of *I* and *III*

Compound *I* has experimentally verified  $M^{Pro}$  binding behavior suggesting that it may prove useful as an inhibitor scaffold for future development. Compound *III* is experimentally untested, but our computational results and the similarity of the underlying thiazole backbone structure with *I* warrants its inclusion for further structural analysis. Within each truncated ensemble of *I* and *III*, an average binding structure within the catalytic site was identified through *cpptraj* analysis. These structures, visualized in Fig. 3 and S17, provide deeper insight to the per-residue decomposition energetics and potential  $\pi$  interactions. Three average binding structures were identified for *I*, while two were identified for *III*. These structures are described as dominant or non-

dominant, with the dominant structure reporting the higher percent occurrence. Collectively, the dominant and non-dominant average structures shown in Fig. 3 and S17 represent 93.16 and 60.99% of frames within the *I* and *III* ensembles, with the remaining frames sampling disordered binding interactions. RMSD and COM graphs visualizing the percent occurrence of these structures over each ensemble are viewable in Figures S18 – S22. MM-GBSA binding free energy analysis of only the frames sampling these average structures was conducted (Fig. 3B and Figure S17B). The binding free energy of the *I* dominant structure closely matches the *I* average  $\Delta\Delta G$  (–26.46 vs. –26.24 kcal/mol). Most notably, both average structures of *III* report  $\Delta\Delta G$ 's more favorable than the *III* average  $\Delta\Delta G$  (–30.93 and –29.27, vs. –27.79 kcal/mol). This suggests increased binding affinity of the *III* average structures over disordered binding interactions.

Differences between the three average structures of *I* are noted in Fig. 4 and S17. RMSD visualization in Figure S18 indicates that sampling of the *I* non-dominant structures (22.20 and 14.62%) occurs near the beginning of each simulation, while sampling of the dominant structure occurs later in the trajectory. This transition to the dominant structure occurs even though the non-dominant structure (14.62%) displays an increased binding affinity (–28.46 for non-dominant *I* vs. –26.46 kcal/mol for dominant *I*). The sampling of these structures suggests the

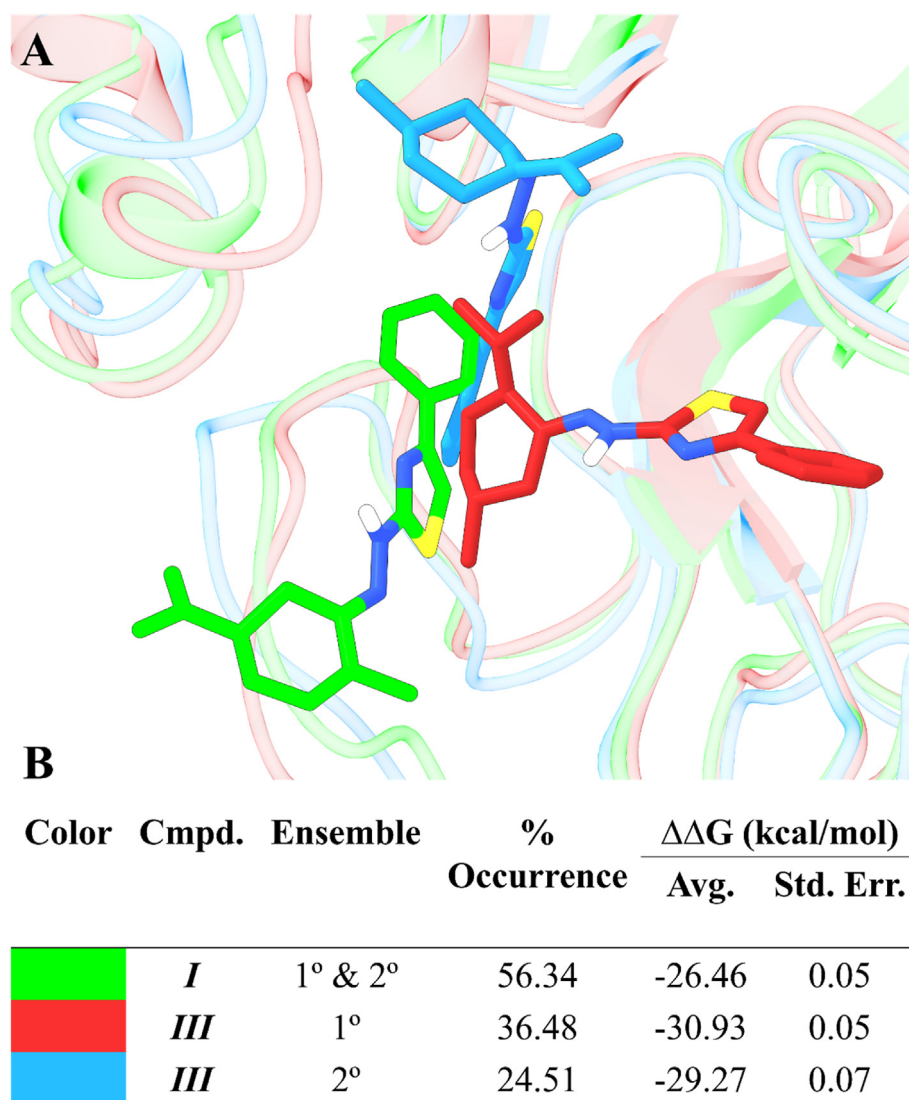


Fig. 3. A. Average binding structures of *I* and *III* within the  $M^{Pro}$  catalytic site. B. A table providing a color key and indicating which ensemble each structure shown in A was found in. The overall % occurrence of each average structure is reported, along with the MM-GBSA binding free energies of only the frames sampling these binding structures. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)



possibility of an energetic barrier separating these important conformations. The repeated occurrence of the *I* dominant structure across individual trajectories suggests that it is entropically favored and explains similarities in per-residue decomposition interaction energies between the two ensembles of *I* despite structural differences in the poses used to initiate simulation (Table 3). Trajectory visualization indicates an RMSD- and COM-indistinguishable variation of the *I* dominant structure where the (*R*)-carvone group is oriented inwards towards His41, Met49, and Met165 (Fig. 4B). As this variation orients the phenylthiazole groups away from the three noted residues, His  $\pi$ -stacking and S/ $\pi$  interactions are not possible in this conformation.

It should be noted that in order to generate a classic-textbook, non-dynamic pharmacophore, Fig. 3A needs to be combined with 3B, i.e. atomistic physics-based simulations capture the thermal motion and energetic variability associated with real molecules, and as such almost never produce rigid pharmacophores. To obtain such a pharmacophore, the conformations in Fig. 3A are combined with the energetics and percent occurrences (or importance weightings) shown in 3B. Such a composite analysis would result in an image for compound *I* with only one orientation (due to the high percent occurrence (56%) and the overall conformer stability ( $-26$  kcal/mol; second most stable, higher than the lowest energy conformer by only 2 kcal/mol and close enough to be considered approximately isoenergetic). Compound *III* does in fact sample two distinct orientations as noted by the two conformations with similar weightings and isoenergetics. Pharmacophores with multiple orientations are not unusual [84,85].

Mapping regions of *I* and *III* to the M<sup>Pro</sup> subsites reveals commonalities amongst the average structures of *I* (Fig. 5A – D). All average structures of *I* occupy the S2 and S4 subsites. This agrees with our per-residue decomposition (Table 3) and COM distance data (Table S12) suggesting interactions with His41, Met49, Met165, and Gln189. Additionally, the reversed *I* dominant structure (Fig. 5B), the *I* non-dominant structure (14.62%) (Fig. 5D), and the *III* dominant structure (Fig. 5E) project their isopropyl and cyclohexane groups into the S2 pocket. Positioning of the isopropyl group into this pocket is expected due to the S2 subsite favoring occupation by a Leu residue. As discussed above, simulations on *V* suggest that the positioning of the isopropyl group affects affinity and so we were not surprised to see it appear as a common binding motif in *I* and *III*; the ability to occupy the S2 subsite is likely related to the conformational preferences of the various isopropyl moieties. Notably, *I* is the only compound to significantly interact with the S4 subsite, while the dominant structure of *III* (Fig. 5E) is the only average structure to bridge both the S1 and S2 subsites. Overall, all average

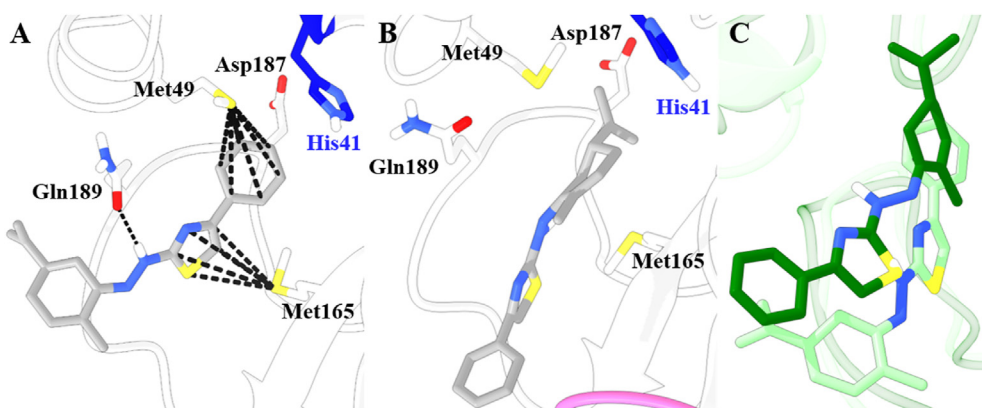
structures demonstrate occupation of the previously noted important subsites for inhibitor binding. Further structural modification and optimization of these scaffolds may allow for additional interactions with various subsites and overall improved binding and inhibition.

### 3.5. Comparison to 5-mer substrates

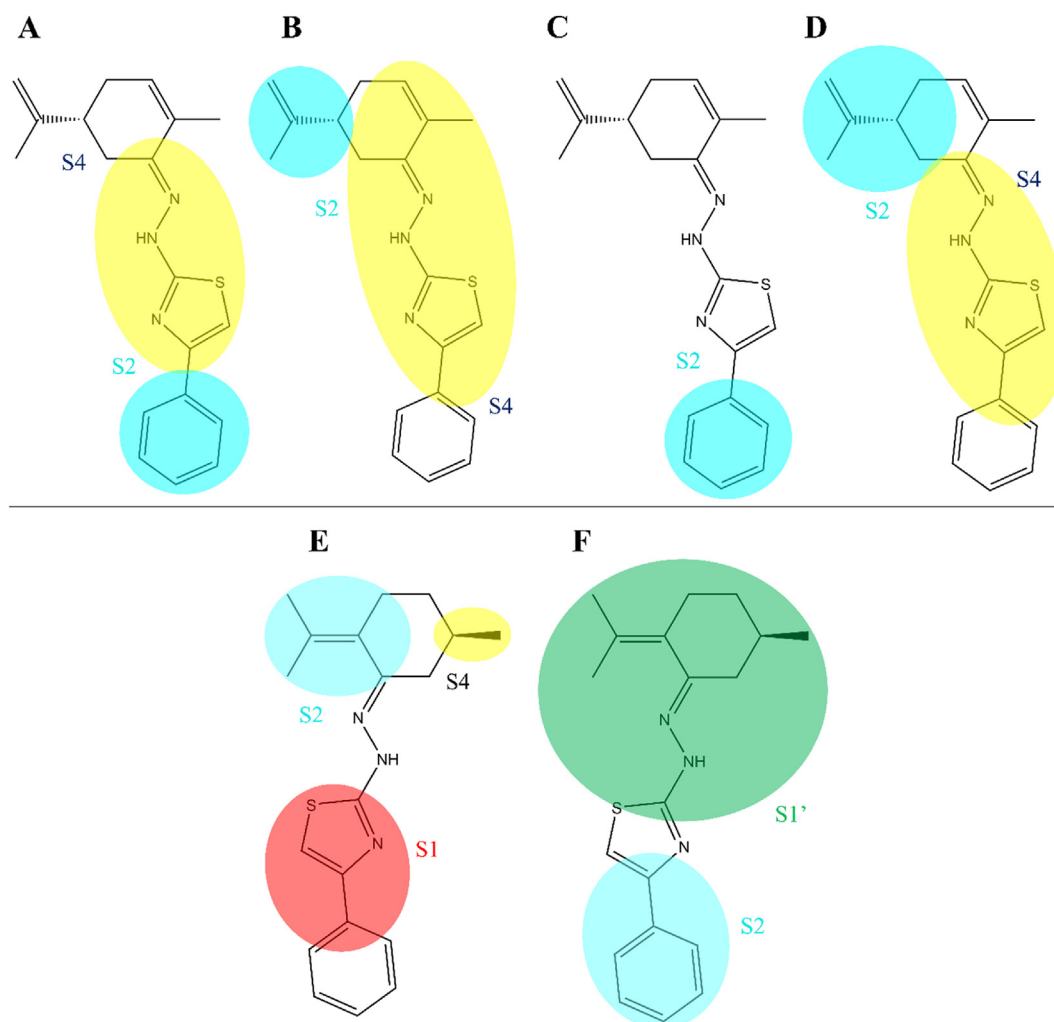
We performed MD simulations on three 5-mer M<sup>Pro</sup> substrates to provide a detailed binding comparison with the thiazole compounds. As above, truncated ensembles were produced and MM-GBSA binding free energy and per-residue decomposition analyses were conducted. Binding affinities between the VKLQA and VKLQS substrates appear highly similar, with both reporting average  $\Delta\Delta G$ 's of  $-34.88$  kcal/mol. The VKLQG substrate reports a notably less favorable binding affinity and a higher percent dissociation from the catalytic site. This is expected, as coronavirus M<sup>Pro</sup> appears to more readily select for Ser and Ala in the P1' position over Gly [86,87]. These energies, reported in Table 4, suggest favorable binding for all three substrates, and on a scale that is relatively comparable to *I*. (Data for each individual seed is viewable in Table S15.)

Normal mode calculations were performed in order to include entropic effects in our binding free energy estimations, and to allow for comparison of *I* and *III* to the 5-mer substrates. These results are displayed in Table 5. Expectedly, the 5-mer substrates report greater average entropic penalties than that of the phenylthiazole compounds. Corrected binding free energies, calculated by assuming the MM-GBSA  $\Delta\Delta G$  to be roughly equivalent to  $\Delta H$  and subtracting the entropic penalty, suggests that *I* and *III* show greater binding affinities than the 5-mer substrates. Compound *III* stands out amongst the phenylthiazole compounds, with the primary and secondary truncated ensembles displaying corrected binding free energies of  $-9.07$  and  $-6.52$  kcal/mol, respectively, while compound *I* displays corrected binding free energies in the range of  $-5.4$  to  $-5.9$  kcal/mol. This further supports the promise of these compounds to serve as scaffolds for further inhibitor development. It should be noted that in these benchmarking tests we are using truncated peptides and as such likely underestimate binding. Also, our simulations do not include any chaperoned folding or other *in vivo* effects.

Per-residue decomposition and hydrogen bonding analyses reveal similar interactions to those described by Chan et al. [27] Many of the same residues noted to interact with their 11-mer substrates also interact favorably with our 5-mer substrates (Table S16). Interactions with these residues are also noted with *I* (Table 3). While the VKLQG substrate does not demonstrate any consistent hydrogen bonds, backbone – backbone interactions between Glu166 and P3 Lys are reported with both VKLQA



**Fig. 4.** Compound *I* dominant structure variations. **A.** *I* dominant structure with phenylthiazole oriented inwards towards His41, Met49, and Met165. Potential S/ $\pi$  interactions are displayed, and while a  $\pi$ -stacking interaction with His41 is not shown in this frame, the phenyl group location allows for these interactions across frames sampling this structure. A different orientation of this binding structure is also shown in Fig. 3A. **B.** *I* dominant structure with (*R*)-carvone oriented inwards towards His41, Met49, and Met165. No potential  $\pi$  interactions are noted. **C.** Overlay of both *I* average structure variations, with A shown in light green and B shown in dark green. While both of the displayed variations occupy the same space and report low RMSD and COM distance values of 0.86 and 1.92 Å respectively, COM analysis reveals that the phenylthiazole groups are located 6.95 Å apart. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)



**Fig. 5.** Compounds *I* and *III* average structure interactions with  $M^{Pro}$  pocket residues. Yellow indicates S4 pocket residues, cyan indicates S2 pocket residues, red indicates S1 pocket residues, and green indicates S1' pocket residues. Each average binding structure is displayed, with the *I* dominant structure and its reversal shown in **A** and **B**, the *I* non-dominant structures shown in **C** and **D**, and the *III* dominant and non-dominant structures shown in **E** and **F**, respectively. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

**Table 4**

5-mer substrate full ensemble percent dissociations and truncated ensemble binding free energies. Frames encompassed by the full ensemble % dissociation [A] were removed, and the remaining frames [B] were used to create truncated ensembles. MM-GBSA binding free energies for the frames of these truncated ensembles and their standard errors are reported.

| Substrate | % Dissociation <sup>[A]</sup> /% Truncated Ensemble <sup>[B]</sup> | MM-GBSA Binding Energies (kcal/mol) |           |
|-----------|--|-------------------------------------|-----------|
|           |  | Avg.                                | Std. Err. |
| VKLQA     | 13.0/87.0  | -34.88                              | 0.14      |
| VKLQS     | 19.2/80.8  | -34.88                              | 0.13      |
| VKLQG     | 30.1/69.9  | -22.64                              | 0.09      |

and VKLQS. This has previously been noted as a primary interaction responsible for holding the substrate in place [27]. Notable interactions are also observed between P2 Leu and His163 as well as P1 Gln and Gly143, Cys145, and His163. Some previously unreported, but less consistent hydrogen bonds are also observed. Most notable of these are interactions between the sidechains of P1 Gln and Ser144 and His41, as well as the P3 Lys backbone and the Gln189 sidechain (Table S17). Peptide: $M^{Pro}$  models are available at <https://github.com/Parish-Lab/sarscov2>.

**Table 5**

Truncated ensemble entropic penalties and entropy-adjusted binding free energies. Normal mode average binding entropies and their standard errors are reported. Entropy-adjusted binding free energies, used to compare binding of our novel compounds to the three  $M^{Pro}$  substrates, are also reported.

| Ensemble         | $\Delta S$ (kcal/mol) |           | MM-GBSA Binding Energy - $\Delta S$ (kcal/mol) |
|------------------|-----------------------|-----------|--|
|                  | Avg.                  | Std. Err. |  |
| <i>I</i> 1°      | -19.69                | 0.58      | -5.87  |
| <i>I</i> 2°      | -21.44                | 0.56      | -5.39  |
| <i>III</i> 1°    | -20.44                | 0.59      | -9.07  |
| <i>III</i> 2°    | -19.57                | 0.52      | -6.52  |
| VKLQA            | -32.55                | 0.82      | -2.32  |
| VKLQS            | -32.91                | 0.76      | -1.98  |
| VKLQG            | -29.67                | 0.83      | 7.04   |
| X77 (Glide pose) | -24.27                | 0.57      | -14.46   |
| X77 (6w63 pose)  | -23.52                | 0.48      | -26.01   |

### 3.6. Comparison to X77

Our benchmarking results on the tight-binder X77 suggests that our methodology is appropriate. An average MM-GBSA  $\Delta\Delta G$  of  $-44.16$  kcal/

mol is reported for **X77** and the simulation did not produce any dissociated frames indicative of a tight binding ligand. The ensemble initiated from the 6w63 binding pose reports a significantly more favorable MM-GBSA binding energy (-49.53 kcal/mol) than that initiated from the top Glide pose (-38.74 kcal/mol) confirming that experimental structures make the best initial poses for subsequent MD, and that there may be a variety of possible active site orientations with a range of favorability. Binding free energy for each individual seed and per-residue decomposition data is reported in Tables S18 – S19. The entropic contribution and adjusted  $\Delta\Delta G$  are reported in Table 5. Our simulations confirm that the enthalpy of **X77** binding is more favorable than any of the phenylthiazole scaffolds but less favorable than the various substrates. When entropy is considered, **X77** is a markedly more favorable binder than either the substrate or the phenylthiazole compounds, underscoring the importance of rigidity imparted by small-molecule drugs, and the need for further development of the thiazole scaffolds.

#### 4. Conclusion

Our experimental binding analysis suggests that previously synthesized phenylthiazole compound **I** is a promising scaffold for future inhibitor development. Computational analysis suggests that **I**, along with the unsynthesized **III**, pass PAINS screenings and display ADME properties in-line with known drug-like molecules. Docking analysis predicts these molecules to bind favorably to the SARS-CoV-2 M<sup>Pro</sup> catalytic site and molecular dynamics simulations suggest prolonged interaction with the enzyme. Subsequent MM-GBSA binding free energy and per-residue decomposition calculations suggest that **I** and **III** experience energetically favorable binding driven predominantly by interactions with residues His41, Met49, Met165, and Gln189. Importance is also demonstrated for residues Ser144, Cys145, Glu166, and Asp187 across some of the ensemble-averaged structures. A detailed structural comparison suggests that the orientation of the isopropyl moiety and conformation of the cyclohexane ring may be important for M<sup>Pro</sup> binding. Trajectory visualization, per-residue decomposition and COM analysis supports the importance of specific  $\pi$  interactions between the phenylthiazole moiety and M<sup>Pro</sup>. Binding is driven by  $\pi$ -stacking interactions with His41 and possibly S/ $\pi$  interactions with Met49 and Met165. Other notable interactions are supported, including OH/ $\pi$  and SH/ $\pi$  interactions with Ser144 and Cys145 respectively and anion/ $\pi$  interactions with Glu166. An entropically corrected binding comparison to 5-mer substrates with demonstrated experimental affinity further suggests that these compounds show promise as scaffolds for future development as SARS-CoV-2 M<sup>Pro</sup> inhibitor drugs.

#### Acknowledgements

This research was supported by funding from the National Science Foundation (Grant CHE-18800014), the Donors of the American Chemical Society Petroleum Research Fund and the Floyd D. and Elisabeth S. Gottwald Endowment. J.A. and C.A.B. were recipients of a summer fellowship from the Puryear-Topham-Pierce-Guption endowment from the Department of Chemistry at the University of Richmond. J.A. was also a recipient of summer support from the University of Richmond Integrated and Inclusive Science program, and the Arts and Sciences Undergraduate Research Committee.

#### Appendix A. Supplementary data

Supplementary data related to this article can be found at <https://doi.org/10.1016/j.ejmc.2022.100034>.

#### References

- [1] N. Zhu, D. Zhang, W. Wang, X. Li, B. Yang, J. Song, X. Zhao, B. Huang, W. Shi, R. Lu, P. Niu, F. Zhan, X. Ma, D. Wang, W. Xu, G. Wu, G.F. Gao, W. Tan, A novel

- coronavirus from patients with pneumonia in China, 2019, *N. Engl. J. Med.* 382 (8) (2020) 727–733, <https://doi.org/10.1056/NEJMoa2001017>.
- [2] Q. Li, X. Guan, P. Wu, X. Wang, L. Zhou, Y. Tong, R. Ren, K.S.M. Leung, E.H.Y. Lau, J.Y. Wong, X. Xing, N. Xiang, Y. Wu, C. Li, Q. Chen, D. Li, T. Liu, J. Zhao, M. Liu, W. Tu, C. Chen, L. Jin, R. Yang, Q. Wang, S. Zhou, R. Wang, H. Liu, Y. Luo, Y. Liu, G. Shao, H. Li, Z. Tao, Y. Yang, Z. Deng, B. Liu, Z. Ma, Y. Zhang, G. Shi, T.T.Y. Lam, J.T. Wu, G.F. Gao, B.J. Cowling, B. Yang, G.M. Leung, Z. Feng, Early transmission dynamics in Wuhan, China, of novel coronavirus-infected pneumonia, *N. Engl. J. Med.* 382 (13) (2020) 1199–1207, <https://doi.org/10.1056/NEJMoa2001316>.
- [3] P. Zhou, X.-L. Yang, X.-G. Wang, B. Hu, L. Zhang, W. Zhang, H.-R. Si, Y. Zhu, B. Li, C.-L. Huang, H.-D. Chen, J. Chen, Y. Luo, H. Guo, R.-D. Jiang, M.-Q. Liu, Y. Chen, X.-R. Shen, X. Wang, X.-S. Zheng, K. Zhao, Q.-J. Chen, F. Deng, L.-L. Liu, B. Yan, F.-X. Zhan, Y.-Y. Wang, G.-F. Xiao, Z.-L. Shi, A pneumonia outbreak associated with a new coronavirus of probable bat origin, *Nature* 579 (7798) (2020) 270–273, <https://doi.org/10.1038/s41586-020-2012-7>.
- [4] 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, M.-L. Yuan, Y.-L. Zhang, F.-H. Dai, Y. Liu, Q.-M. Wang, J.-J. Zheng, L. Xu, E.C. Holmes, Y.-Z. Zhang, A new coronavirus associated with human respiratory disease in China, *Nature* 579 (7798) (2020) 265–269, <https://doi.org/10.1038/s41586-020-2008-3>.
- [5] T.P. Velavan, C.G. Meyer, The COVID-19 epidemic, *Trop. Med. Int. Health* 25 (3) (2020) 278–280, <https://doi.org/10.1111/tmi.13383>.
- [6] M.M. Arons, K.M. Hatfield, S.C. Reddy, A. Kimball, A. James, J.R. Jacobs, J. Taylor, K. Spicer, A.C. Bardossy, L.P. Oakley, S. Tanwar, J.W. Dyal, J. Harney, Z. Chisty, J.M. Bell, M. Methner, P. Paul, C.M. Carlson, H.P. McLaughlin, N. Thornburg, S. Tong, A. Tamin, Y. Tao, A. Uehara, J. Harcourt, S. Clark, C. Brostrom-Smith, L.C. Page, M. Kay, J. Lewis, P. Montgomery, N.D. Stone, T.A. Clark, M.A. Honein, J.S. Duchin, J.A. Jernigan, Presymptomatic SARS-CoV-2 infections and transmission in a skilled nursing facility, *N. Engl. J. Med.* 382 (22) (2020) 2081–2090, <https://doi.org/10.1056/NEJMoa2008457>.
- [7] C. Huang, Y. Wang, X. Li, L. Ren, J. Zhao, Y. Hu, L. Zhang, G. Fan, J. Xu, X. Gu, Z. Cheng, T. Yu, J. Xia, Y. Wei, W. Wu, X. Xie, W. Yin, H. Li, M. Liu, Y. Xiao, H. Gao, L. Guo, J. Xie, G. Wang, R. Jiang, Z. Gao, Q. Jin, J. Wang, B. Cao, Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China, *Lancet* 395 (2020) 497–506, [https://doi.org/10.1016/S0140-6736\(20\)30183-5](https://doi.org/10.1016/S0140-6736(20)30183-5), 10223.
- [8] G. Spinato, C. Fabbri, J. Polesel, D. Cazzador, D. Borsetto, C. Hopkins, P. Boscolo-Rizzo, Alterations in smell or taste in mildly symptomatic outpatients with SARS-CoV-2 infection, *JAMA* 323 (20) (2020) 2089–2090, <https://doi.org/10.1001/jama.2020.6771>.
- [9] W. Yang, S. Kandula, M. Huynh, S.K. Greene, G. Van Wye, W. Li, H.T. Chan, E. McGibbon, A. Yeung, D. Olson, A. Fine, J. Shaman, Estimating the infection-fatality risk of SARS-CoV-2 in New York city during the spring 2020 pandemic Wave: a model-based analysis, *Lancet Infect. Dis.* 21 (2) (2021) 203–212, [https://doi.org/10.1016/S1473-3099\(20\)30769-6](https://doi.org/10.1016/S1473-3099(20)30769-6).
- [10] P. Ssentongo, A.E. Ssentongo, E.S. Heilbrunn, D.M. Ba, V.M. Chinchilli, Association of cardiovascular disease and 10 other pre-existing comorbidities with COVID-19 mortality: a systematic review and meta-analysis, *PLoS One* 15 (8) (2020), e0238215, <https://doi.org/10.1371/journal.pone.0238215>.
- [11] Y. Gao, L. Yan, Y. Huang, F. Liu, Y. Zhao, L. Cao, T. Wang, Q. Sun, Z. Ming, L. Zhang, J. Ge, L. Zheng, Y. Zhang, H. Wang, Y. Zhu, C. Zhu, T. Hu, T. Hua, B. Zhang, X. Yang, J. Li, H. Yang, Z. Liu, W. Xu, L.W. Guddat, Q. Wang, Z. Lou, Z. Rao, Structure of the RNA-dependent RNA polymerase from COVID-19 virus, *Science* 368 (6492) (2020) 779, <https://doi.org/10.1126/science.abb7498>.
- [12] Z. Jin, X. Du, Y. Xu, Y. Deng, M. Liu, Y. Zhao, B. Zhang, X. Li, L. Zhang, C. Peng, Y. Duan, J. Yu, L. Wang, K. Yang, F. Liu, R. Jiang, X. Yang, T. You, X. Liu, X. Yang, F. Bai, H. Liu, X. Liu, L.W. Guddat, W. Xu, G. Xiao, C. Qin, Z. Shi, H. Jiang, Z. Rao, H. Yang, Structure of mpro from SARS-CoV-2 and discovery of its inhibitors, *Nature* 582 (7811) (2020) 289–293, <https://doi.org/10.1038/s41586-020-2223-y>.
- [13] J.H. Beigel, K.M. Tomashek, L.E. Dodd, A.K. Mehta, B.S. Zingman, A.C. Kail, E. Hohmann, H.Y. Chu, A. Luetkemeyer, S. Kline, D. Lopez de Castilla, R.W. Finberg, K. Dierberg, V. Tapson, L. Hsieh, T.F. Patterson, R. Paredes, D.A. Sweeney, W.R. Short, G. Touloumi, D.C. Lye, N. Ohmagari, M.-D. Oh, G.M. Ruiz-Palacios, T. Benfield, G. Fätkenheuer, M.G. Kortepeter, R.L. Atmar, C.B. Creech, J. Lundgren, A.G. Babiker, S. Pett, J.D. Neaton, T.H. Burgess, T. Bonnett, M. Green, M. Makowski, A. Osinusi, S. Nayak, H.C. Lane, ACTT-1 study group members. Remdesivir for the treatment of covid-19 - final report, *N. Engl. J. Med.* 383 (19) (2020) 1813–1826, <https://doi.org/10.1056/NEJMoa2007764>.
- [14] S. Ullrich, C. Nitsche, The SARS-CoV-2 main protease as drug target, *Bioorg. Med. Chem. Lett* 30 (17) (2020) 127377, <https://doi.org/10.1016/j.bmcl.2020.127377>.
- [15] L. Zhang, D. Lin, X. Sun, U. Curth, C. Drosten, L. Sauerhering, S. Becker, K. Rox, R. Hilgenfeld, Crystal structure of SARS-CoV-2 main protease provides a basis for design of improved  $\alpha$ -ketoamide inhibitors, *Science* 368 (6489) (2020) 409, <https://doi.org/10.1126/science.abb3405>.
- [16] S. Seth, J. Batra, S. Srinivasan, COVID-19: targeting proteases in viral invasion and host immune response, *Front Mol Biosci* 7 (2020), <https://doi.org/10.3389/fmolb.2020.00215>, 215–215.
- [17] S.V. Stoddard, S.D. Stoddard, B.K. Oelkers, K. Fitts, K. Whalum, K. Whalum, A.D. Hemphill, J. Manikonda, L.M. Martinez, E.G. Riley, C.M. Roof, N. Sarwar, D.M. Thomas, E. Ulmer, F.E. Wallace, P. Pandey, S. Roy, Optimization rules for SARS-CoV-2 mpro antivirals: ensemble docking and exploration of the coronavirus protease active site, *Viruses* 12 (9) (2020), <https://doi.org/10.3390/v12090942>.
- [18] Sk A. Amin, S. Banerjee, S. Singh, I.A. Qureshi, S. Gayen, T. Jha, First Structure-Activity Relationship Analysis of SARS-CoV-2 Virus Main Protease (Mpro) Inhibitors: an Endeavor on COVID-19 Drug Discovery, *Molecular Diversity*, 2021, <https://doi.org/10.1007/s11030-020-10166-3>.

- [19] Z. Wang, Y. Zhao, Q. Wang, Y. Xing, L. Feng, J. Kong, C. Peng, L. Zhang, H. Yang, M. Lu, Identification of proteasome and caspase inhibitors targeting SARS-CoV-2 mpro, *Signal Transduction and Targeted Therapy* 6 (1) (2021) 214, <https://doi.org/10.1038/s41392-021-00639-8>.
- [20] N.M. Mohamed, E.M.H. Ali, A.M. AboulMagd, Ligand-based design, molecular dynamics and ADMET studies of suggested SARS-CoV-2 mpro inhibitors, *RSC Adv.* 11 (8) (2021) 4523–4538, <https://doi.org/10.1039/D0RA10141A>.
- [21] J. Breidenbach, C. Lemke, T. Pillaiyar, L. Schäkel, G. Al Hamwi, M. Dielt, R. Gedschold, N. Geiger, V. Lopez, S. Mirza, V. Namasivayam, A.C. Schiedel, K. Sylvester, D. Thimm, C. Vielmuth, L. Phuong Vu, M. Zylina, J. Bodem, M. Gütschow, C.E. Müller, Targeting the main protease of SARS-CoV-2: from the establishment of high throughput screening to the design of tailored inhibitors, *Angew. Chem. Int. Ed.* 60 (18) (2021) 10423–10429, <https://doi.org/10.1002/anie.202016961>.
- [22] A. Gimeno, J. Mestres-Truyol, M.J. Ojeda-Montes, G. Macip, B. Saldivar-Espinoza, A. Cereto-Massagué, G. Pujadas, S. Garcia-Vallvé, Prediction of novel inhibitors of the main protease (M-pro) of SARS-CoV-2 through consensus docking and drug reposition, *Int. J. Mol. Sci.* 21 (11) (2020), <https://doi.org/10.3390/ijms21113793>.
- [23] J. Gossen, S. Albani, A. Hanke, B.P. Joseph, C. Bergh, M. Kuzikov, E. Costanzi, C. Manelfi, P. Storic, P. Gribbon, A.R. Beccari, C. Talarico, F. Spyriakis, E. Lindahl, A. Zaliani, P. Carloni, R.C. Wade, F. Musiani, D.B. Kokh, G. Rossetti, A blueprint for high affinity SARS-CoV-2 mpro inhibitors from activity-based compound library screening guided by analysis of protein dynamics, *ACS Pharmacol. Transl. Sci.* 4 (3) (2021) 1079–1095, <https://doi.org/10.1021/acspsci.0c00215>.
- [24] R. Yoshino, N. Yasuo, M. Sekijima, Identification of key interactions between SARS-CoV-2 main protease and inhibitor drug candidates, *Sci. Rep.* 10 (1) (2020) 12493, <https://doi.org/10.1038/s41598-020-69337-9>.
- [25] A.-T. Ton, F. Gentile, M. Hsing, F. Ban, A. Cherkasov, Rapid identification of potential inhibitors of SARS-CoV-2 main protease by deep docking of 1.3 billion compounds, *Molecular Informatics* 39 (8) (2020) 2000028, <https://doi.org/10.1002/minf.202000028>.
- [26] A.M. Andrianov, Y.V. Kornoushenko, A.D. Karpenko, I.P. Bosko, A.V. Tuzikov, Computational discovery of small drug-like compounds as potential inhibitors of SARS-CoV-2 main protease, *J. Biomol. Struct. Dyn.* 39 (15) (2021) 5779–5791, <https://doi.org/10.1080/07391102.2020.1792989>.
- [27] 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, G.M. Morris, Discovery of SARS-CoV-2 mpro peptide inhibitors from modelling substrate and ligand binding, *Chem. Sci.* 12 (41) (2021) 13686–13703, <https://doi.org/10.1039/D1SC03628A>.
- [28] Günther Sebastian, Y.A. Reinke Patrick, Yaiza Fernández-García, Lieske Julia, J. Lane Thomas, M. Ginn Helen, H.M. Koua Faisal, Ehrhart Christiane, Ewert Wiebke, Dominik Oberthuer, Yefanov Oleksandr, Susanne Meier, Lorenzen Kristina, Krichel Boris, Kopicke Janine-Denise, Gelisio Luca, Wolfgang Brehm, Ilona Dunkel, Seychell Brandon, Gieseler Henry, Norton-Baker Brenna, Escudero-Pérez Beatriz, Domaracki Martin, Saouane Sofiane, Tolstikova Alexandra, A. White Thomas, Hänle Anna, Groessler Michael, Fleckenstein Holger, Fabian Trost, Galchenkova Marina, Gevorkov Yaroslav, Chufeng Li, Awel Salah, Ariana Peck, Barthelmeß Miriam, Schlünzen Frank, P. Lourdu Xavier, Nadine Werner, Andaleeb Hina, Ullah Najeeb, Falke Sven, Vasundara Srinivasan, França Bruno Alves, Schwinzer Martin, Brognaro Hévilá, Cromarte Rogers, Diogo Melo, J. Zaitseva-Doyle Joanna, Knoska Juraj, E. Peña-Murillo Gisel, Mashhour Aida Rahmani, Hennicke Vincent, Pontus Fischer, Hakanpää Johanna, Jan Meyer, Gribbon Philip, Ellinger Bernhard, Kuzikov Maria, Wolf Markus, R. Beccari Andrea, Bourenkov Gleb, David von Stetten, Pompoidr Guillaume, Bento Isabel, Panneerselvam Saravanan, Karpics Ivars, R. Schneider Thomas, Marta Garcia-Alai Maria, Niebling Stephan, Günther Christian, Christina Schmidt, Robin Schubert, Huijong Han, Juliane Boger, C.F. Monteiro Diana, Linlin Zhang, Xinyuan Yuan Sun, Jonathan Pletzer-Zelger, Wollenhaupt Jan, G. Feiler Christian, S. Weiss Manfred, Schulz Eike-Christian, Mehrabi Pedram, Karmicar Katarina, Usenik Aleksandra, Loboda Jure, Tidow Henning, Chari Ashwin, Hilgenfeld Rolf, Utrecht Charlotte, Cox Russell, Zaliani Andrea, Tobias Beck, Rarey Matthias, Günther Stephan, Dusan Turk, Hinrichs Winfried, N. Chapman Henry, R. Pearson Arwen, Betzel Christian, Alke Meents, X-ray screening identifies active site and allosteric inhibitors of SARS-CoV-2 main protease, *Science* 372 (6542) (2021) 642–646, <https://doi.org/10.1126/science.abb7945>.
- [29] J. Osipiuk, S.-A. Azizi, S. Dvorkin, M. Endres, R. Jedrzejczak, K.A. Jones, S. Kang, R.S. Kathayat, Y. Kim, V.G. Lisnyak, S.L. Maki, V. Nicolaescu, C.A. Taylor, C. Tesar, Y.-A. Zhang, Z. Zhou, G. Randall, K. Michalska, S.A. Snyder, B.C. Dickinson, A. Joachimiak, Structure of papain-like protease from SARS-CoV-2 and its complexes with non-covalent inhibitors, *Nat. Commun.* 12 (1) (2021) 743, <https://doi.org/10.1038/s41467-021-21060-3>.
- [30] 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, W.L. Jorgensen, Potent noncovalent inhibitors of the main protease of SARS-CoV-2 from molecular sculpting of the drug perampanel guided by free energy perturbation calculations, *ACS Cent. Sci.* 7 (3) (2021) 467–475, <https://doi.org/10.1021/acscentsci.1c00039>.
- [31] A. Shitrit, D. Zaidman, O. Kalid, I. Bloch, D. Doron, T. Yarnitzky, I. Buch, I. Segev, E. Ben-Zeev, E. Segev, O. Kobiler, Conserved interactions required for inhibition of the main protease of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), *Sci. Rep.* 10 (1) (2020) 20808, <https://doi.org/10.1038/s41598-020-77794-5>.
- [32] H.M. Mengist, T. Dilnessa, T. Jin, Structural basis of potential inhibitors targeting SARS-CoV-2 main protease, *Front. Chem.* 9 (2021) 7, <https://doi.org/10.3389/fchem.2021.622898>.
- [33] S. Ikhtani, F. Forouhar, H. Liu, S.J. Hong, F.-Y. Lin, M.S. Nair, A. Zask, Y. Huang, L. Xing, B.R. Stockwell, A. Chavez, D.D. Ho, Lead compounds for the development of SARS-CoV-2 3CL protease inhibitors, *Nat. Commun.* 12 (1) (2021) 2016, <https://doi.org/10.1038/s41467-021-22362-2>.
- [34] M. Jukić, D. Janežič, U. Bren, Ensemble docking coupled to linear interaction energy calculations for identification of coronavirus main protease (3CLpro) non-covalent small-molecule inhibitors, *Molecules* 25 (24) (2020), <https://doi.org/10.3390/molecules25245808>.
- [35] M. Jukić, B. Skrlj, G. Tomšič, S. Plesko, Č. Podlipnik, U. Bren, Prioritisation of compounds for 3CLpro inhibitor development on SARS-CoV-2 variants, *Molecules* 26 (10) (2021), <https://doi.org/10.3390/molecules26103003>.
- [36] H. Yang, W. Xie, X. Xue, K. Yang, J. Ma, W. Liang, Q. Zhao, Z. Zhou, D. Pei, J. Ziebuhr, R. Hilgenfeld, K.Y. Yuen, L. Wong, G. Gao, S. Chen, Z. Chen, D. Ma, M. Bartlam, Z. Rao, Design of wide-spectrum inhibitors targeting coronavirus main proteases, *PLoS Biol.* 3 (10) (2005) e324, <https://doi.org/10.1371/journal.pbio.0030324>.
- [37] Hung Hui-Chen; Ke Yi-Yu; Huang Sheng Yu; Huang Peng-Nien; Kung Yu-An; Chang Teng-Yuan; Yen Kuei-Jung; Peng Tzu-Ting; Chang Shao-En; Huang Chin-Ting; Tsai Ya-Ru; Wu Szu-Huei; Lee Shioh-Ju; Lin Jiunn-Horng; Liu Bing-Sin; Sung Wang-Chou; Shih Shin-Ru; Chen Chiung-Tong; Hsu John Tsu-An. Discovery of M Protease Inhibitors Encoded by SARS-CoV-2. *Antimicrob. Agents Chemother.* 64 (9), e00872-20. <https://doi.org/10.1128/AAC.00872-20>.
- [38] K. Arafat, N. Serrano-Aparicio, A. Lodola, A.J. Mulholland, F.V. González, K. Świderek, V. Moliner, Mechanism of inhibition of SARS-CoV-2 mpro by N3 peptidyl Michael acceptor explained by QM/MM simulations and design of new derivatives with tunable chemical reactivity, *Chem. Sci.* 12 (4) (2021) 1433–1444, <https://doi.org/10.1039/D0SC06195F>.
- [39] L. Fu, F. Ye, Y. Feng, F. Yu, Q. Wang, Y. Wu, C. Zhao, H. Sun, B. Huang, P. Niu, H. Song, Y. Shi, X. Li, W. Tan, J. Qi, G.F. Gao, Both boceprevir and GC376 efficaciously inhibit SARS-CoV-2 by targeting its main protease, *Nat. Commun.* 11 (1) (2020) 4417, <https://doi.org/10.1038/s41467-020-18233-x>.
- [40] C. Ma, M.D. Sacco, B. Hurst, J.A. Townsend, Y. Hu, T. Szeto, X. Zhang, B. Tarbet, M.T. Marty, Y. Chen, J. Boceprevir Wang, GC-376, and calpain inhibitors II, XII inhibit SARS-CoV-2 viral replication by targeting the viral main protease, *Cell Res.* 30 (8) (2020) 678–692, <https://doi.org/10.1038/s41422-020-0356-z>.
- [41] Sheshwar Pola ED1 - Ravi Varala, Significance of thiazole-based heterocycles for bioactive systems, in: *Scope of Selective Heterocycles from Organic and Pharmaceutical Perspective*, IntechOpen, Rijeka, 2016, p. 1, <https://doi.org/10.5772/62077>.
- [42] A. Ayati, S. Emami, A. Asadipour, A. Shafiee, A. Foroumadi, Recent applications of 1,3-thiazole core structure in the identification of new lead compounds and drug discovery, *Eur. J. Med. Chem.* 97 (2015) 699–718, <https://doi.org/10.1016/j.ejmech.2015.04.015>.
- [43] I. Seck, F. Triazole Nguemo, Imidazole, and thiazole-based compounds as potential agents against coronavirus, *Results in Chemistry* 3 (2021) 100132, <https://doi.org/10.1016/j.rechem.2021.100132>.
- [44] M.R. Freidel, R.S. Armen, Mapping major SARS-CoV-2 drug targets and assessment of druggability using computational fragment screening: identification of an allosteric small-molecule binding site on the Nsp13 helicase, *PLoS One* 16 (2) (2021), e0246181, <https://doi.org/10.1371/journal.pone.0246181>.
- [45] S.-G. Zhang, C.-G. Liang, W.-H. Zhang, Recent advances in indazole-containing derivatives: synthesis and biological perspectives, *Molecules* 23 (11) (2018) 2783–2823, <https://doi.org/10.3390/molecules23112783>.
- [46] K. Gao, D.D. Nguyen, J. Chen, R. Wang, G.-W. Wei, Repositioning of 8565 existing drugs for COVID-19, *J. Phys. Chem. Lett.* 11 (13) (2020) 5373–5382, <https://doi.org/10.1021/acs.jpcclett.0c01579>.
- [47] A. N'ait Ousidi, M.Y.A. Itto, A. Auhmani, M. Loubidi, A. Riahi, J.-C. Daran, M. Esseffar, C.A. Parish, Diastereoselective synthesis of new thiazolyl-indazole derivatives from R-carvone: a combined experimental and theoretical study, *Tetrahedron* 78 (2021) 131830, <https://doi.org/10.1016/j.tet.2020.131830>.
- [48] Y. Hu, D. Stumpfe, J. Bajorath, Computational exploration of molecular scaffolds in medicinal Chemistry, *J. Med. Chem.* 59 (9) (2016) 4062–4076, <https://doi.org/10.1021/acs.jmedchem.5b01746>.
- [49] J.B. Baell, G.A. Holloway, New substructure filters for removal of Pan assay interference compounds (PAINS) from screening libraries and for their exclusion in bioassays, *J. Med. Chem.* 53 (7) (2010) 2719–2740, <https://doi.org/10.1021/jm901137j>.
- [50] Schrödinger Release 2019-3, QikProp, Schrödinger, LLC, New York, NY, 2019.
- [51] H.M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T.N. Bhat, H. Weissig, I.N. Shindyalov, P.E. Bourne, The protein Data Bank, *Nucleic Acids Res.* 28 (1) (2000) 235–242, <https://doi.org/10.1093/nar/28.1.235>.
- [52] T. Pillaiyar, M. Manickam, V. Namasivayam, Y. Hayashi, S.-H. Jung, An overview of severe acute respiratory syndrome–coronavirus (SARS-CoV) 3CL protease inhibitors: peptidomimetics and small molecule chemotherapy, *J. Med. Chem.* 59 (14) (2016) 6595–6628, <https://doi.org/10.1021/acs.jmedchem.5b01461>.
- [53] Schrödinger Release 2019-3: Protein Preparation Wizard, Epik, Schrödinger, LLC, New York, NY, 2016. Impact, Schrödinger, LLC, New York, NY, 2016; Prime, Schrödinger, LLC, New York, NY, 2019.
- [54] C.R. Søndergaard, M.H.M. Olsson, M. Rostkowski, J.H. Jensen, Improved treatment of ligands and coupling effects in empirical calculation and rationalization of PKa values, *J. Chem. Theor. Comput.* 7 (7) (2011) 2284–2295, <https://doi.org/10.1021/ct200133y>.

- [55] M.H.M. Olsson, C.R. Søndergaard, M. Rostkowski, J.H. Jensen, PROPKA3: consistent treatment of internal and surface residues in empirical PKa predictions, *J. Chem. Theor. Comput.* 7 (2) (2011) 525–537, <https://doi.org/10.1021/ct100578z>.
- [56] E. Harder, W. Damm, J. Maple, C. Wu, M. Rebol, J.Y. Xiang, L. Wang, D. Lupyan, M.K. Dahlgren, J.L. Knight, J.W. Kaus, D.S. Cerutti, G. Krilov, W.L. Jorgensen, R. Abel, R.A. Friesner, OPLS3: a force field providing broad coverage of drug-like small molecules and proteins, *J. Chem. Theor. Comput.* 12 (1) (2016) 281–296, <https://doi.org/10.1021/acs.jctc.5b00864>.
- [57] J. Wang, R.M. Wolf, J.W. Caldwell, P.A. Kollman, D.A. Case, Development and testing of a general amber force field, *J. Comput. Chem.* 25 (9) (2004) 1157–1174, <https://doi.org/10.1002/jcc.20035>.
- [58] M.D. Hanwell, D.E. Curtis, D.C. Lonie, T. Vandermeersch, E. Zurek, G.R. Hutchison, Avogadro: an advanced semantic chemical editor, visualization, and analysis platform, *J. Cheminf.* 4 (1) (2012) 17, <https://doi.org/10.1186/1758-2946-4-17>.
- [59] Gaussian 16, Revision B.01 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, D.J. Fox, Gaussian, Inc., Wallingford CT, 2016.
- [60] R.A. Friesner, J.L. Banks, R.B. Murphy, T.A. Halgren, J.J. Klicic, D.T. Mainz, M.P. Repasky, E.H. Knoll, M. Shelley, J.K. Perry, D.E. Shaw, P. Francis, P.S. Shenkin, Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy, *J. Med. Chem.* 47 (7) (2004) 1739–1749, <https://doi.org/10.1021/jm0306430>.
- [61] T.A. Halgren, R.B. Murphy, R.A. Friesner, H.S. Beard, L.L. Frye, W.T. Pollard, J.L. Banks, Glide: a new approach for rapid, accurate docking and scoring. 2. Enrichment factors in database screening, *J. Med. Chem.* 47 (7) (2004) 1750–1759, <https://doi.org/10.1021/jm030644s>.
- [62] W. Rut, K. Groborz, L. Zhang, X. Sun, M. Zmudzinski, B. Pawlik, X. Wang, D. Jochmans, J. Neyts, W. Mlynarski, R. Hilgenfeld, M. Drag, SARS-CoV-2 mpro inhibitors and activity-based probes for patient-sample imaging, *Nat. Chem. Biol.* 17 (2) (2021) 222–228, <https://doi.org/10.1038/s41589-020-00689-z>.
- [63] Schrödinger Release 2019-3, Maestro, Schrödinger, LLC, New York, NY, 2021.
- [64] K. Anand, J. Ziebuhr, P. Wadhvani, J.R. Mesters, R. Hilgenfeld, Coronavirus main proteinase (3CL<sup>pro</sup>) structure: basis for design of anti-SARS drugs, *Science* 300 (5626) (2003) 1763, <https://doi.org/10.1126/science.1085658>.
- [65] E.F. Pettersen, T.D. Goddard, C.C. Huang, G.S. Couch, D.M. Greenblatt, E.C. Meng, T.E. Ferrin, UCSF chimera—a visualization system for exploratory research and analysis, *J. Comput. Chem.* 25 (13) (2004) 1605–1612, <https://doi.org/10.1002/jcc.20084>.
- [66] D.A. Case, I.Y. B-S, S.R. Brozell, D.S. Cerutti, T.E. Cheatham III, V.W.D. Cruzeiro, T.A. Darden, R.E. Duke, D. Gohreishi, M.K. Gilson, H. Gohlke, A.W. Goetz, D. Greene, R. Harris, N. Homeyer, S. Izadi, A. Kovalenko, T. Kurtzman, T.S. Lee, S. LeGrand, P. Li, C. Lin, J. Liu, T. Luchko, R. Luo, D.J. Mermelstein, K.M. Merz, Y. Miao, G. Monard, C. Nguyen, H. Nguyen, A. Onufriev, F. Pan, R. Qi, D.R. Roe, A. Roitberg, C. Sagui, S. Schott-Verdugo, J. Shen, C.L. Simmerling, J. Smith, R. Salomon-Ferrer, J. Swails, R.C. Walker, J. Wang, H. Wei, R.M. Wolf, X. Wu, L. Xiao, D.M. York, P.A. Kollman, AMBER 2018, University of California, San Francisco, 2018.
- [67] J.A. Maier, C. Martinez, K. Kasavajhala, L. Wickstrom, K.E. Hauser, C. Simmerling, ff14SB: improving the accuracy of protein side chain and backbone parameters from ff99SB, *J. Chem. Theor. Comput.* 11 (8) (2015) 3696–3713, <https://doi.org/10.1021/acs.jctc.5b00255>.
- [68] J. Wang, W. Wang, P.A. Kollman, D.A. Case, Automatic atom type and bond type perception in molecular mechanical calculations, *J. Mol. Graph. Model.* 25 (2) (2006) 247–260, <https://doi.org/10.1016/j.jmgm.2005.12.005>.
- [69] A. Jakalian, B.L. Bush, D.B. Jack, C.I. Bayly, Fast, efficient generation of high-quality atomic charges. AM1-BCC model: I. Method, *J. Comput. Chem.* 21 (2) (2000) 132–146, [https://doi.org/10.1002/\(SICI\)1096-987X\(20000130\)21:2<132:AID-JCC5>3.0.CO;2-P](https://doi.org/10.1002/(SICI)1096-987X(20000130)21:2<132:AID-JCC5>3.0.CO;2-P).
- [70] A. Jakalian, D.B. Jack, C.I. Bayly, Fast, efficient generation of high-quality atomic charges. AM1-BCC model: II. Parameterization and validation, *J. Comput. Chem.* 23 (16) (2002) 1623–1641, <https://doi.org/10.1002/jcc.10128>.
- [71] W.L. Jorgensen, J.D. Madura, Quantum and statistical mechanical studies of liquids. 25. Solvation and conformation of methanol in water, *J. Am. Chem. Soc.* 105 (6) (1983) 1407–1413, <https://doi.org/10.1021/ja00344a001>.
- [72] A.W. Götz, M.J. Williamson, D. Xu, D. Poole, S. Le Grand, R.C. Walker, Routine microsecond molecular dynamics simulations with AMBER on GPUs. 1. Generalized born, *J. Chem. Theor. Comput.* 8 (5) (2012) 1542–1555, <https://doi.org/10.1021/ct200909j>.
- [73] R. Salomon-Ferrer, A.W. Götz, D. Poole, S. Le Grand, R.C. Walker, Routine microsecond molecular dynamics simulations with AMBER on GPUs. 2. Explicit solvent particle mesh ewald, *J. Chem. Theor. Comput.* 9 (9) (2013) 3878–3888, <https://doi.org/10.1021/ct400314y>.
- [74] T.D. Goddard, C.C. Huang, E.C. Meng, E.F. Pettersen, G.S. Couch, J.H. Morris, T.E. Ferrin, UCSF ChimeraX: meeting modern challenges in visualization and analysis, *Protein Sci.* 27 (1) (2018) 14–25, <https://doi.org/10.1002/pro.3235>.
- [75] B.R. Miller, T.D. McGee, J.M. Swails, N. Homeyer, H. Gohlke, A.E. Roitberg, MMPBSA.py: an efficient program for end-state free energy calculations, *J. Chem. Theor. Comput.* 8 (9) (2012) 3314–3321, <https://doi.org/10.1021/ct300418h>.
- [76] T. Hou, J. Wang, Y. Li, W. Wang, Assessing the performance of the MM/PBSA and MM/GBSA methods. 1. The accuracy of binding free energy calculations based on molecular dynamics simulations, *J. Chem. Inf. Model.* 51 (1) (2011) 69–82, <https://doi.org/10.1021/ci100275a>.
- [77] R. Meurisse, R. Brasseur, A. Thomas, Aromatic side-chain interactions in proteins. Near- and far-sequence his-X pairs, *Biochim. Biophys. Acta Protein Proteomics* 1649 (1) (2003) 85–96, [https://doi.org/10.1016/S1570-9639\(03\)00161-4](https://doi.org/10.1016/S1570-9639(03)00161-4).
- [78] V. Philip, J. Harris, R. Adams, D. Nguyen, J. Spiers, J. Baudry, E.E. Howell, R.J. Hinde, A survey of Aspartate–Phenylalanine and Glutamate–Phenylalanine interactions in the protein Data Bank: searching for Anion– $\pi$  pairs, *Biochemistry* 50 (14) (2011) 2939–2950, <https://doi.org/10.1021/bi200066k>.
- [79] S. Tsuzuki, K. Honda, T. Uchimaru, M. Mikami, K. Tanabe, Origin of the attraction and directionality of the NH/ $\pi$  interaction: comparison with OH/ $\pi$  and CH/ $\pi$  interactions, *J. Am. Chem. Soc.* 122 (46) (2000) 11450–11458, <https://doi.org/10.1021/ja001901a>.
- [80] C.R. Forbes, S.K. Sinha, H.K. Ganguly, S. Bai, G.P.A. Yap, S. Patel, N.J. Zondlo, Insights into thiol–aromatic interactions: a stereoelectronic basis for S–H/ $\pi$  interactions, *J. Am. Chem. Soc.* 139 (5) (2017) 1842–1855, <https://doi.org/10.1021/jacs.6b08415>.
- [81] A.L. Ringer, A. Senenko, C.D. Sherrill, Models of S/ $\pi$  interactions in protein structures: comparison of the H25–benzene complex with PDB data, *Protein Sci.* 16 (10) (2007) 2216–2223, <https://doi.org/10.1110/ps.073002307>.
- [82] K.S.C. Reid, P.F. Lindley, J.M. Thornton, Sulphur-aromatic interactions in proteins, *FEBS (Fed. Eur. Biochem. Soc.) Lett.* 190 (2) (1985) 209–213, [https://doi.org/10.1016/0014-5793\(85\)81285-0](https://doi.org/10.1016/0014-5793(85)81285-0).
- [83] J.F. Ellenbarger, I.V. Krieger, H. Huang, S. Gómez-Coca, T.R. Ioerger, J.C. Sacchettini, S.E. Wheeler, K.R. Dunbar, Anion– $\pi$  interactions in computer-aided drug design: modeling the inhibition of malate synthase by phenyl-diketo acids, *J. Chem. Inf. Model.* 58 (10) (2018) 2085–2091, <https://doi.org/10.1021/acs.jcim.8b00417>.
- [84] L. Mony, L. Krzaczkowski, M. Leonetti, A. Le Goff, K. Alarcon, J. Neyton, H.-O. Bertrand, F. Acher, P. Paoletti, Structural basis of NR2B-selective antagonist recognition by N-Methyl-D-aspartate receptors, *Mol. Pharmacol.* 75 (1) (2009) 60, <https://doi.org/10.1124/mol.108.050971>.
- [85] A. Garrigues, N. Loiseau, M. Delaforge, J. Ferté, M. Garrigues, F. André, S. Orłowski, Characterization of two pharmacophores on the multidrug transporter P-glycoprotein, *Mol. Pharmacol.* 62 (6) (2002) 1288, <https://doi.org/10.1124/mol.62.6.1288>.
- [86] A. Hegyi, J. Ziebuhr, Conservation of substrate specificities among coronavirus main proteases, *J. Gen. Virol.* 83 (2002) 595–599, <https://doi.org/10.1099/0022-1317-83-3-595>.
- [87] J. Ziebuhr, E.J. Snijder, A.E. Gorbalenya, Virus-encoded proteinases and proteolytic processing in the nidovirales, *J. Gen. Virol.* 81 (2000) 853–879, <https://doi.org/10.1099/0022-1317-81-4-853>.