

# Dual-Reporter System for Real-Time Monitoring of SARS-CoV-2 Main Protease Activity in Live Cells Enables Identification of an Allosteric Inhibition Path

Yaron Bram,<sup>¶</sup> Xiaohua Duan,<sup>¶</sup> Benjamin E. Nilsson-Payant,<sup>¶</sup> Vasuretha Chandar, Hao Wu,<sup>¶</sup> Derek Shore,<sup>¶</sup> Alvaro Fajardo, Saloni Sinha, Nora Hassan, Harel Weinstein,\* Benjamin R. TenOever,\* Shuibing Chen,\* and Robert E. Schwartz\*

Cite This: *ACS Bio Med Chem Au* 2022, 2, 627–641

Read Online

ACCESS |

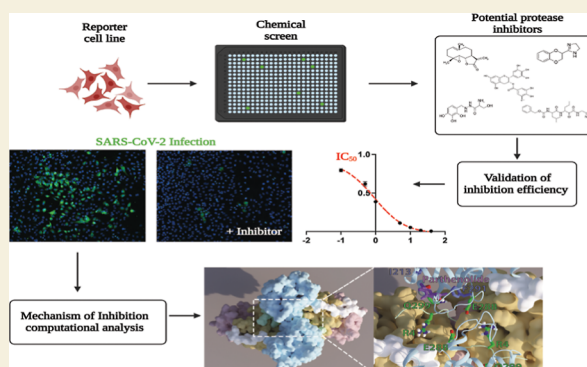
Metrics & More

Article Recommendations

Supporting Information

**ABSTRACT:** The SARS-CoV-2 pandemic is an ongoing threat to global health, and the continuing emergence of contagious variants highlights the urgent need for additional antiviral therapy to attenuate COVID-19 disease. The SARS-CoV-2 main protease (3CL<sup>pro</sup>) presents an attractive target for such therapy due to its high sequence conservation and key role in the viral life cycle. In this study, we designed a fluorescent–luminescent cell-based reporter for the detection and quantification of 3CL<sup>pro</sup> intracellular activity. Employing this platform, we examined the efficiency of known protease inhibitors against 3CL<sup>pro</sup> and further identified potent inhibitors through high-throughput chemical screening. Computational analysis confirmed a direct interaction of the lead compounds with the protease catalytic site and identified a prototype for efficient allosteric inhibition. These developments address a pressing need for a convenient sensor and specific targets for both virus detection and rapid discovery of potential inhibitors.

**KEYWORDS:** SARS-CoV-2, main protease, antiviral, genetic reporter, chemical screen, computational modeling, allosteric inhibition



Coronaviruses (CoVs) are single-stranded positive-sense RNA viruses with a genome size ranging between 27 and 32 kb.<sup>1</sup> The viral replication and maturation processes are highly dependent on a complex cascade of proteolytic cleavages of precursor polypeptides (pp1a and pp1ab) by the viral-encoded papain-like protease (PL<sup>pro</sup>) and the 3-chymotrypsin-like protease (3CL<sup>pro</sup>, main protease, M<sup>pro</sup>, NSP5), which result in the release of 16 non-structural proteins (1–16 NSPs)<sup>2</sup> and the formation of a functional viral RNA replication complex. PL<sup>pro</sup> is responsible for the proteolytic cleavage and the release of NSP1–3. The 3CL<sup>pro</sup> processes at least 11 sites, including its own N- and C-termini, recognizing a general sequence of Leu–Gln↓Ser/Ala/Gly (↓ cleavage site). The ~33 kDa severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) 3CL<sup>pro</sup> consists of three structural domains: Domains I and II (residues 8–101 and 102–184, respectively) fold into an antiparallel  $\beta$ -barrel conformation and contains the substrate binding and catalytic sites (Cys145/His41 catalytic dyad). Domain III (residues 201–303) consists of five  $\alpha$ -helices which are essential for self-assembly to form an enzymatic active homodimer.<sup>3–6</sup> Due to its importance in the viral life cycle,  $\beta$ -coronavirus 3CL<sup>pro</sup> enzymes share this fold with a sequence identity of greater than

80% within the same genus (SARS-CoV and SARS-CoV-2 have 96% sequence identity)<sup>5</sup> and a high degree of sequence conservation in the enzymatic site domain.<sup>7</sup>

Proteases have been the focus of intense research which resulted in the successful development of multiple inhibitors targeting diverse cellular and viral proteases. For example, protease inhibitors are currently being administered as a part of direct-acting antiviral combination therapy which was able to transition human immunodeficiency virus (HIV) infections from an acute fatal disease to a manageable chronic condition and offer above 90% cure rates for long-lasting hepatitis C viral (HCV) infections.<sup>8</sup> Moreover, treatment with the SARS-CoV-2 3CL<sup>pro</sup> inhibitor PF-07321332 (along with the CYP3A4 inhibitor ritonavir) is highly effective in reducing the risk of developing severe illness in high-risk patients.<sup>9</sup> Nonetheless, many potential compounds identified through in vitro assays

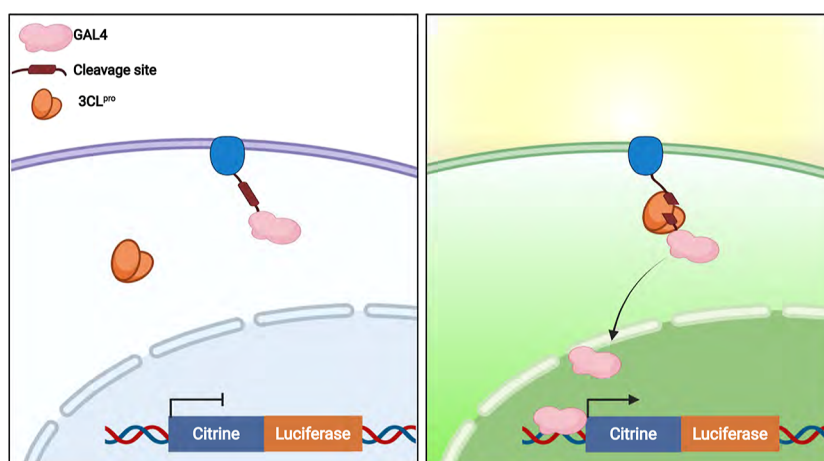
Received: May 29, 2022

Revised: September 21, 2022

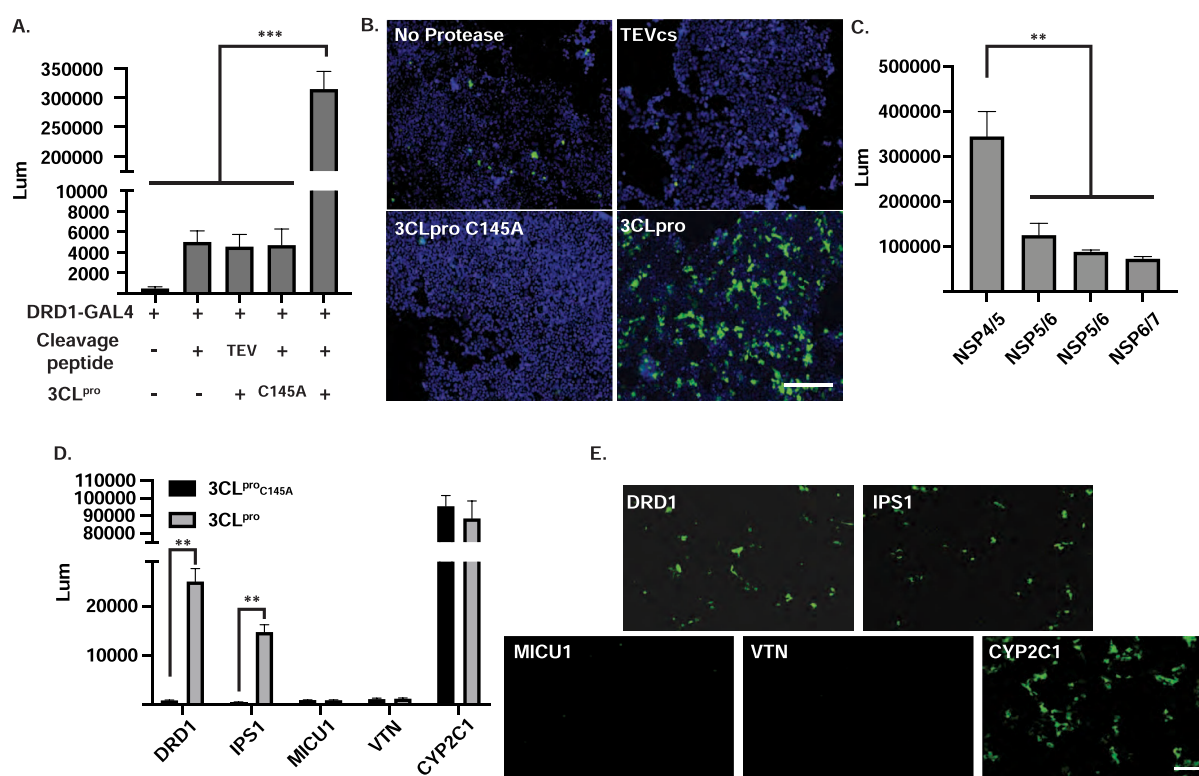
Accepted: September 21, 2022

Published: October 17, 2022





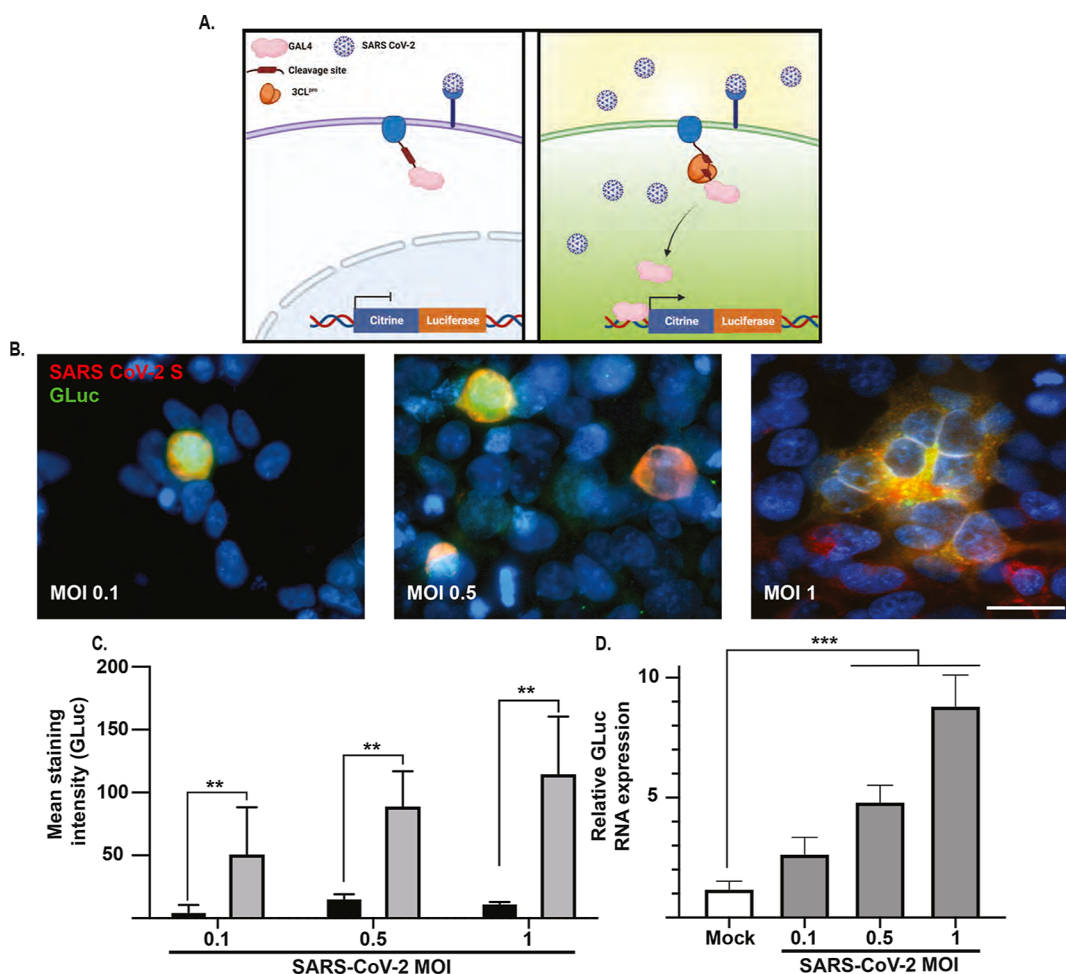
**Figure 1.** Dual live-cell reporter for the detection of SARS-CoV-2 3CL<sup>pro</sup>. Schematic of the live-cell reporter design; GAL4 is fused to membrane-localized receptor DRD1 with a 3CL<sup>pro</sup> cleavable linker and expressed in cells with a genomic integrated UAS::Citrine-T2A-GLuc reporter cassette.



**Figure 2.** Validation of the SARS-CoV-2 3CL<sup>pro</sup> live-cell reporter. (A) HEK293T cells were transfected with reporter components as stated in the panel, and luciferase activity was quantified 48 h after transfection. Significant increase was detected only in cells that expressed catalytically active 3CL<sup>pro</sup> together with the corresponding cleavage site. Data was plotted as mean  $\pm$  SD. Statistical analysis was performed using a two-tailed Student's *t*-test,  $N = 3$ , \*\*\* =  $p \leq 0.005$ . (B) Representative fluorescence images of citrine expression in HEK293T cells 48 h after transfection. Scale bar, 100  $\mu$ m. (C) HEK293T cells were transfected with GAL4-DRD1 fusion with different 3CL<sup>pro</sup> linkers corresponding to the viral NSP's junction sites as stated in the panel. Luciferase activity was quantified 48 h after transfection.  $N = 3$ , data was plotted as mean  $\pm$  SD. (D) 3CL<sup>pro</sup> cleavage efficiency was evaluated with different GAL4 fusion proteins as stated in the graph (NSP4/5 linker was used in all the constructs) and compared to the catalytically inactive 3CL<sup>pro</sup><sub>C145A</sub>.  $N = 3$ , data was plotted as mean  $\pm$  SD. (E) Representative fluorescence images of citrine expression in HEK293T cells 48 h after transfection 3CL<sup>pro</sup> and different GAL4 fusion proteins. Scale bar, 100  $\mu$ m.

are less effective when examined in situ or in vivo either due to poor cell permeability, lack of specificity, short half-life, or cytotoxicity. Moreover, in vitro conditions often fail to fully recapitulate the salt concentration, viscosity, pH, and local protein concentrations in the cellular milieu which can affect protein–protein interactions, stability, and activity. Here, we report the construction of a transcriptional assay which

provides a sensitive fluorescent–luminescent activatable sensor capable of detecting and quantifying intracellular SARS-CoV-2 3CL<sup>pro</sup> activity, enabling the detection of viral infection and providing a platform for detection and validation of potential SARS-CoV-2 3CL<sup>pro</sup> inhibitors. Employing this platform, we examined the efficiency of known protease inhibitors and identified potent 3CL<sup>pro</sup> inhibitors from several chemical



**Figure 3.** SARS-CoV-2 infection activates the live-cell reporter. (A) Schematic of SARS-CoV-2 cell infection activating the reporter signal. Representative immunofluorescence staining for SARS-CoV-2 spike and GLuc in infected cells 24 hpi. Scale bar—20  $\mu$ m. (B) Reporter cells were infected with SARS-CoV-2 using different MOI, and 24 hpi cells were immunostained for SARS-CoV-2 spike and GLuc. (C) The staining intensity of GLuc was compared between SARS-CoV-2 positive and negative cells in four random fields. Data was plotted as mean  $\pm$  SD. Analysis was performed using a two-tailed Student's *t*-test,  $N = 4$ , \*\* =  $p \leq 0.01$ . (D) Relative GLuc expression levels in SARS-CoV-2 infected cells normalized to 18s ribosomal RNA (18s rRNA), 24 hpi. Data was plotted as mean  $\pm$  SD. Statistical analysis was performed using a two-tailed Student's *t*-test,  $N = 3$ , \*\*\* =  $p \leq 0.005$ .

libraries. We conducted a comprehensive computational analysis of the lead compounds' mechanism of inhibition and showed the correlation between the extent of catalytic dyad occlusion and the experimental enzyme inhibition. Amino acid substitutions in positions predicted to interact with specific ligands confirmed the computational simulations and their effects on occlusion of the catalytic site. The occlusion-based mechanistic criterion for predicting inhibition further enabled us to resolve the question of whether some of the inhibitors act by binding at the allosteric site rather than the catalytic site and enabled the identification of a heretofore unrecognized allosteric ligand binding site suitable for inhibitor design.

## RESULTS AND DISCUSSION

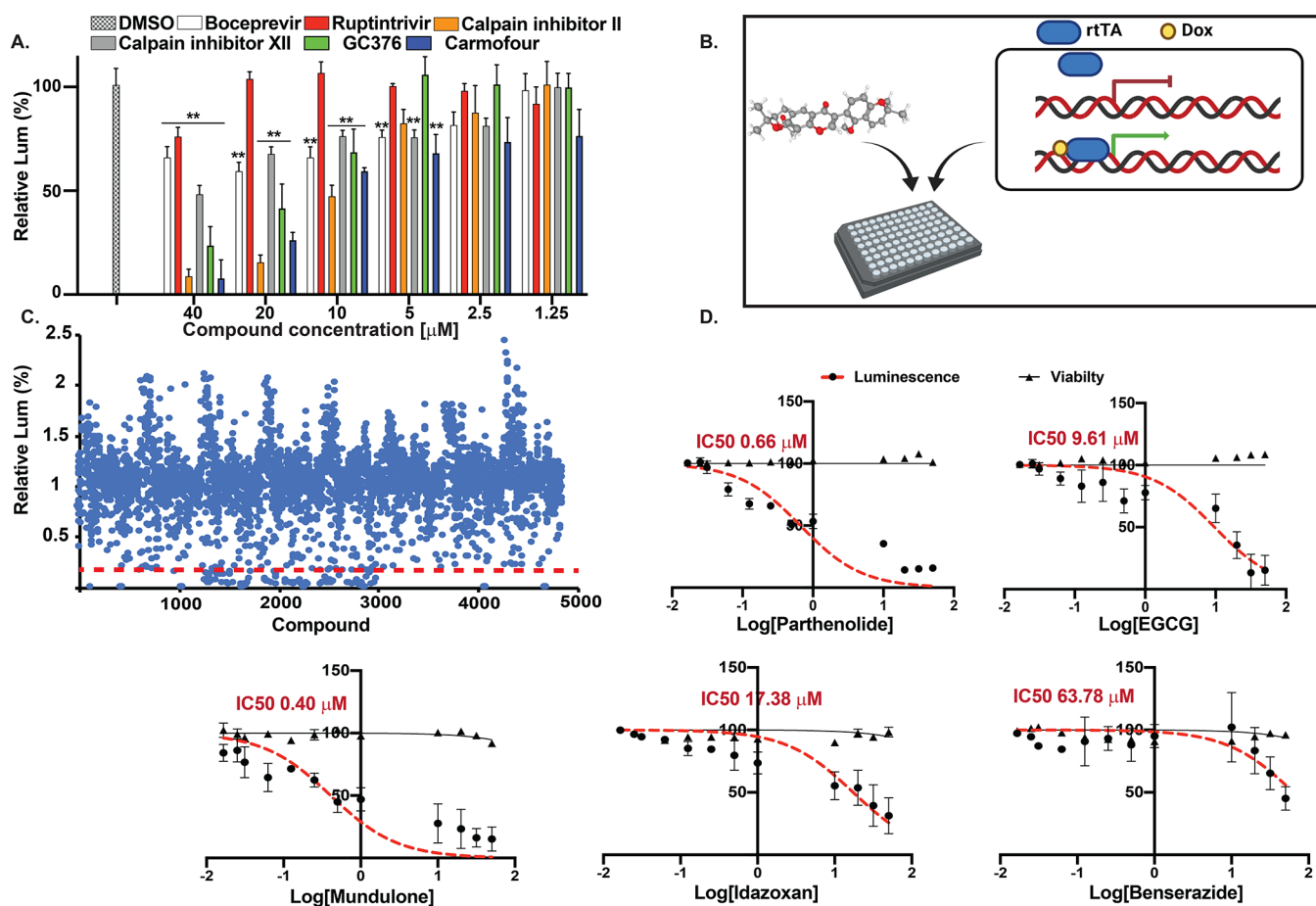
### Construction and Validation of a SARS-CoV-2 3CL<sup>pro</sup> Genetic Reporter

To generate a reproducible cell-based assay, 3CL<sup>pro</sup> protease activity was linked to the release of a membrane-tethered transcription factor (TF) leading to the upregulation of reporter genes. Figure 1 illustrates the general protease reporter design: TF GAL4 was fused to dopamine receptor

D1 (DRD1) with an 18 amino acid linker corresponding to SARS-CoV-2 3CL<sup>pro</sup> protease recognition and cleavage site. Following proteolytic cleavage, GAL4 translocates to the nucleus and activates the transcription of a Citrine-T2A-Gussia luciferase (Gluc) reporter cassette. First, the sensitivity and specificity of the reporter cells was evaluated, and more than a  $\sim$ 70-fold increase in the luminescent signal was detected after 3CL<sup>pro</sup> expression. In contrast, using a catalytically inactive 3CL<sup>pro</sup><sub>CI45A</sub><sup>10</sup> a linker peptide containing the tobacco mosaic virus (TMV), or a protease cleavage site instead did not result in significant luminescence increase (Figure 2A). The fluorescent signal presented a similar trend with significant activation observed during protease expression with the corresponding cleavage site compared to other control samples (Figure 2B), suggesting that transcription activation is strictly dependent on intracellular 3CL<sup>pro</sup> proteolytic activity.

To further evaluate whether the proteolytic efficiency is directly correlated to the reporter cassette transcription, we tested additional linkers corresponding to different SARS-CoV-2 protease cleavage sites (NSP junctions). The NSP4/5 cleavage site exhibited the highest activation with  $\sim$ 70-fold signal increase (Figures 2C and S1), while a lower signal was





**Figure 4.** Live-cell reporter enables high-throughput chemical screening for SARS-CoV-2 3CL<sup>pro</sup> inhibition. (A) 3CL<sup>pro</sup> inhibitor efficiency was evaluated in the reporter cells. Compounds were added at the indicated concentrations and incubated for 48 h, and inhibition was calculated as the percentage of luminescence signal relative to the vehicle-treated group. Data was plotted as mean  $\pm$  SD, and statistical analysis was performed using a two-tailed Student's *t*-test,  $N = 3$ ,  $** = p \leq 0.01$ . (B) Schematic of the chemical screen methodology; cells were plated in 384 poly-lysine coated plates and incubated overnight. The following day, compounds were added at a concentration of 10  $\mu$ M together with 100 ng/mL of doxycycline to induce 3CL<sup>pro</sup> expression. (C) Chemical screen result summary, the luminescence signal of treated cells was measured 48 h post-doxycycline induction, and signal percentage was calculated relative to the vehicle-treated group. Top leads were compounds that decreased the signal by 80% and more. (D) Inhibition efficacy and cell toxicity curves of PRT, EGCG, MND, IDX, and BEN. Each compound was tested at multiple concentrations, and compound toxicity and luminescent signal reduction was evaluated 48 h after addition. Data was plotted as mean  $\pm$  SD,  $N = 3$  for each concentration point, and nonlinear correlation analysis was used to evaluate compounds' IC<sub>50</sub>.

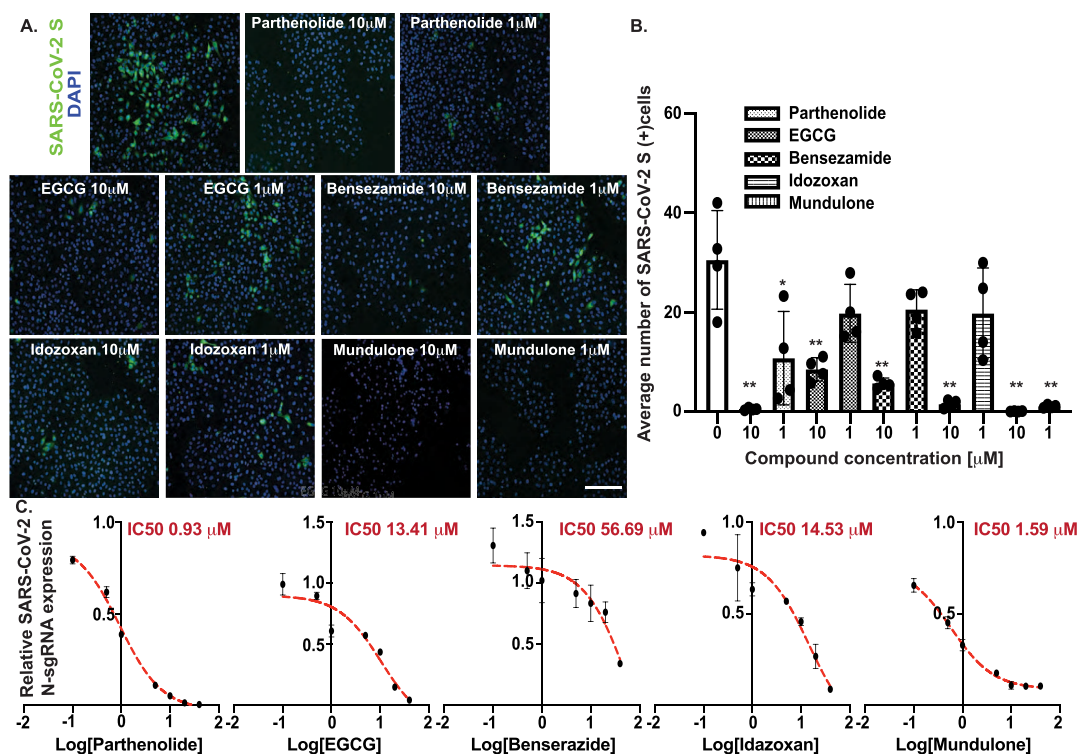
observed by the other cleavage sites (NSP5/6  $\sim$  26 fold, NSP6/7  $\sim$  18 fold, and NSP7/8  $\sim$  15 fold). This is consistent with previous work demonstrating the SARS-CoV 3CL<sup>pro</sup>, which shares a 96% sequence identity with SARS-CoV-2 3CL<sup>pro</sup>,<sup>11</sup> preferentially cleaves the NSP4/5 junction, while NSP5/6, NSP6/7, and NSP7/8 junctions are less efficiently processed.<sup>12,13</sup> Next, we examined the proteolytic cleavage efficiency dependence on intracellular location and fusion partners. To this end, transcription activation was evaluated with different GAL4-fusion partners with and without catalytically active 3CL<sup>pro</sup>. As shown in Figure 2D,E, GAL4 fusion to either vimentin (cytoplasm) or MICU1 (mitochondria) did not result in significant signal activation, while fusion to the N-terminus of cyp2c1 (endoplasmic reticulum) exhibited protease-independent signal activation. 3CL<sup>pro</sup>-dependent activation was observed with GAL4 fusion to the DRD1 receptor (cell membrane) and to the C-terminus of IFN- $\beta$  promoter stimulator protein 1 (IPS-1, MAVS, located in the mitochondria, peroxisomes, and endoplasmic reticulum). We decided to focus on the DRD1 fusion protein since it displayed the highest activation with a  $\sim$ 1.6-fold difference in

luminescent signal compared to the IPS-1 construct. To further validate the 3CL<sup>pro</sup> reporter design, we examined signal change following infection with SARS-CoV-2 (Figure 3). Reporter cells expressing the viral entry receptor ACE2 were infected at different multiplicities of infection (MOI) and examined 24 h post-infection (hpi). Immunostaining showed a detectable activation of Gluc in SARS-CoV-2 spike protein (S) positive cells (Figure 3B). Additionally, the percentage of (S)<sup>+</sup> (Gluc)<sup>+</sup> cells increased in correlation to the virus MOI (Figure 3C). Immunostaining results were further validated by quantitative real-time polymerase chain reaction (qRT-PCR), demonstrating a direct correlation between signal levels and the viral load (Figure 3D).

#### Inhibitor Analysis Using the SARS-CoV-2 3CL<sup>pro</sup> Genetic Reporter

To examine whether the reporter system can be used to evaluate potential viral protease inhibitors, first, we reviewed previously reported 3CL<sup>pro</sup> inhibitors. Compounds were added to the reporter cells at multiple concentrations, and the luminescent signal was measured and normalized to the





**Figure 5.** Identified compounds SARS-CoV-2 inhibition evaluation in lung epithelial cells. (A) A549-ACE2 cells were infected with SARS-CoV-2 (MOI = 0.1) and treated with the identified compounds in the indicated concentrations or with vehicle; 24 hpi cells were immuno-stained for SARS-CoV-2 spike protein (SARS-S). Scale bar = 100 μm. (B) Quantification summary of the percentage SARS-CoV-2 positive cells; images from four random fields were analyzed, and the percentage of SARS-S<sup>+</sup> cells was determined. Data was plotted as mean ± SD, *N* = 3. Analysis was performed using a two-tailed Student's *t*-test, *N* = 3, \*\* = *p* ≤ 0.01. (C) Relative SARS-CoV-2 viral RNA expression in A549-ACE2-infected cells (24 hpi) treated with the identified compounds in the indicated concentrations. Total extracted RNA was evaluated for the presence of SARS-CoV-2 subgenomic *N* transcripts, and the expression level was calculated relative to the vehicle-treated group. Data was plotted as mean ± SD, *N* = 3. Analysis was performed using a two-tailed Student's *t*-test, *N* = 3, \*\* = *p* ≤ 0.01.

vehicle-treated group after 48 h. Rupintrivir, an irreversible inhibitor of human rhinovirus (HRV) 3C protease,<sup>14</sup> showed the lowest impact with a ~20% reduction in GLuc signal measured only at the highest concentration tested (40 μM, Figure 4A) and no detectable effect at lower concentrations. This is in agreement with former reports showing that rupintrivir is an ineffective inhibitor of SARS-CoV<sup>15</sup> or SARS-CoV-2 3CL<sup>pro</sup> activity.<sup>16</sup> The HCV NS3/4A protease inhibitor boceprevir<sup>17</sup> showed a dose-dependent inhibition with a calculated IC<sub>50</sub> of 34.52 μM. Similarly, dose-dependent effects were detected with the broad spectrum 3C protease inhibitor GC373<sup>15</sup> with an IC<sub>50</sub> of 20.5 μM. The antineoplastic compound carmofur was recently shown to directly interact with the catalytic site of SARS-CoV-2 3CL<sup>pro</sup><sup>18</sup> and was also able to inhibit protease activity with a calculated IC<sub>50</sub> of 8.66 μM. Calpain inhibitor II<sup>19</sup> exhibited the strongest inhibition with a reduction of GLuc signal by more than 90% and an IC<sub>50</sub> of 5.81 μM. Considering the strong inhibitory effect of calpain inhibitor II, we further evaluated the effect of other known calpain inhibitors on SARS-CoV-2 3CL<sup>pro</sup> activity (Figure S2). While calpain inhibitor VI did not display a significant effect, we observed a dose-dependent signal reduction by other calpain inhibitors. ALLN, a calpain I inhibitor, was found to be the most effective, inhibiting the protease activity almost completely with a calculated IC<sub>50</sub> of 1.27 μM. Calpain inhibitors have previously been shown to inhibit 3CL<sup>pro</sup> activity of both SARS-CoV<sup>20</sup> and SARS CoV-2,<sup>16</sup> inhibit spike-mediated membrane fusion,<sup>21</sup> and decrease virus-

induced apoptosis.<sup>22</sup> Calpain activation has also been implicated in multiple inflammatory-associated diseases<sup>23</sup> and in a variety of human pathological conditions.<sup>24</sup> However, further chemical and structural optimizations are required due to their relatively low specificity toward different cysteine proteases.

Next, we used the 3CL<sup>pro</sup> reporter cell line to screen three chemical libraries (Prestwick FDA-approved drug library, Lopac library, and MicroSource bioactive chemical library) to identify compound candidates capable of inhibiting SARS-CoV-2 3CL<sup>pro</sup> activity. To this end, a doxycycline-inducible vector was used to control both time and 3CL<sup>pro</sup> expression levels. Library compounds and doxycycline were added to the reporter cells in 384-well plates [10 μM and 100 ng/ml, respectively (Figure 4B)]. 48 h post induction, GLuc activity was measured and compared to vehicle-treated cells (Figure 4C). Compounds that displayed 80% or more signal reduction with low cytotoxicity were considered as primary lead compounds. Five compounds were identified as lead compounds and further tested for their capacity to inhibit 3CL<sup>pro</sup> activity in a dose-dependent manner. As shown in Figure 4D, we observed concentration-dependent protease inhibition without significant impact on cell viability with the following compounds: parthenolide (PRT, IC<sub>50</sub> = 0.66 μM), epigallocatechin gallate (EGCG, IC<sub>50</sub> = 9.61 μM), mundulone (MND, IC<sub>50</sub> = 0.407 μM), idozoxan (IDX, IC<sub>50</sub> = 17.38 μM), and bensezamide (BEN, IC<sub>50</sub> = 63.78 μM). Several of the compounds identified in this screen were previously suggested

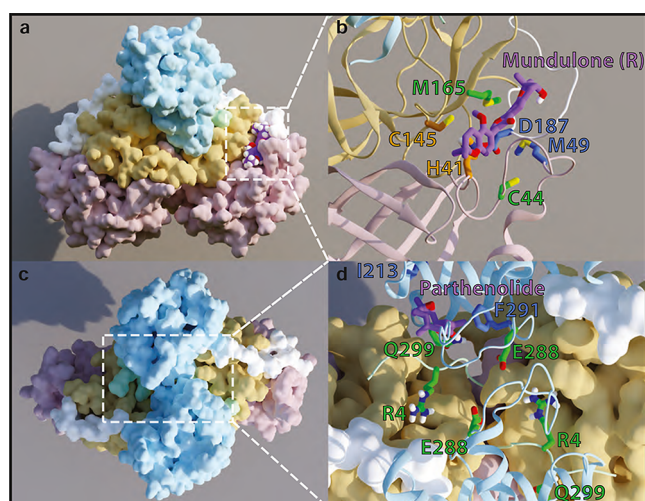
as potential viral inhibitors, for example, EGCG was shown to inhibit SARS-CoV 3CL<sup>pro</sup> in vitro<sup>25,26</sup> and has been suggested as a potential SARS-CoV-2 3CL<sup>pro</sup> inhibitor by in silico molecular docking studies.<sup>27,28</sup> IDX hydrochloride was recently identified as a potential SARS-CoV-2 inhibitor in an in vitro chemical screen.<sup>29</sup> BEN hydrochloride was identified as a potent Coxsackievirus 3C protease (3Cpro) inhibitor,<sup>30</sup> and MND acetate was identified in an in vitro chemical screen targeting Middle East respiratory syndrome coronavirus (MERS-CoV).<sup>31</sup> To further confirm the compounds' specificity, a FRET-based assay was used to quantify the 3CL<sup>pro</sup> activity,<sup>32</sup> recombinant SARS-CoV 3CL<sup>pro</sup> was incubated with the compounds in different concentrations followed by the addition of a FRET peptide substrate, and relative protease activity levels were calculated in comparison to the vehicle-treated protease (Figure S3). We detected a concentration-dependent inhibition with IC<sub>50</sub> values comparable to the levels observed in the 3CL<sup>pro</sup> cell reporter, which suggest the genetic reporter can not only detect potential inhibitors but also anticipate to a certain degree the efficiency of a potential compound. Nonetheless, it is important to highlight that 3CL<sup>pro</sup> enzymatic activity and the levels of a potential inhibitor were previously shown to be sensitive to the condition used in the assay, and effects such as protein conformational changes and monomer–dimer equilibrium can significantly influence the results,<sup>33</sup> which emphasize the importance of examining compounds also in a cellular setting to avoid non-specific effects.

#### Identified Compounds Inhibit SARS-CoV-2 Infection

Next, we examined whether the identified compounds are efficient inhibitors in the context of SARS-CoV-2 infection and can help reduce the viral load. To this end, lung epithelial cells (A549-ACE2) were infected with SARS-CoV-2 (USA-WA1/2020) at an MOI of 0.1 in the presence or absence of chemical compounds (Figure 5A). Immunostaining targeting SARS-CoV-2 spike protein revealed approximately ~30% of infected cells in the vehicle cells, while in the treated group, we observed a statistically significant reduction of infected cells. PRT, idazoxan hydrochloride, and MND acetate were able to inhibit virus propagation almost completely, reducing the number of infected cells by more than 90% compared to the vehicle-treated group at a concentration of 10 μM. Furthermore, MND acetate was able to maintain strong inhibitory effect also at a lower concentration of 1 μM (Figure 5B), in excellent agreement with the ranking by both the SARS-CoV-2 3CL<sup>pro</sup> genetic reporter and the computational findings of the largest occlusion of the catalytic site. Quantification of viral nucleocapsid sgRNA levels (as a measure for active viral genome replication) across multiple concentrations was evaluated by qRT-PCR, and IC<sub>50</sub> values were calculated, showing similar trends with a significant decrease of vRNA in the treated groups compared to the control samples (Figure 5C).

#### Computational Analysis of the Mechanism of Inhibition

Similar to previously reported 3CL<sup>pro</sup> inhibitors, the lead compounds identified in this work can interact with the enzyme active site (orthosteric binding) as shown in Figure 6, blocking the access of substrates to the Cys145/His41 catalytic residues.<sup>5,34–37</sup> However, some repurposed drugs<sup>38–41</sup> and compound fragment hits<sup>42–44</sup> have been reported to favor binding at the interface of the 3CL<sup>pro</sup> protomers (dimerization interface) along domains II and III. Inhibition by binding to a

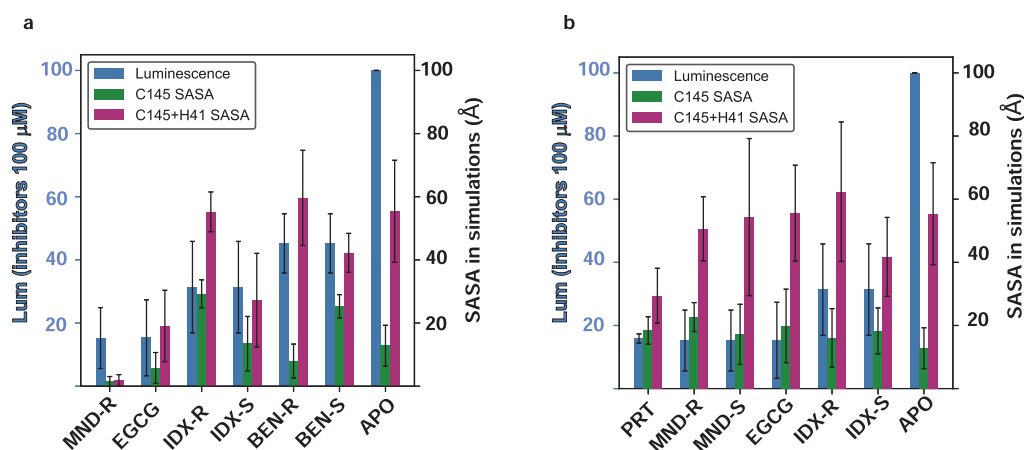


**Figure 6.** Illustration of the mode of inhibitor binding to 3CL<sup>pro</sup> in the catalytic site and the dimerization interface site. The figure panels obtained from representative frames from the MD simulation trajectories show preferred modes of binding for the MND-R that prefers the catalytic site (panels A,B) and for PRT which prefers the dimerization interface site (panels C,D). Panels (A,C) present the protein domains in different colors (domain 1 in pink; 2 in yellow; 3 in light blue; N-finger in mint; and linker loop in white). Panels (B,D) zoom into the binding site regions showing the inhibitors in purple, the residues probed by mutagenesis in green, and residues with the largest contributions to the BFE in blue.

region distal to the active site would be considered to involve an allosteric mechanism<sup>45–47</sup> that has remained unidentified. To reveal the mechanisms of inhibition of the various compounds, we investigated the apo- and ligand-bound forms of the 3CL<sup>pro</sup> dimer with all-atom molecular dynamics (MD) simulations. Computational docking experiments described in Methods were used as the starting points for the analysis, and the MD simulations were carried out for the best docking results for each ligand in the catalytic (orthosteric) site and in the individual docking (putative allosteric) site at the dimerization interface of the enzyme.

#### Inhibitor Binding Modes at the 3CL<sup>pro</sup> Orthosteric and Allosteric Sites

To evaluate the binding of the identified inhibitors at the 3CL<sup>pro</sup> orthosteric and allosteric sites, we modeled the results of the MD simulations of each of the 3CL<sup>pro</sup> inhibitors starting from the docked poses at the two sites explored for each were ranked by the calculated binding free energies (BFEs) from the corresponding MD trajectories using the MM-PB/GBSA approach (see Methods). The protein–ligand total BFE values calculated for each compound (Table S4) indicate the preferences of each ligand for the sites and modes of binding. Thus, we found that both BEN enantiomers have stronger interaction energies with the catalytic site over the dimerization interface region (by ~10 kcal/mol). In contrast, the S enantiomer of MND, (MND)-S, and PRT are more likely to bind at the dimerization site, with a BFE ~ 7 kcal/mol lower than at the catalytic site. The rest of the compounds (EGCG, R and S enantiomers of IDX—(IDX)-R/S, and R-enantiomer of MND—MND-R) have comparable binding energies at the catalytic and dimerization interface sites, indicating no preference between orthosteric and allosteric binding modes. Preferred modes of binding at the two sites are illustrated in Figure 6 for MND-R (catalytic site—Figure 6A,B) and PRT



**Figure 7.** Relation between the extents of occlusion of the catalytic residues by inhibitor binding and the inhibitory efficiency measured by luminescence. The SASA of the catalytic residues ( $C_{145}$  alone, green bars; the combined  $C_{145}/H_{41}$  dyad, maroon bars) was calculated in the trajectories of the  $3CL^{Pro}$  dimer complexes with inhibitor bound at the orthosteric (catalytic) site (left panel) and at a dimerization interface site (right panel). The occlusion was quantified by comparing the SASA calculated in the presence of each inhibitor to the SASA values in the *apo* simulation trajectory. The rank order of occlusion is seen to agree with the ranking of the experimentally measured inhibitory efficiency quantified by the reduction in luminescence produced by each inhibitor.

(dimerization interface—Figure 6C,D), and the poses for all other compounds at the two sites are presented in Figures S3 and S4, respectively. Notably, MND and PRT have the highest calculated binding affinity among the identified inhibitors at the catalytic and dimerization sites, respectively (Table S4). For each inhibitor, we identified the residues that contribute most to the BFE in their corresponding binding sites. These were determined from the decomposition of the free energies obtained with the MM-PB/GBSA approach into contributions from individual residues (see detailed formalism in Methods). The top two contributors to the specific BFE for at least two compounds at one of the sites are highlighted in Table S4. Some of these residues—including C44, M49, and M165 at the catalytic site; and R4, F284, and S291 at the dimerization interface—have been reported to be involved in the binding of a variety of ligands with predicted and measured inhibitory.<sup>5,34,35,38,42,44</sup> As shown in Table S5, the mutagenesis studies have highlighted M165 and R4 as key residues for BEN inhibition, resulting in a complete loss of inhibition in the alanine variants. Moreover, we were able to validate the predictions for the role of M165 in the inhibitory activity of MND, by a significant decrease of MND  $IC_{50}$  in the M165A compared to the WT enzyme. Similarly, the inhibition efficiency of PRT was reduced in the M165A and Q299A mutants compared to WT. The nature of the direct involvement we identified for Q299 in the inhibition of  $3CL^{Pro}$  is consistent with the importance of this region for inhibition by glutathionylation at C300.<sup>48</sup>

#### Inhibitors Binding at the Catalytic Site Occlude the Access to the $C_{145}/H_{41}$ Dyad to Different Extents

The extent of occlusion of the catalytic residues by the binding of the inhibitors was quantified from the results of the MD simulations by comparing the calculated solvent accessible surface area (SASA) of the catalytic residues in the trajectories of the complexes, with the SASA in the *apo*  $3CL^{Pro}$  dimer trajectory. Results in Figure 7A show that the extents of occlusion indicated by this comparison for  $C_{145}$  and for the combined  $C_{145}/H_{41}$  dyad differ among the inhibitors. The largest effect of orthosteric site binding is calculated for MND-R and EGCG, which reduce the SASA of both  $C_{145}$  and  $C_{145}/$

$H_{41}$  SASA to very low values compared to *apo* (see green and purple bars, respectively, relative to the axis on the right). In contrast, IDX and BEN are not as effective. Catalytic site occlusion by each compound was also compared with the corresponding experimental inhibitory efficacy, quantified by the luminescence assay at the highest concentration (100  $\mu M$ ). The comparison of occluding effects with their relative effects on the luminescence-based indicator of inhibitory activity (blue bars in Figure 7A relative to left axis) shows good agreement between these measures. Thus, a 2–3 times higher efficacy in signal reduction by both MND-R and EGCG compared to the other two ligands corresponds to their relative ranking by occlusion of the catalytic residues.

Overall, this analysis shows that all the ligands predicted to prefer binding at the orthosteric site can inhibit  $3CL^{Pro}$  by direct occlusion of the catalytic residues with efficiencies that correspond to their effect measured by the SARS-CoV-2  $3CL^{Pro}$  genetic reporter.

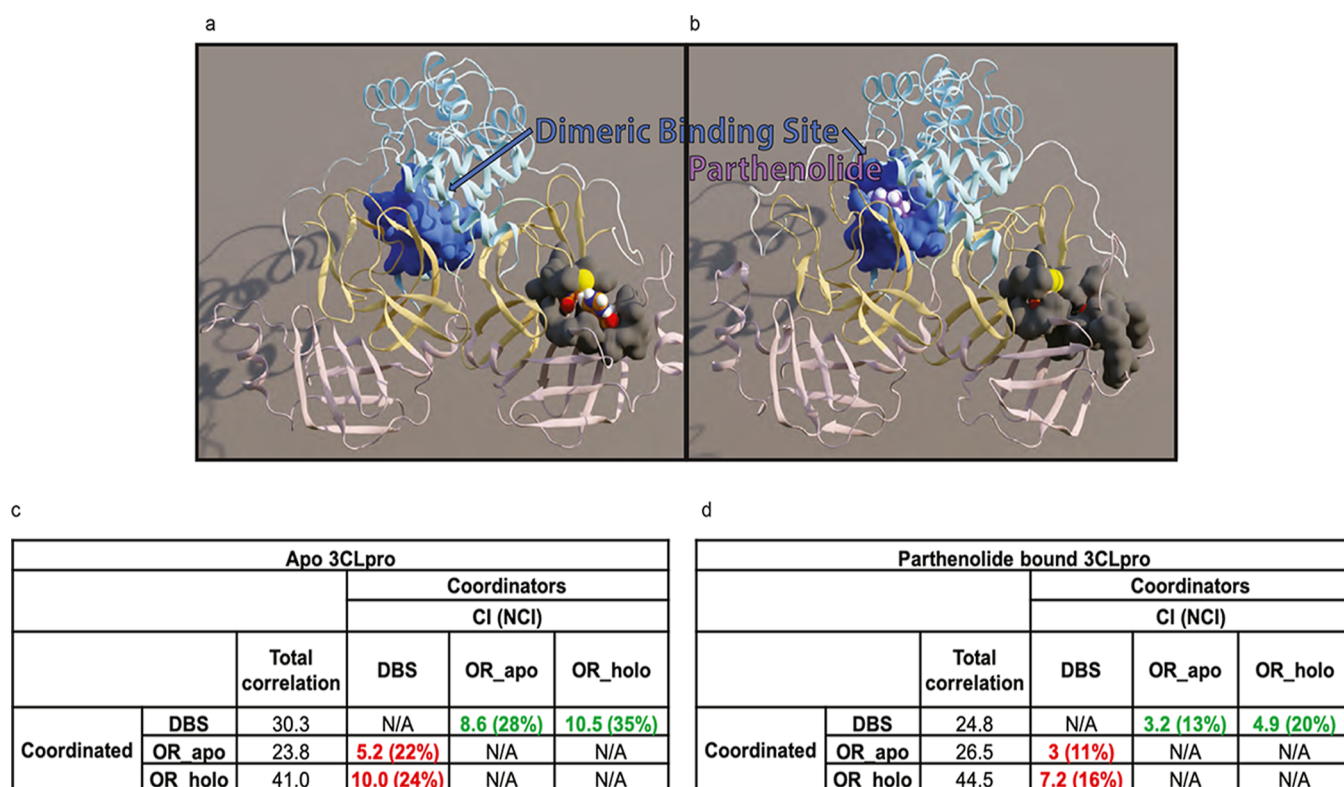
#### Inhibitor Binding at the $3CL^{Pro}$ Dimer Interface Can Occlude Access to the $C_{145}/H_{41}$ Dyad through an Allosteric Mechanism

Unlike the catalytic site binding poses, inhibitor binding at the dimer interface cannot directly occlude the accessibility to the catalytic site. Nevertheless, the same SASA-based indicators show that PRT binding at the dimerization site leads to a large reduction in  $H_{41} + C_{145}$  SASA (29  $\text{\AA}^2$ ) compared to the *apo* simulation result (55  $\text{\AA}^2$ ) (Figure 7B, purple bars). This shows that binding at this allosteric site can occlude access to the catalytic residue pair  $C_{145} + H_{41}$ . Among the other ligands, only IDX-S appears to elicit a SASA reduction relative to *apo* by binding in this region.

#### PRT as a Prototypical Allosteric Inhibitor

Analysis of the MD trajectories for the PRT- $3CL^{Pro}$  dimer complex at the dimerization interface reveals major conformational rearrangements in the catalytic site. We used NbIT<sup>49</sup>—an information theory-based analysis framework for analysis of allosteric pathways (see Methods)—to examine how much information is shared in the dynamics between the binding site of PRT in the dimer interface site (DBS) and the active site residues that occlude  $C_{145}/H_{41}$  (occluding residues—OR).





\* Normalized CI (NCI) by coordinated sites' total correlation are shown in parentheses.

\*\* NCI of one site itself is 100% and not shown. OR\_apo and OR\_holo have many shared residues so their mutual CI/NCI are not shown.

**Figure 8.** PRT disrupts the allosteric communication between the DBS and the catalytic site of 3CL<sup>Prt</sup>. (A) Apo structure, with the residues in the dimeric binding site shown in a blue surface and residues that occlude the catalytic dyad shown in a gray surface. (B) In the complex with PRT (in purple) bound in the DBS, the residue groups are shown in the same colors as in A. Note the dramatic reduction in accessibility to the catalytic dyad. (C) Values of the TC, CI and NCI (in parentheses) are shown for apo 3CL<sup>Prt</sup>, calculated with the DBS as the coordinator and the occluding residues defined from the apo and holo trajectories (OR\_apo/OR\_holo) as coordinated (values highlighted in red). For comparison, values from the corresponding reversed calculation (DBS is coordinated; occluding residue groups are coordinating) are highlighted in green. (D) Results from the same calculations as in (C), from the PRT-bound 3CL<sup>Prt</sup> trajectories.

This is expressed in the coordination information (CI) component calculated with the DBS as a “coordinator” site, and all heavy atoms of the OR as the “coordinated” site. As a control for the significance readout of the calculated CI values for the functional sites, we compare the magnitude of the CI values calculated for the functional sites with values obtained for a set of arbitrarily chosen groups of residues (“decoy” sites) distributed throughout the 3CL<sup>Prt</sup> (as background control groups, we used the set D1–D3 detailed in Supporting Information and Figure S5). The results shown in Figure 8C,D quantify the CI values and the normalized CI (NCI) values between DBS, and the OR\_apo and OR\_holo groups of residues in the active site. The residues in these OR groups were identified to occlude the catalytic dyad in the apo/holo simulations, respectively (Figure 8A,B). In the apo 3CL<sup>Prt</sup> dimer structure, the NCI values for DBS/OR\_apo and DBS/OR\_holo coordination (22 and 24%, respectively, Figure 8C) are much higher than for those with the background control group DBS/D1–D3 (6–12%, see Table S6A in Supporting Information). The high NCI indicates a large mutual information connection between the active site and the allosteric binding pocket in the apo state of the enzyme dimer. By contrast, with PRT bound in this pocket, the NCI values are drastically reduced to low levels (11 and 16%, Figure 8D) that do not differ from the non-significant NCI values of the background controls DBS/D1–D3 (5–11% in Table S6).

The strong allosteric effect of PRT binding is thus found to stabilize the catalytic site in a new conformation that shields sterically the catalytic dyad from interactions with substrates and disrupts the mutual information connection between the binding region and the active site as shown in Figure 8D by the reduction in NCI values. Notably, this major reduction of coordinating information in the PRT-bound 3CL<sup>Prt</sup> pertains to both directions of communication, as the NCI values are drastically reduced even if the occluding residues were considered as coordinators (Figure 8C,D). The occluding rearrangement of the residues surrounding the C<sub>145</sub>/H<sub>41</sub> is evident in the comparison between panels (A,B) in Figure 8. Notably, the same NbIT-based analysis of the MD trajectories for the IDX-S bound in the dimer interface region shows no change in NCI values between the DBS composed of specific IDX-binding residues and the OR sets compared to the apo state, indicating that the allosteric connection is not operative from the IDX binding site in the dimerization interface region (see Table S7). This underscores the specific allosteric interaction capability of binding in the PRT site. Indeed, unlike the substantial occlusion produced by PRT, the small occlusion observed from the SASA calculations for IDX (Figure 6B) is likely due to a local rearrangement because the information connection between the catalytic site and the IDX binding site is not affected (Table S7). Thus, our information theory-based analyses have revealed a site-specific allosteric

mechanism of inhibition by PRT. With the specific identity of the residues involved in the CI between the catalytic site and this locus at the dimer interface, our findings identify a heretofore unrecognized site for allosteric inhibition of 3CL<sup>pro</sup> that should be useful for the development of allosteric SARS-CoV-2 3CL<sup>pro</sup> inhibitors, and a detailed account of the allosteric mechanism of inhibition.

It has become clear that 3CL<sup>pro</sup>, the main protease of SARS-CoV-2, is an attractive target for the development of effective antiviral therapies. Amino acid homology and atomic structural analysis indicates a high degree of similarity and conservation of the active site among CoVs,<sup>50,51</sup> which emphasizes its critical role during the viral life cycle and further suggests that an effective inhibitor potentially could serve as a pan-coronavirus therapeutic. In this study, we constructed a live-cell reporter which is dependent on the cleavage and release of a membrane-tethered TF (a protease substrate) by SARS-CoV-2 3CL<sup>pro</sup>, leading to the transcriptional activation of a fluorescent-luminescent reporter cassette. We observed a direct correlation between the degree of signal activation and the protease activity by comparing different cleavage sites that were previously shown to have different turnover numbers ( $k_{cat}$ ). Moreover, reporter cells infected with SARS-CoV-2 resulted in activation of the reporter cassette corresponding to the viral load, suggesting that it can also be a useful live-cell reporter to study viral infection and propagation without the need for a genetically modified virus. This is especially valuable, since rescue systems for CoVs are challenging to develop due to their large genome size and the risk of altering the virus fitness due to genomic manipulation. Several cell-based assays were previously developed to measure 3CL<sup>pro</sup> intracellular activity, and techniques such as Flip-GFP,<sup>52</sup> Flip-Firefly luciferase,<sup>53</sup> reverse nano-luciferase<sup>54</sup> were shown to be useful to conduct chemical screen and identify potential new inhibitors.<sup>55,56</sup> The ability to measure 3CL<sup>pro</sup> activity in cells using a reporter with high sensitivity and a wide dynamic range is a powerful tool for chemical compound screens. Moreover, other factors such as cell integrity, permeability, and compound toxicity can be simultaneously evaluated, providing a more comprehensive and accurate analysis of potential candidate compounds. Utilizing the live-cell reporter, we evaluated both compounds that were recently reported to inhibit 3CL<sup>pro</sup> and executed a chemical screen to generate new therapeutic leads. Using MD simulations, we were able to characterize the compounds' interactions with 3CL<sup>pro</sup> and the mechanism of inhibition. Moreover, information theory-based analyses identified PRT as an allosteric inhibitor, interacting with the 3CL<sup>pro</sup> dimerization domain which result in occlusion of the enzyme catalytic dyad C<sub>145</sub>/H<sub>41</sub>. This live-cell reporter and the computational analysis presented in this study provide a robust platform to identify, characterize, validate, and further optimize potential compounds targeting the main virus protease, a key enzyme in the viral life cycle, which can hinder viral propagation and possibly mitigate disease severity.

## METHODS

### Plasmid Cloning

Gal4 fusion constructs were generated using the pAAV-DRD1-NNES-eLOV-TEVcs-Flag-Gal4-V5 vector (Addgene 125227, Alice Ting); NSP cleavage sites were generated using a codon-optimized gBlock (IDT) and cloned into the NheI-BamHI digested plasmid to generate pAAV-DRD1-NSPcs-Flag-Gal4-V5. MICU1, vimentin, and cyp2c1 sequences were PCR-amplified from the Addgene plasmids 79057,

66170, and 79055 (Alice Ting) and cloned into HindIII-NheI-digested pAAV-DRD1-NSPcs-Flag-Gal4-V5 replacing the DRD1 gene. The IPS-1 (residues 462–520) fragment was generated using a codon-optimized gBlock (IDT), and Flag-Gal4-V5-NSPcs-IPS1 fusion was PCR-assembled and cloned into the HindIII-PspXI-digested plasmid to generate pAAV-Flag-Gal4-V5-NSPcs-IPS1. The reporter cassette was created using pFPGW-UAS-Citrine (Addgene 125232, Alice Ting), Gaussia luciferase was amplified from the pMCS-Gaussia Luc vector (Thermo Fisher), a T2A sequence was incorporated on the 5' end by PCR, and the fragment was cloned into the BsrGI AscI-digested plasmid to create pFPGW-UAS-Citrine-T2A-GLuc. Inducible NSP5 was made using the pSBtet-Hyg plasmid (Addgene 60508, Eric Kowarz), and NSP5 and the catalytically inactive C145A mutant were amplified using pLVX-EF1alpha-SARS-CoV-2-nsp5-2xStrep-IRES-Puro and pLVX-EF1alpha-SARS-CoV-2-nsp5-C145A-2xStrep-IRES-Puro as templates (Addgene 141370 and 141371, Nevan Krogan) and cloned into the SfiI-digested plasmid to generate pSBtet-NSP5-Hyg and pSBtet-NSP5c145a-Hyg. All the cloning ligations were executed using the In-Fusion method (Takada Bio) and were transformed into Stellar competent cells (Takada Bio) according to the manufacturer's instructions.

### Cell Lines

HEK293T (human [*Homo sapiens*] fetal kidney) and Vero E6 (African green monkey [*Chlorocebus aethiops*] kidney) cells were obtained from ATCC (<https://www.atcc.org/>). A549 lung epithelial cells stably expressing ACE-2 development is described elsewhere.<sup>57</sup> Cells were cultured in Dulbecco's modified Eagle's medium (DMEM) supplemented with 10% FBS and 100 IU/mL penicillin and 100 µg/mL streptomycin. All cell lines were incubated at 37 °C with 5% CO<sub>2</sub>.

### Lentivirus Production

HEK293T cells were grown in a 10 cm plate to ~90% confluency and were transfected with 6.2 µg of pFPGW-UAS-Citrine-T2A-GLuc, 4.6 µg of pPAX2, and 3.1 µg of pMD2.G using Lipofectamine 3000 (Thermo Fisher) according to the manufacturer's instructions; cell media was collected 48 and 72 h after transfection and filtered through a 0.45 µm syringe (Nest Scientific).

### Reporter Imaging and Quantification

Cells were stained in a 1:10,000 diluted Hoescht 33342 (Thermo Fisher) in cell media for 20 min, 37 °C with 5% CO<sub>2</sub>. Images were obtained using an EVOS Cell Imaging System (Thermo Fisher). For Gaussia luciferase measurements, native coelenterazine (GoldBio) was dissolved in acidified methanol (1 mg/mL) and was further diluted 1:100 in GLuc buffer: 0.1% disodium phosphate, 5% glycerol, 150 mM sodium bromide, 1 mM EDTA, 25 mM Tris-HCl pH 8, and 2 mM ascorbic acid. Luciferase activity was measured intracellularly; to this end, media was removed, and cells were lysed in GLuc buffer supplement with 0.1% Triton X-100 for 5 min at room temperature. The cell lysate was added to Gluc buffer with coelenterazine at a ratio of 1:5. Measurements were done in a 96- or 384-well white plate using a Spectramax M5 plate reader (Molecular Devices).

### High-Throughput Chemical Screening

Stable HEK293T UAS:Citrine-T2A-Gaussia cells cultured in 10 cm plates were transfected with pSBtet-NSP5-Hyg and pAAV-DRD1-NSP4/5cs-Flag-Gal4-V5 using Lipofectamine 3000 (Thermo Fisher) according to the manufacturer's instructions. Five hours after transfection, cells were washed and dissociated with trypsin-EDTA solution (Thermo Fisher) and re-plated into 0.01% poly-lysine-coated 384-well plates at 6000 cells/40 µL media/well. After overnight incubation, the cell media was changed, and 100 ng/mL doxycycline was added together with compounds from an in-house library of FDA-approved drugs (Prestwick), Lopac (Sigma Aldrich) and MicroSource (MicroSource discovery systems) at 10 µM concentration. DMSO treatment was used as a vehicle reference. After 48 hours, cells were harvested for luciferase quantification as described above.

## ■ CYTOTOXICITY ASSAY

HEK293T cells were plated in 0.01% poly-lysine-coated 96-clear bottom black plates (Corning) at 20,000 cells per well. After an overnight incubation, cell media was replaced and the examined compounds and vehicle were added in varying concentrations for 48 h. Cell viability was evaluated using PrestoBlue reagent (Thermo Fisher) according to the manufacturer's instructions.

## ■ 3CL<sup>PRO</sup> In Vitro FRET Activity Assay

3CL<sup>PRO</sup> (0.1  $\mu$ M, BPS Bioscience) was incubated with the examined compounds at the indicated concentrations or with vehicle for 1 h at 25 °C (50 mM HEPES pH 6.5). The reaction was initiated by the addition of the peptide. The FRET substrate was added (1.25  $\mu$ M, Anaspec), and each well was incubated at 25 °C for 2 h. Fluorescence was measured using a black 96-well plate on a SpectraMax M5 plate reader (Molecular Devices).

## ■ SARS-COV-2 VIRUS

SARS-CoV-2 isolate USA-WA1/2020 (NR-52281) was deposited at the Center for Disease Control and Prevention and obtained through BEI Resources, NIAID, NIH. SARS-CoV-2 was propagated in Vero E6 cells in DMEM supplemented with 2% FBS, 4.5 g/L D-glucose, 4 mM L-glutamine, 10 mM non-essential amino acids, 1 mM sodium pyruvate, and 10 mM HEPES as described previously.<sup>28</sup> Virus stock was purified using a 100 kDa MWCO Amicon filtration device, as described previously. All work involving live SARS-CoV-2 was performed in the CDC/USDA-approved BSL-3 facility of the Global Health and Emerging Pathogens Institute at the Icahn School of Medicine at Mount Sinai in accordance with institutional biosafety requirement.

## ■ SARS-CoV-2 Virus Infection Inhibition

The viral inhibition efficiency of the identified compounds was evaluated in A549-ACE2 cells. Cells were treated with the examined compounds at the indicated concentrations for 1 h at 37 °C. Subsequently, cells were infected with SARS-CoV-2 for 24 h at an MOI of 0.1 in DMEM supplemented with 2% FBS while maintaining compounds' concentrations. Infected cells were either lysed in TRIzol (Thermo Fisher) for RNA analysis or fixed in 5% formaldehyde for 24 h for immunofluorescence staining, prior to safe removal from the BSL-3 facility.

## ■ IMMUNOSTAINING

Fixed cells were washed three times with PBS and permeabilized using PBS 0.5% Triton X-100. Cells were blocked with 1% donkey serum (Sigma-Aldrich) in PBS 0.1% Triton X-100 for 1 h and stained with primary antibody diluted in 1% donkey serum, PBS 0.1% Triton X-100 at 4 °C overnight and secondary antibodies at RT for 1 h. Information regarding primary and secondary antibodies is provided in Table S3. Nuclei were counterstained using DAPI. Images were obtained using an epifluorescence microscope (Nikon TI-E Eclipse) using a 20 $\times$  objective and a Zyla 5.5 sCMOS camera (Andor, Belfast, United Kingdom).

## ■ QRT-PCR

Total RNA samples were extracted using TRIzol reagent (Thermo Fisher) and the Direct-zol RNA Miniprep Plus kit (Zymo Research) according to the manufacturer's instructions. cDNA was prepared using the LunaScript RT SuperMix Kit

(New England Biolabs), and qRT-PCR was performed using Luna Universal qPCR Master Mix (New England Biolabs) on a cfx384 qPCR instrument (Bio-Rad) with primers specific for the TRS-L and TRS-B sites for the N gene as well as 18s rRNA as an internal reference.<sup>29</sup> Delta–delta-cycle threshold ( $\Delta\Delta CT$ ) was determined relative to the 18s rRNA and mock infected/treated samples. Error bars indicate the standard deviation of the mean from three biological replicates.

## ■ COMPUTATIONAL PROCEDURES

### Molecular Model of 3CL<sup>PRO</sup>

The molecular structure of the SARS-CoV-2 main protease (3CL<sup>PRO</sup>) used in this work is from the joint neutron/X-ray in PDB ID 7JUN, acquired at room temperature (resolution: 2.5 Å).<sup>6</sup> This structure includes all 3CL<sup>PRO</sup> residues, and importantly hydrogen atoms were resolved; thus, additional modeling of the 3CL<sup>PRO</sup> structure was not needed (e.g., no missing loops, no need to predict protonation states, etc.).

### Structural Models of 3CL<sup>PRO</sup> Inhibitors

The initial 3D models of BEN, EGCG, IDX, MND, and PRT were downloaded from PubChem (<https://pubchem.ncbi.nlm.nih.gov/>).<sup>58</sup> Both the R- and S-enantiomers were modeled for BEN, IDX, and MND. Schrodinger's LigPrep protocol was used to assign initial OPLSe3 force field (FF) atom types, FF parameters, FF partial charges, as well as protonation states.<sup>59</sup> These initial OPLS3e partial charges and parameters were refined using Schrodinger's FFBuilder tool. These optimized partial charges and parameters were used in the inhibitor docking calculations. Note: The initial conformation used for each compound is inconsequential, as the modeling procedure (described below) robustly samples inhibitor conformation.

### Identifying the Region of Putative Dimer Interface Binding Sites

A list of residues that compose the putative allosteric inhibitor binding site at the 3CL<sup>PRO</sup> dimer interface (between the domain III regions of the two protomers) were identified from two sources:<sup>6</sup> a collection of residues found in previous studies<sup>38,42,60,61</sup> that bind with the inhibitors or fragments near the 3CL<sup>PRO</sup> dimer interface; and<sup>58</sup> residues predicted to belong to a ligand binding pocket near the 3CL<sup>PRO</sup> dimer interface by two binding site prediction servers—ISMBLAb-LIG<sup>62</sup> and FTSite.<sup>63</sup> The residues from both predictions were included in the list. The resulting list of residues at the DBS includes:

Protomer A: Residue 2–5, 126–127, 131, 137–139, 170, 172, 197, 207, 281–291

Protomer B: Residue 1–5, 126–127, 131, 137–138, 197, 207, 214, 281–291

Protocols for inhibitor docking and MD simulations Schrodinger's Induced Fit protocol<sup>66</sup> was used to dock the inhibitors at both the catalytic site and the putative dimer interface binding site. Briefly, this protocol consists of three stages: (1) flexible docking and scoring of ligands at the proposed binding site, (2) refinement of protein side chain conformations for the poses generated in stage 1, and (3) re-docking of ligands to the side chain-refined structures. To identify binding poses at the catalytic site, the inhibitors were docked (using this three-stage protocol) in a 30 Å box; the center of this box was defined as the centroid of the coordinates of the two catalytic residues, H<sub>41</sub> and C<sub>145</sub>. Similarly, to identify binding poses at the putative DBS, the



inhibitors were docked in a 30 Å box; the center of this box was defined as the centroid of the coordinates of the dimer site residues defined above. Other than the requirement that the inhibitors were docked in the defined boxes described above, no other constraints were used.

MD simulations were performed to improve exploration of inhibitor binding at the two proposed 3CL<sup>Pro</sup> sites. CHARMM-GUI's Solution Builder<sup>67</sup> was used to construct each simulation system, with the best three poses ranked by the Schrodinger Emodel score were used as the starting points. Each 3CL<sup>Pro</sup> inhibitor complex was solvated in a K<sup>+</sup>Cl<sup>-</sup> ionic solution, resulting in a system size of ~180,000 atoms. The optimized partial charges described above were used for the ligands, and the ParamChem server<sup>68</sup> was used to assign CHARMM36 atom types, as well as to generate bonded parameters (e.g., bonds, angles, and dihedrals). Parameters with high penalty values (e.g., >10) were optimized using the Force Field Toolkit<sup>69</sup> within VMD.<sup>70</sup> System equilibration was performed using openMM 7.5.0<sup>71</sup> and the CHARMM-GUI-defined equilibration protocol. The apo-3CL<sup>Pro</sup> system was constructed analogously.

For the production of MD simulations, three replica simulations with openMM 7.5.0 were run for each equilibrated system. The simulations were run under the NPT ensemble ( $T = 310$  K); particle-mesh-Ewald (PME) was used for electrostatics, switched Lennard-Jones interactions with a cutoff of 10–12 Å; additionally, the Monte Carlo barostat and Langevin Middle Integrator (i.e., Langevin dynamics with LFMiddle discretization) were used. A 4 fs time step was enabled by standard procedure of hydrogen mass repartitioning. Each of the 150 replicas (144 3CL<sup>Pro</sup>-inhibitor replicas + 6 apo-3CL<sup>Pro</sup> replicas) was run for 725 ns.

### Calculation of the Free Energy of Inhibitor Binding to 3CL<sup>Pro</sup> with an MM-GBSA Approach

First, solvent and ions were removed from the all-atom MD simulation trajectories, so that only the protein and inhibitor coordinates remain. On each of these frames, we calculated an approximate expression for the free energy of binding,  $G$  defined as

$$G = E_{\text{MM}} + G_{\text{solv}} \quad (1)$$

where  $E_{\text{MM}}$  is the molecular energy calculated by the molecular mechanics algorithm, and  $G_{\text{solv}}$  is the free energy of solvation. Complete formulation also includes an additional term,  $-TS_{\text{SMM}}$ , accounting for the protein conformational entropy, which is neglected in MM-PB/GBSA implementations.<sup>72</sup> This approximation is justified by the use of the calculated values as relative free energy of binding between each frame rather than the absolute free energy of binding. In addition,  $-TS_{\text{SMM}}$  is typically much smaller than either  $E_{\text{MM}}$  or  $G_{\text{solv}}$ <sup>72,75</sup> and the standard error of this term has been reported to be larger than the standard error of  $E_{\text{MM}}$  or  $G_{\text{solv}}$ <sup>74–76</sup> offering only a minor improvement in the correlation between experimental and calculated free energies.<sup>72,75,77,78</sup>

$G_{\text{solv}}$  in eq 1 can be expressed as

$$G_{\text{solv}} = G_{\text{polar}} + G_{\text{nonpolar}} \quad (2)$$

where  $G_{\text{polar}}$  and  $G_{\text{nonpolar}}$  refer to the electrostatic and the nonpolar contributions to the free energy of solvation, respectively.  $G_{\text{polar}}$  is estimated by solving the generalized Born equation; here, we use CHARMM's generalized Born

using Molecular Volume (GBMV) Solvation Energy and Forces Module, model II<sup>79–82</sup> and express  $G_{\text{polar}}$  as

$$G_{\text{polar}} = k \sum_{i,j} \frac{q_i q_j}{\sqrt{r_{ij}^2 + \alpha_i \alpha_j} \times e^{-r_{ij}^2/8\alpha_i \alpha_j}} \quad (3)$$

where  $k = -166.0(\epsilon_{\text{solute}}^{-1} - \epsilon_{\text{solvent}}^{-1})$ ,  $\epsilon$  refers to the dielectric constant (1 and 80 for the solute and solvent, respectively);  $r_{ij}^2$  is the square of the distance between atoms  $i$  and  $j$ ; the charges on atoms  $i$  and  $j$  are given by  $q_i$  and  $q_j$ , respectively. The Born radii of atoms  $i$  and  $j$ , noted  $\alpha_i$  and  $\alpha_j$ , respectively, are calculated by the GBMV method based on the molecular environment of the atoms in each frame.<sup>83,84</sup>

$G_{\text{nonpolar}}$  is calculated using a semi-empirical method, in which  $G_{\text{nonpolar}}$  is a linear function of SASA<sup>29,30</sup>

$$G_{\text{nonpolar}} = (a \times \text{SASA}) - b \quad (4)$$

with  $a$  representing a coefficient related to the solvent's surface tension, and  $b$  is a fitting parameter related to solute–solvent interaction energy; here, we set  $a = 0.0072$  kcal/Å<sup>2</sup> and  $b = 0$ .<sup>82,85</sup>

Here, we adapted an implementation previously used for protein–protein interactions<sup>73,86,87</sup> to treat protein–inhibitor interactions. Starting from an MD trajectory, the code uses CHARMM modules to calculate  $E_{\text{MM}}$ , the Born radii, and the SASA. With the formalism mentioned above, the free energy of binding for the 3CL<sup>Pro</sup> inhibitor complex is given by

$$\Delta G = G_{\text{complex}} - (G_{\text{NSPS}} + G_{\text{inhibitor}}) \quad (5)$$

where  $G_{\text{complex}}$  is the free energy of the 3CL<sup>Pro</sup> inhibitor complex;  $G_{\text{NSPS}}$  is the free energy of 3CL<sup>Pro</sup> alone; and  $G_{\text{inhibitor}}$  is the free energy of the inhibitor alone. We used a one-trajectory approach, in which each term in eq 5 was calculated using subsets of atoms from the same all-atom simulation of the 3CL<sup>Pro</sup> inhibitor complex; this results in the cancelation of many intramolecular terms in  $G$ , and consequently more stable results overall.<sup>87</sup>

### Residue-Based Decomposition of the Free Energy

The residue-based decomposition of the free energy (on selected frames, see Clustering on the MM-GBSA Free Energy of Binding) was performed as described in refs 86 and 87. Briefly, each individual atom's contribution to  $E_{\text{electrostatic}}$  was defined as

$$E_{\text{electrostatic}}^i = \frac{1}{2} \sum_{j \neq i} \frac{q_i q_j}{r_{ij}} \quad (6)$$

Likewise, each individual atom's contribution to  $E_{\text{vdW}}$  was defined as

$$E_{\text{vdW}}^i = \frac{1}{2} \sum_{j \neq i} \left[ \left( \frac{r_{ij}^{\text{min}}}{r_{ij}} \right)^{12} - 2 \left( \frac{r_{ij}^{\text{min}}}{r_{ij}} \right)^6 \right] \quad (7)$$

The decomposition of  $G_{\text{nonpolar}}$  is given by

$$\Delta G_{\text{nonpolar}}^i = a \times (\text{SASA}_{\text{complex}}^i - \text{SASA}_{\text{NSPS}}^i - \text{SASA}_{\text{inhibitor}}^i) - b \quad (8)$$

where  $\text{SASA}_{\text{complex}}^i$  refers to atom  $i$ 's contribution to the SASA of the entire complex, whereas  $\text{SASA}_{\text{NSPS}}^i$  and  $\text{SASA}_{\text{inhibitor}}^i$  refer

to atom  $i$ 's contribution to the SASA of 3CL<sup>Pto</sup> and the inhibitor, respectively. The decomposition of  $G_{\text{polar}}$  is given by

$$\Delta G_{\text{polar}}^i = k \frac{q_i^2}{\alpha_i} + \frac{1}{2} k \sum_{j \neq i} \frac{q_i q_j}{\sqrt{r_{ij}^2 + \alpha_i \alpha_j} \times e^{-r_{ij}^2/8\alpha_i \alpha_j}} \quad (9)$$

The individual contributions of each atom within each side chain were summed to calculate each side chain's contribution to  $\Delta G$ .

### Clustering on the MM-GBSA Free Energy of Binding

To identify sets of frames with the lowest free energy of binding (which are used in the residue-based energy decomposition), we first calculated the free energy of binding for every simulation frame as described above. All frames were aligned based on 3CL<sup>Pto</sup> backbone atoms, to the frame with the lowest free energy of binding (highest affinity). The RMSD of the inhibitor atoms was calculated for the aligned trajectories using the inhibitor coordinates from the lowest free energy frame as reference. Finally, the subset of frames with inhibitor RMSD  $\leq 4.0$  Å was selected for the subsequent residue-based decomposition of the free energy of binding. This energy decomposition identifies residues that are important for the inhibitor interaction in the energetically most probable pose.

### NbIT Analysis of MD Trajectories

**Definition of Functional Sites for NbIT Analysis.** The following sites on 3CL<sup>Pto</sup> were defined for NbIT analyses. The residues in the PRT binding site at the dimer interface, termed DBS, were defined as those placing an atom within 4 Å from the ligand in at least 40% of frames in the MD trajectories of the complex. The same definition criteria were applied to define the residues surrounding the catalytic residue dyad, C<sub>145</sub>/H<sub>41</sub> in the active site. These residue groups are termed "Apo Occluding Residues (OR\_apo)" and "Holo Occluding Residues (OR\_holo)" when identified from the respective MD simulation trajectories. Additionally, three "decoy" sites that do not overlap with any of the other sites were chosen arbitrarily on both protomers and termed D1\_A-D3\_A and D1\_B-D3\_B, respectively. The residue lists for the sites defined in this manner are

DBS: Protomer A Residue 2–5, 210, 213–214, 282, 291, 295, 296, 299, 300

OR\_apo: Protomer A Residues 27–28, 38–40, 42–44, 143–144, 146–147, 163–164

OR\_holo: Protomer A Residues 27–28, 38–40, 42–43, 52–55, 58, 60, 143–144, 146–147, 163–164. In each protomer, D1–D3 residues include D1 Residues 69–75; D2 Residues 151–157; D3 Residues 235–241. Note that all NbIT analyses were performed on the heavy atoms.

### Calculation of CI

The CI is defined in NbIT<sup>64</sup> from the total correlation (TC) which is the total amount of information shared among all residues in a "coordinated site" S1 (TC(S1)) that is not shared with another "coordinator site" S2. In the definition of TC(S1)

$$\text{TC}(S1) = \sum_i^{S1} H(r_i) - H(S1)$$

The  $H(r_i)$  term is the configurational entropy of residue  $r_i$ , defined and calculated as described together with other detailed procedures in NbIT analyses in the original paper.<sup>49</sup>

Denoting the coordinated site as S1 and the coordinator site as S2, the CI(S1,S2) is defined as

$$\text{CI}(S1, S2) = \text{TC}(S1) - \text{TC}(S1|S2)$$

where TC(S1|S2) is the TC of S1 conditioned by S2

$$\text{TC}(S1|S2) = \sum_i^{S1} (H(r_i + S2) - H(S2)) - (H(S1 + S2) - H(S2))$$

The normalization of CI values with respect to the TC in S1 is defined as

$$\overline{\text{CI}(S1, S2)} = \frac{\text{CI}(S1, S2)}{\text{TC}(S1)} \times 100\%$$

### 3CI<sup>Pto</sup> Alanine Substitutions

Specific residue substitutions were accomplished by inverse PCR. Primers carrying the specific exchanges were used to amplify the entire 3CI<sup>Pto</sup> plasmid (Q5, New England Biolabs). PCR products were treated with FastDigest DpnI (Thermo Fisher) at 37 °C overnight to remove template plasmid and separated by electrophoresis. Correct size amplicons were purified (Zymo Research) and treated with T4 Polynucleotide Kinase and T4 Ligase (New England Biolabs) and transformed into Stellar competent cells (Takada Bio), according to the manufacturer's instructions. All plasmids were sequenced to validate correct substitution.<sup>65</sup>

## ■ QUANTIFICATION AND STATISTICAL ANALYSIS

$N = 3$  independent biological replicates were used for all experiments unless otherwise indicated.  $P$ -values were calculated by unpaired two-tailed Student's  $t$ -test unless otherwise indicated. \* $p < 0.05$ , \*\* $p < 0.005$ .

## ■ ASSOCIATED CONTENT

### Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/acsbiochemau.2c00034>.

Protease inhibitors used in this study; 3CI<sup>Pto</sup> inhibitors identified in the cell reporter chemical screen; sequences of primers used for qRT-PCR in this study; inhibitors' BFE; 3CI<sup>Pto</sup> alanine substitution inhibition analysis; CI of decoy sites; IDX-S' CI; fluorescence images of the 3CL<sup>Pto</sup> cell-reporter with different NSP junction sequences; 3CL<sup>Pto</sup> cell reporter activity quantification with calpain inhibitors; chemical screen lead compounds' IC<sub>50</sub> quantification using an in vitro FRET assay; inhibitors' interactions with the 3CL<sup>Pto</sup> catalytic site; inhibitors' interactions with the 3CL<sup>Pto</sup> dimerization site; and structural definition of 3CL<sup>Pto</sup> arbitrary sites used for comparison (PDF)

## ■ AUTHOR INFORMATION

### Corresponding Authors

Harel Weinstein – Department of Physiology, Biophysics, Weill Cornell Medicine, New York, New York 10065, United States; [orcid.org/0000-0003-3473-9818](https://orcid.org/0000-0003-3473-9818); Email: [haw2002@med.cornell.edu](mailto:haw2002@med.cornell.edu)

**Benjamin R. TenOever** – Department of Microbiology, Icahn School of Medicine at Mount Sinai, New York, New York 10029, United States; Present Address: Department of Microbiology, New York University, New York, NY 10016, USA; Present Address: TWINCORE Centre for Experimental and Clinical Infection Research, Institute for Experimental Virology, 30625 Hannover, Germany.; Email: [Benjamin.tenOever@NYUlangone.org](mailto:Benjamin.tenOever@NYUlangone.org)

**Shuibing Chen** – Department of Surgery, Weill Cornell Medicine, New York, New York 10065, United States; Email: [shc2034@med.cornell.edu](mailto:shc2034@med.cornell.edu)

**Robert E. Schwartz** – Division of Gastroenterology and Hepatology, Department of Medicine, Weill Cornell Medicine, New York, New York 10065, United States; Department of Physiology, Biophysics, Weill Cornell Medicine, New York, New York 10065, United States; Email: [res2025@med.cornell.edu](mailto:res2025@med.cornell.edu)

## Authors

**Yaron Bram** – Division of Gastroenterology and Hepatology, Department of Medicine, Weill Cornell Medicine, New York, New York 10065, United States; [orcid.org/0000-0001-9711-575X](https://orcid.org/0000-0001-9711-575X)

**Xiaohua Duan** – Department of Surgery, Weill Cornell Medicine, New York, New York 10065, United States

**Benjamin E. Nilsson-Payant** – Department of Microbiology, Icahn School of Medicine at Mount Sinai, New York, New York 10029, United States; Present Address: TWINCORE Centre for Experimental and Clinical Infection Research, Institute for Experimental Virology, 30625 Hannover, Germany.

**Vasuretha Chandar** – Division of Gastroenterology and Hepatology, Department of Medicine, Weill Cornell Medicine, New York, New York 10065, United States

**Hao Wu** – Department of Physiology, Biophysics, Weill Cornell Medicine, New York, New York 10065, United States

**Derek Shore** – Department of Physiology, Biophysics, Weill Cornell Medicine, New York, New York 10065, United States

**Alvaro Fajardo** – Division of Gastroenterology and Hepatology, Department of Medicine, Weill Cornell Medicine, New York, New York 10065, United States; [orcid.org/0000-0003-3998-6703](https://orcid.org/0000-0003-3998-6703)

**Saloni Sinha** – Division of Gastroenterology and Hepatology, Department of Medicine, Weill Cornell Medicine, New York, New York 10065, United States

**Nora Hassan** – Division of Gastroenterology and Hepatology, Department of Medicine, Weill Cornell Medicine, New York, New York 10065, United States

Complete contact information is available at:

<https://pubs.acs.org/10.1021/acsbiomedchemau.2c00034>

## Author Contributions

<sup>†</sup>Y.B., X.D., B.E.N., H.W., and D.S. contributed equally. CRediT: **Yaron Bram** conceptualization (lead), data curation (lead), formal analysis (lead), investigation (lead), methodology (lead), project administration (lead), software (lead), supervision (lead), validation (lead), visualization (lead), writing-original draft (lead), writing-review & editing (lead); **Xiaohua Duan** investigation (equal); **Benjamin E. Nilsson-Payant** investigation (equal), validation (equal), writing-review & editing (equal); **Vasuretha Chandar** data curation (supporting), investigation (supporting); **Hao Wu** conceptualization

(equal), formal analysis (equal), investigation (equal), methodology (equal); **Derek M Shore** conceptualization (equal), data curation (equal), investigation (equal), methodology (equal); **Alvaro Fajardo** investigation (supporting); **Saloni Sinha** investigation (supporting), writing-review & editing (supporting); **Nora Hassan** investigation (supporting), writing-review & editing (supporting); **Harel Weinstein** data curation (equal), funding acquisition (equal), investigation (equal), methodology (equal), resources (equal), software (equal), supervision (equal), writing-original draft (equal), writing-review & editing (equal); **Shuibing Chen** funding acquisition (equal), investigation (equal), methodology (equal), resources (equal), supervision (equal); **Robert E. Schwartz** conceptualization (equal), data curation (equal), funding acquisition (equal), investigation (equal), methodology (equal), project administration (equal), resources (equal), supervision (equal), validation (equal), writing-original draft (equal), writing-review & editing (equal).

## Notes

The authors declare the following competing financial interest(s): R.E.S. is on the scientific advisory board of Miromatrix Inc, and a speaker and consultant for Alnylam Pharmaceuticals.

<sup>‡</sup>Lead contact.

## ACKNOWLEDGMENTS

This work was supported by the National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK) (R01DK121072 and 1R03DK117252), NIAID (2R01AI107301), Department of Medicine, Weill Cornell Medicine (R.E.S.), (DP3DK111907, R01DK116075, R01DK119667, and R01DK124463 to S.C.), Department of Surgery, Weill Cornell Medicine (S.C.), the Defense Advanced Research Projects Agency (DARPA-16-35-INTERCEPT-FP-006 to B.R.T.), and NYSTEM (DOH01-STEM5-2016-00300-C32S99GG to S.C.). S.C. and R.E.S. are supported as Irma Hirsch Trust Research Award Scholars. The computational work was performed using the resources, services, and support provided at the RPI Artificial Intelligence Multiprocessing Optimized System (AiMOS) accessed through an award from the COVID-19 HPC Consortium (<https://covid19-hpc-consortium.org/>), which is a unique private-public effort to bring together government, industry, and academic leaders volunteering free compute time and resources in support of COVID-19 research. The authors gratefully acknowledge, as well, the use of COVID-19 High Performance Computing Consortium resources of the Oak Ridge Leadership Computing Facility, which is a DOE Office of Science User Facility supported under contract DE-AC05-00OR22725, and of the in-house computational resources of the David A. Cofrin Center for Biomedical Information in the Institute for Computational Biomedicine at Weill Cornell Medical College.

## REFERENCES

- (1) Cui, J.; Li, F.; Shi, Z. L. Origin and evolution of pathogenic coronaviruses. *Nat. Rev. Microbiol.* **2019**, *17*, 181–192.
- (2) V'kovski, P.; Kratzel, A.; Steiner, S.; Stalder, H.; Thiel, V. Coronavirus biology and replication: implications for SARS-CoV-2. *Nat. Rev. Microbiol.* **2020**, *19*, 155–170.
- (3) Anand, K.; Ziebuhr, J.; Wadhwani, P.; Mesters, J. R.; Hilgenfeld, R. Coronavirus main proteinase (3CLpro) Structure: Basis for design of anti-SARS drugs. *Science* **2003**, *300*, 1763–1767.



- (4) Yang, H.; et al. The crystal structures of severe acute respiratory syndrome virus main protease and its complex with an inhibitor. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 13190–13195.
- (5) Zhang, L.; et al. Crystal structure of SARS-CoV-2 main protease provides a basis for design of improved  $\alpha$ -ketoamide inhibitors. *Science* **2020**, *368*, 409–412.
- (6) Kneller, D. W.; et al. Unusual zwitterionic catalytic site of SARS-CoV-2 main protease revealed by neutron crystallography. *J. Biol. Chem.* **2020**, *295*, 17365–17373.
- (7) Roe, M. K.; Junod, N. A.; Young, A. R.; Beachboard, D. C.; Stobart, C. C. Targeting novel structural and functional features of coronavirus protease nsp5 (3CLpro, Mpro) in the age of COVID-19. *J. Gen. Virol.* **2021**, *102*, 001558.
- (8) Zeuzem, S.; et al. Glecaprevir-Pibrentasvir for 8 or 12 Weeks in HCV Genotype 1 or 3 Infection. *N. Engl. J. Med.* **2018**, *378*, 354–369.
- (9) Hammond, J.; et al. Oral Nirmatrelvir for High-Risk, Nonhospitalized Adults with Covid-19. *N. Engl. J. Med.* **2022**, *386*, 1397–1408.
- (10) Muramatsu, T.; et al. 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.
- (11) Dömling, A.; Gao, L. Chemistry and Biology of SARS-CoV-2. *Chem* **2020**, *6*, 1283–1295.
- (12) Fan, K.; et al. The substrate specificity of SARS coronavirus 3C-like proteinase. *Biochem. Biophys. Res. Commun.* **2005**, *329*, 934–940.
- (13) Grum-Tokars, V.; Ratia, K.; Begaye, A.; Baker, S. C.; Mesecar, A. D. Evaluating the 3C-like protease activity of SARS-Coronavirus: Recommendations for standardized assays for drug discovery. *Virus Res.* **2008**, *133*, 63–73.
- (14) Dragovich, P. S.; et al. Structure-based design, synthesis, and biological evaluation of irreversible human rhinovirus 3C protease inhibitors. 4. Incorporation of P1 lactam moieties as L-glutamine replacements. *J. Med. Chem.* **1999**, *42*, 1213–1224.
- (15) Kim, Y.; et al. Broad-Spectrum Antivirals against 3C or 3C-Like Proteases of Picornaviruses, Noroviruses, and Coronaviruses. *J. Virol.* **2012**, *86*, 11754–11762.
- (16) Ma, C.; et al. Boceprevir, GC-376, and calpain inhibitors II, XII inhibit SARS-CoV-2 viral replication by targeting the viral main protease. *Cell Res.* **2020**, *30*, 678–692.
- (17) Howe, A. Y. M. The Discovery and Development of Boceprevir: A Novel, First-generation Inhibitor of the Hepatitis C Virus NS3/4A Serine Protease. *J. Clin. Transl. Hepatol.* **2013**, *1*, 22–32.
- (18) Jin, Z.; et al. Structural basis for the inhibition of SARS-CoV-2 main protease by antineoplastic drug carmofur. *Nat. Struct. Mol. Biol.* **2020**, *27*, 529–532.
- (19) Sacco, M. D.; et al. Structure and inhibition of the SARS-CoV-2 main protease reveal strategy for developing dual inhibitors against Mpro and cathepsin L. *Sci. Adv.* **2020**, *6*, 50.
- (20) Schneider, M.; et al. Severe Acute Respiratory Syndrome Coronavirus Replication Is Severely Impaired by MG132 due to Proteasome-Independent Inhibition of M-Calpain. *J. Virol.* **2012**, *86*, 10112–10122.
- (21) Simmons, G.; et al. Inhibitors of cathepsin L prevent severe acute respiratory syndrome coronavirus entry. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 11876–11881.
- (22) DeBiasi, R. L.; Edelstein, C. L.; Sherry, B.; Tyler, K. L. Calpain Inhibition Protects against Virus-Induced Apoptotic Myocardial Injury. *J. Virol.* **2001**, *75*, 351–361.
- (23) Ji, J.; Su, L.; Liu, Z. Critical role of calpain in inflammation. *Biomed. Rep.* **2016**, *5*, 647–652.
- (24) Potz, B. A.; Abid, M. R.; Sellke, F. W. Role of Calpain in Pathogenesis of Human Disease Processes. *J. Nat. Sci.* **2016**, *2*, 9.
- (25) Chen, C. N.; et al. Inhibition of SARS-CoV 3C-like protease activity by theaflavin-3,3'-digallate (TF3). Evidence-based Complement. *Altern. Med.* **2005**, *2*, 209–215.
- (26) Nguyen, T. T. H.; et al. Flavonoid-mediated inhibition of SARS coronavirus 3C-like protease expressed in *Pichia pastoris*. *Biotechnol. Lett.* **2012**, *34*, 831–838.
- (27) Das, S.; Sarmah, S.; Lyndem, S.; Singha Roy, A. An investigation into the identification of potential inhibitors of SARS-CoV-2 main protease using molecular docking study. *J. Biomol. Struct. Dyn.* **2021**, *39*, 3347–3357.
- (28) Ghosh, R.; Chakraborty, A.; Biswas, A.; Chowdhuri, S. Evaluation of green tea polyphenols as novel corona virus (SARS CoV-2) main protease (Mpro) inhibitors - an in silico docking and molecular dynamics simulation study. *J. Biomol. Struct. Dyn.* **2020**, *39*, 4362–4374.
- (29) Touret, F.; Gilles, M.; Barral, K.; Nougairède, A.; Van Helden, J.; Decrolyde Lamballerie, E. X.; Coutard, B. In Vitro Screening of a FDA Approved Chemical Library Reveals Potential Inhibitors of SARS-CoV-2 Replication. *Sci. Rep.* **2020**, *10*, 13093.
- (30) Kim, B. K.; et al. Benserazide, the first allosteric inhibitor of Coxsackievirus B3 3C protease. *FEBS Lett.* **2015**, *589*, 1795–1801.
- (31) Ko, M.; et al. Screening of FDA-approved drugs using a MERS-CoV clinical isolate from South Korea identifies potential therapeutic options for COVID-19. *Viruses* **2021**, *13*, 651.
- (32) Fu, L.; et al. Both Boceprevir and GC376 efficaciously inhibit SARS-CoV-2 by targeting its main protease. *Nat. Commun.* **2020**, *11*, 4417.
- (33) Behnem, M. A. M. Inhibitor potency and assay conditions: A case study on SARS-CoV-2 main protease. *Proc. Natl. Acad. Sci. U.S.A.* **2021**, *118*, No. e2106095118.
- (34) Dai, W.; et al. Structure-based design of antiviral drug candidates targeting the SARS-CoV-2 main protease. *Science* **2020**, *368*, 1331–1335.
- (35) Jin, Z.; et al. Structure of Mpro from SARS-CoV-2 and discovery of its inhibitors. *Nature* **2020**, *582*, 289–293.
- (36) Qiao, J.; et al. SARS-CoV-2 M pro inhibitors with antiviral activity in a transgenic mouse model. *Science* **2021**, *371*, 1374–1378.
- (37) Drayman, N.; et al. Masitinib is a broad coronavirus 3CL inhibitor that blocks replication of SARS-CoV-2. *Science* **2021**, *373*, 931–936.
- (38) Liang, J.; et al. Site mapping and small molecule blind docking reveal a possible target site on the SARS-CoV-2 main protease dimer interface. *Comput. Biol.* **2020**, *89*, 107372.
- (39) Menéndez, C. A.; Byléhn, F.; Perez-Lemus, G. R.; Alvarado, W.; Pablo, J. J. de. Molecular characterization of ebsele binding activity to SARS-CoV-2 main protease. *Sci. Adv.* **2020**, *6*, No. eabd0345.
- (40) Günther, S.; et al. X-ray screening identifies active site and allosteric inhibitors of SARS-CoV-2 main protease. *Science* **2021**, *372*, 642–646.
- (41) Jiménez-Avalos, G.; et al. Comprehensive virtual screening of 4.8 k flavonoids reveals novel insights into allosteric inhibition of SARS-CoV-2 MPRO. *Sci. Rep.* **2021**, *11*, 15452.
- (42) Douangamath, A.; et al. Crystallographic and electrophilic fragment screening of the SARS-CoV-2 main protease. *Nat. Commun.* **2020**, *11*, 5047.
- (43) El-Baba, T. J.; et al. Allosteric Inhibition of the SARS-CoV-2 Main Protease: Insights from Mass Spectrometry Based Assays. *Angew. Chem., Int. Ed.* **2020**, *59*, 23544–23548.
- (44) Sencanski, M.; et al. Drug Repurposing for Candidate SARS-CoV-2 Main Protease Inhibitors by a Novel In Silico Method. *Molecules* **2020**, *25*, 3830.
- (45) Dubanevics, I.; McLeish, T. C. B. Computational analysis of dynamic allostery and control in the SARS-CoV-2 main protease. *J. R. Soc., Interface* **2021**, *18*, 20200591.
- (46) Sztain, T.; Amaro, R.; McCammon, J. A. Elucidation of Cryptic and Allosteric Pockets within the SARS-CoV-2 Main Protease. *J. Chem. Inf. Model.* **2021**, *61*, 3495–3501.
- (47) Carli, M.; Sormani, G.; Rodriguez, A.; Laio, A. Candidate Binding Sites for Allosteric Inhibition of the SARS-CoV-2 Main Protease from the Analysis of Large-Scale Molecular Dynamics Simulations. *J. Phys. Chem. Lett.* **2020**, *12*, 65–72.
- (48) Davis, D. A.; et al. Regulation of the Dimerization and Activity of SARS-CoV-2 Main Protease through Reversible Glutathionylation of Cysteine 300. *mBio* **2021**, *12*, No. e0209421.

- (49) LeVine, M. V.; Weinstein, H. NbIT - A New Information Theory-Based Analysis of Allosteric Mechanisms Reveals Residues that Underlie Function in the Leucine Transporter LeuT. *PLoS Comput. Biol.* **2014**, *10*, No. e1003603.
- (50) Pathak, N.; et al. Uncovering Flexible Active Site Conformations of SARS-CoV-2 3CL Proteases through Protease Pharmacophore Clusters and COVID-19 Drug Repurposing. *ACS Nano* **2021**, *15*, 857–872.
- (51) Ullrich, S.; Nitsche, C. The SARS-CoV-2 main protease as drug target. *Bioorg. Med. Chem. Lett.* **2020**, *30*, 127377.
- (52) Froggatt, H. M.; Heaton, B. E.; Heaton, N. S. Development of a Fluorescence-Based, High-Throughput SARS-CoV-2 3CLpro Reporter Assay. *J. Virol.* **2020**, *94*, 22.
- (53) O'Brien, A.; et al. Detecting SARS-CoV-2 3CLpro expression and activity using a polyclonal antiserum and a luciferase-based biosensor. *Virology* **2021**, *556*, 73.
- (54) Chen, K. Y.; et al. A highly sensitive cell-based luciferase assay for high-throughput automated screening of SARS-CoV-2 nsp5/3CLpro inhibitors. *Antiviral Res.* **2022**, *201*, 105272.
- (55) Ma, C.; Tan, H.; Choza, J.; Wang, Y.; Wang, J. Validation and invalidation of SARS-CoV-2 main protease inhibitors using the Flip-GFP and Protease-Glo luciferase assays. *Acta Pharm. Sin. B* **2022**, *12*, 1636–1651.
- (56) Li, X.; et al. Ethacridine inhibits SARS-CoV-2 by inactivating viral particles. *PLoS Pathog.* **2021**, *17*, No. e1009898.
- (57) Blanco-Melo, D.; et al. Imbalanced Host Response to SARS-CoV-2 Drives Development of COVID-19. *Cell* **2020**, *181*, 1036–1045.
- (58) Kim, S.; et al. PubChem in 2021: new data content and improved web interfaces. *Nucleic Acids Res.* **2021**, *49*, D1388–D1395.
- (59) Harder, E.; et al. OPLS3: A Force Field Providing Broad Coverage of Drug-like Small Molecules and Proteins. *J. Chem. Theory Comput.* **2016**, *12*, 281–296.
- (60) Sencanski, M.; et al. Drug repurposing for candidate SARS-CoV-2 main protease inhibitors by a novel in silico method. *Molecules* **2020**, *25*, 3830.
- (61) Menéndez, C. A.; Byléhn, F.; Perez-Lemus, G. R.; Alvarado, W.; de Pablo, J. J. Molecular characterization of ebselen binding activity to SARS-CoV-2 main protease. *Sci. Adv.* **2020**, *6*, No. eabd0345.
- (62) Jian, J. W.; et al. Predicting Ligand Binding Sites on Protein Surfaces by 3-Dimensional Probability Density Distributions of Interacting Atoms. *PLoS One* **2016**, *11*, No. e0160315.
- (63) Kozakov, D.; et al. The FTMap family of web servers for determining and characterizing ligand-binding hot spots of proteins. *Nat. Protoc.* **2015**, *10*, 733–755.
- (64) LeVine, M.; Perez-Aguilar, J. M.; Weinstein, H. N-Body Information Theory (NbIT) Analysis of Rigid-Body Dynamics in Intracellular Loop 2 of the 5-HT<sub>2A</sub> Receptor. *Proceedings Iwbbio—International Work-Conference on Bioinformatics and Biomedical Engineering*, 2014; Vol. 1–2, pp 1190–1201.
- (65) 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.
- (66) Jo, S.; Kim, T.; Iyer, V. G.; Im, W. CHARMM-GUI: a web-based graphical user interface for CHARMM. *J. Comput. Chem.* **2008**, *29*, 1859–1865.
- (67) Vanommeslaeghe, K.; Raman, E. P.; MacKerell, A. D., Jr. Automation of the CHARMM General Force Field (CGenFF) II: assignment of bonded parameters and partial atomic charges. *J. Chem. Inf. Model.* **2012**, *52*, 3155–3168.
- (68) Mayne, C. G.; Saam, J.; Schulten, K.; Tajkhorshid, E.; Gumbart, J. C. Rapid Parameterization of Small Molecules Using the Force Field Toolkit. *J. Comput. Chem.* **2013**, *34*, 2757–2770.
- (69) Humphrey, W.; Dalke, A.; Schulten, K. VMD: visual molecular dynamics. *J. Mol. Graphics* **1996**, *14*, 33–38.
- (70) Eastman, P.; et al. OpenMM 7: Rapid development of high performance algorithms for molecular dynamics. *PLoS Comput. Biol.* **2017**, *13*, No. e1005659.
- (71) Kumari, R.; Kumar, R.; Lynn, A.; Open Source Drug Discovery Consortium. g\_mmpbsa-A GROMACS Tool for High-Throughput MM-PBSA Calculations. *J. Chem. Inf. Model.* **2014**, *54*, 1951–1962.
- (72) Kollman, P. A.; et al. Calculating structures and free energies of complex molecules: Combining molecular mechanics and continuum models. *Acc. Chem. Res.* **2000**, *33*, 889–897.
- (73) Homeyer, N.; Gohlke, H. Free Energy Calculations by the Molecular Mechanics Poisson–Boltzmann Surface Area Method. *Mol. Inf.* **2012**, *31*, 114–122.
- (74) Yang, T. Y.; et al. Virtual screening using molecular simulations. *Proteins* **2011**, *79*, 1940–1951.
- (75) Spiliotopoulos, D.; Spitaleri, A.; Musco, G. Exploring PHD Fingers and H3K4me0 Interactions with Molecular Dynamics Simulations and Binding Free Energy Calculations: AIRE-PHD1, a Comparative Study. *PLoS One* **2012**, *7*, No. e46902.
- (76) Rastelli, G.; Del Rio, A.; Degliesposti, G.; Sgobba, M. Fast and Accurate Predictions of Binding Free Energies Using MM-PBSA and MM-GBSA. *J. Comput. Chem.* **2010**, *31*, 797–810.
- (77) Genheden, S.; Ryde, U. The MM/PBSA and MM/GBSA methods to estimate ligand-binding affinities. *Expert Opin. Drug Discovery* **2015**, *10*, 449–461.
- (78) Brown, S. P.; Muchmore, S. W. Corrections to Large-Scale Application of High-Throughput Molecular Mechanics with Poisson–Boltzmann Surface Area for Routine Physics-Based Scoring of Protein–Ligand Complexes. *J. Med. Chem.* **2009**, *52*, 4549.
- (79) Brooks, B. R.; et al. CHARMM: The biomolecular simulation program. *J. Comput. Chem.* **2009**, *30*, 1545–1614.
- (80) Lee, M. S.; Feig, M.; Salsbury, F. R.; Brooks, C. L. New analytic approximation to the standard molecular volume definition and its application to generalized born calculations. *J. Comput. Chem.* **2003**, *24*, 1821.
- (81) Lee, M. S.; Salsbury, F. R.; Brooks, C. L. Novel generalized Born methods. *J. Chem. Phys.* **2002**, *116*, 10606–10614.
- (82) Still, W. C.; Tempczyk, A.; Hawley, R. C.; Hendrickson, T. Semianalytical Treatment of Solvation for Molecular Mechanics and Dynamics. *J. Am. Chem. Soc.* **1990**, *112*, 6127–6129.
- (83) Hermann, R. B. Theory of Hydrophobic Bonding .2. Correlation of Hydrocarbon Solubility in Water with Solvent Cavity Surface-Area. *J. Phys. Chem.* **1972**, *76*, 2754–2759.
- (84) Amidon, G. L.; Yalkowsky, S. H.; Anik, S. T.; Valvani, S. C. Solubility of Nonelectrolytes in Polar-Solvents .5. Estimation of Solubility of Aliphatic Monofunctional Compounds in Water Using a Molecular Surface-Area Approach. *J. Phys. Chem.* **1975**, *79*, 2239–2246.
- (85) Jayaram, B.; Sprous, D.; Beveridge, D. L. Solvation free energy of biomacromolecules: Parameters for a modified generalized born model consistent with the AMBER force field. *J. Phys. Chem. B* **1998**, *102*, 9571–9576.
- (86) Zoete, V.; Irving, M. B.; Michielin, O. MM-GBSA binding free energy decomposition and T cell receptor engineering. *J. Mol. Recognit.* **2010**, *23*, 142–152.
- (87) Zoete, V.; Meuwly, M.; Karplus, M. Study of the insulin dimerization: Binding free energy calculations and per-residue free energy decomposition. *Proteins* **2005**, *61*, 79–93.