Crystal-Structures-Guided Design of Fragment-Based Drugs for Inhibiting the Main Protease of SARS-CoV-2

Binquan Luan* and Tien Huynh

Computational Biological Center, IBM Thomas J. Watson Research, Yorktown Heights, New York 10598, USA

E-mail: bluan@us.ibm.com

Abstract

Since the beginning of the COVID-19 pandemic, researchers and scientists across the globe are racing to find a cure for the highly contagious infectious disease caused by the SARS-CoV-2 virus. Despite many promising ongoing progress, there are currently no FDA approved drugs to treat infected patients. Among the various protein targets of SARS-CoV-2 virus, the main protease (Mpro) has attracted most interests. Recently, the crowdsourcing of drug discovery for inhibiting Mpro have yielded a plenty of drug fragments resolved inside the active site of Mpro via the crystallography method. Following the principle of fragment-based drug design (FBDD), we are motivated to design a potent drug molecule through merging several of these newly discovered drug fragments. Among various designed ligands, we found that B19 by merging three fragments JFM, U0P and HWH is the most stable one, evidenced through extensive (∼10 µs totally) all-atom molecular dynamics simulation. We further estimated that the binding free energy of B19 is comparable or even a little better than that of a native protein ligand processed by Mpro. Our promising results suggest that B19 can potentially be an efficacious drug molecule for inhibiting Mpro of SARS-CoV-2.
Introduction

The ongoing coronavirus 2019 (COVID-19) pandemic is caused by an enveloped, positive-sense, single stranded RNA virus, that is known as the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2).\(^1\) SARS-CoV-2 was first identified in China and is the third pathogenic novel coronavirus to emerge over the last two decades. The first one was discovered in 2003 named SARS-CoV-1 which caused the severe acute respiratory syndrome (SARS), a serious and atypical pneumonia.\(^2\) The second one emerged a decade later and was named MERS-CoV which caused a similar respiratory ailment called Middle East respiratory syndrome (MERS).\(^3\) These three coronaviruses are all zoonotic pathogens which begin in infected animals and are transmitted from animals to humans. Compared with the other two, SARS-CoV-2 has significant person-to-person transmission. Indeed, it took just a few months for COVID-19 to spread to almost every corner of the world, resulting in detrimental effect on global healthcare systems with a ripple effect on every aspect of human life. To combat this disease, dozens of drugs have been tested in clinical as possible treatments for SARS-CoV-2 infections. These drugs fall into two general groups: repurposed antiviral drugs (\textit{e.g.} remdesivir\(^4\)) to help stop the virus from reproducing, and immune-based therapies (\textit{e.g.} dexamethasone\(^5\)) to calm the “cytokine storm” from overreaction of the immune system which can lead to organ failure and death. However, at this time there are still no FDA-approved drugs to cure or prevent COVID-19 except treatments to ease the symptoms from a supportive care perspective.

Alternatively, new antiviral drugs have been explored simultaneously, targeting mainly at the main protease (Mpro) of SARS-CoV-2 that is highly conserved\(^6\) when compared with Mpros of SARS-CoV-1 and MERS-CoV. It is well known that Mpro processes many polyproteins translated from the viral RNA, such as the RNA-dependent RNA polymerase (RdRp, Nsp12) and the helicase (Nsp13), playing an essential role in viral maturation. Targeting at the key residue cysteine 145 (C145) in the catalytic site of Mpro, many mechanism-based ligands such as N3,\(^7\) O6K\(^8\) and carmofur\(^9\) were designed to irreversibly (or covalently)
bind C145, which yields inhibited or nonfunctional Mpro. Additionally, dozens of fragment drugs were shown to be covallently bound to C145 inside Mpro, as listed on the website of the COVID Moonshot open initiative (https://postera.ai/covid), a project aiming to accelerate the development of antiviral drugs that can inhibit the SARS-CoV-2 main protease with joined forces from scientists around the world.\(^\text{10}\) Besides thecovallently bound ligands, dozens of noncovallently bound drug fragments that span the entire active site of the SARS-CoV-2 Mpro have also been identified using large crystallographic and mass-spectrometry screening.\(^\text{11}\) These structures (listed on COVID Moonshot website and deposited in Protein Data Bank) reveal an exceptionally rich set of information, with extensive opportunities for fragment-based drug discovery (FBDD) for inhibiting the SARS-CoV-2 Mpro.

Generally, FBDD provides an effective alternative to explore the chemical space for binding to a specific target protein. Since its inception two decades ago, FBDD has attracted a lot of interest from researchers and emerged as a mainstream approach in the pharmaceutical industry for reducing attrition and finding high-quality lead candidates as a part of the drug discovery process.\(^\text{12}\) FBDD is based on screening very small chemical compounds (fragments) with low molecular weight which could bind only weakly to the biological target as starting points. The binding affinity between the initial fragment hits to their target is usually in a \(\mu\text{M–mM}\) range.\(^\text{13}\) This is due to the fact that there are fewer heavy atoms in the fragments to form multiple attractive interactions with the protein when compared to larger molecules.\(^\text{14}\) Once the fragment hits are identified, their potency can be enhanced by applying fragment linking, merging or growing strategies to form drug-like leads optimized for potential clinical candidates.

Given a plethora of drug fragments resolved in the Mpro’s crystal structure\(^\text{11}\) and the established protocol for FBDD, in this work, we are motivated to design a potent drug molecule structurally merged from several of these drug fragments for inhibiting Mpro. Complimentary to experimental efforts, \textit{in silico} approaches such as the all-atom molecular dynamics (MD)\(^\text{15,16}\) and docking\(^\text{17}\) methods have been widely used to investigate the molecular mech-
anism of proteins as well as the ligand-protein interaction, and proven to produce results consistent with experimental ones. Here, we carried out MD simulations to explore a total of 19 designed ligands based on the FBDD approach and compare them with a native ligand (a protein peptide) previously resolved in the crystal structure of Mpro of SARS-CoV-1. Out of these 19 designed ligands, we found that the binding affinity of a designed drug molecule (named B19) to Mpro is comparable or even slightly better than that of the native ligand, suggesting that B19 is highly promising for inhibiting Mpro of SARS-CoV-2.

Results

Among the 50 known non-covalently bound fragments inside the active site, we selected three ligands JFM, U0P and HWH (Fig. 1a) that bind the active site at different locations (Figs. 1b-1d) and contain their own specific features as described in the following. Namely, from head to tail, JFM contains a methanesulfonamide and a benzene groups; U0P contains a cyclopropane and a carbonyl groups in the head part and an isoxazole group in the tail part; HWH contains an indole group with an attached fluorine atom (pink, Fig. 1a) in the head part and an acetamide in the tail part. These chemical groups form either hydrogen bonds or hydrophobic interactions with Mpro’s residues inside the active site, stabilizing these bound fragments.

Following previous studies, the active site of Mpro is composed of four key subsites that are respectively labeled as S1’, S1, S2, and S4 (Fig. 1b). As shown in Figs. 1b-1d, the fragment U0P binds the S1 and S1’ subsites, and the fragments JFM and HWH bind the S2 and S4 subsites respectively. Interestingly, a co-crystallization agent dimethyl sulfoxide (DMSO) is also present inside the active site (S1) along with HWH as shown in Fig. 1d. Due to their proximity (Fig. 1e), with an added linker molecule we combined JFM, U0P and HWH into one drug molecule (named B19 with the molecular formula C_{26}H_{30}N_{7}O_{6}SF ), occupying all four subsites. B19 has a molecular weight of 587.6, which is comparable to
Figure 1: Scheme of fragment-based drug design of B19. a) Three selected drug-fragments: JFM, U0P and HWH. For each of three fragment drugs, the head and tail parts are on the left and right sides, respectively. b) The crystal structure of JFM in the ligand-binding pocket of Mpro (PDB ID: 5R7Y). c) The crystal structure of U0P in the ligand-binding pocket of Mpro (PDB ID: 5RGI). d) The crystal structure of HWH in the ligand-binding pocket of Mpro (PDB ID: 5R7Z). A DMSO molecule was co-crystallized inside the subsite S1. e) The assembly of JFM, U0P and HWH into B19 with a linker (shaded).
O6K (595.7) and less than N3 (680.8). Totally, we designed 19 drug molecules using various linker molecules, with B19 shown in Fig. 1e being the most stable one and the rest (from B1 to B18) shown in Fig. S1 (Supporting Information). In our design of B19, we first merged JFM and HWH together because their benzene rings overlap with each other. After that, we employed the linker molecule shown in Fig. 1e to bridge the merged compound and U0P. In B19, the carbon atom attached to the benzene ring (originally in JFM) was replaced by a nitrogen atom, which we found can form a stable three-way junction, maintaining the overall conformation of B19 inside the active site of Mpro (see below). Importantly, the catalytic dyad comprising H41 and C145 is completely covered by the designed B19 molecule.

Note that the Mpro’s active site is formed by two protein domains, residues 1-100 (I) and residues 101-200 (II). The subsites S1, S2 are respectively in domains II and I, while the subsites S1’ and S4 are at the interface of domains I and II. Therefore, the active site of Mpro is not rigid and can deform slightly to harbor a ligand (such as B19) bridging two Mpro domains. Generally, the host’s flexibility is amenable to the FBDD method, i.e. capable to accommodate a merged complex containing slightly displaced fragments.

We used all-atom molecular dynamics (MD) simulation to verify the stability of each designed drug molecule (from B1 to B19) bound inside the active site. Figure 2a illustrates the simulation system: a Mpro dimer with B19 bound inside the active site of one monomer was solvated in a 0.15 M electrolyte. Details are provided in the Methods section. We carried out two independent MD simulations, each of which lasted about 190 ns. Figures 2b and 2c illustrate a representative conformation of B19 in Mpro shown in the molecular surface and cartoon representations, respectively.

Overall, during the 190 ns simulation, B19 maintained its initial conformation in Mpro, with slight local adjustments. For example, compared with its initial pose (Figs. 1b and 1e) the SO2 group turned itself away from the Mpro surface and faced the water environment, because of its hydrophilicity. Meanwhile, the hydrophobic methyl (CH3) group changed its initial unfavorable pose facing water to the one contacting the hydrophobic M49 and T25 on
Figure 2: MD simulation of B19 in Mpro. a) Illustration of the MD system. The two monomers (grey and purple) in Mpro are in the cartoon representation. B19 is in the van der Waals sphere representation. Water is shown transparently. K$^+$ and Cl$^-$ ions are colored in tan and cyan, respectively. b, c) The equilibrated conformation of B19 in the active site of Mpro. Mpro is in the surface representation (b) and in the cartoon representation (c).
the Mpro surface. Additionally, the isoxazole group (present in the S1 subsite) could turn 180° from time to time. Thus, different from the one shown in Fig. 2b and Fig. 2c, the N and O atoms in the five-membered isoxazole ring could also be in contact with the Mpro surface (Fig. 1c and Movie B19.mpg in the Supporting Information).

From the simulation trajectories, we calculated the root mean square deviations (RMSD) for backbone atoms in the monomer (of Mpro) that harbors the B19. Figure 3a shows that after about 75 ns RMSDs from two independent MD simulations saturate around 1.7 Å, suggesting that the Mpro’s crystal structure was properly equilibrated in the physiology-like (a 0.15 M electrolyte) environment. The saturated RMSDs for B19 inside the active site of Mpro are about 1.2 Å, corroborating the stable conformation of B19 in the active site as shown in the Movie B19.mpg (Supporting Information).

Energetically, we calculated the pair-wise interaction energy between B19 and Mpro, including the van der Waals (vdW, Fig. 3c) and electrostatic (Fig. 3d) ones. Over the entire 190 ns, both the vdW and electrostatic potential energies are nearly constant for both simulations (Sim-1 and Sim-2), with the electrostatic ones (Fig. 3d) showing larger fluctuations partly due to the omission of interfacial water molecules in these calculations. In comparison, results for Sim-2 are slightly better (more negative) than those for Sim-1, with the vdW energy about -68.3 kcal/mol and the electrostatic one about -12.3 kcal/mol. Besides demonstrating the binding stability of B19 in the active site, these results will be further used in the MM/GBSA calculations to obtain the binding free energy of B19 (see below).

In the other 18 designs (B1 to B18) as shown in Fig. S1 (Supporting Information), these drug molecules with unsuccessfully designed linker molecules moved more or less away from their initial positions during MD simulations, i.e. at lease one fragment in each designed drug molecule either rotated or shifted away from the one as seen in the crystal environment. For example, observed in two independent MD simulations, the U0P fragment in B15 was not stable and moved out of the S1 and S1’ subsites, however the other two fragments of
Figure 3: Dynamics and energetics of B19 inside the Mpro’s pocket. a) Time-dependent RMSDs of the backbone atoms in the protein monomer that harbors B19. b) Time-dependent RMSDs of B19. c) The van der Waals interaction energies (between B19 and Mpro) vs. time. d) The electrostatic interaction energies (between B19 and Mpro) vs. time. The dielectric constant used in electrostatic interaction is 4.
B15 remained inside the subsites S2 and S4 respectively, indicating the problematic design of the link molecule in B15 (Fig. S1 in Supporting Information).

In Fig. 4, we highlight key coordinations between B19 and Mpro to unveil the molecular mechanism of the binding. In Fig. 4a, the carbonyl group in the original drug fragment U0P formed two hydrogen bonds with the backbones (N-H groups) of residues G143 and C145, which pins the entire U0P fragment in the subsites S1 and S1' (Fig. 2b). Remarkably, this same local coordination can also be found in the complex of Mpro and O6K (PDB code: 6LU7). Furthermore, through hydrophobic interactions the cyclopropane group (originally in U0P) was stabilized by hydrophobic interactions with residues C145, L27 and T25 of Mpro.

Figure 4a also shows the hydrophobic interaction between T25 and the methyl group in the methanesulfonamide group (originally in JFM). Notably, in Fig. 4b the N-H group in the methanesulfonamide group (in the subsite S2) formed a hydrogen bond with the residue H41 at the bottom of the active site. Furthermore, as shown in Fig. 4c, the methyl group in the methanesulfonamide group also interacts hydrophobically with the residue M49. Overall, the residues T25 (Fig. 4a), H41 (Fig. 4b) and M49 (Fig. 4c) stabilize the methanesulfonamide group in the subsite S2.

Illustrated in Fig. 4c, the indole group (originally in HWH) insert itself deeply into the gap between the two protein domains I and II. Namely, the indole group interacts hydrophobically with the residue M49 in the protein domain I (of Mpro) on one side and interacts also hydrophobically with the residue M165 in the protein doamin II (of Mpro) on the other side. In addition, the acetamide group (originally in the tail part of HWH) can form a hydrogen bond with the residue Q192 and also have the hydrophobic interaction with the residue P168 (Fig. 4c).

The isoxazole group (originally in U0P) occupies the subsite S1, a pocket formed by residues H163, N142, L141, F140, E166 as well as the residue S1 in the N-terminal of the other protein monomer (purple, Fig. 4d). There is no hydrogen bond formed in this local
interaction, which is consistent with the observation that the isoxazole group can rotate from time to time (see Movie B19.mpg). Taking all together, the isoxazole group and the methanesulfonamide group are respectively hosted by the domains II and I, while the cyclo-propane group and the indole group lie between the domains I and II (see Figs. 2b and 2c). Thus, fitting nicely inside the active site, B19 appears to act like a glue bonding the two protein domains together.

Figure 4: Illustrations of local interactions between B19 and Mpro. a) The head part of the U0P fragment. b) The head part of the JFM fragment. c) The HWH fragment. d) The tail part of the U0P fragment.

One restriction of the above MD simulation is that we cannot determine whether the discovered binding pose of B19 inside the active site is correspondent with the lowest energy state (or the best pose), due to the limited simulation time. Here, we resort to the docking method to explore the ligand-receptor conformation space for all plausible poses of B19
as well as their binding affinities. Among many docking tools, Autodock Vina\textsuperscript{20} has been used as a powerful one for \textit{in silico} drug design as it can help understand the structural determinants for protein-ligand complex at the atomic level which is crucial for designing ligands with high specificity and affinity to a target protein. The detailed docking protocol can be found in the Methods section.

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{docking_results.png}
\caption{The docking study of top 200 poses of B19 inside the pocket of Mpro. The insets illustrate four different docking poses (1st, 20th, 40th and 60th) of B19.}
\end{figure}

Figure 5 shows the docking results for the top 200 poses. Remarkable, from a total of more than 700 docked poses of B19 inside the active site, the best one (Fig. 5 \textit{inset}) indeed agrees with the one that discovered in MD simulations (Figs. 2b and 2c). The binding affinity for the best pose of B19 is -10.1 kcal/mol, which is significantly better than many repurposed drug molecules (typically \textgreater -7.0 kcal/mol), N3 (-7.1 kcal/mol) and O6K (-7.4 kcal/mol) without their covalent bonds with C145.\textsuperscript{21} In \textit{insets} of Fig. 5, we also show several other binding poses ranked as the 20th (-8.8 kcal/mol), 40th (-8.4 kcal/mol) and 60th (-8.2 kcal/mol).
kcal/mol) ones. Therefore, the docking study reasonably confirmed that the B19’s pose from MD simulation is the best one, residing in the global minimum on the binding free energy surface.

Next, we compare the binding affinities between B19 and a native ligand. So far, there is no crystal structure for Mpro of SARS-CoV-2 bound with a native ligand. However, due to highly conserved Mpro structures among all coronavirus, we refer to the crystal structure (PDB code: 2Q6G) of SARS-CoV-1’s Mpro bound with a native ligand (peptide sequence: TSAVLQSGFRK). Note that residues in the active site are identical for both Mpros in SARS-CoV-1 and SARS-CoV-2. Thus, we carried out two independent MD simulations (Sim-3 and Sim-4) of Mpro with the bound native ligand whose binding pose is same as that in the crystal structure (PDB code: 2Q6G), as shown in Fig. S2a in Supporting Information. After about 190 ns, the equilibrated structure of the native ligand is shown in Figs. 6a and 6b. The RMSD of protein backbones saturated around 2.2 and 1.7 Å for Sim-3 and Sim-4 respectively (Fig. S2b). The difference in RMSDs resulted from the different thermal fluctuations of some disordered cords in the secondary structure of Mpro. The saturated RMSDs for the native ligand are about 1.5 Å (Fig. S2c), which is slightly larger than the one for B19 due to a few flexible resides outside the active site. Overall, the native ligand is stable inside the active site, which is also demonstrated in Movie native.mpg (Supporting Information).

As shown in Fig. 6a, the native ligand is stabilized by forming two antiparallel beta sheets with residues (e.g., T26) in the protein domain I and with residues (e.g., E166) in the protein domain II. Inside the active site, Q6 (of the native ligand) occupies the S1 subsite; S7 and G8 are in the subsite S1’; A3 and S2 are in the subsite S4; and L5 occupies the subsite S2. The side chain of V4 is oriented toward outside the active site (Fig. 6b). The other residues T1, F9, R10 and K11 are all outside of the active site.

Figure 6c highlights detailed interactions between the native ligand and Mpro. We found that the C=O group in the backbone of Q6, similar to that of B19 as shown in Fig. 4a, also
Figure 6: MD simulation of the native ligand inside the active site of Mpro. a) Illustration of the complex in the cartoon representation. The native ligand is colored in cyan. b) Illustration of the complex, with residues 2 to 8 of the native ligand in the stick representation and Mpro in the molecular surface representation. c) Illustration of essential interactions between the native ligand and Mpro. The Mpro residues in contact with the native ligand are in the van der Waals sphere representation.
formed two hydrogen bonds with G143 and C145. It is worth noting that Q6 is very specific to the subsite S1, as shown in many other native protein strands for Mpro of SARS-CoV-1.\textsuperscript{22} However, the side chain of Q6 (also the isoxazole group in B19) neither forms any hydrogen bond nor has any hydrophobic interactions with the nearby Mpro residues. Presumably, the orientation of the Q6’s side chain and the isoxazole group in B19 inside the subsite S1 enables the neighboring C=O group to form strong hydrogen bonds with G143 and C145. Notably, S7 forms a hydrogen bond with H41. Additionally, G8 forms a hydrogen bond with T26 in Mpro \textit{via} their backbones, forming antiparallel $\beta$-sheets. All these hydrogen bonds tightly positioned the native ligand in the front of the catalytic center C145, illustrating the molecular mechanism of catalytic processing of the bound peptide.

The binding of L5 inside the S2 subsite is also highly conserved among all native protein strands in Mpro of SARS-CoV-1.\textsuperscript{22} As shown in Fig. 6c, L5 is sandwiched between two hydrophobic residues M49 and M165 in Mpro. Although the side chain of V4 is outside the active site, its backbone forms two hydrogen bonds with backbone atoms in E166 of Mpro, forming antiparallel $\beta$-sheets locally. The side chain of A3 of the native ligand interacts hydrophobically with L167 of Mpro, and the oxygen atom in the A3’s backbone forms a hydrogen bond with the NH2 group in Q189. Overall, S2, A3, V4 and Q6 mainly bind the protein domain II, while L5, S7 and G8 mainly bind the protein domain I (of Mpro), demonstrating a sophisticated binding mechanism of the native ligand with Mpro.

To compare the binding affinities between B19 and the native ligand, we utilized the MM/GBSA method that is generally applied to estimate the ligand-protein binding free energy. The detailed description of the calculation procedures are summarized in the Methods section. Briefly, the MM/GBSA binding free energy contains three parts: (1) the molecular mechanics (MM); (2) the Generalized Born (GB) for solvation energy and (3) the nonpolar energy proportional to the solvent accessible surface area (SA). Using the NAMD package, the MM part was calculated from the trajectory analysis on a pair-wise interaction between a ligand and Mpro. Here, the MM part for B19 is the addition of vdW and electrostatic
interactions as shown in Figs. 3c and 3d, which is -80.6 kcal/mol (Table I). Similarly, we obtained the MM energy for the native ligand in the active site of Mpro, which is -83.5 kcal/mol (Table I). Here, for the MM part only, the binding is stronger for the native ligand, mainly due to more hydrogen bonds (\textit{i.e.} stronger electrostatic interaction) formed by the native ligand and Mpro.

Table 1: Comparisons of MM/GBSA binding free energies for the native ligand and B19 inside the active site of Mpro.

| ligand | MM (kcal/mol) | GBSA (kcal/mol) | total (kcal/mol) |
|--------|---------------|----------------|-----------------|
| native | -83.5         | -44.7          | -128.2          |
| B19    | -80.6         | -54.6          | -135.2          |

With the NAMD package, the GBSA part can be calculated together during the trajectory analysis\textsuperscript{23} (see Methods section). The GBSA binding free energies for B19 and the native ligand are -54.6 and -44.7 kcal/mol (Table I), respectively. For the GBSA part, because B19 is less hydrophilic than the native ligand, the desolvation penalty for B19 after its entry into the active site of Mpro is less than that for the native ligand. Thus, it is energetically more beneficial for B19 to move from the solution into the Mpro’s active site. Adding together the MM and GBSA parts, the total binding free energies for B19 and the native ligand are -135.2 kcal/mol and -128.2 kcal/mol, respectively, suggesting that B19 is slightly (\textasciitilde 5\%) better than the native ligand when competing for Mpro. Here, the entropy contribution that typically yields noisy data is ignored. Qualitatively, the native protein ligand is more flexible in the solution (due to peptide bonds) than B19, causing more entropy loss (\textit{i.e.} unfavorable) for the native ligand when entering the active site and thus an even larger difference in binding free energies between B19 and the native ligand. Note that typically the MM/GBSA binding free energy is about 5-10 times larger than the experimental value. Nevertheless, the MM/GBSA method provides good ranking results that are generally consistent with experimental ones.
Discussion and Conclusions

To evaluate the synthetic accessibility of the drug-like molecule B19, we employed the recently debuted synthetic analysis provided by the IBM-RXN for Chemistry (https://rxn.res.ibm.com) that predicts chemical reactions and the development of molecules. The newly added retrosynthesis module is based on an extension of the artificial intelligence (AI) model called the Molecular Transformer combined with a hyper-graph exploration strategy to automate the planning of the retrosynthesis route without human supervision.\textsuperscript{24,25} When compared with other works, statistical analysis indicates that the retrosynthetic architecture of RXN performs very well for a broad class of disconnections.\textsuperscript{25} Here, due to the complexity of B19, we applied the software to delineate the steps for synthesizing B19 without the methanesulfonamide group, \textit{i.e.} the complex including fragments U0P, HWH and the linker. The result shows that the complex can be synthesized in merely eight steps (see Supporting Information). The synthesized complex can be further linked with the methanesulfonamide group (terminated with a bromine atom) through the Bromo N-alkylation reaction in the NaOH solution.

In summary, by merging drug fragments occupying different subsites of the Mpro’s active site, we designed the drug-like molecule B19 that was confirmed to bind the active site stably from two independent MD simulations. Similar to the previously designed mechanism-based N3 and O6K ligands, B19 occupies four subsites (S1, S1’, S2 and S4), through various hydrogen bonds and hydrophobic interactions. Using the docking method with a large exhaustiveness number, we screened over 700 different conformations of B19 inside the active site of Mpro and verified that the equilibrated B19 structure in MD simulation is the most stable one (with the highest binding affinity). Additionally, we compared the binding free energies between B19 and a native protein ligand processed by Mpro, which indicates that B19 can bind Mpro more strongly than the native ligand. Taking all these encouraging results together, B19 is a promising candidate for inhibiting Mpro of SARS-CoV-2 and deserves further investigation \textit{in vitro/in vivo}. With the unveiled molecular mechanism of
B19’s binding in the active site of Mpro, it is possible to further optimize B19 through the in silico alchemical free-energy perturbation (FEP) method.\textsuperscript{26,27} Last but not least, the in silico approach described in this work might yield other efficacious drug molecules with the ongoing crowdsourcing efforts, to fight collectively and efficiently against COVID-19.

**Methods**

**MD simulations** We performed all-atom MD simulations for designed ligand (drug) molecules inside Mpro, using the NAMD2.13 package\textsuperscript{28} running on the IBM Power Cluster. To model the ligand-Mpro complex, we obtained the previously resolved crystal structure for Mpro (PDB code: 5R7Y)\textsuperscript{11} from the Protein Data Bank (PDB) and assembled together two-three fragments (PDB codes: 5R7Y, 5RGI and 5R7Z) with a short linker molecule into a ligand in Mpro. Overall, we designed 19 ligands with various linker molecules. The position of each fragment in a ligand was kept as close to the crystallized one in Mpro as possible. The complex was solvated in a cubic water box measuring about $97.2 \times 97.2 \times 97.2$ Å$^3$. 88 K$^+$ and 80 Cl$^-$ were added in order to neutralize the entire simulation system and set the ion concentration to be 0.15 M (Fig. 2a). Because the active sites of Mpro among all coronavirus Mpros are highly conserved, we also modeled a native ligand (with the sequence TSAVLQS-GFRK) in Mpro of SARS-CoV-2 (Fig. 6a), with the ligand’s position adopted from the crystal structure (PDB code: 2Q6G) for Mpro in SARS-CoV-1. The final system containing about 94 K atoms was first minimized for 10 ps and further equilibrated for 500 ps in the NPT ensemble ($P \sim 1$ bar and $T \sim 300$ K), with atoms in the Mpro’s backbone and non-hydrogen atoms in the ligand harmonically restrained (spring constant $k=1$ kcal/mol/Å$^2$). After removing all restraints, the production run was carried out in the NVT ensemble.

The CHARMM36 force field\textsuperscript{29} was chosen for proteins (Mpro and the native ligand); the TIP3P model\textsuperscript{30,31} was chosen for water; the standard force field\textsuperscript{32} was used for ions. The force field for the designed ligands were obtained from SwissParam.\textsuperscript{33} We provided the .psf and
.pdb files for B19 in the Supporting Information. The periodic boundary conditions (PBC) were applied in all three dimensions. Long-range Coulomb interactions were calculated using particle-mesh Ewald (PME) full electrostatics with the grid size about 1 Å in each dimension. The van der Waals (vdW) energies between a pair of atoms were calculated using a smooth (10-12 Å) cutoff. The temperature $T$ was maintained at 300 K by applying the Langevin thermostat, while the pressure was kept constant at 1 bar using the Nosé-Hoover method. With the SETTLE algorithm applied to keep all bonds rigid, the simulation time-step was set to be 2 fs for bonded and non-bonded (e.g. vdW, angle and dihedral) interactions, and electric interactions were calculated every 4 fs, with the multiple time-step algorithm.

**MM/GBSA Free energy calculations** We used the GBIS module in NAMD to run postprocessing of MD simulation trajectories, for obtaining the GBSA binding free energy that comprises the polar solvation free energy estimated from the Generalized Born (GB) method and the nonpolar one obtained as a function of the solvent accessible surface area (SASA), $-\gamma \cdot \text{SASA}$ where $\gamma = 0.00542 \text{ kcal/molÅ}^2$ is the surface tension. The solvent dielectric constant was set to be 78.5 and the cutoff distance was set to be 16 Å. Overall, the GBSA free energy change $\Delta G = \Delta G_{\text{complex}} - \Delta G_{\text{protein}} - \Delta G_{\text{ligand}}$. For the MM part, we used the NAMD-energy module for calculating the pair-wise energy between a ligand and Mpro, with the solute dielectric constant set to be 4.

**Docking method** we employed Autodock Vina, one of the most popular and highly cited open source docking software applications in the research community, to identify the conformation of the studied drug molecules and the corresponding binding affinity with Mpro. Autodock Vina was shown to have the highest scoring power in a study that performed a comprehensive evaluation of ten famous currently available docking programs, including five commercial and five academic ones. To prepare the input for the docking simulations we used the set of commands provided by AutoDockTools (ADT) specifically developed to support the AutoDock users downloaded from The Center for Computational Structural
Biology (CCSB), and followed the protocol in our previous study.\textsuperscript{21}

**Competing Interests**

T. H. and B. L. declare no conflicts of interest.

**Acknowledgement**

T. H. and B. L. gratefully acknowledge the help from Wendy Cornell and the computing resource from the IBM Cognitive Computing Program.

**Supporting Information Available**

Fig. S1 shows the other 18 designed ligands (B1 to B18) that are less stable than B19 in the active site of Mpro; Fig. S2 shows simulation system of Mpro with a bound native ligand and RMSD values from two independent MD simulations; Recommended synthesis steps for B19; the .psf and .pdb files for B19; Movies for B19 in Mpro (B19.mpg) and the native ligand in Mpro (native.mpg).

**References**

1. Wu, F.; Zhao, S.; Yu, B.; Chen, Y.-M.; Wang, W.; Song, Z.-G.; Hu, Y.; Tao, Z.-W.; Tian, J.-H.; Pei, Y.-Y. \textit{et al.} A new coronavirus associated with human respiratory disease in China. \textit{Nature} \textbf{2020}, \textit{579}, 265–269.

2. Kuiken, T.; Fouchier, R. A.; Schutten, M.; Rimmelzwaan, G. F.; Van Amerongen, G.; van Riel, D.; Laman, J. D.; de Jong, T.; van Doornum, G.; Lim, W. \textit{et al.} Newly discovered coronavirus as the primary cause of severe acute respiratory syndrome. \textit{The Lancet} \textbf{2003}, \textit{362}, 263–270.
3. Bermingham, A.; Chand, M.; Brown, C.; Aarons, E.; Tong, C.; Langrish, C.; Hoschler, K.; Brown, K.; Galiano, M.; Myers, R. et al. Severe respiratory illness caused by a novel coronavirus, in a patient transferred to the United Kingdom from the Middle East, September 2012. *Eurosurveillance* 2012, 17, 20290.

4. Beigel, J. H.; Tomashek, K. M.; Dodd, L. E.; Mehta, A. K.; Zingman, B. S.; Kalil, A. C.; Hohmann, E.; Chu, H. Y.; Luetkemeyer, A.; Kline, S. et al. Remdesivir for the treatment of Covid-19—preliminary report. *New England Journal of Medicine* 2020,

5. Villar, J.; Ferrando, C.; Martínez, D.; Ambrós, A.; Muñoz, T.; Soler, J. A.; Aguilar, G.; Alba, F.; González-Higueras, E.; Conesa, L. A. et al. Dexamethasone treatment for the acute respiratory distress syndrome: a multicentre, randomised controlled trial. *The Lancet Respiratory Medicine* 2020, 8, 267–276.

6. Ullrich, S.; Nitsche, C. The SARS-CoV-2 main protease as drug target. *Bioorganic Med. Chem. Lett.* 2020, 30, 127377.

7. Jin, Z.; Du, X.; Xu, Y.; Deng, Y.; Liu, M.; Zhao, Y.; Zhang, B.; Li, X.; Zhang, L.; Peng, C. et al. Structure of Mpro from COVID-19 virus and discovery of its inhibitors. *Nature* 2020, 582, 289–293.

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

9. Jin, Z.; Zhao, Y.; Sun, Y.; Zhang, B.; Wang, H.; Wu, Y.; Zhu, Y.; Zhu, C.; Hu, T.; Du, X. 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.

10. Chodera, J.; Lee, A. A.; London, N.; von Delft, F. Crowdsourcing drug discovery for pandemics. *Nat. Chem.* 2020, 12, 581–581.
11. Douangamath, A.; Fearon, D.; Gehrtz, P.; Krojer, T.; Lukacik, P.; Owen, C. D.; Resnick, E.; Strain-Damerell, C.; Ábrányi-Balogh, P.; Brandaő-Neto, J. et al. Crystallographic and electrophilic fragment screening of the SARS-CoV-2 main protease. bioRxiv 2020,

12. Erlanson, D. A. *Fragment-based drug discovery and X-ray crystallography*; Springer, 2011; pp 1–32.

13. Murray, C. W.; Rees, D. C. The rise of fragment-based drug discovery. *Nat. Chem.* 2009, 1, 187–192.

14. Jhoti, H.; Williams, G.; Rees, D. C.; Murray, C. W. The ‘rule of three’ for fragment-based drug discovery: where are we now? *Nat. Rev. Drug Discov.* 2013, 12, 644–644.

15. Karplus, M.; McCammon, J. Molecular dynamics simulations of biomolecules. *Nature Struct. Biol.* 2002, 265, 654–652.

16. Freddolino, P. L.; Harrison, C. B.; Liu, Y.; Schulten, K. Challenges in protein-folding simulations. *Nature Phys.* 2010, 6, 751–758.

17. Wang, Z.; Sun, H.; Yao, X.; Li, D.; Xu, L.; Li, Y.; Tian, S.; Hou, T. Comprehensive evaluation of ten docking programs on a diverse set of protein–ligand complexes: the prediction accuracy of sampling power and scoring power. *Phys. Chem. Chem. Phys.* 2016, 18, 12964–12975.

18. Xue, X.; Yu, H.; Yang, H.; Xue, F.; Wu, Z.; Shen, W.; Li, J.; Zhou, Z.; Ding, Y.; Zhao, Q. et al. Structures of two coronavirus main proteases: implications for substrate binding and antiviral drug design. *J. Virol.* 2008, 82, 2515–2527.

19. Dai, W.; Zhang, B.; Su, H.; Li, J.; Zhao, Y.; Xie, X.; Jin, Z.; Liu, F.; Li, C.; Li, Y. et al. Structure-based design of antiviral drug candidates targeting the SARS-CoV-2 main protease. *Science* 2020, 368, 1331–1335.
20. Trott, O.; Olson, A. J. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J. Comput. Chem.* **2010**, *31*, 455–461.

21. Huynh, T.; Wang, H.; Luan, B. *in silico* Exploration of Molecular Mechanism of Clinically Oriented Drugs for Inhibiting SARS-CoV-2’s Main Protease. *J. Phys. Chem. Letts* **2020**, *11*, 4413–4420.

22. Muramatsu, T.; Takemoto, C.; Kim, Y.-T.; Wang, H.; Nishii, W.; Terada, T.; Shirouzu, M.; Yokoyama, S. SARS-CoV 3CL protease cleaves its C-terminal autoprocessing site by novel subsite cooperativity. *Proc. Natl. Acad. Sci. USA* **2016**, *113*, 12997–13002.

23. Vergara-Jaque, A.; Comer, J.; Monsalve, L.; González-Nilo, F. D.; Sandoval, C. Computationally efficient methodology for atomic-level characterization of dendrimer–drug complexes: a comparison of amine-and acetyl-terminated PAMAM. *The Journal of Physical Chemistry B* **2013**, *117*, 6801–6813.

24. Vaucher, A. C.; Zipoli, F.; Geluykens, J.; Nair, V. H.; Schwaller, P.; Laino, T. Automated extraction of chemical synthesis actions from experimental procedures. *Nat. Commun.* **2020**, *11*, 3601.

25. Schwaller, P.; Petraglia, R.; Zullo, V.; Nair, V. H.; Haeuselmann, R. A.; Pisoni, R.; Bekas, C.; Iuliano, A.; Laino, T. Predicting retrosynthetic pathways using transformer-based models and a hyper-graph exploration strategy. *Chem. Sci.* **2020**, *11*, 3316–3325.

26. Chipot, C.; Pohorille, A. *Free energy calculations*; Springer, 2007.

27. Luan, B.; Xu, G.; Feng, M.; Cong, L.; Zhou, R. Combined Computational–Experimental Approach to Explore the Molecular Mechanism of SaCas9 with a Broadened DNA Targeting Range. *J. Am. Chem. Soc.* **2019**, *141*, 6545–6552.
28. Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kale, L.; Schulten, K. Scalable Molecular Dynamics with NAMD. *J. Comp. Chem.* **2005**, *26*, 1781–1802.

29. MacKerell, A., Jr.; Bashford, D.; Bellott, M.; Dunbrack, R. L., Jr.; Evanseck, J.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S. *et al.* All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J. Phys. Chem. B* **1998**, *102*, 3586–3616.

30. Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of Simple Potential Functions for Simulating Liquid Water. *J. Chem. Phys.* **1983**, *79*, 926–935.

31. Neria, E.; Fischer, S.; Karplus, M. Simulation of Activation Free Energies in Molecular Systems. *J. Chem. Phys.* **1996**, *105*, 1902–1921.

32. Beglov, D.; Roux, B. Finite representation of an infinite bulk system: Solvent boundary potential for computer simulations. *J. Chem. Phys.* **1994**, *100*, 9050–9063.

33. Zoete, V.; Cuendet, M. A.; Grosdidier, A.; Michielin, O. SwissParam: a fast force field generation tool for small organic molecules. *J. Comput. Chem.* **2011**, *32*, 2359–2368.

34. Allen, M. P.; Tildesley, D. J. *Computer Simulation of Liquids*; Oxford University Press: New York, 1987.

35. Martinetz, T.; Schulten, K. Topology Representing Networks. *Neur. Netw.* **1994**, *7*, 507–522.

36. Miyamoto, S.; Kollman, P. A. SETTLE: An Analytical Version of the SHAKE and RATTLE Algorithm for Rigid Water Molecules. *J. Comp. Chem.* **1992**, *13*, 952–962.

37. Tuckerman, M.; Berne, B. J.; Martyna, G. J. Reversible multiple time scale molecular dynamics. *The Journal of chemical physics* **1992**, *97*, 1990–2001.
38. Tsui, V.; Case, D. A. Theory and applications of the generalized Born solvation model in macromolecular simulations. *Biopolymers: Original Research on Biomolecules* **2000**, *56*, 275–291.