

# Computational Evaluation of the COVID-19 3c-like Protease Inhibition Mechanism, and Drug Repurposing Screening

*Hao Liu,<sup>†</sup> Tao Jiang,<sup>‡</sup> Wenlang Liu,<sup>‡</sup> Zheng Zheng<sup>\*,‡</sup>*

<sup>†</sup>School of Mechanical and Electronic Engineering, Wuhan University of Technology, 122  
Luoshi Road, Wuhan 430070, PR China

<sup>‡</sup>School of Chemistry, Chemical Engineering and Life Science, Wuhan University of Technology,  
122 Luoshi Road, Wuhan 430070, PR China

Email: laozhengzz@gmail.com

## **Abstract**

The rapid spread of the COVID-19 outbreak is now a global threat with over a million diagnosed cases and more than 70 thousand deaths. Specific treatments and effective drugs regarding such disease are in urgent need. To contribute to the drug discovery against COVID-19, we performed computational study to understand the inhibition mechanism of the COVID-19 3c-like protease, and search for possible drug candidates from approved or experimental drugs through drug repurposing screening against the DrugBank database. Two novel computational methods were applied in this study. We applied the “Consecutive Histogram Monte Carlo” (CHMC) sampling method for understanding the inhibition mechanism from studying the 2-D binding free energy landscape. We also applied the “Movable Type” (MT) free energy method for the lead compound screening by evaluating the binding free energies of the COVID-19 3c-like protease – inhibitor complexes. Lead

compounds from the DrugBank database were first filtered using ligand similarity comparison to 19 published SARS 3c-like protease inhibitors. 70 selected compounds were then evaluated for protein-ligand binding affinities using the MT free energy method. 4 drug candidates with strong binding affinities and reasonable protein-ligand binding modes were selected from this study, *i.e.* Enalkiren (DB03395), Rupintrivir (DB05102), Saralasin (DB06763) and TRV-120027 (DB12199).

## **Introduction**

In 2003, the outbreak of severe acute respiratory syndrome (SARS) terrified the world by lethally infecting thousands of people and taking hundreds of lives. This pandemic was caused by SARS-CoV, a member from the big coronaviruses' family. Although the epidemic of SARS died down soon, we had to admit the fact that we were still far from winning the battle against coronaviruses. In 2012, another human coronavirus responsible for the Middle East Respiratory Syndrome (MERS) emerged, which again led to thousands of confirmed diagnosis globally and killed more than 500 people.<sup>1,2</sup> 7 years later in the end of 2019, COVID-19, another acute respiratory syndrome which was very similar to SARS was rapidly spread global wise. The cause was soon identified as SARS-CoV-2, a coronavirus closely related to the SARS-CoV.<sup>3</sup> It is the largest global coronavirus outbreak event in history, with over 700 thousand diagnosed cases and 34 thousand deaths and is still evolving rapidly at this moment. Research groups and pharmaceutical companies had been racing against time to develop treating agents against COVID-19, yet neither a vaccine nor an efficacious therapy has been available to date. Therefore, it is urgently

necessary to design the potential therapeutic agents against COVID-19 and other mass-pathogenic coronaviruses.

Coronaviruses are positive-sense, single-stranded RNA viruses that contain the largest RNA virus genomes as known so far.<sup>4,5,6</sup> They are the cause of many highly prevalent diseases in humans and animals and thus pose a great threat to public health. The main protease that is responsible for maturation of the polyproteins in the viral replication cycle is a chymotrypsin-like protease that uses a Cys instead of a Ser as the nucleophile in the active site, so it is often referred to as the 3C-like protease (3CLpro).<sup>7</sup> The fact that 3CLpro plays such a significant role in the virus life cycle and is found in most coronaviruses made this protein a prominent target to develop anti-CoV drugs.<sup>4,8-12</sup> Harris T. *et. al.* and Rao Z. *et. al.* reported the crystal structure of SARS-CoV 3CLpro in 2003,<sup>13,14</sup> with the coordinates deposited in the Protein Data Bank (<http://www.rcsb.org>, PDB IDs: 1q2w, 1uj1, 1uk2, 1uk3, 1uk4). The COVID-19 3CLpro crystal structure was released recently with PDB ID 6lu7<sup>15</sup> having 87.1% sequence similarity, and identical residue alignments at the active site compared to SARS-CoV 3CLpro.<sup>16</sup>

The crystal structure of COVID-19 3CLpro reveals a protein dimer containing three domains at each protomers. Domain I and II forms a chymotrypsin fold, rich in  $\beta$ -sheets and is responsible for the catalytic reaction with a Cys-His catalytic dyad (Cys-145 and His-41). Domain III contains five  $\alpha$  helix strands with unclear biological functions, and is connected to domain II via a long loop region. The substrate-binding site is located in a cleft between domains I and II.<sup>13,17,18</sup> So far, research groups have discovered a number of SARS-CoV 3CLpro inhibitors, which could potentially be further developed into anti-SARS drugs. Among the various inhibitors that have been reported, the majority was found

to target at His-41, His-163, His-164, Cys-145, Glu-166, Met-49, and Met-165 residues via aromatic  $\pi$ - $\pi$  stackings and hydrogen bonds.<sup>2, 13, 19-29</sup> Possible COVID-19 treatment could be provided through the development of the COVID-19 3CLpro inhibitors, which can be hugely encouraged from the understanding of the inhibition mechanism of SARS-CoV 3CLpro, due to the structural similarity between the two target proteins.

Two computational methods were employed in this study for drug candidate virtual screening and binding mechanism analysis respectively. (1) The Movable Type (MT) method was first developed in our lab in 2013.<sup>30</sup> Further refinement was later on published in 2018.<sup>31</sup> This method is now commercialized through QuantumBio Inc. Please contact <http://www.quantumbioinc.com> for requesting the MT method software product. This method is a molecular energy-state sampling algorithm using Monte Carlo integration and numerical integral for fast free energy estimation. In this study, we apply the MT method to perform virtual screenings, and free energy evaluations aiming to discover drug candidates for COVID-19 3CLpro inhibition referencing a free-energy level computational protocol. (2) The Consecutive Histogram Monte Carlo (CHMC) sampling method was a novel simulation technique developed in our group. The CHMC protocol employs independent energy-state sampling for calculating the ensemble energy within a series of consecutively distributed ligand sampling volumes across the protein surface centered at the binding active site. Evenly distributed ligand's energy-state ensembles on the protein surface were generated using a Monte Carlo sampling protocol representing a reversible work pathway for the ligand transition. We applied this protocol for the binding mechanism analysis and possible binding pathway study against the selected good binders from the virtual screening procedure.

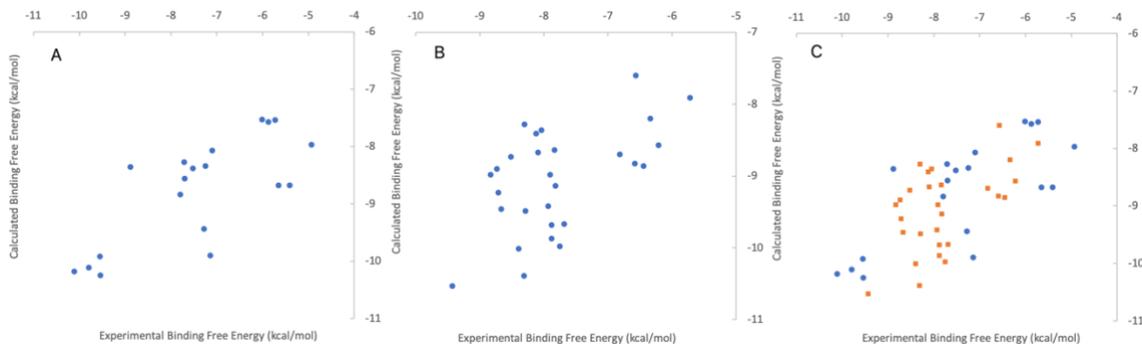
## Results and Discussion

In this study, we first examined the validity of applying the MT method to the 3CLpro – inhibitor binding free energy estimation by analyzing the binding affinity reproduction regarding the 3CLpro – inhibitor systems with published experimental binding affinities using 19 complexes for SARS-CoV and 28 complexes for other viruses *e.g.* Enterovirus 71 and Norovirus. Apart from that we also explored the structure of the COVID-19 3CLpro binding active site using the 6lu7 crystal structure, and the binding mode between COVID-19 3CLpro and a peptide-like inhibitor suggested by the 6lu7 crystal structure deposition authors, in order to study the contact map and locate the important residues at the active site for effective inhibitor discovery. The CHMC sampling protocol was employed to analyze the free energy surface for the peptide-like inhibitor's transition on the 3CLpro surface around the binding active site. Understanding the general performance of the MT method against the 3CLpro – inhibitors binding affinity predictions, as well as the inhibition mechanism of COVID-19 3CLpro on a molecular level, we performed a MT-based virtual screening against the Drugbank database, through ligand similarity comparison, binding affinity evaluation, and binding mode analysis, aimed to select druggable inhibitor for COVID-19 3CLpro. Employing this database, our research focused more on drug repurposing, to look for COVID-19 3CLpro inhibitors from approved and experimental drugs, in order to increase the drug discovery efficiency, or find ready-to-use drugs.

### **(1) Method Validation Against Published 3CLpro – Inhibitor Complexes**

We first examined the binding affinity reproduction performance of the MT method. Yet there is no proven inhibitor published for COVID-19 3CLpro for our free energy

estimation method validation, we first introduced validation cases targeting the SARS-CoV 3CLpro receptor, which closely relates to the COVID-19 3CLpro with only 14 out of a total number of 306 residue disagreement in a protomer, and 100% identical residue alignments at the binding active area. 19 complexes with experimental binding affinity data for the SARS-CoV 3CLpro – inhibitor systems were collected from the Protein Data Bank. The 19 published ligands in this validation set have a broad binding affinity distribution range from -5.41 to -10.11 kcal/mol. Statistics of this calculation approach showed an RMSE as 1.72 kcal/mol, the Pearson's correlation as 0.74 and Kendall's tau correlation as 0.48, revealing a good prediction accuracy and ranking capability of the MT method against the SARS-CoV 3CLpro – inhibitor validation set. We then introduced 28 more test cases using 3CLpro receptors from other viruses to broaden the validation and examine the robustness of the MT method against different 3CLpro receptor – ligand complex systems. This validation set contained 3CLpro from different organisms including TGEV, feline coronavirus, bat coronavirus HKU4, norovirus, coxsackievirus A16, B3, B4, Human enterovirus 71, 72, 73, 74, 75, 76, 77, 78, and 79. Against this validation set, the MT calculation generated an RMSE as 1.49 kcal/mol, the Pearson's correlation as 0.56 and Kendall's tau correlation as 0.29, illustrating an acceptable error and moderate correlation compared to the experimental binding free energies. Dropping of prediction accuracy could be caused by the increase of chemical variety of the target structures in this validation set. Against all the validation cases when merging the two sets, the MT calculation generated an RMSE as 1.59 kcal/mol, the Pearson's correlation as 0.67 and Kendall's tau correlation as 0.44, showing a generally good reproduction accuracy and correlation regarding the overall 47 3CLpro – inhibitor complexes.



**Figure 1.** Scattered plots comparing the binding free energy calculated by Moveable Type to experimental data for the 3CLpro – inhibitor test sets: (A) SARS-CoV 3CLpro test set (19 test cases) (B) expanded 3CLpro test set (28 test cases) (C) all 47 3CLpro – inhibitor test cases (blue dots for the SARS-CoV 3CLpro test cases and yellow squares for the expanded 3CLpro test cases).

## (2) COVID-19 3CLpro Binding Active Site Analysis

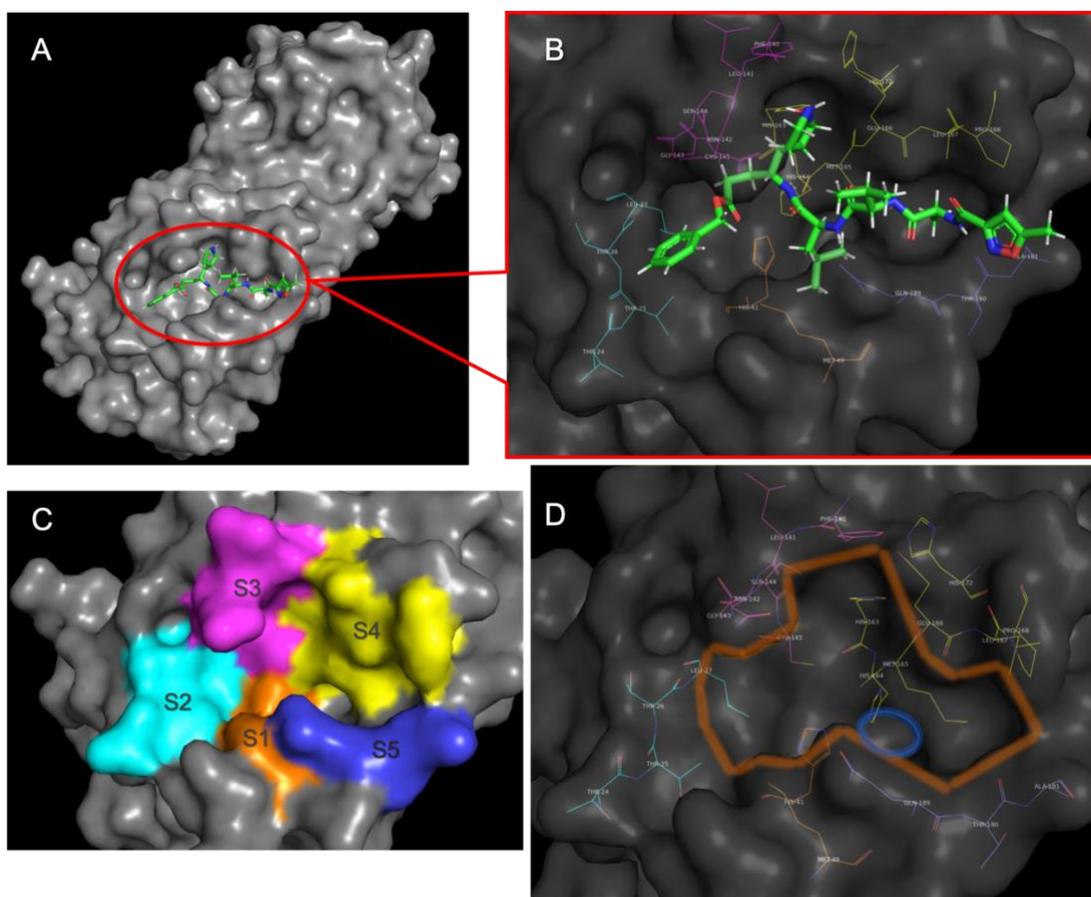
Introducing these test sets revealed a general picture of the binding affinity prediction using the MT method against 3CLpro – inhibitor complexing, with diverse structural features. On the other hand, drug discovery for the COVID-19 3CLpro inhibition required in-depth understanding of the inhibition mechanism of the COVID-19 3CLpro and the structural feature at the active binding site. 6lu7 was the first published COVID-19 3CLpro crystal structure revealing important information regarding the active binding site structure and valuable protein-ligand contact map using a competitive protease inhibitor (ligand ID: PRD\_002214). Residues with close contacts to the inhibitor can be discovered at 5 active regions in the binding site:

- (1) S1 region included residue His41 and Met49;
- (2) S2 region included residue Thr24, Thr25, Thr26 and Leu27;

- (3) S3 region included residue Phe140, Leu141, Asn142, Gly143, Ser144, and Cys145;
- (4) S4 region included residue His163, His164, Met165, Glu166, Leu167, and Pro168;
- (5) S5 region included residue Glu189, Thr190, and Ala191.

The S4 and S5 regions acted as a pair of “tweezers” to grasp and stabilize the polypeptide substrate by forming backbone – backbone hydrogen bonds, and receptor sidechain – substrate backbone hydrogen bond using Glu189. His41 at S1 and Cys145 at S3 formed a His ... Cys catalytic dyad. Other residues at these active regions formed cavities for containing substrate residues or deploy close-range contacts for further stabilizing the substrate at the binding site.

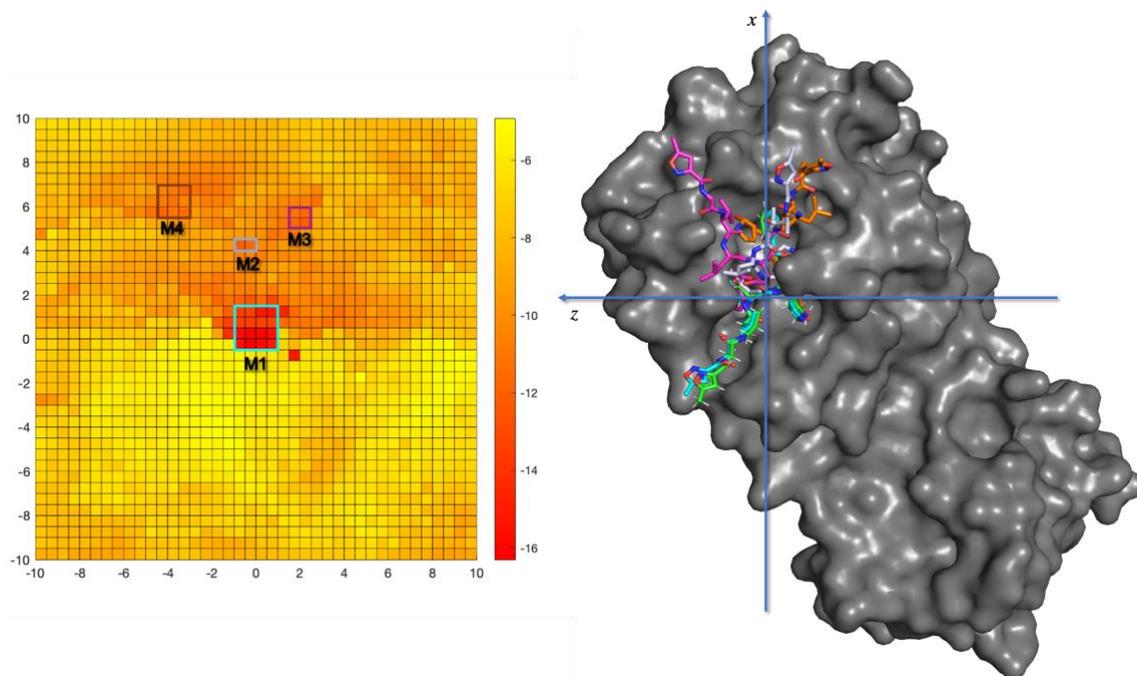
Other studies also reported matching residues and active regions at the binding site playing important roles to the substrate-3CLpro complexing, by analyzing the SARS-CoV 3CLpro structure. Song J. X. *et al.* reported 6 binding subsites binding correspondingly to the P1-P6 residues of the substrates. Those 6 binding subsites generally agreed with the aforementioned S1 to S5 active regions in this work, with Glu166, Gly-143 and Cys-145 as in the S1 subsite specifically recognizing the side-chain NE2 of Gln-P1; Thr-25, Leu-27, Val-42, Cys-44, Thr-47, Asp-48, Met-49, Tyr-54, Leu-164 and Met-165 as in the S2 subsite binding in particular with the Leu-P2.<sup>19,20</sup> Yokoyama, S. *et al.* also reported a similar group of residues holding the Leu-P2 of substrates in position<sup>21</sup> during the catalyzation. All these studies provided us important information and inspiring perspectives for the anti-CoV drug design in this work.



**Figure 2.** Illustration of the active regions at the substrate binding site of the COVID-19 3CLpro. (A) Crystal structure of COVID-19 3CLpro in complex with the inhibitor PRD\_002214 (PDB ID: 6lu7). (B) A peptide-like inhibitor suggested in the crystal structure 6lu7 as a competitive inhibitor at the substrate binding site, important residues are marked in different colors representing different active regions at the binding site. (C) Five active regions marked as S1~S5 with different colors at the substrate binding site. (D) The five active regions sieged to form a reversed “T-shaped” pocket (circled with brown curve) with a small cavity at the “T” junction (circled with blue curve).

We applied the CHMC protocol for generating the ligand transition free energy landscape near the binding site of the COVID-19 3CLpro using the inhibitor provided in

6lu7. A 2-D free energy heatmap was generated, where the (0,0) coordinate represented the location of the inhibitor's centroid in the 6lu7 crystal structure, the x and z axis represents the inhibitor's centroid deviation from its original location. The heatmap is separated into  $0.5 \text{ \AA} \times 0.5 \text{ \AA}$  evenly distributed grids. Each grid represents a histogram of volume on the protein surface, stretching for  $\pm 20 \text{ \AA}$  in both direction along the y axis with  $0.5 \text{ \AA}$  width in both x and z axis. Ligand conformational state sampling in each histogram was performed with the ligand's centroid restricted in the histogram. Converged ensemble energies were then collected for all the histograms distributed from  $-10 \text{ \AA}$  to  $10 \text{ \AA}$  along both x and z axis centered at the (0,0) coordinate. We hence generated the 2-D free energy heatmap representing the protein-ligand free energy differences by moving the ligand's centroid on the protein's surface. The free energy heatmap demonstrated that the global minimum state M1 was at the (0,0) histogram. The global minimal binding mode generated from CHMC had a conformational RMSD as  $0.80 \text{ \AA}$  compared to the crystal structure ligand (shown in Figure 3). the M2 local minimum located at the (4,-1) histogram, with  $3.92 \text{ kcal/mol}$  as the free energy change for the ligand transition from the M1 state to the M2 state, *i.e.*  $\Delta G_{M1-M2} = 3.92 \text{ kcal/mol}$ . The M3 and M4 local minima located further along the positive direction of the x axis while different on the z axis direction, at (5,2) histogram and (5.5,-4.5) histogram respectively, with  $\Delta G_{M1-M3} = 3.91 \text{ kcal/mol}$ , and  $\Delta G_{M1-M4} = 4.86 \text{ kcal/mol}$ .

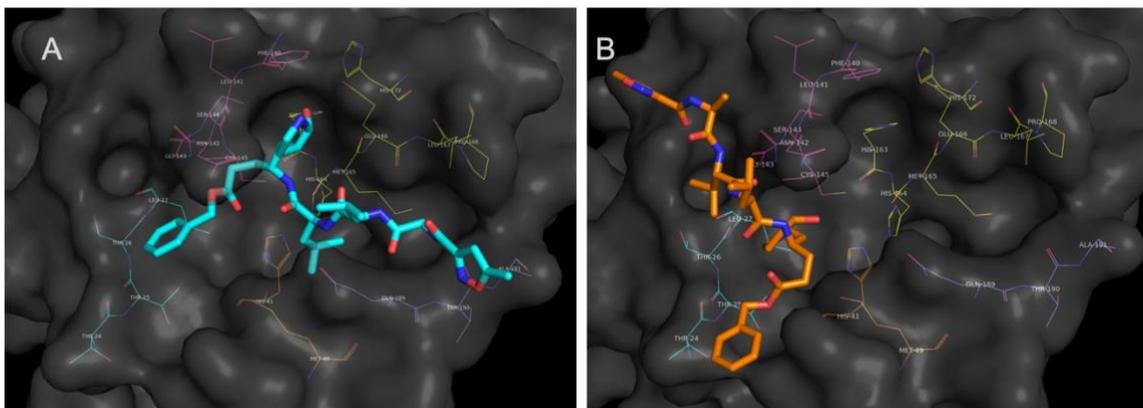


**Figure 3.** The ligand transition free energy surface and the energy minimal states generated using the CHMC 2-D protocol. The free energy surface is generated on the  $xz$  plane centered at the global minimal histogram (0,0) and ranges  $\pm 10$  Å along both the  $x$  and  $z$  axis. Global and local minimum energy histogram regions are circled in squares. The global minimum state (M1) and three metastable states (M2~M4) were discovered according to the free energy surface. The M1 to M4 histogram regions are marked in cyan, orchid, magentas, and brown respectively. The best ligand binding mode for the four energy minimal histograms are shown on the right-hand side of this figure in the corresponding colors. The crystal structure ligand (green) was shown together with the sampled ligand poses with a RMSD = 0.80 Å compared to the global minimal binding mode at the (0,0) histogram.

The inhibitor at M1 state bound with the “tweezers residues”: Glu166, His164, Gln189, and Thr190 through hydrogen bonds; blocked the catalytic dyad by forming a NH  $\cdots$  S

hydrogen bond with residue Cys145, and inserted a hydrophobic group into the “T junction” cavity to restrain the flexibility of His41.

The free energy heatmap generated from our calculation indicated that the binding site’s entrance/exit for the ligand located in the region on the positive  $x$ -axis half of the heatmap. Two possible dissociation/binding pathways were suggested following M1-M2-M3 route or M1-M2-M4 route. M2 is an intermediate state connecting the global minimum state and the other two local minima, M3 and M4 states settled along the  $x$  axis. Free energy landscape around the M4 state region was smoother compared to the M3 state region indicating that it was easier for the inhibitor to move further away from the binding site through the M4 state region. At the M4 state histogram, the ligand passed through the gateway surrounded by Thr24, Thr25, Thr26, Asn142, Gly143, and Cys145 before it departed the binding site of COVID-19 3CLpro.



**Figure 4.** Inhibitor PRD\_002214 bound to the COVID-19 3CLpro at the binding site. (A) The M1 global minimum binding mode of the inhibitor (B) The M4 local minimum binding mode of the inhibitor, which is at the edge of the COVID-19 3CLpro substrate binding site.

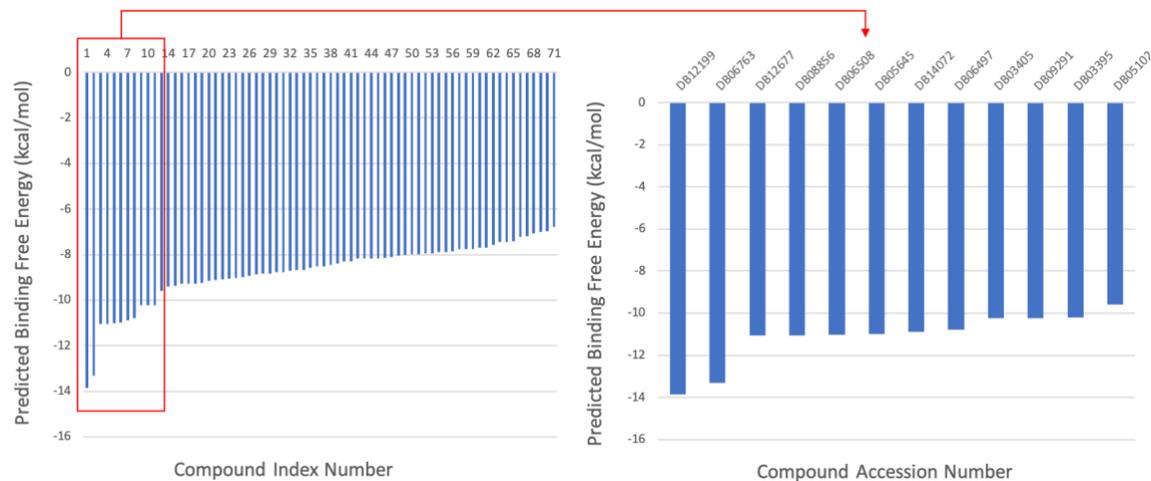
### **(3) Virtual Screening and Drug Repurposing Screening Targeting COVID-19 3CLpro**

After looking at the binding site structure features and possible inhibitor binding mechanism regarding the COVID-19 3CLpro receptor, we performed the MT-method based virtual screening aiming to discover possible drug candidates for this protein target. Initial drug discovery using massive virtual screening against large bioactive compound database is inefficient concerning the severe outbreak of the current COVID-2019 epidemic. In this study, we focused on approved or experimental drug repurposing screening aiming to increase the drug discovery efficiency or hopefully to find ready-to-use drugs.

We first performed ligand similarity search using the 19 aforementioned published SARS-CoV 3CLpro inhibitors. 70 Compounds from the Drugbank database with 70% or higher ligand similarity were selected for binding free energy evaluation. Using the MT free energy method, we generated the binding free energies for the 70 chosen ligand compounds complexing with COVID-19 3CLpro. The results are shown in Figure 5. On the other hand, ligand druggability is not necessarily correlated to the ligand-target binding affinity. From a structural point of view, the ligand-target binding modes with reasonable active site blocking and effective functional residue interference are more important to determine good drug candidates from the strong binders.

12 strong binders with -9.5 kcal/mol or higher calculated binding free energy were selected for further binding mode analysis. As is mentioned above, in COVID-19 3CLpro, the S4 and S5 “tweezers” regions, and the His41 ... Cys145 catalytic dyad need to be effectively blocked or interfered. Besides, due to the receptor’s structural similarity, those

catalysis-related residues in the SARS-CoV 3CLpro also need to be restrained by the drug candidate in the 3CLpro-inhibitor complex, *e.g.* the Gln-P1 and Leu-P2 substrate binding residues according to Song J. X. *et al.* and Yokoyama, S. *et al.* We then categorized those catalysis-related residues and merged them to the S1-S5 active in our study and performed the binding mode analysis against the 12 selected strong binders.



**Figure 5.** MT-predicted binding free energies for the 70 compounds selected from Drugbank using ligand similarity filtering. 12 lead compounds with high binding affinities (stronger than -9.5 kcal/mol) were further chosen for binding modes analysis.

Knowing that binding affinity values are not the most important criterion for selecting good drug candidates, instead, strong binders could end up quite differently due to their different ADMET properties and other criteria. From a molecular level point of view, we looked at the significant binding modes of the 12 lead compounds and studied the receptor-ligand critical contacts, to narrow down the 12 lead compounds to the final set of drug candidates.

Binding enthalpies and configurational entropies both contributed to the protein-ligand binding free energy, meaning that a strong binding affinity for a protein-ligand complex can be generated not only by having a significant global minimum binding mode, but also by forming a series of strong and stable local minima. For all the 12 lead compounds, we looked at their top 100 binding modes and collected the critical contact numbers including all the protein-ligand hydrogen bonds, close-ranged  $\pi$ - $\pi$  stackings and C-H  $\cdots$   $\pi$  interactions concerning the residues S1~S5 active regions. The critical contact numbers counted from all the 100 binding modes illustrated how frequently each ligand binding to the active regions in the complex state.

**Table 1.** Predicted binding free energies, molecular weights and protein-ligand critical contact numbers (CCN) at the active regions for the 12 selected lead compounds.

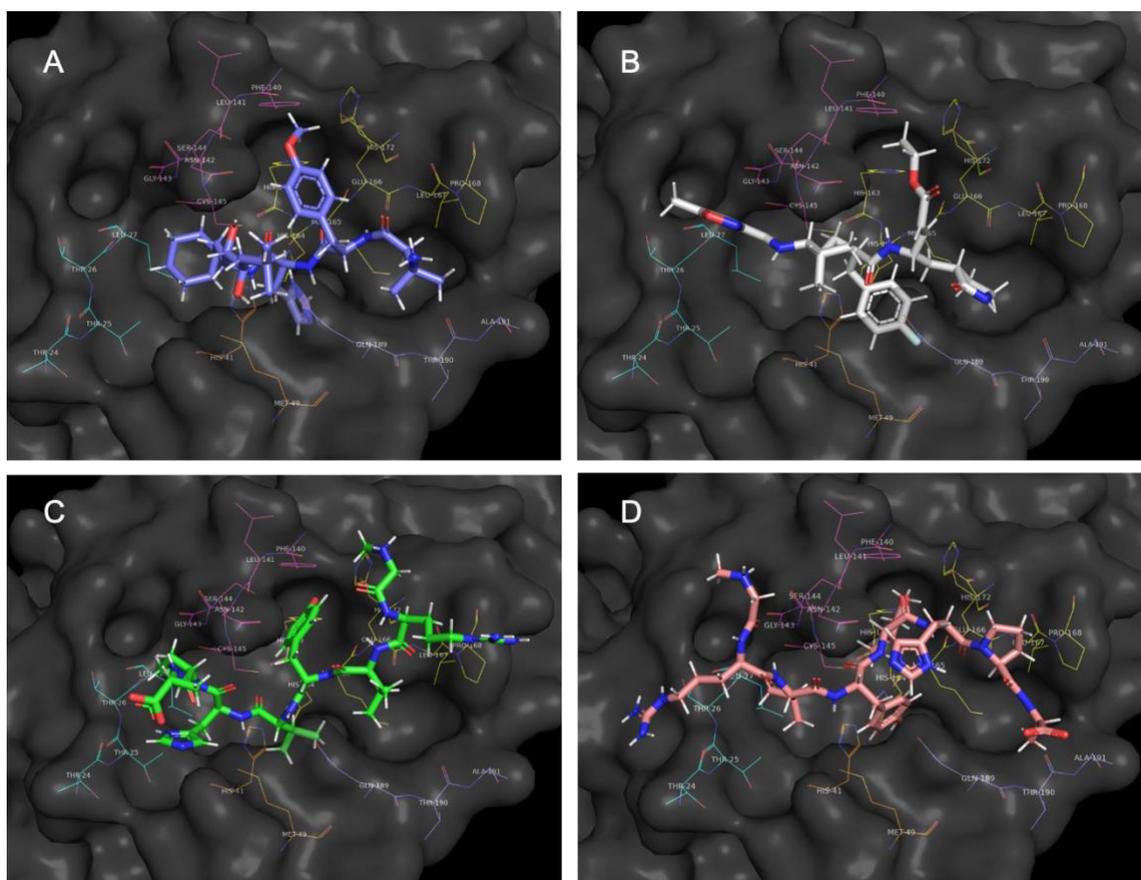
Accession Number	Drug Name	Original Target or Purpose	Predicted Binding Free Energy (kcal/mol)	Molecular Weight	CCN for S1	CCN for S2	CCN for S3	CCN for S4	CCN for S5	Total CCN
DB03395	Enalkiren	Human Renin	-10.21	656.85	16	2	64	18	53	153
DB03405	-	Human Cathepsin K	-10.24	594.69	20	1	21	15	29	86
DB05102*	Rupintrivir	HRV-14 Genome Polyprotein	-9.58	598.66	34	4	36	4	26	104
DB05645	Larazotide	Human Zonulin Receptor Antagonist	-10.98	725.84	9	7	20	2	18	56
DB06497	Aplaviroc	Human C-C Chemokine Receptor Type 5	-10.78	577.72	12	3	18	5	11	49
DB06508	GW-559090	Human Integrin	-11.03	596.68	15	6	24	7	17	69
DB06763	Saralasin	Human Angiotensin II Receptors	-13.32	912.06	3	9	47	11	37	107
DB08856	DADLE	Human Opioid Receptor	-11.05	569.65	5	5	8	3	7	28
DB09291	Rolapitant	Human Neurokinin 1 Receptor	-10.24	500.48	23	0	19	3	24	69
DB12199	TRV-120027	Human Angiotensin II Receptor Type 1	-13.86	926.09	0	3	49	13	50	115
DB12677	Soblidotin	Human Tubulin Polymerization Inhibitor	-11.06	701.99	16	2	29	11	21	79
DB14072	Reversin 121	Human P-glycoprotein	-10.90	641.76	2	7	15	2	23	49

\* compound that has been studied for COVID-19 treatment.

Enalkiren (DB03395), Rupintrivir (DB05102), Saralasin (DB06763) and TRV-120027 (DB12199) were found with significantly higher CCNs compared to the other lead compounds. It suggested that compared to the other strong binders, these four lead compounds bound more frequently with the catalysis-related active regions of COVID-19 3CLpro. Potentially making them possible drug candidates for the COVID-19 treatment. Concerning their CCN for each active region, all four drug candidates preferred to bind with the S3 and S5 regions. Enalkiren had the highest CCN at the S4 region while Rupintrivir bound more frequently with the S1 region. Saralasin and TRV-120027 both bound to human angiotensin II receptors as their original targets and had the same binding preference against the S3 and S5 regions over other regions of COVID-19 3CLpro. These two bulky compounds blocked larger area at the binding site and having more contacts with the S2 region compared to enalkiren and rupintrivir. On the other hand, they were facing the difficulty to reach deep inside the binding pocket and interact with the His41 ... Cys145 catalytic dyad.

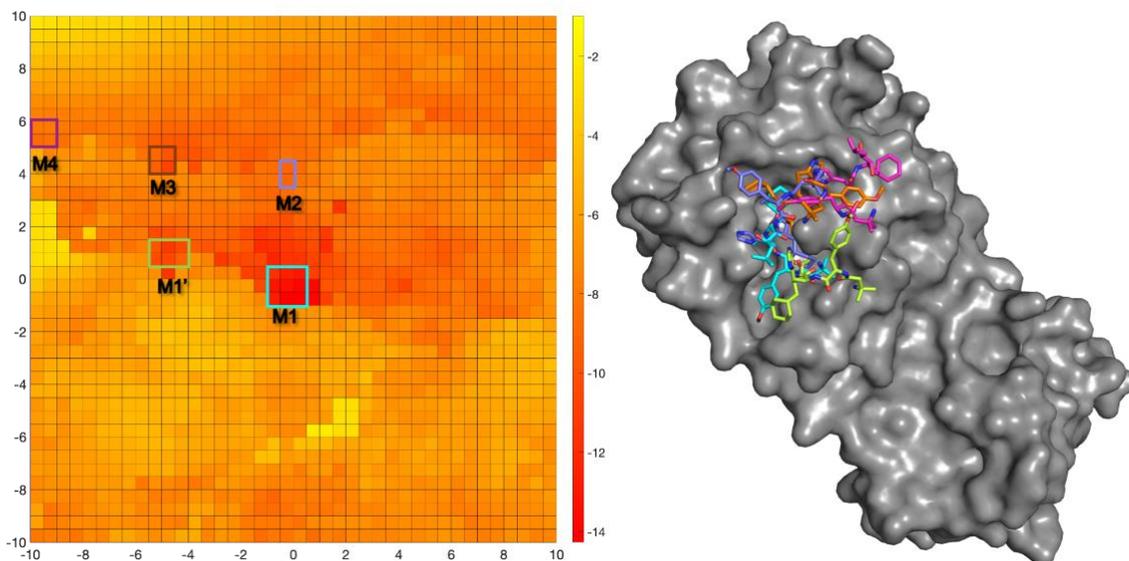
The enalkiren ligand poses fitted in the “T” shaped binding site well. Its global minimum state formed multiple critical contacts with different active regions of COVID-19 3CLpro. It formed a  $\pi$ - $\pi$  stacking with the His41 sidechain at the S1 region by inserting its imidazole group to the “T junction hole”, one hydrogen bond with the Asn142 sidechain at the S3 region, one hydrogen bond with His164 and two hydrogen bonds with Glu166 at the S4 region, as well as one hydrogen bond with Gln189 at the S5 region. Rupintrivir was investigated for the treatment of SARS<sub>34</sub> and COVID-19<sub>35</sub>. In this study, we also discovered its potential for treating COVID-19. Its global minimum state formed a  $\pi$ - $\pi$  stacking with the His41 sidechain at the S1 region using a benzene ring, one hydrogen bond

with the Gly143 backbone amide at the S3 region, and one hydrogen bond with the Gln189 sidechain. The saralasin global minimum state formed a hydrogen bond with Thr25 at S2 region, a C-H  $\cdots \pi$  interactions with His41 at the S1 region, one hydrogen bond with the His163 sidechain at the S4 region and one hydrogen bond with Gln189 sidechain at the S5 region. The TRV-120027 global minimum state bound to the S2 region by forming a hydrogen bond with the Thr25 sidechain, bound to the S3 region by forming a hydrogen bond with the Asn142 sidechain, and bound to the S4 region by forming a hydrogen bond with the His164 backbone.



**Figure 6.** The global minimum binding modes for (A) enalkiren (B) rupintrivir (C) saralasin and (D) TRV-120027 at the COVID-19 3CLpro binding site.

By analyzing the global minimum binding modes and the critical contacts formed in the top 100 binding modes, we suggested that enalkiren was the best drug candidate among the 4 drug candidates discovered through the virtual screening process. It had significantly more critical contacts number interacting with the COVID-19 3CLpro active regions compared from the top 100 binding modes and formed a more reasonable global minimum complex state compared to the other drug candidate. We further studied the binding and dissociation mechanism during the enalkiren – COVID-19 3CLpro complexing process. A 2-D free energy heatmap was generated using the CHMC protocol, showing that the enalkiren had a local minimum state M1' at the binding site near the M1 global minimum state. At the M1' state, enalkiren leaned against the S3 and S4 active region by forming two hydrogen bond with Glu166 and one hydrogen bond with His172, while broke the  $\pi$ - $\pi$  stacking interaction and exposed the His41 residue, with  $\Delta G_{M1-M1'} = 1.88$  kcal/mol. The 2-D free energy heatmap also suggested a possible dissociation path from the M1 state passing through the M2, M3, M4 metastable states. The M2 state was the intermediate state with  $\Delta G_{M1-M2} = 3.57$  kcal/mol, where enalkiren moved to the substrate's Leu-P2 binding region, surrounded by Thr24, Thr25, Thr26, Leu27, Met49, Asn142 and Cys 145 residues. The M3 state was at the edge of the binding site, where enalkiren formed a hydrogen bond with the Asn142 sidechain and Thr26 backbone, and  $\Delta G_{M1-M3} = 2.78$  kcal/mol. At the M4 state, enalkiren was fully out of the binding pocket, only forming a hydrogen bond with the Asn142 backbone, with  $\Delta G_{M1-M4} = 4.86$  kcal/mol.



**Figure 7.** Enalkiren transition free energy surface and the energy minimal states generated using the CHMC 2-D protocol. The free energy surface is generated on the  $xz$  plane centered at the global minimal histogram (0,0) and ranges  $\pm 10$  Å along both the  $x$  and  $z$  axis. Global and local minimum energy histogram regions are circled in squares. The global minimum state M1 (marked in cyan) and a close-by local minimum state M1' (marked in lime) were discovered at the binding site. Three metastable states (M2~M4, marked in orchid, brown and magentas respectively) were discovered to illustrate a possible dissociation path for enalkiren to move outside the binding site.

## Conclusion

In this work, we studied the binding site structural features of the COVID-19 3C-like Protease, and performed the drug repurposing screening against the Drugbank database looking for possible drug candidates for the COVID-19 3CLpro target. First, 70 lead compounds were selected using the ligand similarity search referencing 19 published

SARS 3CLpro inhibitors. Binding affinities for these lead compounds complexing with COVID-19 3CLpro were then evaluated using the MT free energy method, with 12 strong binders further chosen for the binding mode analysis. Eeankiren, rupintrivir, saralasin and TRV-120027 were finally selected due to their effective functional residue interference and binding site blocking in their complex states. With significantly more critical contacts number interacting with the COVID-19 3CLpro active regions and more reasonable global minimum complex state compared to the other drug candidate, enalkiren was the best drug candidate for COVID-19 3CLpro suggested by our study.

## **Method**

### **“Movable Type” Free Energy Method**

Sampling the configurational energy states for a macromolecule is quite computationally expensive given the all-atom flexibility in computation. The coupling of all atom pairwise interactions results in a huge molecular configurational sampling space. Regarding the computational cost for the molecular energy states sampling, the Movable Type (MT) method introduces the assumption that a molecular conformational energy can be approximated as a linear combination of all the atomic energies in that molecular system, so that each atom possesses independent configurational degree of freedom. The idea of the MT method is to separate the sampling of the molecular energy states into the atomic local partition function calculation of all atoms in the studied molecular system. Since the energy-state samplings for different atoms are independent from each other, the MT method uses numerical procedures instead of explicit configuration sampling to calculate the local partition function of every atom for an input molecular conformation given a

certain movement range for all atoms from their initial position, before grouping the atomic partition functions to derive the molecular local partition function for the local free energy calculation.

First, using the end-state conformations of a molecule, the MT method allows the same space of movement for all atoms, and generates atom pairwise Boltzmann factors using an group of evenly distributed discrete pairwise distance values within the movement ranges. In the MT-based atomic energy state sampling, the energy of an atom, say atom  $\alpha$ , is separated into all pairwise interactions regarding this atom  $\alpha$ . Given the identical movement space for all the atoms, every pairwise contact has the same sampling range between the studied atom  $\alpha$  and every other atom in the molecular system, say atom  $i$ . All the atom  $\alpha$ - $i$  pairwise energy states within such range can be expressed using an Boltzmann-factor vector modeled in equation 1, where  $\tau_{\alpha i}^0$  represents the relative initial coordinate of the atom pair  $\alpha$ - $i$  in the input structure, and  $\Delta\tau$  is their geometric deviation step unit with a sampling range ( $\pm n\Delta\tau$ ).

$$V_{\alpha i} = \begin{bmatrix} e^{-\beta E_{\alpha i}^n} \\ e^{-\beta E_{\alpha i}^{n-1}} \\ \vdots \\ e^{-\beta E_{\alpha i}^1} \\ e^{-\beta E_{\alpha i}^0} \\ e^{-\beta E_{\alpha i}^{-1}} \\ \vdots \\ e^{-\beta E_{\alpha i}^{-(n-1)}} \\ e^{-\beta E_{\alpha i}^{-n}} \end{bmatrix} = \begin{bmatrix} e^{-\beta E(\tau_{\alpha i}^0 + n\Delta\tau)} \\ e^{-\beta E(\tau_{\alpha i}^0 + (n-1)\Delta\tau)} \\ \vdots \\ e^{-\beta E(\tau_{\alpha i}^0 + \Delta\tau)} \\ e^{-\beta E(\tau_{\alpha i}^0)} \\ e^{-\beta E(\tau_{\alpha i}^0 - \Delta\tau)} \\ \vdots \\ e^{-\beta E(\tau_{\alpha i}^0 - (n-1)\Delta\tau)} \\ e^{-\beta E(\tau_{\alpha i}^0 - n\Delta\tau)} \end{bmatrix} \quad (1)$$

$V_{\alpha i}$  is a collection of Boltzmann factors of the atom pair  $\alpha$ - $i$ . For every pairwise contact regarding atom  $\alpha$ , its Boltzmann-factor collection can be modeled likewise, as  $V_{\alpha j}$ ,  $V_{\alpha k}$ , or  $V_{\alpha l}$ , etc. The local partition function for atom  $\alpha$  in such case contains a huge number of

energy states by merging all the pairwise Boltzmann factors matching all the sampled atomic pairwise configurations.

Generating all the configurations with respect to the local energy states for atom  $\alpha$  is quite time consuming, not to mention the molecular free energy calculation requires the energy states for all the atoms in the molecule. However, due to that Boltzmann factors for different atom pairwise contacts are independent, using the following steps, we can numerically derive the local partition function for every atom without generating these configurations.

Equation 2 shows the sum of the energy states of the atom pair  $\alpha$ -i. The multiplication of  $\mathcal{E}_{\alpha i}$  and  $\mathcal{E}_{\alpha j}$  represents the sum of all energy states for atom  $\alpha$  merging  $\alpha$ -i and  $\alpha$ -j contact potential energies according to the distributive property of multiplication.

$$\mathcal{E}_{\alpha i} = \left( e^{-\beta E(\tau_{\alpha i}^0 - n\Delta\tau)} + \dots + e^{-\beta E(\tau_{\alpha i}^0)} + \dots + e^{-\beta E(\tau_{\alpha i}^0 + n\Delta\tau)} \right) \quad (2)$$

$$\begin{aligned} \mathcal{E}_{\alpha i} \times \mathcal{E}_{\alpha j} = & \left( e^{-\beta(E_{\alpha i}^{-n} + E_{\alpha j}^{-n})} + e^{-\beta(E_{\alpha i}^{-n} + E_{\alpha j}^{-(n-1)})} + \dots + e^{-\beta(E_{\alpha i}^0 + E_{\alpha j}^0)} + e^{-\beta(E_{\alpha i}^0 + E_{\alpha j}^1)} + \right. \\ & \left. \dots + e^{-\beta(E_{\alpha i}^n + E_{\alpha j}^{n-1})} + e^{-\beta(E_{\alpha i}^n + E_{\alpha j}^n)} \right) \end{aligned} \quad (3)$$

The right-hand side of Equation 3 contains all conformational energy states of atom  $\alpha$  for the molecular system with only  $\alpha$ , i, j atoms, when atom pairwise contact energies are considered independent. Hence in a N-atom molecular system, the multiplication through the sums of N-1 pairwise contacts regarding the atom  $\alpha$  gives the conformational energy ensemble of atom  $\alpha$  for the N-atom molecular system including atom  $\alpha$  (equation 4).

$$\mathcal{E}_{\alpha} = \prod_i^{N-1} \mathcal{E}_{\alpha i} \quad (4)$$

$\mathcal{E}_\alpha$  is the local partition function within the defined movement ranges for all atoms in the molecular system. By multiplying through the local partition functions of all atoms in the molecular system, we can generate the molecular energy ensemble given the corresponding movement range for each atom (equation 5).

$$\mathcal{Z}_M = \mathcal{E}_\alpha \times \mathcal{E}_\beta \times \dots \times \mathcal{E}_n = \prod_{\xi}^N \mathcal{E}_\xi \quad (5)$$

Through equation 1-5, the MT protocol first assigns the atom pairwise configurational sampling range as  $(\pm n\Delta\tau)$  for every atom pair in the molecular system and calculates the ensemble of every atomic energy states using equation 1-4. The molecular local partition function is then calculated by merging all atomic energy ensembles using equation 5.

$\mathcal{Z}_M$  in equation 5 is the molecular local partition function regarding an initial molecular conformation given the identical atomic pairwise contact sampling range as  $(\pm n\Delta\tau)$  for every atom pair. Given that every pairwise contact has  $2n\Delta\tau$  uniformly distributed pairwise distance values, and an N-atom molecular produces a number of  $N(N-1)/2$  atom pairwise contacts,  $\mathcal{Z}_M$  contains a total number of  $(2n\Delta\tau) N(N-1)/2$  uniformly sampled molecular energy states in the aforementioned calculation.

Using this technique, a local energy ensemble centered on one initial end-state configuration can be fast collected and transformed into the local partition function calculation.

By feeding the MT protocol with multiple molecular configurations, local molecular partition functions  $\mathcal{Z}_M$  can be calculated using equation 5 for estimation of the free energy. Regarding the protein-ligand binding affinity study, conformations for both free and bound states are generated using the Monte Carlo sampling protocols followed by local minimizations. The free state molecular system includes unbound ligand and protein in the

solution phase.  $\mathbb{Z}_L$  and  $\mathbb{Z}_P$  are their corresponding partition functions which are necessary for the binding free energy calculation. On the other hand, the bound state molecular system includes the protein ligand molecules in the complex form in the solution.  $\mathbb{Z}_{PL}$  is the bound state partition function containing all the protein-ligand binding mode energy states. In the present study we only performed the ligand conformational sampling and the protein-ligand binding mode sampling by considering the flexibility of the ligand structures and the protein binding site residues while keeping the rest of the protein geometry fixed. The protein conformational sampling is skipped because (1) the massive degrees of freedom associated with inclusion of protein flexibility will significantly increase the computational burden, while (2) having limited contributions to the computational accuracies regarding relative binding affinities studies using identical or similar protein target, due to that the  $\mathbb{Z}_{PL}$  values are very similar among all the test cases.

In-house programs developed in our group are introduced to perform such tasks. For the free-state calculation, the MT-CS conformational search program<sup>32</sup> was introduced to generate significant free-state molecular conformations with reference to the molecular flexibility. The MT-CS conformational search program generated ligand conformers using a torsion library with pre-calculated torsion energies using the Amber ff14 force field, the solvation free energy was calculated using the KMTISM model.<sup>33</sup> The MT protocol was then applied to each ligand conformer to estimate the local partition function  $\mathbb{Z}_L$ . The ligand's total partition function was then generated using all the MT-CS sampled configurational ensemble energies in equation 6.

$$\mathbb{Z}_L = \sum_{\alpha}^{N_L \text{ Conformers}} \mathcal{Z}_L^{\alpha} = \mathcal{Z}_L^1 + \mathcal{Z}_L^2 + \dots \mathcal{Z}_L^N \quad (6)$$

The Heatmap docking program<sup>31</sup> was employed for the bound state configuration sampling in this work. The bound-state protein-ligand complex ensemble energy is calculated using the same protocol by summing all the local partition functions.

$$\mathbb{Z}_{PL} = \sum_{\alpha}^{N_{PL} \text{ Poses}} \mathcal{Z}_{PL}^{\alpha} = \mathcal{Z}_{PL}^1 + \mathcal{Z}_{PL}^2 + \dots + \mathcal{Z}_{PL}^N \quad (7)$$

The binding free energy change was then estimated by using the ratio of partition functions in bound and free states.

$$\Delta G_{binding} \approx -RT \log \left( \frac{\mathbb{Z}_{PL}}{\mathbb{Z}_L} \right) \quad (8)$$

All related codes and data can be obtained by contacting the authors for validation and review purpose only.

## **The “Consecutive Histogram Monte Carlo” Sampling Method**

The Consecutive Histogram Monte Carlo (CHMC) protocol generates evenly distributed receptor-ligand state ensembles that represent a reversible work pathway for the receptor-ligand dissociation. The CHMC protocol restrains the degrees of freedom of the configurational sampling during the receptor-ligand dissociation by partitioning the cavity of the receptor’s binding site into evenly distributed Monte Carlo sampling units, which then have their free energies estimated. In this section, we explain in detail the CHMC protocol that generates the 2-dimensional free energy surface with respect to the ligand transition on the receptor’s surface. This is done in a series of steps including: (1) generation of multiple Monte Carlo sampling units with reference to the receptor-ligand dissociation coordinates, (2) sampling the receptor-ligand energy states within each “sampling unit”, and (3) calculation of the ensemble energies regarding the sampling units

and generation of the free energy curve/surface. The following paragraphs describe each of these steps in detail.

### **Generation of the Multiple Monte Carlo “Sampling Units”**

The CHMC protocol first sets up a series of thin volumes as individual sampling units distributed layer-by-layer with equal thickness from a starting sampling point inside the receptor binding site to the edge of the binding site. Without applying a biasing potential to limit the ligand’s mobility, the CHMC protocol uses a regular energy function and tracks the location of the ligand’s geometric center during the sampling. In each sampling window, only the ligands’ final poses whose geometric centers are inside the sampling unit will be taken into account for the NVT local partition function calculation for the corresponding sampling unit. Hence independent configurational samplings can be performed simultaneously for all sampling units to calculate evenly distributed NVT ensembles, so as to generate the PMF curve or surface representing a reversible work pathway for the receptor-ligand dissociation process.

First, the receptor’s binding site needs to be located to start generating the sampling units. Two options are provided for binding site residue selection: (1) manual selection based on, for example, a known crystal structure of a protein-ligand complex, or (2) ligand is docked into the putative binding pocket and any residues 6 Å from ligand atoms are then defined as the binding site, or the active “pocket”.

The CHMC method is able to generate consecutive sampling bins with different shapes and volumes to satisfy different sampling needs. In this work, we apply a CHMC 2-dimensional protocol to accomplish the free energy surface calculation regarding the ligand’s transition on a receptor surface. The CHMC method creates evenly distributed 0.5



away from the protein surface are excluded to ensure the sampling of the surface region. The zoomed-in figure at the bottom left demonstrates a  $2 \text{ \AA} \times 2 \text{ \AA}$  area of the sampling histogram distribution centered at the geometric center of the receptor binding site.

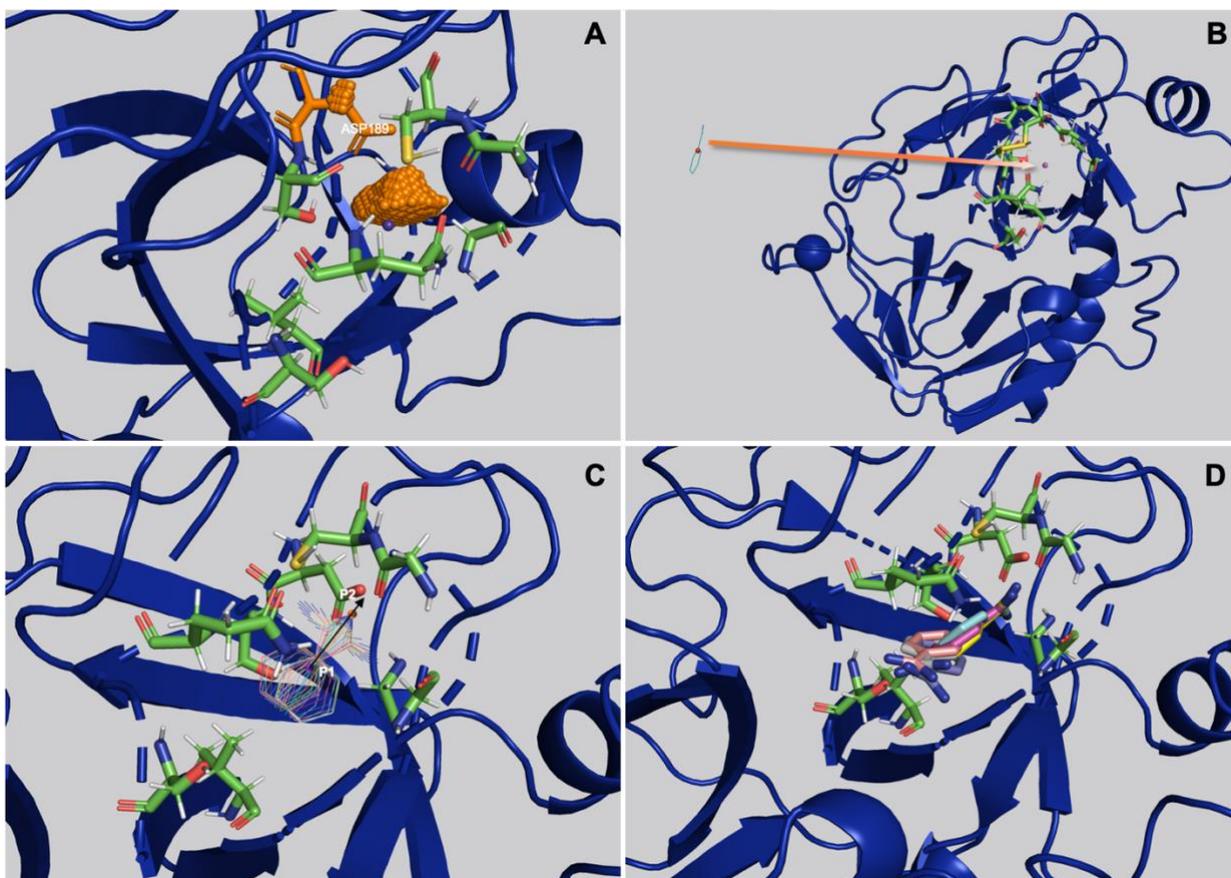
### **Placement and Refinement of the Ligand within the “Sampling Units”**

We employ a docking-like procedure for the receptor-ligand complex configurational sampling within each Monte Carlo sampling unit. The ligand placement followed by a local geometric optimization is performed to generate configurational state ensembles within all the sampling units. We have developed a heatmap-based docking protocol in this work to satisfy the need for the ligand’s local configurational search within each Monte Carlo sampling unit at the binding site in this study.

The heatmap-based docking program performs a two-step docking procedure, including (1) an initial ligand-to-pocket placement using a point-to-point coordinate superimposition coupled with a rotational search, and (2) local optimizations for the final binding mode generation.

The first step deposits the ligand onto the selected receptor binding site grid point using point-to-point coordinate superimposition, which places the coordinate of the geometric center of the ligand onto every pre-placed grid point in each sampling shell. Rotational search of the ligand is then performed to optimize the ligands interactions with the receptor referencing a set of pre-generated heatmap grid points surrounding every pocket residue. Next the geometric center of the ligand (labelled P1), and select a heatmap grid point (labelled P2) are used to define a p1->p2 vector around which a 360 degree rotational search is carried out, yielding initial poses for the subsequent optimization procedure. The

top 50 poses, as defined by their interaction energy with the receptor, are kept and passed onto the next computational stage. The last step of the calculation employs the Newton-Raphson method for binding mode optimization against all the initial poses. The goal here is to generate an appropriate ensemble of receptor-ligand complex states for each sampling unit. We employ the GARF energy function recently developed in our lab<sup>43</sup> to calculate the interatomic energies between the probe atoms with the receptor and the geometric optimization. The heatmap docking protocol is illustrated in Figure 9.



**Figure 9.** Illustration of the mechanism of the Heatmap Docking protocol using the 1bty trypsin-benzamidine complex structure as an example. (A) the active residues at the receptor's binding site (shown as licorice) are selected using a pre-docked ligand. The purple dot inside the binding pocket is a grid point from one of the sampling units serving

as an “anchor point” to place the ligand geometric center. The orange dots are the “heatmap” grid points near the carboxyl group of the Asp189 residue (marked in orange color) showing regions where hydrogen bond with the ligand might be possible. (B) The first step of the docking protocol is to deposit the ligand (structure shown as thin lines) to the binding site by placing the ligand’s geometric center (the red dot) on the anchor point (the purple dot) (C) The ligand is rotated by placing one nitrogen on the benzamidine group to one of the heatmap grid points near the Asp189 residue. Then the ligand energy states are evaluated by rotating it around the vector from P1, the geometric center of the binding site (the purple dot) to P2, the selected heatmap grid point (the orange dot). (D) The top 50 initial receptor-ligand poses selected from step C are optimized to generate the final binding modes (6 binding modes are generated and shown in different colors).

### **Ensemble Energy Calculation**

For every receptor’s binding site structure, the CHMC protocol sets up the sampling units with the same volume, creating a set of evenly spaced NVT ensembles to estimate the Helmholtz free energy changes along the dissociation pathway. In this work, we calculate the Helmholtz free energy changes between sampling units using the ratio of partition functions:

$$\Delta A_{\xi \rightarrow \xi+1} = -RT \log \left( \frac{Q_{\xi+1}}{Q_{\xi}} \right) \quad (9)$$

where  $Q_{\xi}$  and  $Q_{\xi+1}$  are the calculated Boltzmann factor ensembles for the sampling unit  $\xi$  and  $\xi+1$ .

$$Q_{\xi} = \sum_i^{n_i} e^{-\beta U_i(\xi)} \quad (10)$$

$U_i(\xi)$  is the potential energy of each sampled complex configurational state. The docking protocol generates and keeps the top 50 receptor-ligand poses for each sampling unit according to the calculated binding energy. Each of the ensembles is calculated using the sum of the Boltzmann weighted energies for all the sampled configurational states.

The sampling starts from the sampling unit at the geometric center of the binding pocket across the binding site until the Helmholtz free energy change converges ( $\Delta A_{\xi \rightarrow \xi+1} < 0.1$  kcal/mol for 3 consecutive sampling units). In this work the ligand is recognized as the unbound state.

$$\Delta A_{\xi \rightarrow \xi+1} < 0.1 \text{ kcal/mol} \Rightarrow Q_u = Q_{\xi} \quad (11)$$

Then the binding free energy is calculated using the ratio of partition function for the energy minimum histogram to the partition function for the unbound state.

$$\Delta A_{binding} = -RT \log \left( \frac{Q_{min}}{Q_u} \right) \quad (12)$$

Similar to the umbrella sampling method, the CHMC method defines a reaction coordinate prior to the simulation. On the other hand, there are no weighting functions applied to the configurations along the reaction coordinate in the CHMC method, while it requires that the geometric centers of the sampled ligand configurations are limited to staying inside each of the sampling units during the placement and optimization procedures. Ligand configurational sampling for each sampling unit is performed and  $Q_{\xi}$  is hence calculated using equation 10 until its value converges. Calculations are performed within all the sampling units separately to generate the final energy landscape.

## Reference

- (1) Hilgenfeld, R. From SARS to MERS: crystallographic studies on coronaviral proteases enable antiviral drug design, *FEBS Journal*, **2014**, 281, 4085–4096.
- (2) Kumar, V., Tan, K., Wang, Y., Lin, S., Liang, P. Identification, synthesis and evaluation of SARS-CoV and MERS-CoV 3C-like protease inhibitors. *Bioorganic & Medicinal Chemistry*, **2016**, 24, 3035–3042.
- (3) Yan, R. H., Zhang, Y. Y., Li Y. N., Xia, L., Guo, Y. Y., Zhou, Q. Structural basis for the recognition of SARS-CoV-2 by full-length human ACE2. *Science*. **2020**, 367, 1444–1448.
- (4) Anand, K., Ziebuhr, J., Wadhvani, P., Mesters, J. R., Hilgenfeld, R. Coronavirus Main Proteinase (3CLpro) Structure: Basis for Design of anti-SARS Drugs. *Science*. **2003**, 300, 1763–176.
- (5) Luhua Lai, L. H., Han, X. F., Chen, H., Wei, P., Huang, C. K., Liu, S. Y., Fan, K. Q., Zhou, L., Liu, Z. M., Pei, J. F., Liu, Y. Quaternary Structure, Substrate Selectivity and Inhibitor Design for SARS 3C-Like Proteinase. *Current Pharmaceutical Design*, **2006**, 12, 4555–4564.
- (6) Wei, P., Fan, K. Q., Chen, H., Ma, L., Huang, C. K., Tan, L., Xi, D., Li, C. M., Liu, Y., Cao, A. N., Lai, L. H. The N-terminal octapeptide acts as a dimerization inhibitor of SARS coronavirus 3C-like proteinase. *Biochemical and Biophysical Research Communications*, **2006**, 339, 865–872.
- (7) Hsu, J. T. -A., Kuo, C., Hsieh, H., Wang, Y., Huang, K., Lin, C. P. -C., Huang, P., Chen, X., Liang, P. *FEBS Letters*, **2004**, 574, 116–120.

- (8) Chuck, C., Chow, H., Wan, D. C., Wong, K. Profiling of Substrate Specificities of 3C-Like Proteases from Group 1, 2a, 2b, and 3 Coronaviruses. *PLoS One*. **2011**, 6, e27228.
- (9) Ren, Z. L., Yan, L. M., Zhang, N., Guo, Y., Yang, C., Lou, Z. Y., Rao, Z. H. The newly emerged SARS-Like coronavirus HCoV-EMC also has an “Achilles’ heel”: current effective inhibitor targeting a 3C-like protease. *Protein & Cell*, **2013**, 4, 248–250.
- (10) Abuhammad, A., Al-Aqtash, R. A., Anson, B. J., Taha, M. O. Computational modeling of the bat HKU4 coronavirus 3CLpro inhibitors as a tool for the development of antivirals against the emerging Middle East respiratory syndrome (MERS) coronavirus. *J. Mol. Recognit.* **2017**, 30, e2644.
- (11) Regniera, T., Sarma, D., Hidaka, K., Bacha, U., Freire, E., Hayashi, Y., Kiso, Y. New developments for the design, synthesis and biological evaluation of potent SARS-CoV 3CLpro inhibitors. *Bioorganic & Medicinal Chemistry Letters*. **2009**, 19, 2722–2727.
- (12) Kumara, V., Shin, J. S., Shie, J., Ku, K. B., Kim, C. S., Go, Y. Y., Huang, K., Kim, M. H., Liang, P. Identification and evaluation of potent Middle East respiratory syndrome coronavirus (MERS-CoV) 3CLPro inhibitors. *Antiviral Research*. **2017**, 141, 101–106.
- (13) Yang, H. T., Yang, M. J., Ding Y., Liu, Y. W., Lou, Z. Y., Zhou, Z., Sun L., Mo, L. J., Ye, S., Pang, H., Gao, G. F., Anand, K., Bartlam, M., Hilgenfeld, R., Rao, Z. H. The crystal structures of severe acute respiratory syndrome virus main protease and its complex with an inhibitor. *PNAS*, **2003**, 100, 13190–13195.

- (14) Bonanno, J.B., Fowler, R., Gupta, S., Hendle, J., Lorimer, D., Romero, R., Sauder, J.M., Wei, C.L., Liu, E.T., Burley, S.K., Harris, T. X-Ray Crystal Structure of the SARS Coronavirus Main Protease. **2003**.
- (15) Jin, Z. M., Du, X. Y., Xu, Y. C., Deng, Y. Q., Liu, M., Q., Zhao, Y., Zhang, B., Li, X. F., Zhang, L. K., Peng, C., Duan, Y. K., Yu, J., Wang. L., Yang, K. L., Liu, F. J., Jiang, R. D., Yang, X. L., You, T., Liu, X. C., Yang, X. N., Bai, F., Liu, H., Liu, X., Guddat, L. W., Xu, W. Q., Xiao, G. F., Qin, C. F., Shi, Z. L., Jiang, H. L., Rao, Z. H., Yang, H. T. Structure of Mpro from COVID-19 virus and discovery of its inhibitors. *BioXRiv*, **2020**, <https://doi.org/10.1101/2020.02.26.964882>.
- (16) Boratyn, G. M., Schäffer, A. A., Agarwala, R., Altschul, S. F., Lipman, D. J., Madden, T. L. Domain enhanced lookup time accelerated BLAST. *Biol. Direct.* **2012**, 7, 12.
- (17) Bacha, U., Barrila, J., Velazquez-Campoy, A., Leavitt, S. A., Freire, E. Identification of Novel Inhibitors of the SARS Coronavirus Main Protease 3Clpro. *Biochemistry*, **2004**, 43, 4906–4912.
- (18) Xia, B., Kang, X. Activation and maturation of SARS-CoV main protease. *Protein Cell*, **2011**, 2, 282–290.
- (19) Lim, L. Z., Shi, J. H., Mu, Y. G., Song, J. X. Dynamically-Driven Enhancement of the Catalytic Machinery of the SARS 3C-Like Protease by the S284- T285-I286/A Mutations on the Extra Domain. *PLoS ONE*, **2014**, 9, e101941.
- (20) Shi, J. H., Han, N. Y., Lim, L. Z., Lua, S. X., Sivaraman, J., Wang, L. S., Mu, Y. G., Song, J. X. Dynamically-Driven Inactivation of the Catalytic Machinery of the

SARS 3C-Like Protease by the N214A Mutation on the Extra Domain. *PLoS Computational Biology*. **2011**, 7, e1001084.

- (21) Muramatsu, T., Takemoto, C., Kim, Y., Wang, H. F., Nishii, W., Terada, T., Shirouzu, M., Yokoyama, S. SARS-CoV 3CL protease cleaves its C-terminal autoprocessing site by novel subsite cooperativity. *PNAS*, **2016**, 113, 12997–13002.
- (22) Liu, Z. M., Huang, C. K., Fan, K. Q., Wei, P., Chen, H., Liu, S. Y., Pei, J. F., Shi, L., Li, B., Yang, K., Liu, Y., Lai, L. H. Virtual Screening of Novel Noncovalent Inhibitors for SARS-CoV 3C-like Proteinase. *J. Chem. Inf. Model.* **2005**, 45, 10–17.
- (23) Kim, Y. J., Lovell, S., Tiew, K., Mandadapu, S. R., Alliston, K. R., Battaile, K. P., Groutas, W. C., Changa K. Broad-Spectrum Antivirals against 3C or 3C-Like Proteases of Picornaviruses, Noroviruses, and Coronaviruses. *Journal of Virology*, **2012**, 86, 11754–11762.
- (24) Tsai, K., Chen, S., Liang, P., Lu, I., Mahindroo, N., Hsieh, H., Chao, Y., Liu, L., Liu, D., Lien, W., Lin, T., Wu, S. Discovery of a Novel Family of SARS-CoV Protease Inhibitors by Virtual Screening and 3D-QSAR Studies. *J. Med. Chem.* **2006**, 49, 3485–3495.
- (25) Zhou, L., Liu, Y., Zhang, W. L., Wei, P., Huang, C. K., Pei, J. F., Yaxia Yuan, Y. X., Lai, L. H. Isatin Compounds as Noncovalent SARS Coronavirus 3C-like Protease Inhibitors. *J. Med. Chem.* **2006**, 49, 3440–3443.
- (26) Chen, L., Wang, Y., Lin, Y. W., Chou, S., Chen, S., Liu, L. T., Wu, Y., Kuo, C., Chen, T. S., Juanga, S. Synthesis and evaluation of isatin derivatives as effective SARS coronavirus 3CL protease inhibitors. *Bioorganic & Medicinal Chemistry Letters*, **2005**, 15, 3058–3062.

- (27) Mukherjee, P., Shah, F., Desai, P., Avery, M. Inhibitors of SARS-3CLpro: Virtual Screening, Biological Evaluation and Molecular Dynamics Simulation Studies. *J. Chem. Inf. Model.* **2011**, 51, 1376–1392.
- (28) Yang, S., Chen, S., Hsu, M., Wu, J., Tseng, C. K., Liu, Y., Chen, H., Kuo, C., Wu, C., Chang, L., Chen, W., Liao, S., Chang, T., Hung, H., Shr, H., Liu, C., Huang, Y., Chang, L., Hsu, J., Peters, C. J., Wang, A. H. -J., Hsu, M. Synthesis, Crystal Structure, Structure-Activity Relationships, and Antiviral Activity of a Potent SARS Coronavirus 3CL Protease Inhibitor. *J. Med. Chem.* **2006**, 49, 4971–4980.
- (29) Shie, J., Fang, J., Kuo, C., Kuo, T., Liang, P., Huang, H., Yang, W., Lin, C., Chen, J., Wu, Y., Wong, C. Discovery of Potent Anilide Inhibitors against the Severe Acute Respiratory Syndrome 3CL Protease. *J. Med. Chem.* **2005**, 48, 4469–4473.
- (30) Zheng, Z., Ucisik, M.N., Merz, K.M., Jr. The Movable Type Method Applied to Protein–Ligand Binding. *J. Chem. Theory Comput.* **2013**, 9, 5526–5538.
- (31) Bansal, N., Zheng, Z., Song, L.F., Pei, J., Merz, K.M., Jr. The Role of the Active Site Flap in Streptavidin/Biotin Complex Formation. *J. Am. Chem. Soc.* **2018**, 140, 5434–5446.
- (32) Pan, L., Zheng, Z., Wang, T., Merz Jr., K. M., A Free Energy Based Conformational Search Algorithm Using the “Movable Type” Sampling Method. *J. Chem. Theory Comput.* **2015**, 11, 5853.
- (33) Zheng, Z.; Wang, T.; Li, P.; Merz, K. M. Jr. KECSA-Movable Type Implicit Solvation Model (KMTISM). *J. Chem. Theory Comput.* **2014**, 11, 667–682.

- (34) 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.
- (35) Liu, C., Zhou, Q., Li, Y., Garner, L. V., Watkins, S. P., Carter, L. J., Smoot, J., Gregg, A. C., Daniels, A. D., Jervey, S., Albaiu, D. Research and Development on Therapeutic Agents and Vaccines for COVID-19 and Related Human Coronavirus Diseases. *ACS Central Science*. **2020**, 6, 315–331.