

Received  
Accepted

DOI:

# Binding of Inhibitors to the Monomeric and Dimeric SARS-CoV-2 Mpro

Nguyen Minh Tam,<sup>ab</sup> Pham Cam Nam,<sup>c</sup> Duong Tuan Quang,<sup>d</sup> Nguyen Thanh Tung,<sup>ef</sup> and Van V. Vu,<sup>g</sup> and Son Tung Ngo<sup>bh\*</sup>

SARS-CoV-2 rapidly infects millions of people worldwide since December 2019. There is still no effective treatment for the virus, resulting in the death of more than one million of patients. Inhibiting the activity of SARS-CoV-2 main protease (Mpro), 3C-like protease (3CLP), is able to block the viral replication and proliferation. In this context, our study has revealed that in silico screening for inhibitors of SARS-CoV-2 Mpro can be reliably done using the monomeric structure of the Mpro instead of the dimeric one. Docking and fast pulling of ligand (FPL) simulations for both monomeric and dimeric forms correlate well with the corresponding experimental binding affinity data of 30 compounds. The obtained results were also confirmed via binding pose and noncovalent contact analyses. Our study results show that it is possible to speed up computer-aided drug design for SARS-CoV-2 Mpro by focusing on the monomeric form instead of the larger dimeric one.

## Introduction

The novel coronavirus (2019-nCoV or SARS-CoV-2), a member of the Coronaviridae virus family, has been reported to be able to spread among humans.<sup>1</sup> The virus initially appeared the first case since December 2019 in Wuhan, Hubei province, China.<sup>2-4</sup> It shares more than 82% identical RNA genome to the SARS-CoV, SARS-CoV-2 severe cases of respiratory syndromes.<sup>5</sup> Although the bat has been thought of as the original reservoir, the intermediate host is still unknown.<sup>6</sup> Moreover, it is known that the SARS-CoV-2 can endure in aerosol for more than 3 hours,<sup>7</sup> which may be a major factor behind the outbreak of COVID-19 pandemic, which has caused several hundred thousands of deaths worldwide.<sup>5</sup> Therefore, the COVID-19 pandemic becomes an urgency for community health, which requires to develop an effective treatment or vaccine immediately.

Coronaviruses genomes occupy ca. 26-32 kb in length that is the largest sequence among RNA viruses.<sup>8, 9</sup> The SARS-CoV-2 genome encodes more than 20 various structural and non-structural proteins. Particularly, the SARS-CoV-2 main protease (Mpro), 3C-like protease (3CLP), is one of the most important viral enzymes, having more than

96% similarity with SARS-CoV 3CLP.<sup>9, 10</sup> SARS-CoV-2 Mpro cleaves nascent polypeptides, which are produced by the translation of the viral RNA. During this process, 11 non-structural polypeptides are auto-cleaved to become polypeptides, which are required for the viral replication and transcription.<sup>9</sup> Therefore, SARS-CoV-2 Mpro turns out to be an attractive target for antiviral drug aiming since blocking viral protease can inhibit viral replication and proliferation.<sup>10, 11</sup> Numerous investigations following this strategy have been carried out and shown some initial success.<sup>12-18</sup> However, unfortunately, an effective drug for COVID-19 is still unavailable until the date.

Currently, it should be noted that the time and cost to advance a drug has been significantly decreased by using the power of computational approaches.<sup>19-22</sup> Normally, the binding free energy,  $\Delta G$ , between a ligand and an enzyme can be probed via computational approaches. The  $\Delta G$  is associated with experimental inhibition constant,  $k_i$ , via formula  $\Delta G_{\text{bind}} = RT \ln(k_i)$ , where  $R$  is gas constant,  $T$  is absolute temperature, and  $k_i$  is a critical metric revealing the nature of binding between two biomolecules.<sup>19</sup> Accurate assessment of the ligand-binding free energy is very important in computer-aided drug design (CADD) problem.<sup>23</sup>

In addition, it should be noted that the dimer was shown to be the biologically active form of the SARS-CoV-2 Mpro but the interface does not contain a ligand-binding pocket.<sup>12</sup> An important question which arise is that can we use monomeric form of SARS-CoV-2 Mpro as inhibitor-screening target instead of the dimeric one to reduce CPU time consumption? Therefore, in this context, the binding free energy of 30 available inhibitors<sup>12-18</sup> to the monomeric and dimeric SARS-CoV-2 Mpro was examined via docking and FPL schemes. The similar of correlation coefficients between computational and experimental values of monomeric and dimeric systems suggests that we can use the monomeric form of SARS-CoV-2 Mpro as CADD target instead of the dimeric form. The obtained results can be beneficial to the COVID-19 therapy by speeding up CADD progression.

<sup>a</sup> Computational Chemistry Research Group, Ton Duc Thang University, Ho Chi Minh City, Vietnam

<sup>b</sup> Faculty of Applied Sciences, Ton Duc Thang University, Ho Chi Minh City, Vietnam

<sup>c</sup> Department of Chemistry, The University of Danang, University of Science and Technology, Danang, Vietnam

<sup>d</sup> University of Education, Hue University, Vietnam

<sup>e</sup> Institute of Materials Science, Vietnam Academy of Science and Technology, Hanoi, Vietnam

<sup>f</sup> Graduate University of Science and Technology, Vietnam Academy of Science and Technology, Hanoi, Vietnam

<sup>g</sup> NTT Hi-Tech Institute, Nguyen Tat Thanh University, Ho Chi Minh City, Vietnam

<sup>h</sup> Laboratory of Theoretical and Computational Biophysics, Ton Duc Thang University, Ho Chi Minh City, Vietnam. Email: [ngosontung@tdtu.edu.vn](mailto:ngosontung@tdtu.edu.vn)

Electronic Supplementary Information (ESI) available include the protein-ligand interaction diagram, pulling forces and works over 8 independent trajectories of SMD simulations

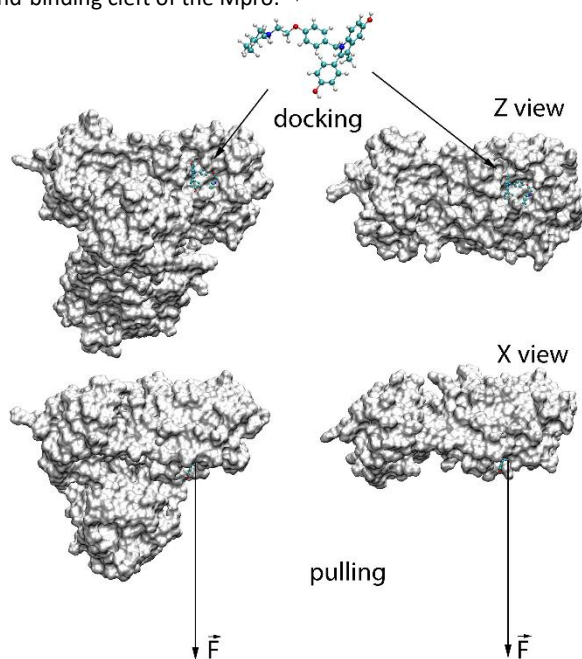
## Materials and Methods

### Structure of Inhibitors and SARS-CoV-2 Mpro

Three-dimensional structures of the monomeric and dimeric SARS-CoV-2 Mpro were copied from the Protein Data Bank with ID 6Y2F<sup>12</sup> and 6XBG,<sup>24</sup> respectively. Inhibitor shapes were taken from the PubChem database.<sup>25</sup> The ligand protonation state was assessed by using a webserver, [www.chemicalize.com](http://www.chemicalize.com), which is an tool of ChemAxon. The ligand structure was first optimized using quantum mechanics (QM) simulation with the B3LYP functional at 6-31G(d) level of basis set.

### Molecular Docking Simulations

The binding position and affinity of ligands to the monomeric and dimeric SARS-CoV-2 Mpro were probed via the Autodock Vina package (cf. Figure 1).<sup>26</sup> The docking parameter was selected referring to the previous study,<sup>27-29</sup> in which the exhaustiveness is of 8. The obtained-docking result was chosen as the highest binding affinity conformations. The grid center was selected as the geometric center of the  $\alpha$ -ketoamide **13b** and UAW246 compounds, which correspond to the monomeric and dimeric Mpro, respectively.<sup>12, 24</sup> The grid size was chosen as  $24 \times 24 \times 24$  Å, which entirely cover the ligand-binding cleft of the Mpro.<sup>28, 29</sup>



**Figure 1.** Computational scheme for evaluation of the ligand-binding affinity to the monomeric and dimeric SARS-CoV-2 Mpro.

### Steered-Molecular Dynamics Simulations

GROMACS version 5.1.3<sup>30</sup> was appointed to imitate the solvated complex involving the ligand and monomeric/dimeric SARS-CoV-2 Mpro. The protease and inhibitor were topologized via the Amber99SB-ILDN<sup>31</sup> and general Amber force field (GAFF),<sup>32</sup> correspondingly. It should be recorded that the ligand parameterization was completed by using AmberTools18 and ACPYPE approaches.<sup>33, 34</sup> In particular, the inhibitor atomic charges were assigned via the Restrained Electrostatic Potential (RESP) method<sup>32</sup> through QM investigation at the level of B3LYP/6-31G(d,p). During QM simulation, the implicit water model,  $\epsilon = 78.4$ , was involved. The

monomeric and dimeric SARS-CoV-2 Mpro + inhibitor were inserted into a rectangular periodic boundary condition (rPBC) box with a dimension of (9.8, 5.9, 8.7) and (9.4, 9.0, 12.1) nm, respectively. The corresponding box volumes of the monomeric and dimeric systems are 506.28 and 1013.82 nm<sup>3</sup>, respectively. Therefore, the total atoms of these systems approximately are 50 000 and 100 000 atoms, respectively.

The atomistic simulation was realized by utilizing the parameters signified to the prior appraisals.<sup>28, 29</sup> Particularly, the MD time step is 2 fs. The noncovalent pair was affected within a radius of 0.9 nm. The electrostatics interaction was assessed implementing the fast Particle-Mesh Ewald electrostatics scheme.<sup>35</sup> The SARS-CoV-2 + inhibitor was then optimized and equalized throughout the EM, NVT, and NPT imitations. The NVT and NPT imitations were operated during intervals of 0.1 and 2.0 ns, correspondingly. Moreover, the SARS-CoV-2 Mpro  $C_{\alpha}$  atoms were restrained during these imitations via a small harmonic force with a value of 1000 kJ mol<sup>-1</sup> nm<sup>-2</sup> per proportions. The relaxed conformation of the SARS-CoV-2 Mpro + inhibitor was then employed as initial structure of FPL simulation. During which, the inhibitor was pulled out of the binding cleft under effect of an externally harmonic force with parameters of  $k = 0.005$  nm ps<sup>-1</sup> and  $v = 600$  kJ mol<sup>-1</sup> nm<sup>-2</sup> for pulling speed and cantilever spring constant (cf. Figure 1), respectively.<sup>29</sup> Totally, 8 independent trajectories were carried out to assess the ligand-binding affinity.

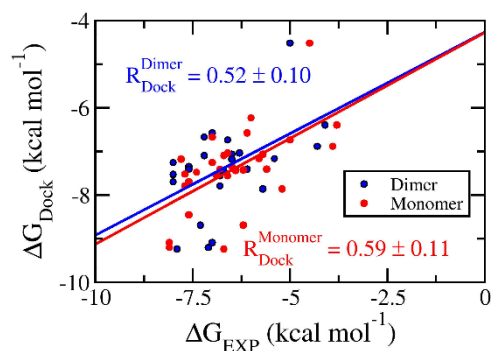
### Analyzed Tools

A intermolecular nonbonded contact was enumerated when the minimum distance of nonhydrogen atoms of a residue to the inhibitor was smaller than 0.45 nm. A hydrogen bond was enumerated when the angle between donor, D,-hydrogen, H,- acceptor, A, is larger than 135° and the distance between D and A is lesser than 0.35 nm. The error of computations was computed through 1000 rounds of the bootstrapping method.<sup>36</sup>

## Results and Discussion

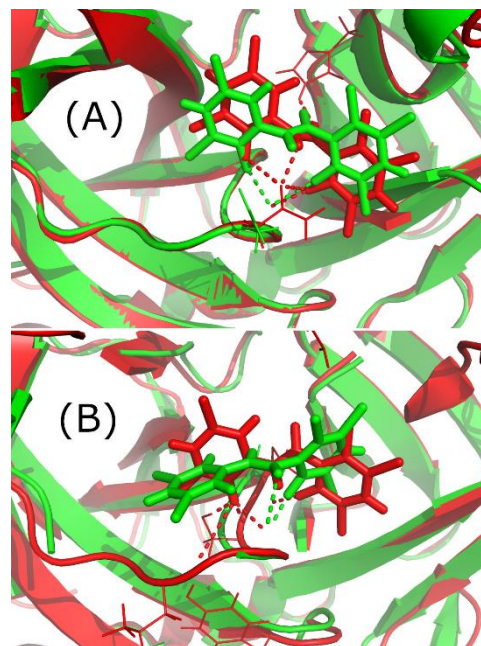
### Docking Calculations

The preliminary investigations of binding pose and affinity of the trial inhibitors to the monomeric and dimeric SARS-CoV-2 Mpro were initially estimated by a molecular docking method. Autodock Vina,<sup>26</sup> a very efficient molecular docking approach with a successful-docking rate up to 81 %, <sup>27</sup> would be able to complete this task. We have thus docked 30 available inhibitors to the monomeric and dimeric SARS-CoV-2 Mpro using Autodock Vina referring to the previous study.<sup>28, 29</sup> By using exhaustiveness 8 as suggested in the previous work,<sup>27</sup> the results were rapidly obtained in few hours (Table 1 and Table S1 of the Supplementary – ESI file). Interestingly, the correlation coefficient between docking and experimental affinities of the monomeric target,  $R_{\text{Dock}}^{\text{Monomer}} = 0.59 \pm 0.11$ , is slightly larger than that of the dimeric target,  $R_{\text{Dock}}^{\text{Dimer}} = 0.52 \pm 0.10$  (cf. Figure 2). The root mean square error (RMSE) between calculated and experimental values also indicates good consistency between monomer and dimer docking results. In particular, the monomeric system gives a value of  $RMSE_{\text{Dock}}^{\text{Monomer}} = 1.17 \pm 0.15$  kcal mol<sup>-1</sup> and the dimeric system adopts a metric of  $RMSE_{\text{Dock}}^{\text{Dimer}} = 1.35 \pm 0.18$  kcal mol<sup>-1</sup>. It should be noted that the computed error bars was obtained via 1000 rounds of the bootstrapping method.<sup>36</sup>



**Figure 2.** Correlation between docking and experimental binding free energy. Computational results were obtained using Autodock Vina. The experimental binding free energies were estimated using IC50 value<sup>12-18</sup> as an approximation for the inhibition constant  $k_i$ . The computed error was attained via 1000 rounds of the bootstrapping method.<sup>36</sup>

Furthermore, the binding pose of inhibitors to the monomeric and dimeric SARS-CoV-2 Mpro is in good agreement together since espousing the root-mean-square deviation (RMSD) of  $0.19 \pm 0.02$  nm (cf. Figure 3A and Table S1 of the ESI). It should be noted that the RMSD of the ligand-binding poses, which is smaller than 0.20 nm, normally counted as the conformations locating in the same cluster. The structural observation is thus confirmed the obtained docking energy above.



**Figure 3.** The superposition of the Oxyclozanide in the binding mode with the monomeric and dimeric SARS-CoV-2 Mpro. (A) the docking pose. (B) the MD-refined binding pose. In particular, the green and red colors mentioned the monomeric and dimeric complexes, correspondingly.

**Table 1.** Computed values of docking energy in comparison with experiments.

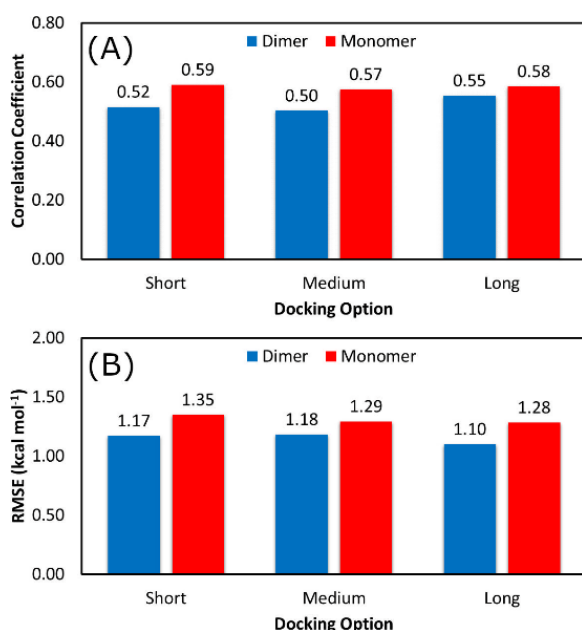
Nº	Name	$\Delta G_{Dock}^{Monomer}$			$\Delta G_{Dock}^{Dimer}$			$\Delta G_{EXP}^a$
		short	medium	long	short	medium	long	
1	11r	-6.7	-6.4	-6.3	-7.9	-8.1	-8.3	-9.23
2	13a	-7.6	-7.6	-7.6	-8.0	-7.8	-7.8	-7.70
3	13b	-7.6	-7.8	-7.8	-7.6	-7.1	-7.8	-8.45
4	Bazedoxifene	-7.4	-7.5	-7.4	-7.4	-7.4	-7.5	-7.48
5	Calpain inhibitor XII	-6.2	-6.3	-6.3	-7.3	-7.3	-7.2	-8.69
6	Carmofur	-5.2	-5.5	-5.6	-5.7	-5.8	-6.1	-7.86
7	Chloroquine	-5.0	-5.3	-5.1	-6.6	-6.6	-6.6	-6.74
8	Cyclosporine	-5.8	-5.7	-5.7	-5.4	-5.4	-5.4	-7.17
9	Digitoxin	-8.1	-8.1	-8.2	-7.0	-7.0	-7.2	-9.09
10	Digoxin	-8.1	-8.1	-8.1	-7.1	-7.2	-7.2	-9.20
11	Dihydrogambogic Acid	-7.0	-7.0	-7.0	-7.2	-7.2	-7.2	-6.67
12	Disulfiram	-3.9	-3.8	-3.9	-4.3	-4.1	-4.1	-6.89
13	Ebastine	-5.7	-6.5	-6.1	-6.5	-6.3	-6.4	-7.06
14	Favipiravir	-4.5	-4.8	-4.8	-5.0	-5.0	-5.0	-4.52
15	Fluspirilene	-6.9	-7.2	-7.3	-8.0	-7.7	-7.6	-7.53
16	Isoosajin	-7.7	-7.7	-7.7	-8.0	-8.0	-8.0	-7.52
17	Ivacaftor	-6.7	-6.7	-6.7	-7.2	-7.6	-7.5	-7.10
18	Lusutrombopag	-6.2	-6.1	-6.8	-6.4	-6.5	-6.3	-7.42
19	Mefloquine	-6.5	-6.5	-6.5	-7.6	-7.7	-7.6	-7.34
20	Mequitazine	-6.6	-6.6	-6.6	-6.3	-6.3	-6.3	-7.03
21	MG-132	-5.6	-6.2	-6.2	-6.1	-5.8	-6.2	-7.41
22	Narlaprevir	-7.8	-7.5	-7.4	-6.5	-6.9	-6.8	-7.18
23	Osajin	-6.8	-6.9	-6.8	-7.6	-8.0	-8.0	-7.41
24	Oxyclozanide	-6.4	-6.4	-6.4	-6.7	-6.7	-6.7	-7.44
25	Penfluridol	-7.0	-6.9	-6.9	-8.0	-8.2	-8.2	-7.26
26	Phenazopyridine	-6.0	-6.0	-6.0	-6.0	-6.0	-6.0	-6.23
27	Proscarlardin	-7.7	-7.7	-7.7	-6.8	-7.3	-7.3	-7.79
28	PX-12	-3.8	-3.8	-3.8	-4.1	-4.2	-4.5	-6.39
29	Shikonin	-6.1	-6.1	-6.1	-7.0	-6.9	-6.9	-6.58
30	Tetrandrine	-6.6	-6.6	-6.6	-6.8	-6.8	-6.8	-7.56

<sup>a</sup>The experimental binding free energies were gained based on IC50 value,<sup>12-18</sup> approximating that the one equals to the inhibition constant  $k_i$ . The unit is of kcal mol<sup>-1</sup>.

The molecular docking with larger exhaustiveness, which selected as 56 and 400 according to the previous study,<sup>27</sup> were also performed in order to validate the convergence of the docking scheme. In total

we used three different values of exhaustiveness including 400, 56, and 8 which are denoted as *long*, *medium*, and *short* options, respectively. The accuracies of the docking simulations for monomer

and dimer with respect to experiment are shown in Figure 3 Figure 4. Interestingly, changing the docking exhaustiveness parameter from *short* to *medium* and/or *long* does not have a significant impact on the correlation coefficient and RMSE, which is consistent with the prior benchmark<sup>27</sup>. In particular, the correlation coefficients slightly change to  $R_{\text{Dock}}^{\text{Monomer}} = 0.57 \pm 0.11$  and  $R_{\text{Dock}}^{\text{Dimer}} = 0.50 \pm 0.11$  matching with the *medium* option Figure 3A(cf. Figure 4A). The metrics are of  $R_{\text{Dock}}^{\text{Monomer}} = 0.58 \pm 0.12$  and  $R_{\text{Dock}}^{\text{Dimer}} = 0.55 \pm 0.10$  resembling the *long* option (Figure 4A). Moreover, the calculated accuracy is also associated with the RMSE value. Absolutely, within computed error, the RMSE was unchanged over the docking options *short*, *medium*, and *long* with amounts of  $RMSE_{\text{Dock}}^{\text{Dimer}} = 1.17 \pm 0.15$ ,  $RMSE_{\text{Dock}}^{\text{Dimer}} = 1.18 \pm 0.15$ , and  $RMSE_{\text{Dock}}^{\text{Dimer}} = 1.10 \pm 0.15$  kcal mol<sup>-1</sup> for dimeric systems and  $RMSE_{\text{Dock}}^{\text{Monomer}} = 1.35 \pm 0.18$ ,  $RMSE_{\text{Dock}}^{\text{Monomer}} = 1.29 \pm 0.19$ , and  $RMSE_{\text{Dock}}^{\text{Monomer}} = 1.28 \pm 0.19$  kcal mol<sup>-1</sup> for monomeric systems, respectively (cf. Figure 4B). Overall, the docking simulations provide slightly accurate results for monomeric systems than for the dimeric systems.



**Figure 4.** Correlation and RMSE values between calculated and experimental binding affinity.

### MD-Refined Investigations

As mentioned above, the binding affinity of 30 available inhibitors<sup>12-18</sup> to the monomeric and dimeric SARS-CoV-2 Mpro was appropriately probed using molecular docking calculations. However, it should be noted that the dynamics of receptors were not considered in docking simulations, and the number of trial docking poses was small. To overcome this limitation we have operated the atomistic simulations which serve as a validation for the docking results<sup>37-39</sup>. Moreover, FPL is an efficient computational approach to assess ligand-binding affinity with a suitable time-consuming calculation.<sup>40, 41</sup> Furthermore, the scheme was successfully applied to the monomeric SARS-CoV-2 Mpro system recently.<sup>28, 29</sup> The FPL approach is thus used to probe the binding affinity of 30 available inhibitors to the monomeric and dimeric SARS-CoV-2 Mpro. In the simulations, the ligand binding pose was optimized over short

canonical and isothermal-isobaric simulations. The equilibrated ligand was then pulled to translocate from *bound* to *unbound* states. The maximum of pulling force, called rupture force, and pulling work are typically assumed to correlate with ligand-binding affinity. It should be noted that the rupture force corresponds to the point that the non-covalent bond between a ligand and a receptor was terminated.

The computed values of the rupture force and pulling work were shown in

Table 2. The denoted pulling force and work profiles were described in Tables S2 and S3 of the ESI file. The shape of both pulling force and work appear reliable when compared to the previous exertion.<sup>40, 41</sup> In particular, starting at zero, the pulling force quickly increases to the maximum value, then suddenly drops to zero due to the loss the non-covalent bond to the receptor. During this process, recorded-pulling work speedily rises from zero value to a stable value, corresponding to the distance at which the contact between protein and inhibitor is vanished. Moreover, the rupture force  $F_{\text{Max}}^{\text{Monomer}}$  of monomeric Mpros diffuses in the range from 295.0 to 977.6 pN corresponding with the spreading of pulling work  $W^{\text{Monomer}}$  from 13.7 to 106.1 kcal mol<sup>-1</sup>. Besides that, the matching metrics of dimeric Mpros forms in the range from 336.1 to 769.6 pN and 20.5 to 84.7

kcal mol<sup>-1</sup>, correspondingly. It should be noted that the computed works are significantly larger than the magnitude of experimental binding affinity, which diffuses in the range from 4.52 to 9.23 kcal mol<sup>-1</sup>, since applied large cantilever and high pulling velocity.<sup>40</sup> Although the discrepancy can be reduced to zero by using a small cantilever and an extremely low pulling velocity, it is not appropriate since it requires to perform several trajectories with hundred nanoseconds each.<sup>42</sup> Furthermore, previous investigations revealed that although reducing the magnitude of cantilever spring constant and pulling velocity was able to enlarge the accuracy of the estimations, the observed results are approximately the equivalent as those at high pulling velocity.<sup>40</sup>

**Table 2.** Computed values of rupture force and pulling work in comparison with experiments.

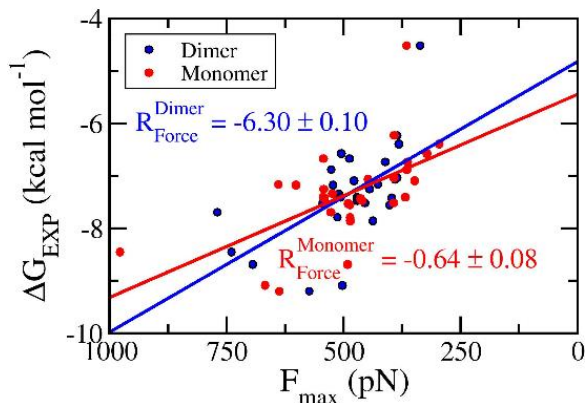
N <sup>0</sup>	Name	$F_{\text{Max}}^{\text{Monomer}}$	$W^{\text{Monomer}}$	$F_{\text{Max}}^{\text{Dimer}}$	$W^{\text{Dimer}}$	$\Delta G_{\text{EXP}}^a$
1	<b>11r</b>	724.8 ± 57.7	77.6 ± 7.1	636.6 ± 28.2	71.5 ± 2.9	-9.23
2	<b>13a</b>	526.9 ± 56.4	54.4 ± 7.3	769.6 ± 16.3	84.7 ± 3.2	-7.70
3	<b>13b</b>	977.6 ± 33.7	106.1 ± 4.6	739.1 ± 28.4	81.6 ± 3.0	-8.45
4	Bazedoxifene	460.3 ± 26	41.2 ± 3.1	471.1 ± 20.0	47.5 ± 3.6	-7.48
5	Calpain inhibitor XII	491.6 ± 20.5	46 ± 2.3	693.6 ± 50.7	63.5 ± 4.8	-8.69
6	Carmofur	485.5 ± 34.2	36.2 ± 2.7	436.9 ± 16.3	33.6 ± 1.8	-7.86
7	Chloroquine	363.4 ± 32.1	28.5 ± 2.8	410.9 ± 12.5	36.0 ± 1.6	-6.74
8	Cyclosporine	638.8 ± 33.4	67.7 ± 5.4	426.5 ± 41.6	44.1 ± 4.7	-7.17
9	Digitoxin	667.4 ± 17.7	70.9 ± 2.1	502.6 ± 65	55.3 ± 8.3	-9.09
10	Digoxin	637.0 ± 30.3	75.0 ± 2.5	573.1 ± 42.3	59.4 ± 4.9	-9.20
11	Dihydrogambogic Acid	542.8 ± 37.7	59.6 ± 3.2	487.5 ± 29.9	44.0 ± 3.3	-6.67
12	Disulfiram	364.7 ± 24.7	22.7 ± 1.9	526.2 ± 30.3	40.1 ± 1.9	-6.89
13	Ebastine	447.5 ± 40.1	40.2 ± 3.5	389.8 ± 25.0	32.8 ± 2.8	-7.06
14	Favipiravir	364.9 ± 26.2	21.3 ± 2.9	336.1 ± 19.1	20.5 ± 2.5	-4.52
15	Fluspirilene	490.1 ± 23.6	43.8 ± 2.0	544.6 ± 36.3	58.0 ± 3.2	-7.53
16	Isoosajin	393.1 ± 32.8	28.9 ± 3.2	454.4 ± 19.7	40.4 ± 2.5	-7.52
17	Ivacaftor	347.9 ± 34.8	22.3 ± 4.4	477.5 ± 22.1	41.0 ± 2.1	-7.10
18	Lusutrombopag	540.6 ± 37.5	59.1 ± 3.7	396.8 ± 24.3	41.8 ± 2.2	-7.42
19	Mefloquine	523.7 ± 23.5	41.5 ± 2.3	509.6 ± 43.3	46.3 ± 3.3	-7.34
20	Mequitazine	392.5 ± 51.3	29.5 ± 4.0	384.9 ± 24.4	29.0 ± 2.2	-7.03
21	MG-132	543.2 ± 22.2	49.8 ± 2.1	505.7 ± 41.1	47.5 ± 6.0	-7.41
22	Narlaprevir	601.8 ± 31.9	64.8 ± 2.8	522.0 ± 38.3	54.7 ± 4.3	-7.18
23	Osajin	367.9 ± 20.4	30.8 ± 2.9	471.4 ± 23.9	39.8 ± 1.8	-7.41
24	Oxyclozanide	463.7 ± 31.7	33.6 ± 3.2	468.1 ± 13.3	39.2 ± 3.5	-7.44
25	Penfluridol	542.3 ± 33.1	53.3 ± 2.7	444.5 ± 25.0	48.0 ± 3.9	-7.26
26	Phenazopyridine	391.7 ± 36.2	25.6 ± 2.8	384.8 ± 22.7	32.4 ± 1.4	-6.23
27	Proscillaridin	485.6 ± 37.2	45.8 ± 3.3	512.8 ± 18.9	58.0 ± 1.6	-7.79
28	PX-12	295.0 ± 17.4	13.7 ± 1.2	382.0 ± 25.5	27.2 ± 2.0	-6.39
29	Shikonin	321.8 ± 29.7	19.7 ± 3.0	504.5 ± 22.8	39.1 ± 1.2	-6.58
30	Tetrandrine	485.6 ± 37.2	45.8 ± 3.3	401.5 ± 18.5	31.6 ± 1.8	-7.56

<sup>a</sup>The experimental binding free energies were gained based on IC50 value,<sup>12-18</sup> approximating that the one equals to the inhibition constant  $k_i$ . The unit of force and energy/work are in pN and kcal mol<sup>-1</sup>, respectively.

In practice, the rupture force has been used as a predictor of ligand-binding affinity based on the assumption that a ligand binds with a higher affinity requires a stronger pulling force to dissociate it from binding cleft.<sup>43</sup> Using the rupture force as a proxy to ligand-binding affinity, numerous investigations were successful in predicting the ligand-binding affinity to various targets.<sup>43, 44</sup> Here, the average of rupture forces were estimated over 8 independent FPL trajectories (cf.

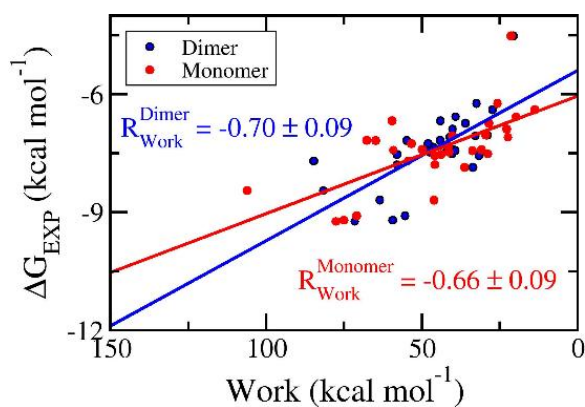


Table 2). The correlation coefficient, obtained results of monomeric systems, is  $R_{\text{Force}}^{\text{Monomer}} = -0.64 \pm 0.08$ ; while the analogous value of dimeric forms is  $R_{\text{Force}}^{\text{Dimer}} = -0.63 \pm 0.10$  as sketched in Figure 5. Clearly, the accuracy of the FPL technique is significantly larger than that of molecular docking calculation. Moreover, because the correlation coefficients appear to be the same within the error range, we may conclude that there is no difference when using monomer or dimer as a CADD target.



**Figure 5.** Relationship between rupture force and experimental binding free energy. Rupture forces were obtained via FPL calculations. The binding free energies were gained based on IC50 value,<sup>12-18</sup> approximating that the one equals to the inhibition constant  $k_i$ . The computed error was attained via 1000 rounds of the bootstrapping method.<sup>36</sup>

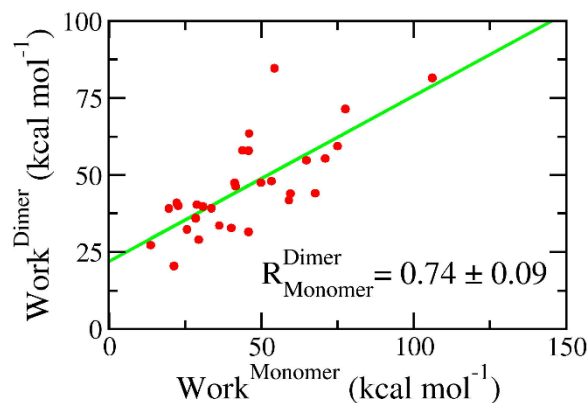
The work of pulling force was assessed via formula  $W = v \int_0^t F(t) dt$ , where  $v$  is pulling velocity and  $F(t)$  is pulling force. In isothermal-isobaric simulations,  $W$  is related to the experimental binding affinity via Jarzynski equality.<sup>45</sup> Therefore, utilizing  $W$  to estimate the ligand-binding affinity commonly acquires a better accurate result in comparison to rupture force.<sup>37, 40, 44</sup> The obtained results reaffirmed this statement. The correlation coefficients of the monomeric and dimeric SARS-CoV-2 Mpro are  $R_{\text{Work}}^{\text{Monomer}} = -0.66 \pm 0.09$  and  $R_{\text{Work}}^{\text{Dimer}} = -0.70 \pm 0.09$  as shown in Figure 6, respectively. Although, the computational accuracy targeting the SARS-CoV-2 Mpro dimer is slightly larger than that of the monomeric system, the difference in correlation coefficients is small implying that the monomeric form of SARS-CoV-2 Mpro can be used as CADD target instead of the dimeric one.



**Figure 6.** Association between pulling work and experimental binding free energy. Pulling works were obtained via FPL calculations. The binding free energies were gained based

on IC50 value,<sup>12-18</sup> approximating that the one equals to the inhibition constant  $k_i$ . The computed error was attained via 1000 rounds of the bootstrapping method.<sup>36</sup>

In addition, the association of computed pulling works of the monomeric and dimeric SARS-CoV-2 Mpro was probed and shown in Figure 7. Over the bootstrapping examination, the correlation coefficient is  $R_{\text{Monomer}}^{\text{Dimer}} = 0.74 \pm 0.09$  confirming the observation above. We can manipulate the inhibitor screening for SARS-CoV-2 Mpro with smaller computing resources since targeting the monomeric form.



**Figure 7.** Association between calculated pulling work of the monomeric and dimeric SARS-CoV-2 Mpro. The computed error was attained via 1000 rounds of the bootstrapping method.<sup>36</sup>

In addition, the MD-refined ligand-binding affinity results are confirmed since the RMSD between ligand-binding poses to the monomeric and dimeric forms is of  $0.22 \pm 0.02$  nm only as an example in Figure 3B. It should be noted that the RMSD metrics were calculated based on the last snapshot of NPT simulations, which were utilized for the binding free energy prediction via the FPL scheme. Moreover, the RMSD of MD-refined structure was slightly larger than docking results due to the effects of the conformational entropy. Moreover, the intermolecular hydrogen bond analyses suggests that three residues including Asn142, Gly143, and Glu166 are critical residues controlling the binding mechanism of the inhibitors to both monomeric and dimeric SARS-CoV-2 Mpro.

## Conclusions

Both of Autodock Vina and FPL simulations were confirmed to be able to appropriately estimate the ligand-binding affinity of the SARS-CoV-2 Mpro in both monomeric and dimeric forms. The assessed results suggested that the monomeric form of SARS-CoV-2 Mpro can be used as a CADD target instead of the dimeric form. In particular, the correlation coefficients between computational and experimental binding free energy of the monomeric SARS-CoV-2 Mpro are  $R_{\text{Dock}}^{\text{Monomer}} = 0.59 \pm 0.11$  and  $R_{\text{Work}}^{\text{Monomer}} = -0.66 \pm 0.08$ . The metrics are approximately similar to the dimeric target with the coefficients of  $R_{\text{Dock}}^{\text{Dimer}} = 0.52 \pm 0.10$  and  $R_{\text{Work}}^{\text{Dimer}} = -0.70 \pm 0.09$ . Moreover, the correlation coefficient between the rupture forces to binding free energy are roughly the same since  $R_{\text{Force}}^{\text{Monomer}} = -0.64 \pm 0.08$  and  $R_{\text{Force}}^{\text{Dimer}} = -0.63 \pm 0.10$ . Furthermore, the correlation coefficient between the calculated metrics of the monomeric and dimeric SARS-CoV-2 Mpro is  $R_{\text{Monomer}}^{\text{Dimer}} = 0.74 \pm 0.09$ . It should be

noted that the observation is in good agreement with structure analyses with the RMSD between ligand-binding pose to the monomeric and dimeric form of  $0.19 \pm 0.02$  and  $0.22 \pm 0.02$  nm for docking and MD-refined structures, respectively. In addition, in good agreement with the previous observation,<sup>27</sup> the molecular docking by Vina package rapidly converged since the correlation coefficient between computed and experimental values did not change when the docking option was altered. The RMSE of docking results also unchanged upon these alterations. Finally, it may be concluded that for SARS-CoV-2 Mpro system the pulling work is better than rupture force in predicting the ligand-binding affinity. It is well compatible with earlier probe various protein-ligand complexes.<sup>37, 40, 44</sup>

## Acknowledgements

This work was supported by Vietnam National Foundation for Science & Technology Development (NAFOSTED) grant #104.99-2019.57.

## References

- J. F. W. Chan, S. F. Yuan, K. H. Kok, K. K. W. To, H. Chu, J. Yang, F. F. Xing, J. L. Liu, C. C. Y. Yip, R. W. S. Poon, H. W. Tsoi, S. K. F. Lo, K. H. Chan, V. K. M. Poon, W. M. Chan, J. D. Ip, J. P. Cai, V. C. C. Cheng, H. L. Chen, C. K. M. Hui and K. Y. Yuen, *Lancet*, 2020, **395**, 514-523.
- C. L. Huang, Y. M. Wang, X. W. Li, L. L. Ren, J. P. Zhao, Y. Hu, L. Zhang, G. H. Fan, J. Y. Xu, X. Y. Gu, Z. S. Cheng, T. Yu, J. A. Xia, Y. Wei, W. J. Wu, X. L. Xie, W. Yin, H. Li, M. Liu, Y. Xiao, H. Gao, L. Guo, J. G. Xie, G. F. Wang, R. M. Jiang, Z. C. Gao, Q. Jin, J. W. Wang and B. Cao, *Lancet*, 2020, **395**, 497-506.
- C. Wang, P. W. Horby, F. G. Hayden and G. F. Gao, *Lancet*, 2020, **395**, 470-473.
- C. Yu Wai, Y. Chin-Pang and W. Kwok-Yin, *F1000Res*, 2020, **9**, 129.
- WHO, Coronavirus disease 2019 (COVID-19) Situation Report - 52).
- P. Zhou, X.-L. Yang, X.-G. Wang, B. Hu, L. Zhang, W. Zhang, H.-R. Si, Y. Zhu, B. Li, C.-L. Huang, H.-D. Chen, J. Chen, Y. Luo, H. Guo, R.-D. Jiang, M.-Q. Liu, Y. Chen, X.-R. Shen, X. Wang, X.-S. Zheng, K. Zhao, Q.-J. Chen, F. Deng, L.-L. Liu, B. Yan, F.-X. Zhan, Y.-Y. Wang, G.-F. Xiao and Z.-L. Shi, *Nature*, 2020, **579**, 270-273.
- N. van Doremalen, T. Bushmaker, D. H. Morris, M. G. Holbrook, A. Gamble, B. N. Williamson, A. Tamin, J. L. Harcourt, N. J. Thornburg, S. I. Gerber, J. O. Lloyd-Smith, E. de Wit and V. J. Munster, *N Engl J Med*, 2020, **382**, 1564-1567.
- D. Schoeman and B. C. Fielding, *Virology*, 2019, **16**, 69.
- C. M. Fauquet and D. Fargette, *Virology*, 2005, **2**, 64.
- Z. Alex, A. Vladimir, Z. Alexander, Z. Bogdan, T. Victor, B. Dmitry S., P. Daniil, S. Rim, F. Andrey, O. Philipp, Y. Yilin, P. Olga, V. Quentin, A. Alex and I. Yan, *Potential COVID-2019 3C-like Protease Inhibitors Designed Using Generative Deep Learning Approaches*, 2020.
- V. Nukoolkarn, V. S. Lee, M. Malaisree, O. Aruksakulwong and S. Hannongbua, *J. Theor. Biol.*, 2008, **254**, 861-867.
- L. Zhang, D. Lin, X. Sun, U. Curth, C. Drosten, L. Sauerhering, S. Becker, K. Rox and R. Hilgenfeld, *Science*, 2020, **368**, 409-412.
- Z. Jin, Y. Zhao, Y. Sun, B. Zhang, H. Wang, Y. Wu, Y. Zhu, C. Zhu, T. Hu, X. Du, Y. Duan, J. Yu, X. Yang, X. Yang, K. Yang, X. Liu, L. W. Guddat, G. Xiao, L. Zhang, H. Yang and Z. Rao, *Nat Struct Mol Biol*, 2020, **27**, 529-532.
- W. Dai, B. Zhang, H. Su, J. Li, Y. Zhao, X. Xie, Z. Jin, F. Liu, C. Li, Y. Li, F. Bai, H. Wang, X. Cheng, X. Cen, S. Hu, X. Yang, J. Wang, X. Liu, G. Xiao, H. Jiang, Z. Rao, L.-K. Zhang, Y. Xu, H. Yang and H. Liu, *Science*, 2020, **368**, 1331-1335.
- K. Gao, D. D. Nguyen, J. Chen, R. Wang and G.-W. Wei, *J. Phys. Chem. Lett*, 2020, **11**, 5373-5382.
- C. Ma, M. D. Sacco, B. Hurst, J. A. Townsend, Y. Hu, T. Szeto, X. Zhang, B. Tarbet, M. T. Marty, Y. Chen and J. Wang, *Cell Res*, 2020, **30**, 678-692.
- S. Weston, C. M. Coleman, R. Haupt, J. Logue, K. Matthews, Y. Li, H. M. Reyes, S. R. Weiss and M. B. Frieman, *J. Virol.*, 2020, DOI: 10.1128/jvi.01218-20, JVI.01218-01220.
- S. Jeon, M. Ko, J. Lee, I. Choi, S. Y. Byun, S. Park, D. Shum and S. Kim, *Antimicrob. Agents Chemother.*, 2020, **64**, e00819-00820.
- G. R. Marshall, *Ann. Rev. Pharmacol. Toxicol.*, 1987, **27**, 193-213.
- N. Homeyer, F. Stoll, A. Hillisch and H. Gohlke, *J. Chem. Theory Comput.*, 2014, **10**, 3331-3344.
- S. Decherchi and A. Cavalli, *Chem. Rev.*, 2020, DOI: 10.1021/acs.chemrev.0c00534.
- S. T. Ngo, T. H. Nguyen, N. T. Tung, P. C. Nam, K. B. Vu and V. V. Vu, *J. Comput. Chem*, 2020, **41**, 611-618.
- W. Yu and A. D. MacKerell, in *Antibiotics: Methods and Protocols*, ed. P. Sass, Springer New York, New York, NY, 2017, DOI: 10.1007/978-1-4939-6634-9\_5, pp. 85-106.
- C. Ma, M. Sacco, Y. Chen and J. Wang, 2020.
- S. Kim, P. A. Thiessen, E. E. Bolton, J. Chen, G. Fu, A. Gindulyte, L. Han, J. He, S. He, B. A. Shoemaker, J. Wang, B. Yu, J. Zhang and S. H. Bryant, *Nucleic Acids Res.*, 2016, **44**, D1202-D1213.
- O. Trott and A. J. Olson, *J. Comput. Chem.*, 2010, **31**, 455-461.
- N. T. Nguyen, T. H. Nguyen, T. N. H. Pham, N. T. Huy, M. V. Bay, M. Q. Pham, P. C. Nam, V. V. Vu and S. T. Ngo, *J. Chem. Inf. Model.*, 2020, **60**, 204-211.
- S. T. Ngo, N. Quynh Anh Pham, L. Thi Le, D.-H. Pham and V. V. Vu, *J. Chem. Inf. Model.*, 2020, DOI: 10.1021/acs.jcim.0c00491.
- M. Q. Pham, K. B. Vu, T. N. Han Pham, L. T. Thuy Huong, L. H. Tran, N. T. Tung, V. V. Vu, T. H. Nguyen and S. T. Ngo, *RSC Adv*, 2020, **10**, 31991-31996.
- M. J. Abraham, T. Murtola, R. Schulz, S. Páll, J. C. Smith, B. Hess and E. Lindahl, *SoftwareX*, 2015, **1-2**, 19-25.
- A. E. Aliev, M. Kulke, H. S. Khaneja, V. Chudasama, T. D. Sheppard and R. M. Lanigan, *Proteins: Struct., Funct., Bioinf.*, 2014, **82**, 195-215.
- J. Wang, R. M. Wolf, J. W. Caldwell, P. A. Kollman and D. A. Case, *J. Comput. Chem.*, 2004, **25**, 1157-1174.
- D. A. Case, I. Y. Ben-Shalom, S. R. Brozell, D. S. Cerutti, T. E. C. Cheatham, III, V.W.D. , T. A. Darden, R. E. Duke, D. Ghoreishi, M. K. Gilson, H. Gohlke, A. W. Goetz, D. Greene, R. Harris, N. Homeyer, Y. Huang, S. Izadi, A. Kovalenko, T. Kurtzman, T. S. Lee, S. LeGrand, P. Li, C. Lin, J. Liu, T. Luchko, R. Luo, D. J. Mermelstein, K. M. Merz, Y. Miao, G. Monard, C. Nguyen, H. Nguyen, I. Omelyan, A. Onufriev, F. Pan, R. Qi, D. R. Roe, A. Roitberg, C. Sagui, S. Schott-Verdugo, J. Shen, C. L. Simmerling, J. Smith, R. SalomonFerrer, J. Swails, R. C. Walker, J. Wang, H. Wei, R. M. Wolf, X. Wu, L. Xiao, Y.



D.M. and a. K. P.A., *University of California, San Francisco*, 2018.

34. A. W. Sousa da Silva and W. F. Vranken, *BMC Research Notes*, 2012, **5**, 1-8.
35. T. Darden, D. York and L. Pedersen, *J. Chem. Phys.*, 1993, **98**, 10089-10092.
36. B. Efron, *Ann. Stat.*, 1979, **7**, 1-26.
37. N. T. Mai, N. T. Lan, T. Y. Vu, P. T. Mai Duong, N. T. Tung and H. T. Thu Phung, *J. Mol. Graph. Modell.*, 2020, **100**, 107648.
38. M. A. Mansour, A. M. AboulMagd and H. M. Abdel-Rahman, *RSC Adv*, 2020, **10**, 34033-34045.
39. N. T. Dan, H. D. Quang, V. Van Truong, D. Huu Nghi, N. M. Cuong, T. D. Cuong, T. Q. Toan, L. G. Bach, N. H. T. Anh, N. T. Mai, N. T. Lan, L. Van Chinh and P. M. Quan, *Sci Rep*, 2020, **10**, 11429.
40. S. T. Ngo, H. M. Hung and M. T. Nguyen, *J. Comput. Chem.*, 2016, **37**, 2734-2742.
41. S. T. Ngo, M. T. Nguyen and M. T. Nguyen, *Chem. Phys. Lett.*, 2017, **676**, 12-17.
42. D.-C. Li and B.-H. Ji, *Acta Mech. Sin.*, 2012, **28**, 891-903.
43. B. K. Mai, M. H. Viet and M. S. Li, *J. Chem. Inf. Model.*, 2010, **50**, 2236-2247.
44. N. M. Tam, K. B. Vu, V. V. Vu and S. T. Ngo, *Chem. Phys. Lett.*, 2018, **701**, 65-71.
45. S. Park and K. Schulten, *J. Chem. Phys.*, 2004, **120**, 5946-5961.