Unbinding ligands from SARS-CoV-2 Mpro via umbrella sampling simulations

Nguyen Minh Tam\textsuperscript{1,2}, Trung Hai Nguyen\textsuperscript{2,3}, Vu Thi Ngan\textsuperscript{4}, Nguyen Thanh Tung\textsuperscript{5,6} and Son Tung Ngo\textsuperscript{2,3}

\textsuperscript{1}Computational Chemistry Research Group, Ton Duc Thang University, Ho Chi Minh City, Vietnam  
\textsuperscript{2}Faculty of Applied Sciences, Ton Duc Thang University, Ho Chi Minh City, Vietnam  
\textsuperscript{3}Laboratory of Theoretical and Computational Biophysics, Ton Duc Thang University, Ho Chi Minh City, Vietnam  
\textsuperscript{4}Laboratory of Computational Chemistry and Modelling, Department of Chemistry, Quy Nhon University, Quy Nhon, Vietnam  
\textsuperscript{5}Institute of Materials Science, Vietnam Academy of Science and Technology, Hanoi, Vietnam  
\textsuperscript{6}Graduate University of Science and Technology, Vietnam Academy of Science and Technology, Hanoi, Vietnam

The umbrella sampling (US) simulation is demonstrated to be an efficient approach for determining the unbinding pathway and binding affinity to the SARS-CoV-2 Mpro of small molecule inhibitors. The accuracy of US is in the same range as the linear interaction energy (LIE) and fast pulling of ligand (FPL) methods. In detail, the correlation coefficient between US and experiments does not differ from FPL and is slightly smaller than LIE. The root mean square error of US simulations is smaller than that of LIE. Moreover, US is better than FPL and poorer than LIE in classifying SARS-CoV-2 Mpro inhibitors owing to the receiver operating characteristic–area under the curve analysis. Furthermore, the US simulations also provide detailed insights on unbinding pathways of ligands from the binding cleft of SARS-CoV-2 Mpro. The residues Cys44, Thr45, Ser46, Leu141, Asn142, Gly143, Glu166, Leu167, Pro168, Ala191, Glu192 and Ala193 probably play an important role in the ligand dissociation. Therefore, substitutions at these points may change the mechanism of binding of inhibitors to SARS-CoV-2 Mpro.

1. Introduction

SARS-CoV-2 effectuates the COVID-19 pandemic. The first human infection with severe acute respiratory syndromes was reported in...
Wuhan, Hubei Province, China in December 2019 [1–4]. Since then, it has caused millions of deaths worldwide. Bats have been suggested to be the carrier of SARS-CoV-2 before it jumped to and quickly transfected among humans [5]. The spreading rate is extremely high because it can be airborne [6]. In just 1 year, more than 130 million people have been infected despite international effort to contain the spread of the virus. Therefore, it is of great urgency to develop an effective therapy for community health. The approval of remdesivir as an anti-viral treatment for COVID-19 [7] may be considered as a rather controversial decision [8] because it showed poor efficacy in clinical trials [9,10]. Moreover, although two vaccines for SARS-CoV-2 have recently been approved for emergency use by the US Food and Drug Administration [11], using the vaccines to create COVID-19 herd immunity may not be possible [12]. Furthermore, especially, recent reports on new variants of SARS-CoV-2 appearing in the UK (B1.1.7) and South Africa (B1.351) can escape the neutralizing antibodies [13,14].

The main protease (Mpro) of SARS-CoV-2, a 3-chymotrypsin-like protease (3CLpro), is characterized as a pivotal enzyme associated with viral proliferation and replication [15,16]. The protease cleaves the polyproteins to polypeptides corresponding to the formation of 13 non-structural proteins (nsp) [17]. Therefore, the SARS-CoV-2 Mpro is indicated as the ultra-potent target for drug design against SARS-CoV-2 because inhibiting enzymatic activity of the protease can stop the virus from replicating.

The SARS-CoV-2 Mpro is a homodimer enzyme comprising of two monomers, each monomer consists of 306 residues separating into three domains I, II and III (cf figure 1a) [18]. In particular, the substrate binding site of the SARS-CoV-2 Mpro contains two residues, His41 and Cys145 located in the cleft formed by domains I and II [19,20]. Strong binding inhibitors to the protease normally adopt rigid contacts to these residues [21]. Furthermore, the residues Thr26, Ser46, Asn142, Gly143, His164, Glu166 and Gln189 were found to be crucial for the ligand-binding process [21]. Furthermore, although the two monomers interact with each other via domain II and crucially associate with the catalytic activity since the active site conformation strongly depends on the dimerization [18], the ligand-binding affinity of ligands to the Mpro in silico could reliably obtained from calculations performed with only the monomer instead of the dimer [22].

Computer-aided drug design (CADD) has emerged as a promising approach for screening and optimizing potential inhibitors for the biological activity of an enzyme [23–25]. In CADD, many computational approaches have been employed to estimate the binding free energy, \( \Delta G_{\text{bind}} \), of ligands to proteins [26–48], which is associated with the experimental inhibition constant/half-maximal inhibitory concentration metrics [23,49–51]. Determination of the \( \Delta G_{\text{bind}} \) thus is a very important issue in CADD [52,53]. In molecular dynamics (MD) simulations, a computational free energy approach is normally selected based on the balance between the computation accuracy and the consumed central processing unit resources [54–58]. Popular calculation approaches include end-to-end schemes such as the linear interaction energy (LIE) [59], molecular mechanism/Poisson-Boltzmann surface area (MM/PBSA) [34], the free energy perturbation [41], and thermodynamic integration [60] methods. These methods directly provide the binding free energy. However, the binding/unbinding pathways also provide crucial insights into the binding process of ligands to proteins. Therefore, in this work, a combination of the fast pulling of ligand (FPL) [36] and umbrella sampling (US) [61,62] simulations were employed to unbind a ligand from SARS-CoV-2 Mpro. During the simulations, the \( \Delta G_{\text{bind}} \) value and details of unbinding pathway are characterized. In particular, the metric \( \Delta G_{\text{bind}} \) is calculated as

![Figure 1. (a) The monomeric SARS-CoV-2 Mpro + Narlaprevir (PDB ID: 7JYC), in which, the protease was highlighted with three domains I, II and III; (b) starting shapes of SARS-CoV-2 Mpro + 11b in SMD/US simulations.](image-url)
the difference in the potential of mean force (PMF) between bound and unbound states of the complexes. The PMF value is obtained by the weighted histogram analysis method (WHAM) [63] calculations. The unbinding pathway is able to estimate via the collective-variable free energy landscape (FEL) [64]. The representative structures of the complexes can be determined using the clustering method [65]. The obtained results revealed that the combination of FPL/US simulations was useful for accurately evaluating potential inhibitors for SARS-CoV-2 Mpro.

2. Material and methods

2.1. Structure of receptor and ligand

Twenty four three-dimensional structures of the solvated SARS-CoV-2 Mpro + inhibitor were reported in the previous work [22], in which the structure of SARS-CoV-2 Mpro was downloaded from the protein data bank with identification of 6Y2F [66] and structure of ligands downloaded from the PubChem database [67]. Details of the ligands are reported in the electronic supplementary material, table S1.

2.2. Atomistic simulations

GROMACS v. 5.1.5 [68] was used to carry out atomistic simulations. As mentioned above, the initial conformations of complexes (cf. figure 1b) were obtained from previous work [22]. Force field parameters for the protein + ions, water molecules and ligand were taken from the Amber99SB-ILDN force field [69], TIP3P water model [70] and general Amber force field [71], respectively. The MD parameters were chosen according to the previous work [21]. Particularly, the integration was carried out with a timestep of 2 fs. Cut-off for non-bonded interaction was set at a distance of 0.9 nm. The electrostatic interactions were computed using the Particle Mesh Ewald method [72], whereas the van der Waals interactions were calculated with the cut-off scheme. The solvated complexes were relaxed using energy minimization, constant number of particles, volume and temperature (NVT) and constant number of particles, volume and pressure (NPT) simulations before the steered-MD was applied.

2.2.1. Fast pulling of ligand simulations

The NPT-equilibrated complex conformation was used as a starting structure of FPL calculations. During an FPL simulation the centre of mass of the ligand was pulled along the reaction coordinate, ξ, via an external-harmonic force. In particular, the spring constant of pulling force is $k = 600 \text{ kJ mol}^{-1} \text{nm}^{-2}$ and the pulling velocity is $v = 0.005 \text{ nm ps}^{-1}$. An FPL calculation was during an interval of 0.7 ns. The pulling parameters were referred to in the previous works [36,73]. The systemic coordinate, pulling force and pulling work were recorded every 0.1 ps.

2.2.2. Umbrella sampling simulations

The optimized FPL trajectory, whereas the deviation of the pulling work to the mean is the smallest value, was used to generate US windows. The displacement of ligands between neighbouring windows is ca 0.1 nm. The PMF was estimated from US simulations with 21 windows and length of 2 ns per window. It should be noted that NPT simulations were performed with length of 0.1 ns to evade any staring bias. Moreover, the $C_{α}$ atoms were positionally restrained via a weak harmonic force to prevent the enzymic reorientation and translation during both FPL and US simulations. It helped reduce the computational effort during US simulations because we didn’t have to thoroughly sample the whole protein conformational space which may be huge. The approach is reasonable because the protein conformation does not change significantly upon ligand binding.

2.3. Analysed tools

The two-dimensional FEL was constructed using the number of contacts within 0.45 nm between protein-ligand and ligand displacement as two reaction coordinates [64,65]. Representative structures of complexes localizing in the FEL minima were predicted using the clustering methods [65]. The WHAM [74] was used to estimate the PMF. The computed error was estimated using a bootstrapping method [75]. The Scikit-Learn library [76] was used to compute the receiver operating characteristic--
area under the curve (ROC-AUC). In particular, the set of ligands were sorted and split into two groups according to their experimental binding free energy, \( \Delta G_{\text{EXP}} \). Namely, strong binders are ligands having \( \Delta G_{\text{EXP}} < -7.60 \text{ kcal mol}^{-1} \) whereas weak binders are those with \( \Delta G_{\text{EXP}} \geq -7.60 \). The Python code for estimating errors using the bootstrapping method and for calculating ROC-AUC is included in the electronic supplementary material.

3. Results

We can roughly classify proteins into two groups based on the geomorphology of their active sites. The first group includes proteins having a ‘close’ active site which requires protein structural change to allow a ligand to bind/unbind. The second one consists of proteins having an ‘open’ active site permitting its ligand to move in/out without protein conformational change. SARS-CoV-2 Mpro belongs to the second group with an ‘open’ active site that we can use pathway-based binding affinity approaches, such as FPL [36], US [62] and non-equilibrium molecular dynamics (NEMD) [39,40,77] simulations, to evaluate the ligand-binding affinity besides end-to-end free energy approaches. FPL was shown to be suitable for rapidly estimating the binding affinity of inhibitors to SARS-CoV-2 Mpro [22,58,73]. However, the obtained values cannot directly estimate the binding free energy, \( \Delta G \). The approach also cannot evaluate the unbinding pathway of ligands owing to the limited sampling that was generated. \( \Delta G \) and details of the unbinding pathway can be calculated via NEMD, but the approach requires huge computing resources. US simulations, an enhanced sampling method, emerges as a potential technique to complete this task force.

Initial conformations used for US simulations were generated by FPL simulations with a window spacing of ca 0.1 nm. During FPL, the ligand displacement was recorded as shown in figure 2. Inhibitors remain in the binding cavity until the pulling force reaches a maximum value, called rupture force. Inhibitors then constantly dissociate from the binding cleft. In particular, more details about FPL results are described in the electronic supplementary material, tables S2–S4. The results obtained in this study are consistent with the recent works [21,22].

Coordinates of solvated complexes were recorded during FPL simulations every 0.1 ps that the US windows were then extracted over the optimized FPL trajectory. The PMF values along the reaction coordinate, \( \xi \), (cf. figure 3; electronic supplementary material, table S5) were calculated using the WHAM [74], whereas the US histograms are reported in the electronic supplementary material, table S6. The shape of the free energy profile is very consistent with previous works [48,78]. In particular, the PMF features a deep minimum at short distance and a plateau region at large distance where the non-bonded interactions between protein and ligand vanish. The observation is in good agreement with the FEL analysis as mentioned below. A free energy profile is given in figure 3 and the values of calculated binding free energy \( \Delta G_{\text{US}} \) are given in table 1.

The calculated binding free energy ranges from \(-8.63\) to \(-1.80\) kcal mol\(^{-1}\), with the mean value of \( \Delta G_{\text{US}} = -4.59 \pm 0.41 \text{ kcal mol}^{-1} \). The calculated mean underestimates the experimental mean, which
of $\Delta G_{\text{EXP}} = -7.81 \pm 0.21 \text{ kcal mol}^{-1}$. The computed error bars were estimated by the standard error of the mean. The underestimation of US calculations was similar to the results when the US approach was applied to the cathepsin K (CTSK) system [62]. The inaccuracy of the force field in describing the interaction between biomolecules [83,84], which included the SARS-CoV-2 Mpro, inhibitor, water molecule, and counter ion, may cause the computed binding free energies to deviate from experimental values. The US calculations adopted an appropriate Pearson correlation with $R_{\text{US}} = 0.66 \pm 0.13$ (figure 4), which is similar to the FPL results, $R_W = 0.65$, but is lower than that of the LIE, $R_{\text{LIE}} = 0.73 \pm 0.09$ [21,22]. Moreover, the accuracy of computational approaches was also investigated via root mean square error (RMSE) calculation. The obtained result of US is of $\text{RMSE}_{\text{US}} = 3.58 \pm 0.28 \text{ kcal mol}^{-1}$. The $\text{RMSE}_{\text{US}}$ is slightly smaller than that from LIE calculation, $\text{RMSE}_{\text{LIE}} = 4.12 \pm 0.40 \text{ kcal mol}^{-1}$ and is significantly smaller than that from MM/PBSA, $\text{RMSE}_{\text{MM/PBSA}} = 10.15 \pm 1.92 \text{ kcal mol}^{-1}$ [21]. Furthermore, because the initial complexes obtained via the docking approach is probably far from the native structure, the US accuracy is not very high. Therefore, the systems need to relax in the MD simulations for long enough for the system to overcome local minima and reach equilibrium states [21]. The accuracy of the FPL approach was increased according to the strategy [21]. Otherwise, increasing the US window interval would probably resolve the issues since the complex in the bound state can reach equilibrium states. The issues should be carefully investigated in further work.

In order to characterize the insights into the unbinding pathway of ligand dissociation, the collective-variable FEL was constructed. Two coordinates are the number of protein-ligand contacts at a cut-off distance of 0.45 nm and the ligand displacement. The two-dimensional FEL is given in figure 5 and the electronic supplementary material, table S7. Consistent with the PMF curve (figure 3a), the number of contacts between protein-ligand rapidly decreases when the ligand is displaced from 1.1 to 2.5 nm. Along the dissociation pathway, numerous local minima were observed. The representative structures of SARS-CoV-2 Mpro + inhibitor, which corresponded to these minima, were characterized using a data clustering algorithm with all-atom root mean square deviation chosen as pair distance and a cut-off of 0.2 nm. A representation of these conformations is described in figure 5. In particular, the SARS-CoV-2 Mpro + 11b formed six minima, which were denoted as B, D1, D2, D3, D4 and D5. Structure B responds to the bound state of the complex, while structures D1–4 correspond to the transition states during the dissociated process of the ligand. Shape D5 responds to the completed dissociation state of the protein-ligand complex. Besides, the two-dimensional interaction diagram of SARS-CoV-2 Mpro + 11b is described in the electronic supplementary material, figure S1 mentioning more details of critical residues controlling the binding process of the compound 11b. The residues Cys44, Thr45, Ser46, Leu141, Asn142, Gly143, Glu166, Leu167, Pro168, Ala191, Gln192 and Ala193

![Figure 3.](https://example.com/figure3.png)
Figure 4. Linear correlation between $\Delta G_{US}$ and $\Delta G_{\text{EXP}}$. The computing error was estimated using the bootstrapping method.

Table 1. US binding free energy in comparison with experiments.

| \(n^0\) | name                  | $\Delta G_{US}$    | $\Delta G_{\text{EXP}}$ |
|---------|-----------------------|--------------------|--------------------------|
| 1       | 7J                    | $-6.08 \pm 0.61$   | $-8.69$                  |
| 2       | 11a                   | $-7.81 \pm 0.91$   | $-9.96$                  |
| 3       | 11b                   | $-8.07 \pm 1.29$   | $-10.13$                 |
| 4       | 11r                   | $-8.63 \pm 0.78$   | $-9.23$                  |
| 5       | 13a                   | $-6.22 \pm 0.58$   | $-7.70$                  |
| 6       | 13b                   | $-4.56 \pm 0.66$   | $-8.45$                  |
| 7       | calpain inhibitor I   | $-4.51 \pm 1.01$   | $-6.94$                  |
| 8       | calpain inhibitor II  | $-4.88 \pm 0.50$   | $-8.23$                  |
| 9       | calpain inhibitor XII | $-5.20 \pm 0.70$   | $-8.69$                  |
| 10      | calpeptin             | $-5.43 \pm 1.11$   | $-6.81$                  |
| 11      | candesartan cilexetil | $-5.61 \pm 0.80$   | $-7.60$                  |
| 12      | carmustin             | $-1.89 \pm 1.07$   | $-7.86$                  |
| 13      | chloroquine           | $-2.94 \pm 0.43$   | $-7.41$                  |
| 14      | dipyridamole          | $-5.60 \pm 0.81$   | $-8.52$                  |
| 15      | disulfiram            | $-2.68 \pm 0.43$   | $-6.89$                  |
| 16      | GC-373                | $-3.74 \pm 0.78$   | $-8.76$                  |
| 17      | hydroxychloroquine    | $-1.80 \pm 0.65$   | $-7.58$                  |
| 18      | MG-115                | $-3.14 \pm 0.71$   | $-7.53$                  |
| 19      | MG-132                | $-2.57 \pm 0.88$   | $-7.41$                  |
| 20      | narlaprevir           | $-2.41 \pm 1.13$   | $-6.40$                  |
| 21      | omeprazole            | $-3.61 \pm 1.23$   | $-6.60$                  |
| 22      | oxytetracycline       | $-7.29 \pm 0.75$   | $-7.18$                  |
| 23      | PX-12                 | $-2.45 \pm 0.70$   | $-6.39$                  |
| 24      | shikonin              | $-2.96 \pm 0.62$   | $-6.58$                  |

*The experimental binding affinities were calculated from the IC50 value, [19,20,66,79–82] assuming that IC50 equals the inhibition constant $k_i$. The unit is kcal mol$^{-1}$. 
probably are important elements over the ligand dissociation. These residues contribute a large number of contacts to the inhibitor over the dissociated process. The potential mutations at these points would probably alter the ligand binding pathway to SARS-CoV-2 Mpro, especially, the change in binding kinetics appeared.

In addition, ROC-AUC is commonly used to assess the predictive power of a binary classifier in separating a dataset into two classes. If we separate the set of ligands into two sets according to their experimental binding free energy as being strong binders (low binding free energy) and weak binders (high binding free energy) using some arbitrary cut-off, we can use ROC-AUC, to assess the ability of our computed free energy in distinguishing strong binders against weak binders. More specifically, if we randomly select a ligand from the strong binder set and a ligand from the weak binder set, ROC-AUC gives us the probability that the selected strong binder will have lower calculated binding free energy than the weak binder. Therefore, ROC-AUC for SARS-CoV-2 Mpro was calculated. The obtained result is $\text{ROC} = \text{AUC}_{\text{US}} = 0.83 \pm 0.09$, indicating that US is weaker than LIE, $\text{ROC} = \text{AUC}_{\text{LIE}} = 0.86 \pm 0.08$, and better than FPL, $\text{ROC} = \text{AUC}_{\text{FPL}} = 0.76 \pm 0.11$, in classifying SARS-CoV-2 Mpro inhibitors [21].

4. Conclusion

We have demonstrated that the US approach provides a reliable estimate of binding free energies $\Delta G_{\text{US}}$ when compared with experimental data [19,20,66,79–82]. The correlation between US binding free energy and experiments is rather high with a value of $R = 0.66 \pm 0.13$, which is in the same range of correlation with FPL [22]. Moreover, the obtained RMSE of US simulations, $\text{RMSE}_{\text{US}} = 3.58 \pm 0.28$ kcal mol$^{-1}$, is slightly smaller than that of LIE [21]. Furthermore, US is better than FPL and poorer than LIE in classifying SARS-CoV-2 Mpro inhibitors owing to ROC-AUC analysis, $\text{ROC} = \text{AUC}_{\text{US}} = 0.83 \pm 0.09$. Clustering conformations found on local minima of the two-dimensional FEL revealed important insights into the unbinding pathway of the ligand from SARS-CoV-2 Mpro binding cavity. The obtained results imply that the residues Cys44, Thr45, Ser46, Leu141, Asn142, Gly143, Glu166, Leu167, Pro168, Ala191, Gln192 and Ala193 probably play an important role in the ligand dissociation. Therefore, substitutions appearing at these points possibly change the ligand binding mechanism of SARS-CoV-2 Mpro.

Figure 5. The collective-variable FEL and representative structures of SARS-CoV-2 Mpro + 11b over US simulations.
References

1. WHO. 2019 Coronavirus disease 2019 (COVID-19) Situation Report – 52. 2020. Geneva, Switzerland: World Health Organisation.

2. Huang CL et al. 2020 Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. Lancet 395, 497–506. (doi:10.1016/S0140-6736(20)30183-5)

3. Wang C, Horby PW, Hayden FG, Gao GF. 2020 A novel coronavirus outbreak of global health concern. Lancet 395, 470–473. (doi:10.1016/S0140-6736(20)30189-5)

4. Yu Wai C, Chin-Pang Y, Kwok-Yin W. 2020 Prediction of the SARS-CoV-2 (2019-nCoV) 3C-like protease (3CLpro) structure: virtual screening reveals velpatasvir, ledipasvir, and other drug repurposing candidates. F1000Res. 9, 129. (doi:10.12688/f1000research.22457.2)

5. Chan JFW et al. 2020 A familial cluster of pneumonia associated with the 2019 novel coronavirus indicating person-to-person transmission: a study of a family cluster. Lancet 395, 514–523. (doi:10.1016/S0140-6736(20)30154-9)

6. van Doremalen N et al. 2020 Aerosol and surface stability of SARS-CoV-2 as compared with SARS-CoV-1. N. Engl. J. Med. 382, 1564–1567. (doi:10.1056/NEJMc2004973)

7. FDA Approves First Treatment for COVID-19. 2020 (cited 27 October 2020). See https://www.fda.gov/news-events/press-announcements/fda-approves-first-treatment-covid-19.

8. Cohen J, Kupersmith K. 2020 The ‘very, very bad look’ of remdesivir, the first FDA-approved COVID-19 drug. Sci. News. See https://pubcenter.org/stories/very-very-bad-look-remdesivir-first-fda-approved-covid-19-drug.

9. Holshue ML et al. 2020 First case of 2019 novel coronavirus in the United States. N. Engl. J. Med. 382, 1299–1322. (doi:10.1056/NEJMc2001191)

10. Beigel JH et al. 2020 Remdesivir for the treatment of COVID-19—final report. N. Engl. J. Med. 382, 1813–1826. (doi:10.1056/NEJMc2007764)

11. COVID-19 Vaccines. 2021 (cited 8 January 2022). See https://www.fda.gov/emergency-preparedness-and-response/coronavirus-disease-2019-covid-19/covid-19-vaccines.

12. Aschwanden C. 2021 Five reasons why COVID herd immunity is probably impossible. Nature 591, 520–522.

13. Wang P et al. 2021 Antibody resistance of SARS-CoV-2 variants B.1.351 and B.1.1.7. Nature 593, 130–135. (doi:10.1038/s41586-021-03198-2)

14. Hoffmann M et al. 2021 SARS-CoV-2 variants B.1.351 and P.1 escape from neutralizing antibodies. Cell 184, 2384–2393. (doi:10.1016/j.cell.2021.03.036)

15. Faquet CM, Fargette D. 2005 International committee on taxonomy of viruses and the committee on taxonomy of viruses and the domain Virology 2, 64. (doi:10.1186/1743-422X-2-64)

16. Alex Z et al. 2020 Potential COVID-19 3C-like protease inhibitors designed using generative deep learning approaches. Preprint. (doi:10.26434/chemrxiv.11829102.x2)

17. Freitas BT, Durie IA, Murray J, Longo JE, Miller KA, Bandeira VS, Vandrift TR, Huang B, Collins W, Geller JE, Huang Y, Desai M, Johnson SA, Winn KT, Markert WC, Welter J, Ferrone SR, Heslop HE, Crumpler MD, Gribben JG, Gmeiner S, Bagdasarian M, Sciamanna E, Duarte J, Banerjee S, Waller J, Monje MC, Zhu Y, Quigley ML, Heslop H, Collins W, Desai M, Miller KA. 2021 SARS-CoV-2 main protease. EMBO J. 40, e112125. (doi:10.15252/embj.2021125)

18. Anand K, Palm GJ, Mesters JR, Siddgel SG, Ziebuhr J, Hilgenfeld R. 2002 Structure of human coronavirus main proteinase reveals combination of a chymotrypsin fold with an extra α helical domain. EMBO J. 21, 3213–3224. (doi:10.1093/emboj/cvf227)

19. Jin Z et al. 2020 Structure of Mpro from SARS-CoV-2 and discovery of its inhibitors. Nature 582, 289–293. (doi:10.1038/s41586-020-2223-y)

20. Dai W et al. 2020 Structure-based design of antiviral drug candidates targeting the SARS-CoV-2 main protease. Science 368, 1331–1335. (doi:10.1126/science.abb4489)

21. Ngo ST, Tam NM, Pham MQ, Nguyen TH. 2021 Benchmark of popular free energy approaches revealing the inhibitors binding to SARS-CoV-2 Mpro. J. Chem. Inf. Model. 61, 2302–2312. (doi:10.1021/acs.jcim.1c00199)

22. Tam NM, Nam PC, Quang DT, Tung NT, Vu VV, Ngo ST. 2021 Binding of Inhibitors to the main protease of SARS-CoV-2 variants B.1.351 and B.1.1.7. RSC Adv. 11, 2926–2994. (doi:10.1039/D0RA09858B)

23. Marshall GR. 1987 Computer-aided drug design. Ann. Rev. Pharmacol. Toxicol. 27, 193–213.

24. Amam RE, Bann R, McCammon JA. 2008 An improved relaxed complex scheme for receptor flexibility in computer-aided drug design. J. Comput.-Aided Mol. Des. 22, 693–705.

25. Sliwoski G, Kothiwale S, Meiler J, Lowe EW. 2014 Computational methods in drug discovery. Pharmacol. Rev. 66, 334–395. (doi:10.1124/pr.111.007336)

26. Holliday JD, Randan SS, Willett P. 1995 A fast algorithm for selecting sets of dissimilar molecules from large chemical databases. Quant. Struct.-Act Rel. 14, 501–506. (doi:10.1002/qsar.19950140602)

27. Morris GM, Huey R, Lindstrom W, Sanner MF, Belew RK, Goodsell DS, Olson AJ. 2009 AutoDock4 and AutoDockTools: automated docking with selective receptor flexibility. J. Comput. Chem. 30, 2785–2791. (doi:10.1002/jcc.21256)

28. Forli S, Huey R, Pique ME, Sanner MF, Goodsell DS, Olson AJ. 2016 Computational protein–lignad docking and virtual drug screening with the AutoDock suite. Nat. Protoc. 11, 905. (doi:10.1038/nprot.2016.651)

29. Vrontaki E, Melagakis G, Mavromoustakis T, Mantfis A. 2016 Searching for anarhichin acid-based thumb pocket 2 HCV NS5B polymerase inhibitors through a combination of molecular docking, 3D-QSAR and virtual screening. J. Enzyme Inhib. Med. Chem. 31, 38–52. (doi:10.3109/14756366.2014.1030925)

30. Nguyen TH, Zhou H-X, Minh DL. 2018 Using the fast Fourier transform in binding free energy calculations. J. Comput. Chem. 39, 621–636. (doi:10.1002/jcc.25139)

31. Arviniti J, Medina C, Samuelsson J-E. 1994 A new method for predicting binding affinity in computer-aided drug design. Protein Eng. 7, 385–391. (doi:10.1093/proteins/7.3.385)

32. Miranda WE, Noskov SY, Valiente PA. 2015 Improving the LIE method for binding free energy calculations of protein–ligand complexes. J. Chem. Inf. Model. 55, 1867–1877. (doi:10.1021/acs.jicim.5b00012)
74. Hub JS, de Groot BL, van der Spoel D. 2010 g_wham—a free weighted histogram analysis implementation including robust error and autocorrelation estimates. J. Chem. Theory Comput. 6, 3713–3720. (doi:10.1021/ct100494z)

75. Efron B. 1979 Bootstrap methods: another look at the jackknife. Ann. Stat. 7, 1–26. (doi:10.1214/aos/1176344552)

76. Pedregosa F et al. 2011 Scikit-learn: machine learning in python. J. Mach. Learn. Res. 12, 2825–2830.

77. Jarzynski C. 1997 Equilibrium free-energy differences from nonequilibrium measurements: a master-equation approach. Phys. Rev. E 56, 5018–5035.

78. Lan NT, Vu KB, Dao Ngoc MK, Tran P-T, Hiep DM, Tung NT, Ngo ST. 2019 Prediction of AChE-ligand affinity using the umbrella sampling simulation. J. Mol. Graph. Model. 93, 107441. (doi:10.1016/j.jmgm.2019.107441)

79. Rathnayake AD et al. 2020 3C-like protease inhibitors block coronavirus replication in vitro and improve survival in MERS-CoV–infected mice. Sci. Transl. Med. 12, eabc5332. (doi:10.1126/scitranslmed.abc5332)

80. Ma C et al. 2020 Boceprevir, GC-376, and calpain inhibitors II, XII inhibit SARS-CoV-2 viral replication by targeting the viral main protease. Cell Res. 30, 678–692. (doi:10.1038/s41422-020-0556-z)

81. Li Z et al. 2020 Identify potent SARS-CoV-2 main protease inhibitors via accelerated free energy perturbation-based virtual screening of existing drugs. Proc. Natl Acad. Sci. USA 117, 27381–27387. (doi:10.1073/pnas.2010470117)

82. Vuong W et al. 2020 Feline coronavirus drug inhibits the main protease of SARS-CoV-2 and blocks virus replication. Nat. Commun. 11, 4282. (doi:10.1038/s41467-020-18096-2)

83. Zhang H, Yin C, Jiang Y, van der Spoel D. 2018 Force field benchmark of amino acids: I. Hydration and diffusion in different water models. J. Chem. Inf. Model. 58, 1057–1052. (doi:10.1021/acs.jcim.8b00026)

84. Zhang H, Jiang Y, Cui Z, Yin C. 2018 Force field benchmark of amino acids. 2. Partition coefficients between water and organic solvents. J. Chem. Inf. Model. 58, 1669–1681. (doi:10.1021/acs.jcim.8b00493)