FEP-based screening prompts drug repositioning against COVID-19

Zhe Li[1#], Xin Li[2,3#], Yi-You Huang[1#], Yaoxing Wu[4#], Lingli Zhou[4], Runduo Liu[1], Deyan Wu[1], Lei Zhang[4], Hao Liu[5], Ximing Xu[2,7], Yuxia Zhang[6], Jun Cui[4*], Xin Wang[2,7*], and Hai-Bin Luo[1*]

[1]Guangdong Provincial Key Laboratory of New Drug Design and Evaluation, School of Pharmaceutical Sciences, Sun Yat-Sen University, Guangzhou 510006, P. R. China

[2]Center for Innovative Marine Drug Screening & Evaluation (QNLM), School of Medicine and Pharmacy, Ocean University of China, 23 Xianggang E Road, Qingdao, 266100, China

[3]School of Life Sciences, Lanzhou University, 220 Tianshui S Road, Lanzhou, 734000, China

[4] MOE Key Laboratory of Gene Function and Regulation, State Key Laboratory of Biocontrol, School of Life Sciences, Sun Yat-sen University, Guangzhou, 510006, China

[5]High Performance Computing Center, Pilot National Laboratory for Marine Science and Technology (QNLM), 1 Wenhai Road, Aoshanwei, Qingdao, 266237, China

[6]Guangzhou Institute of Pediatrics, Guangzhou Women and Children's Medical Center, State Key Laboratory of Respiratory Diseases, Guangzhou Medical University, Guangzhou, 510623, China

[7]Marine Biomedical Research Institute of Qingdao, Qingdao, 266100, China

Abstract: The new coronavirus COVID-19, also known as SARS-CoV-2, has infected more than 300,000 patients and become a global health emergency due to the very high risk of spread and impact of COVID-19. There are no specific drugs or vaccines against COVID-19, thus effective antiviral agents are still urgently needed to combat this virus. Herein, the FEP (free energy perturbation)-based screening strategy is newly derived as a rapid protocol to accurately reposition potential agents against COVID-19 by targeting viral proteinase Mpro. Restrain energy distribution (RED) function was derived to optimize the alchemical pathway of FEP, which greatly accelerated the calculations and first made FEP possible in the virtual screening of the FDA-approved drugs database. As a result, fifteen out of twenty-five drugs validated *in vitro* exhibited considerable inhibitory potencies towards Mpro. Among them, the most potent Mpro inhibitor dipyridamole potentially inhibited NF-κB signaling pathway and inflammatory responses, and has just finished the first round clinical trials. Our result demonstrated that the FEP-based screening showed remarkable advantages in prompting drug repositioning against COVID-19.

Keywords: SARS-CoV-2, FEP, repositioning, Mpro, dipyridamole, and montelukast sodium

1. Introduction

The novel coronavirus 2019-nCoV (also known as HCoV-19 or SARS-CoV-2) outbreak had emerged from Wuhan, Hubei Province, China in December 2019[1, 2]. On March 22, there were 813,00 confirmed COVID-19 cases including 3,253 deaths in China. This virus has also infected more than 220,000 patients in all of the continents and over 180 other countries, such as Italy, Spain, U.S.A, Germany, France, and Iran gradually became a global pandemic due to the very high risk of spread and impact of COVID-19. To date, there is no specific treatment or vaccine against COVID-19, thus it is urgently need to repositioning potential agents against COVID-19.[3]

The COVID-19's replicase gene encodes two over-lapping translation products, polyproteins 1a and 1ab (pp1a and pp1ab), which mediate all of the functions required for viral replication. Mpro, as the key enzyme in proteolytic processing of viral replication, is initially released by the auto-cleavage of pp1a and pp1ab. Then Mpro in turn cleaves pp1a and pp1ab to release functional proteins necessary for viral replication.[4] In the view of essential functions of Mpro in viral life cycle and its high conservatism, it is an attractive target for the discovery of anti-COVID-19 agents.

Great efforts from various research groups have been done to discover new agents from several databases by targeting the target Mpro via several virtual screening strategy,[5,6] which consists of pharmacophore, molecule docking, and molecular simulations approaches. As a result, six drugs inhibited Mpro with IC_{50} values ranging from 0.67 to 21.4 μM.[5] These drug design methods contributed considerably to the lead discovery, but the computational accuracy and efficiency need to be improved especially when dealing with emergency situations such as the COVID-19 outbreak. Free energy perturbation (FEP) method is a promising method with satisfactory accuracy[7-14], but their actual applications to drug design are still limited to simulate minor structural changes of the ligands, thus predicting the relative binding free energy (RBFE).[7, 14] In order to perform virtual screening of a large

molecule database, the absolute binding free energy (ABFE) calculation must be performed for each ligand without using of a reference ligand structure. The FEP approach has an advantage in predicting the affinities more precisely between drugs and their targets than conventional methods, such as pharmacophore, molecule docking, and molecular simulations. However, the FEP-ABFE approaches are extremely expensive/time-consumpting and therefore not used for virtual screening purpose.[15, 16]

To accelerate the discovery of Mpro inhibitors from the small molecule database to combat COVID-19, we represent a newly derived FEP-ABFE-accelerated screening strategy together with bioassay validation to rapidly reposition potential agents against COVID-19 by targeting viral proteinase Mpro. As a result, fifteen of twenty-five drugs were validated *in vitro* to exhibit considerable inhibitory potencies towards Mpro. Among them, the most potent and representative Mpro inhibitor dipyridamole just finished its first-round clinical trials, and showed significant clinical outcomes.[17] In short, this is the first report to screen the FDA-approved database by using the FEP-ABFE approach, and this FEP-based method showed significant advantages by means of improving the hit rates and repositioning more potent leads.

2. Methods

2.1 Molecular docking

The crystal structure of viral proteinase Mpro (PDB ID: 6LU7)[18] for COVID-19 was used for the molecule docking purpose. Based on the crystal structure, more than 2500 small molecules in the FDA-approved drug database were first screened by using molecular docking program Glide[19]. Considering Mpro being a protease, Cys145-His41/Ser144-His163 can act as the nucleophilic agent and acid that assist the hydrolysis reaction of the substrate proteins, and Gly143 and Gln166 can form hydrogen bonds with the “CO-NH-Cα-CO-NH-Cα” structure of the backbone of the substrate protein. Thus, these 6 residues were considered as the key residues of Mpro in the screening. After docking, the

binding modes of all the ligands were carefully checked, and 100 molecules with specific interaction with the key residues and relatively high docking scores were selected for further FEP studies to evaluate their ABFE.

2.2 Free energy perturbation (FEP)

2.2.1 Preliminary MD simulations. All the 100 ligands selected by molecular docking were further evaluated by FEP calculations carried out in Gromacs-2019[20,21]. Before FEP calculations, 4 ns preliminary MD simulations were performed for each receptor-ligand complex to improve the fit of the ligand into the binding pocket. All the ligands are parameterized by the general AMBER force field (GAFF)[22]. Restrained electrostatic potential (RESP) charges calculation of relative ligand was performed with Gaussian 03 program[23] at the HF/6-31G* level. The parameters of protein were described by the AMBER FF14SB force field[24]. The TIP3P model[25] was used for water molecules. The systems were neutralized by adding counter ions (either Na^+ or Cl^- ions). The systems were first minimized by using steepest descent method for 5000 cycles and then heated from 0 to 298 K in an NVT ensemble within 100 ps. The systems were then equilibrated in an NPT ensemble with weak restraints of 1000 kJ/mol/nm^2 for 500 ps followed by a 4 ns unconstrained production simulation. The last snapshot of the MD simulations was used for the following FEP calculations, and the trajectory of the last 2 ns was analyzed to get the parameters for adding restraints between receptors and ligands.

2.2.3 Protocol for automatically adding restraints. Based on the preliminary MD simulations results, the FEP-ABFE calculations were carried based on the thermodynamic cycle given in Figure 1. As shown in the thermodynamic cycle, a restraint should be added to the receptor and ligand for each FEP calculation. The strategy of adding restraints first reported by Boresch *et al.*[26] was used in this study, which consists of one distance, two angles, and three dihedrals harmonic potentials with a force constant of 10 kcal/mol/Å^2 [rad^2]. The contribution of the restraints to the Lig (ΔA^L_{restr}) system was

calculated analytically, and the contribution of the restraint to the Rec-Lig system (ΔA_{restr}^{RL}) was calculated by FEP. According to the strategy, three atoms of the ligand and three atoms of the receptor will be selected to add the restraint. In order to add the restraints at the equilibrium position, a program was designed to automatically detect the required parameters and select the three ligand atoms and the three receptors atoms. For ligand, the heavy atom which is closest to the geometry center was selected as the first atom; the heavy atom which is most distant from the first atom was selected as the second atom; the heavy atom forms an angle that is larger than 90 degrees with the first two atoms and most distant from the first atom is selected as the third atom. For the receptor, based on the last 2 ns trajectory of the 4 ns preliminary MD simulations, the distances, angles, dihedrals between the three ligand atoms and C_α, C_c (carbon of the carboxyl group) and N (N atom of the amino group) atoms of all the residues within 5 Å of the ligand were calculated along the MD trajectory. C_α, C_c and N atoms from the same residue with the most stable (lowest standard deviation) distances, angles, and dihedrals values were selected as the three receptor atoms. The determined mean values of distance, angles, and dihedrals were used for adding the restraints between the three ligand atoms and the three receptor atoms.

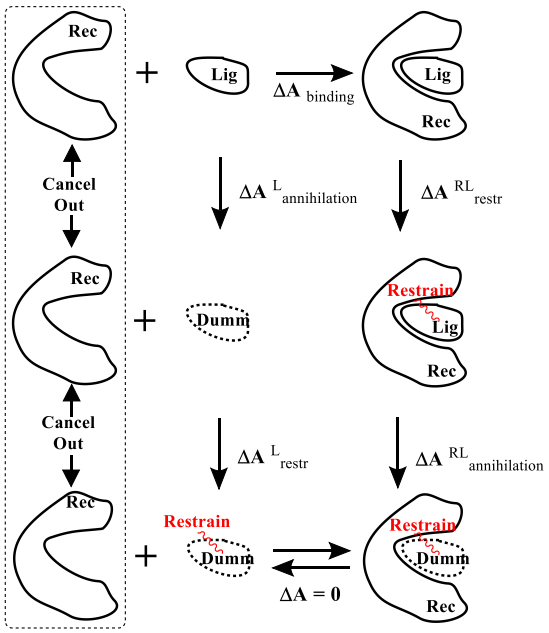

Figure 1. The thermodynamic cycle used for the FEP-ABFE calculations derived from our previous version[28].

2.2.3. ABFE calculation. In order to calculate the ABFE value between a ligand and its receptor, the ligand electrostatic and van der Waals interactions will be decoupled. In the recently published works [15, 16], 12 λ were used for adding restraints, 10 λ were used for decoupling electrostatic interactions, and 20 λ were used for decoupling vdW interactions, respectively, which need a lot of computation resources and thus is not applicable for the virtual screening purpose.

In order to make it possible to use FEP-ABFE in rapid discovery of inhibitors against COVID-19, the alchemical pathway was optimized to find one that can significantly not only improve the calculation speed but also keep the accuracy. According to the FEP theory, in order to calculate the free energy difference ΔA, probability distribution of potential energy difference between adjacent λ, denoted as $P(\Delta U)$, will be sampled. In the reported studies[27, 28], $P(\Delta U)$'s are considered as Gaussian-like distributions, which is true for the steps of electrostatic interactions and vdW interactions decoupling. However, for the restraint addition steps, the $P(\Delta U)$'s are far from Gaussian distributions. In this study, we first derived and introduce the restrain energy distribution (RED) function which can be used to correctly describe the $P(\Delta U)$ of the restraint addition steps. By using the automatic restrain adding program described above and fitting $P(\Delta U)$ by the RED function, restraint energy can be accurately estimated by just using single-step perturbation (λ from 0.0 directly to 1.0), which greatly decreased the calculation resources needed for the restraint addition. For the decoupling of electrostatic and vdW interactions, the alchemical pathways which can significantly decrease the number of λ needed and keep the accuracy were also studied. After all these procedures have been done, the FEP-ABFE calculations were greatly accelerated, which made the application of FEP-ABFE for virtual screening possible.

After the alchemical pathway is determined, for each window, 5000 cycles of steepest descent energy minimization were first carried out, and then 100 ps simulations in the NVT ensemble along with Langevin dynamics[29, 30] for temperature coupling was performed to heat the system to 298 K with

weak position restraints of 1000 kJ/mol/nm^2 applied to the receptor and the ligand heavy atoms. The simulation system was subsequently equilibrated in an NPT ensemble for 500 ps with the position restraints still applied followed by a 4 ns unconstrained production simulation. Pressure was coupled using Parrinello–Rahman pressure coupling scheme[31]. The LINCS constraint algorithm[32] was used only on H-bonds. In all simulations, the particle mesh Ewald (PME) algorithm[33] was used for calculation of long-range electrostatic interactions. ΔU values were sampled along the unconstrained simulation, and the free energy differences between each window were calculated by using the Bennet acceptance ratio (BAR) method[34, 35].

Several researches[36-38] have reported that the FEP-ABFE method is relatively accurate for electrically neutral ligands, but when the net charge of the ligand is not 0, there will be a systematic error. Considering many of the FDA approved drugs are not electrically neutral, all the molecules evaluated via the FEP-ABFE calculations were grouped by their net charges, and in each group, the molecules with highest binding free energies were selected for further bioassay validation.

2.3 Enzymatic assay for Mpro inhibitors. The pGEX4T1-Mpro plasmid was constructed (Atagenix, Wuhan) and transfected the E. coli strain BL21 (Codonplus, Stratagene). GST-tagged protein was purified by GST-glutathione affinity chromatography and cleaved with FXa. The purity of recombinant protein was greater than 95% as judged by SDS–PAGE. The binding of DIP to Mpro was measured by the Biacore 8K system (GE healthcare) at 25 °C. Mpro was immobilized on a CM5 chip surface via covalent linkage to Mpro N-terminus. First, the CM5 chip was activated using 1:4 *N*-hydroxysuccinimide (NHS)/1-ethyl-3-(3-dimethylaminopropyl) carbodiimide (EDC) at the flow rate of 10 μL/min for 7 min. Then, Mpro (90 μg/mL) in 10 mM acetate buffer (pH 4.5) was passed over separate flow cells at 10 μL/min for 3 min (1800 response units), which followed by a blocking step using ethanolamine (1 M, pH 8.5) at 10 μL/min for 7 min. Binding studies were performed by passing

5-80 μM of DIP over the immobilized Mpro at the flow rate of 30 μL/min and the contact time was set to be 200 s. A sample volume of 120 μL DIP in running buffer was injected into the flow cell and the bound ligand was washed by running buffer which contained 10 mM phosphate, 137 mM NaCl, 2.7 mM KCl, pH 7.4, and 0.5% DMSO.

2.4 Surface plasmon resonance (SPR) between ligands and Mpro. Surface plasmon resonance experiments were performed with a BIAcore T200 (GE Healthcare). The running buffer contained 1.05×PBS, 0.005% (vol/vol) surfactant P20, pH 7.4, and 1% DMSO. The purified DHODH, which was diluted in sodium acetate solution (pH 5.5) with a final concentration of 30 μg/mL, was immobilized on a CM5 sensor chip by amine coupling. All analyte measurements were performed at a flow rate of 30 μL/min. The analyte was diluted in the running buffer from the top concentration. Data processing and analysis were performed using BIAevaluation 1.1 software.

2.5. Inhibition of NF-κB signaling pathway and inflammatory responses. Luciferase reporter assay was applied to determine the function of indicated drugs in NF-κB signaling pathway. 293T cells were transfected with a mixture of luciferase reporter (firefly luciferase), TK-luc (renilla luciferase plasmid), then the cells were transfected with the active form of viral receptor RIG-I (RIG-I-CARD), or infected with Sendai virus (SeV). Luciferase activity was measured at 24 hrs after transfection or infection using a luminometer (Thermo scientific) with a dual-luciferase reporter assay system according to the manufacturer's instructions (Promega). Data represent relative firefly luciferase activity, normalized to renilla luciferase activity. qRT-PCR and ELISA assay was applied to determine the expression of pro-inflammatory cytokines. Total RNA was extracted from the cells using TriZol reagent (Invitrogen) according to the manufacturer's instructions. For RT-PCR analysis, cDNA was generated with HiScript II Q RT SuperMix for qPCR (+gDNA wiper) (Vazyme, R223-01) and analyzed by real-time

qRT-PCR using 2 × RealStar SYBR Mixture (Genestar). All data were normalized to GAPDH expression. Primer sequences were as follows: GAPDH, sense: 50-CGGAGTCAACGGATTTGGTC-30, anti-sense: 50-GACAAGCTTCCCGTTCTCAG-30; TNF-a, sense: 50-CCAGACCAAGGTCAACCTCC-30, and anti-sense: 50-CAGACTCGGCAAAGTCGAGA-30. For measurement of cytokines, human IL-1β in cell culture supernatants was detected with an ELISA kit (BD Biosciences, No. 557953) according to the manufacturer's protocols.

3. Result and Discussion

3.1 The RED function accelerated the FEP-ABFE approach

The probability distribution of energy difference between different windows $P(\Delta U)$ for calculating ΔA_{restr}^{RL} can be described by the following restrain energy distribution (RED) function (Eq. 1), and the derivation and detailed discussion of the RED function is given Supporting Information (SI) section S1.

$$P(\Delta U) = \frac{1}{\exp(c\Delta U)^{n_1}} a \cdot b^{\frac{3}{2}} \exp(-b\Delta U) \cdot (\Delta U)^2 + \frac{1}{1+\left(\frac{d}{\Delta U}\right)^{n_2}} \left(h \frac{1}{\sqrt{\pi}\mu_i} \exp\left(\frac{-(\Delta U - \mu)^2}{2\sigma} \right) \right) \quad (1),$$

The RED function can correctly describe the sampled restrain energy distribution $P(\Delta U)$ and greatly increase the convergence for calculating ΔA_{restr}^{RL}. As shown in Figure 2(a), the sampled $P(\Delta U)$ (yellow dots) for adding restraint can be fitted quite well by the RED function (red line). By using 3 targets and 28 ligands as a test set, with the use of the automatic restrain adding program and fitting the $P(\Delta U)$ by the RED function, the ΔA_{restr}^{RL} calculated by using one-step perturbation (2 λ with values of 0.0 and 1.0) has good correlation with that calculated by using the previously reported 12-λ perturbation with R^2 larger than 0.97 (results are shown in Figure 2(b)). The energy difference between the two

alchemical pathways for all the tested systems are less than 0.5 kcal/mol. Besides using one-step perturbation for the calculation of $\Delta A_{\mathrm{restr}}^{\mathrm{RL}}$, the alchemical pathway for calculating $\Delta A_{\mathrm{annihilation}}$ is also optimized which further decreased the number of λ values needed. Detailed discussion about the correctness and rationality of the RED function are given in SI sections S2 and S3; detailed results about the comparison between one-step perturbation and 12-λ perturbation are given in SI section S2; and detailed strategy for further optimizing the alchemical pathway for calculating $\Delta A_{\mathrm{annihilation}}$ are given in SI section S4.After all the above efforts, the FEP-ABFE can be accurately calculated by using just 16 λ values which need only 38 % computation resources compared with the reported FEP protocols, and the calculation is accelerated by more than 2.5 times. With such acceleration, FEP-ABFE calculation was first made possible to be applied in virtual screening to rapidly discover inhibitors against the coronavirus COVID-19.

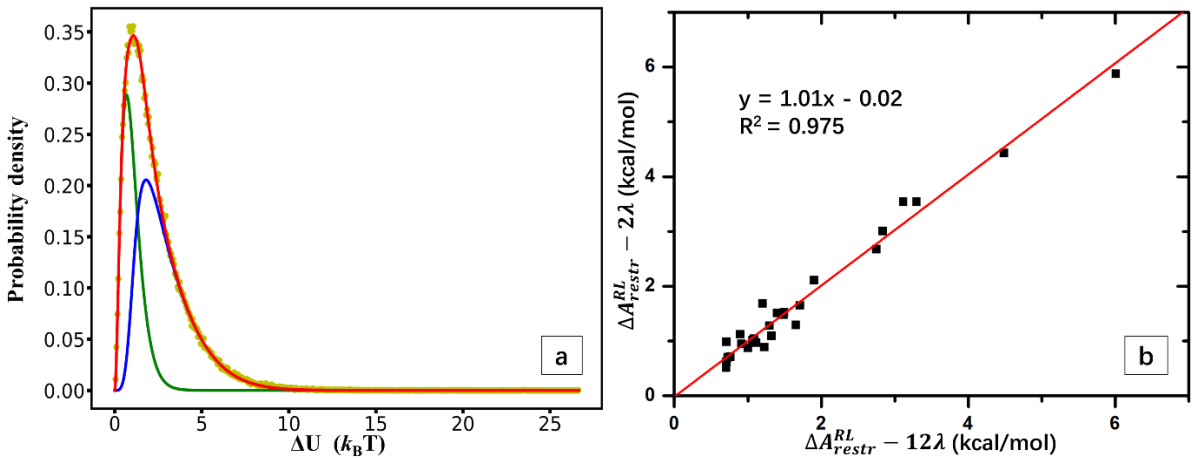

Figure 2. (a) P(Δ*U*) can be fitted by RED function quite well. The yellow dots are the corresponding sampled P(Δ*U*), the red line is the fitted RED function, the green line is the first term in the RED function, and the blue line is the second term in the RED function. (b) The linear regression between the results of $\Delta A_{\mathrm{restr}}^{\mathrm{RL}}$ calculated from 12λ perturbation and one-step perturbation (denoted as 2λ), which showed quite an excellent correlation with slope ~1.0, interception ~0.0, and R^2 larger than 0.97.

3.2 Remarkable accuracy of FEP-ABFE calculations based on test results from 28 receptor-ligand systems

The accuracy of the accelerated 16-λ-FEP-ABFE calculation was first tested against 3 targets with 28 receptor-ligand systems. One of the targets is BRD4 which is also used in the study of the FEP-ABFE method using 42 λ values previously reported by Aldeghi *et al.*,[15, 16] and since FEP-ABFE calculations have systematic bias when the net charge of the ligand is not 0, all the 8 neutral ligands from Aldeghi's work are used in this study.[15] Besides the target of BRD4, the other two targets include HIV-1 protease (10 electrically neutral ligands) and Human factor Xa (10 electrically neutral ligands). All the binding modes of the 28 calculated test systems are known from the protein data bank (www.rcsb.org), which ensured the correctness of the starting structures. As given in details in Table S3, the ligands are quite diverse to each other with the following properties: molecular weight from 241 to 662 Da; number of atoms from 22 to 89; number of rotatable bonds from 0 to 21; number of hydrogen bond acceptors from 1 to 7; number of hydrogen bond donors from 0 to 6; calculated log*P* from 1.17 to 4.85. The FEP-ABFE calculation results for all the 28 systems are summarized in Table S4, and the linear regression statistics between the calculation and experimental results are given in Table 1 and Figure 3. For the HIV-1 protease and Human factor Xa, most of the predicted binding free energies ΔG_{pred} are consistent with the experimental binding free energies ΔG_{exp} with the average prediction errors less than 2.0 kcal/mol. For BRD4, although the calculations showed considerable systematic error with all the calculation results shifting to the negative direction, the calculation results still have a linear correlation with the experimental ones. The test results based on the 28 ligands with diverse chemical scaffolds indicate the FEP-ABFE method achieves a remarkable accuracy, which encourage us to perform further virtual screening target the FDA-approved drug database via the FEP-ABFE method.

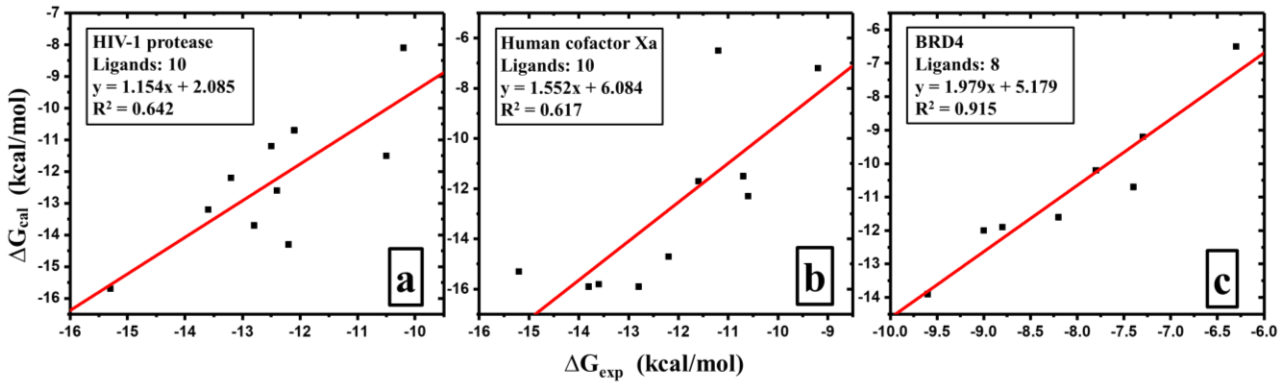

Figure 3. The linear regression models between the experimental ΔG_{exp} and the FEP-ABFE predicted ΔG_{cal} for the three targets and 28 ligands.

Table 1. The statistical results of the FEP-ABFE calculations for the three tested targets.

Target	R^2	slope	MAE*	RMSE#	Pearson's *r*	Spearsman's ρ
HIV-1 protease	0.642	1.154	1.073	1.239	0.794	0.673
Human factor Xa	0.617	1.552	1.923	2.334	0.788	0.784
BRD4	0.915	1.979	2.703	2.938	0.956	0.976

* MAE refers to mean absolute error. # RMSE refers to root mean square error.

3.3 FEP-ABFE screening led to drug repositioning of potent Mpro inhibitors

After all the FDA-approved drugs were docked into the binding site of Mpro, 100 molecules which have specific interactions with the six key amino acid residues including Cys145, His41, Ser144, His163, Gly143, and Gln166, were subjected to perform further FEP-ABFE calculations. Among these 100 drugs, 49, 46, and 5 of them are electrically neutral, with negative charges, and with positive charges, respectively. Since the FEP method is known to have systematic error when the ligands are not electrically neutral, to select molecules based on FEP-ABFE results, the molecules were grouped by their charges in order to make sure that there is an error cancellation inside each group. From each group, top 20% to 40% molecules were selected based on their ABFE values and stability along MD simulations. As a result, 25 molecules were bought for further experimental activity tests. According to the bioassay results, 15 out of the 25 molecules exhibited considerable inhibitory potencies against

Mpro (Figure 4). Since candesartan cilextil is a prodrug, its metabolite candesartan was also proved to be an Mpro inhibitor and showed an IC_{50} of 9.45 μM. Thus, totally 16 active molecules were found in this study, and their structures and activities are shown in Figure 4. The inhibitory curves of the most potent leads are given in Figure 5.

Among these 16 active molecules, ten leads (names shown in black in Figure 4) are originally discovered in this study, while other six compounds including disulfiram, chloroquine, montelukast sodium, atazanavir, indinavir, and maribavir are also reported in other previous studies[5, 39, 40] (Figure 4). Dipyridamole was identified to be the most potent Mpro inhibitor with an IC_{50} of 0.55 μM, which was first discovered by this FEP-ABFE method. In addition, the SPR assay was performed to determine the binding affinity between dipyridamole and Mpro *in vitro*. This has revealed that DIP bound to Mpro with an experimentally confirmed affinity of 34 μM ($K_{D,eq}$) (**Figure 6**). In short, the relative hit rate of 60% and the ten originally discovered leads including dipyridamole indicated the significant advantage of the FEP-ABFE method in discovering biologically active compounds over traditional methods such as molecular docking and molecular dynamics simulations.

The putative binding pattern between dipyridamole and Mpro is given in Figure 7a, and as shown in this figure, hydrogen bond interactions and hydrophobic interactions are the main driving forces for the binding between dipyridamole and Mpro. The protein-ligand interaction fingerprint (PLIF) is given in Figure 7b, which shows the key residues (Thr25, Asn142, Gly143, Ser144, His163 and Glu166) that played key roles in the binding between the 16 active molecules and the target Mpro. Recently, dipyridamole has finished the first-round clinical trials including 31 patients with COVID-19 and showed significant clinical improvement.[17] The second-round clinical trials including more than 130

patients with COVID-19 are subjected to further investigation into its therapeutic use in COVID-19, particularly in the shadow of its rapid spread into the world.

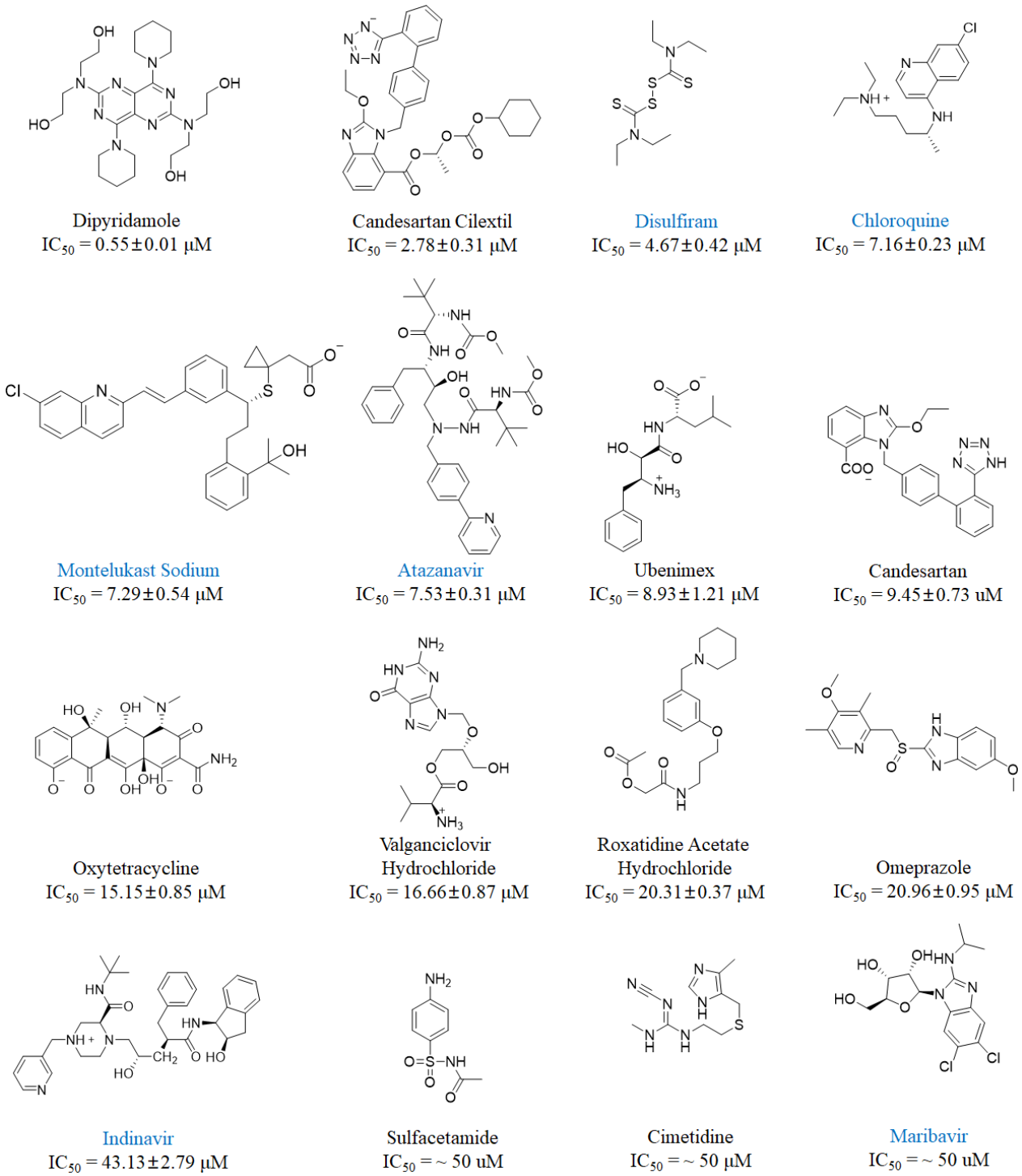

Figure 4. Chemical structures and IC_{50} values of 16 Mpro inhibitors (six compounds in blue color previously reported by other groups[5,39-40]) from the FDA-approved drug database. Disulfiram worked as the positive control for the bioassay (the IC_{50} value of 5.72 μM in the literature).

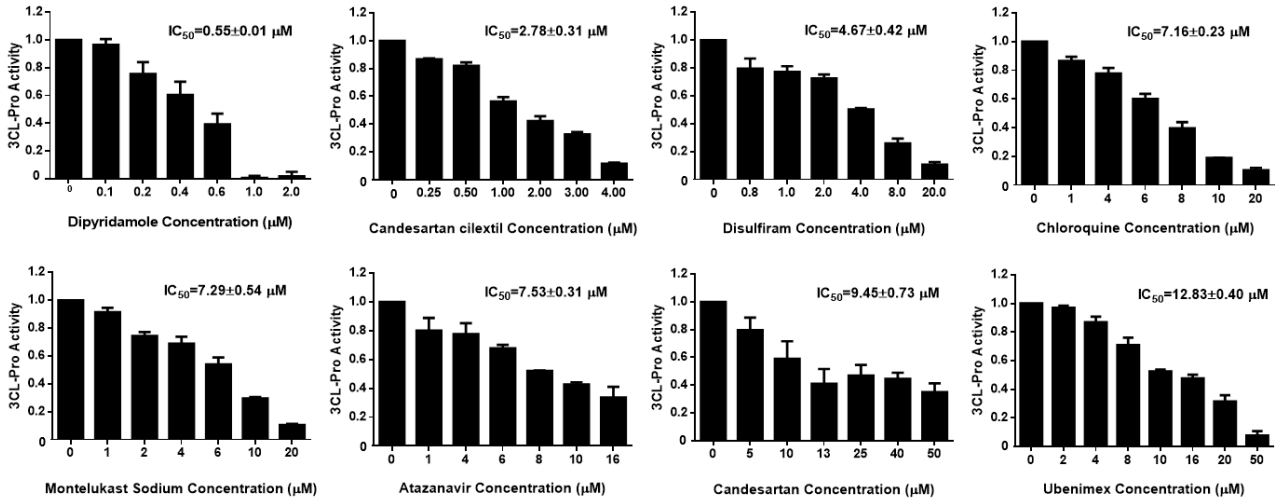

Figure 5. The inhibitory curves for the most potent Mpro inhibitors.

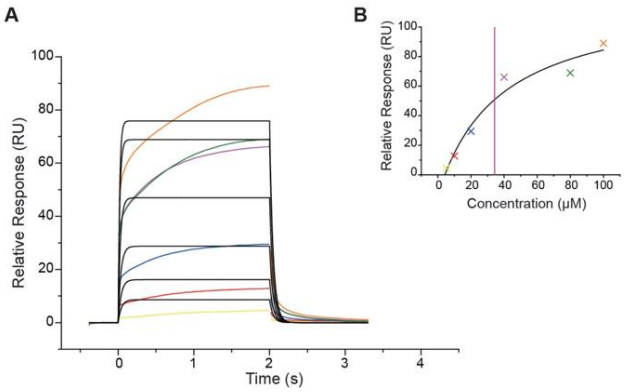

Figure 6. Dipyridamole suppresses Mpro activity by the SPR assay. Experimental SPR sensorgram from Biacore 8K between dipyridamole and Mpro, color points overlaid with a 1:1 Langmuir binding model kinetic fit (black). (A) Relative response from injection of DIP at 5, 10, 20, 40, 80, and 100 μM. Global analysis of the shape of the response curves yielded the parameters of on-rate (K_{on} = 3.20E+3), off-rate (K_{off} = 2.19E-1) and the ratio of K_{off} divided by K_{on} (K_D = 68 μM). （B） Equilibrium binding responses plotted versus dipyridamole concentration and fitted to a simple binding isotherm to yield an affinity of 34 μM ($K_{D,eq}$).

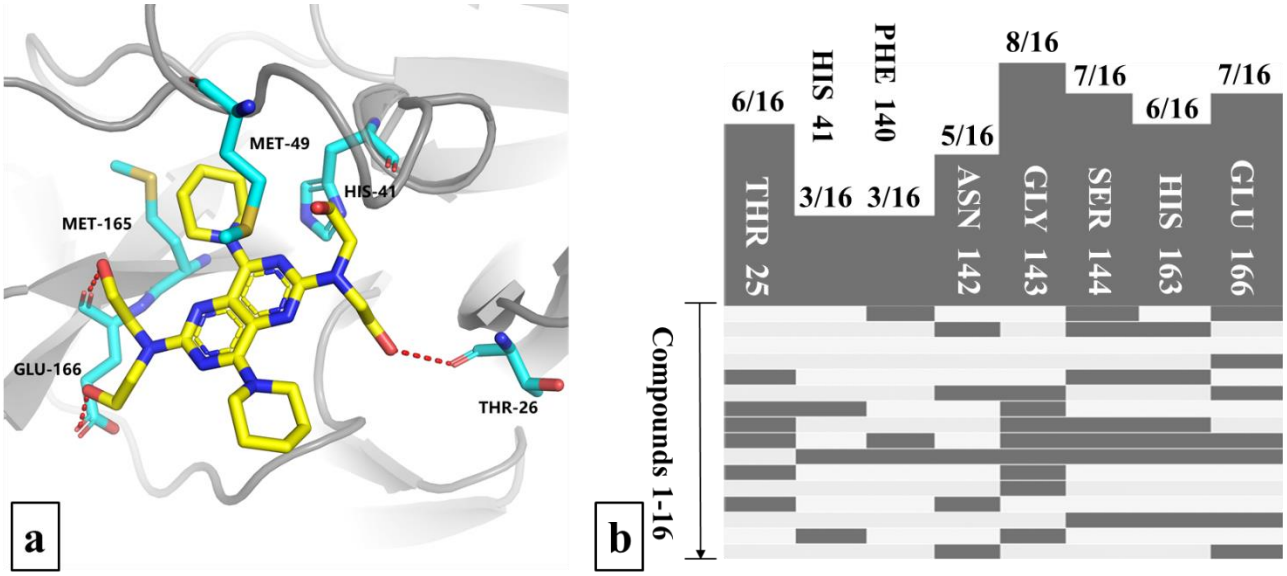

Figure 7. (a) Binding mode between Mpro and dipyridamole after molecular dynamics simulations. Dipyridamole is shown in yellow stick model, key amino acid residues of Mpro are shown in cyan stick model, and hydrogen-bonds are shown as red dashed lines. (b) The protein-ligand interaction fingerprint (PLIF) between the 16 active molecules and Mpro.

3.4 The effect of dipyridamole and montelukast sodium on NF-kB signaling and inflammation activation

It has been reported COVID-19 led to the increased amounts of pro-inflammatory cytokines in serum, which may be associated with pulmonary inflammation and lung damage in patients.[3] During viral infection, host pattern recognition receptors (PRRs) such as RIG-I, sense viral RNA and hierarchically trigger signaling cascades, which include adaptor proteins (MAVS and TRAF), a set of well-studied serine/threonine kinases (IKK complex and TBK1/IKKi) and several transcription factors (NF-κB and IRF3/IRF7) , to turn on the transcription of antiviral type I interferon (IFN) and pro-inflammatory cytokine genes, including TNFα and pro-IL-1β. Subsequently, pro-IL-1β can be further processed by active caspase-1 from inflammasome, to become mature IL-1β and initiates inflammatory responses (Figure 8A). Here we investigated the function of active molecules in virus-induced pro-inflammatory cytokines. Among the active molecules in Figure 5, two molecules dipyridamole and montelukast

sodium, inhibited the NF-κB activation induced by SeV infection or RIG-I (CARD) (an active form of RNA viral receptor RIG-I) overexpression (Figure 8B). In addition, both dipyridamole and montelukast sodium decreased SeV-induced mRNA expression of pro-inflammatory cytokine TNFα (Figure 8C). Interestingly, dipyridamole, but not montelukast sodium could also inhibit the activity of NLRP3 inflammasome and reduced IL-1β secretion (Figure 8D). Collectively, we found both dipyridamole and montelukast sodium showed a global inhibitory function on NF-κB signaling and inflammatory responses during viral infection.

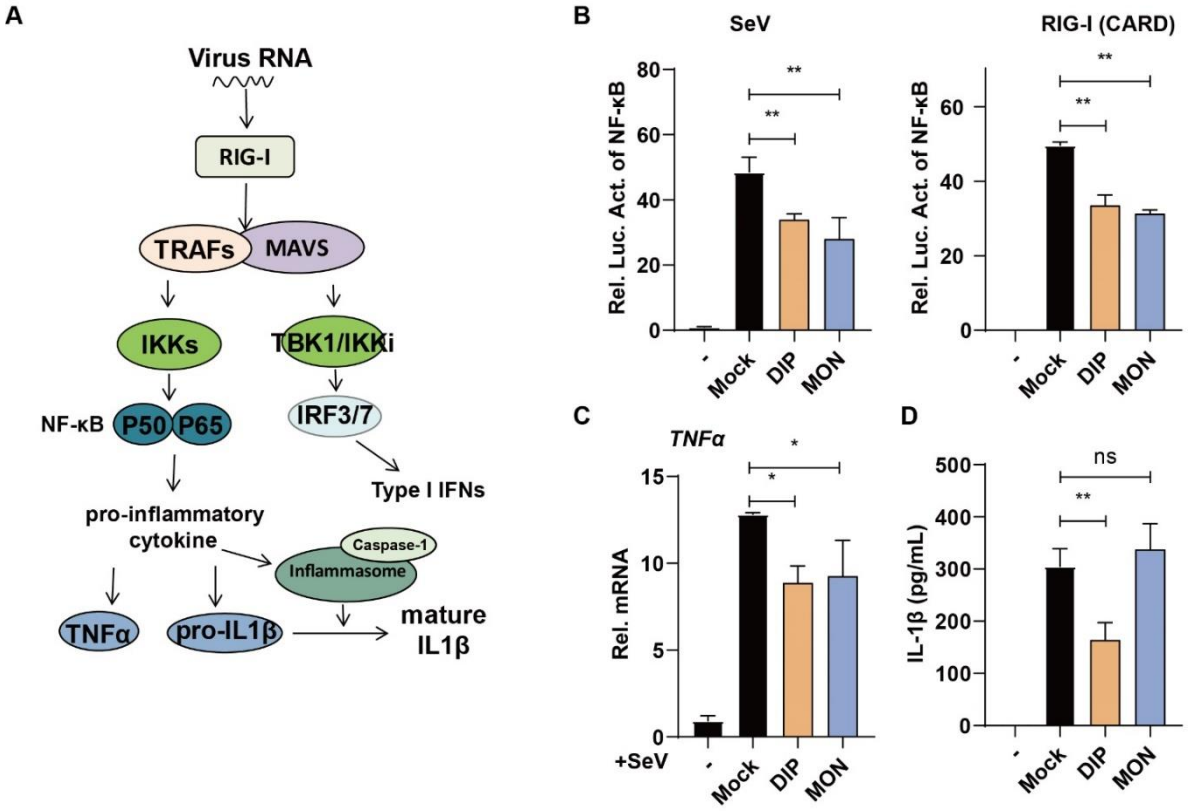

Figure 8. Dipyridamole and montelukast sodium inhibit NF-κB signaling and inflammation. (A) Illustration of virus-induced activation of NF-κB and inflammasome. (B) Dipyridamole (DIP) and montelukast sodium (MON) inhibited SeV- and RIG-I-mediated activation of NF-κB signaling. 293T cells were transfected with the NF-κB luciferase reporter, together with TK-luc, followed by SeV infection and dipyridamole or montelukast sodium treatment. (C) Real-time PCR analysis of *TNFα* mRNA in A549 cells that were treated with dipyridamole or montelukast sodium, followed by SeV infection or TNFα treatment. (D) HEK293T cells were transiently transfected with NLRP3, ASC, pro-

caspase-1, and pro-IL-1β plasmid, followed by dipyridamole or montelukast sodium treatment. Cell culture supernatants were collected 24 hr post-transfection to measure IL-1β release by ELISA.

Encouraged by the above results, dipyridamole has been approved for clinical studies by Dawu County People's Hospital and the First Affiliated Hospital of Guangzhou Medical University. Clinical trial (ChiCTR2000030055) was registered. On March 08, it has finished the first-round clinical trials including 31 patients with COVID-19 and showed significant clinical outcomes. Its second round of clinical trials are on-going with over 130 enrolled patients with COVID-19.

4. Conclusion

In this study, an accelerated FEP-ABFE based virtual screening strategy with bioassay validation is introduced as a rapid protocol to accurately discover anti-COVID-19 proteinase Mpro inhibitors. The RED function was derived and introduced to optimize the alchemical pathway of FEP-ABFE, which greatly accelerated the calculation, and FEP-ABFE was first made possible in the application of screening a molecule database. As a result, 15 out of 25 drugs tested by bioassay exhibited considerable inhibitory affinities against Mpro. Besides candesartan cilextil's metabolite candesartan, totally 16 active molecules were found in this study. Among them, we found that the most potent Mpro inhibitors (dipyridamole and montelukast) sodium showed a global inhibitory function on NF-κB signaling and inflammatory responses during viral infection. Dipyridamole has been approved for clinical studies by Dawu County People's Hospital and the First Affiliated Hospital of Guangzhou Medical University. Clinical trial (ChiCTR2000030055) was registered. As a result, we has finished its first-round clinical trials including 31 patients with COVID-19 and it demonstrated markedly improved clinical outcomes. In summary, the FEP-ABFE-based method showed significant advantages by means of improving the hit rates and repositioning potent agents, which is especially suitable in prompting drug repositioning

when not only against emergency situations such as the COVID-19 outbreak and virus mutations, but also against other important drug targets.

Acknowledgments: We cordially acknowledge Tencent Cloud and National Supercomputing centers in Shenzhen and Guangzhou for providing HPC resources for virtual screening and free energy perturbation calculations. We also acknowledge the Beijing Super Cloud Computing Center (BSCC) for providing HPC resources that have contributed to the research results reported within this paper. URL: http://www.blsc.cn/.

Funding: National Natural Science Foundation of China (8152204), Science Foundation of Guangzhou City (201904020023), Guangdong Provincial Key Laboratory of Construction Foundation (2017B030314030), Local Innovative and Research Teams Project of Guangdong Pearl River Talents Program (2017BT01Y093), Sun Yat-Sen University and Zhejiang University special scientific research fund for COVID-19 prevention and control, and philanthropy donation from individuals. The funders had no roles in the design and execution of the study.

References

1. Chen, N.; Zhou, M.; Dong, X.; Qu, J.; Gong, F.; Han, Y.; Qiu, Y.; Wang, J.; Liu, Y.; Wei, Y.; Xia, J.; Yu, T.; Zhang, X.; Zhang, L., Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: a descriptive study. *Lancet* **2020**.
2. Huang, C.; Wang, Y.; Li, X.; Ren, L.; Zhao, J.; Hu, Y.; Zhang, L.; Fan, G.; Xu, J.; Gu, X.; Cheng, Z.; Yu, T.; Xia, J.; Wei, Y.; Wu, W.; Xie, X.; Yin, W.; Li, H.; Liu, M.; Xiao, Y.; Gao, H.; Guo, L.; Xie, J.; Wang, G.; Jiang, R.; Gao, Z.; Jin, Q.; Wang, J.; Cao, B., Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *Lancet* **2020**.
3. Holshue, M. L.; DeBolt, C.; Lindquist, S.; Lofy, K. H.; Wiesman, J.; Bruce, H.; Spitters, C.; Ericson, K.; Wilkerson, S.; Tural, A.; Diaz, G.; Cohn, A.; Fox, L.; Patel, A.; Gerber, S. I.; Kim, L.; Tong, S.; Lu, X.; Lindstrom, S.; Pallansch, M. A.; Weldon, W. C.; Biggs, H. M.; Uyeki, T. M.; Pillai, S. K.; Washington State -nCo, V. C. I. T., First Case of 2019 Novel Coronavirus in the United States. *N Engl J Med* **2020**.
4. Adedeji, A. O.; Sarafianos, S. G., Antiviral drugs specific for coronaviruses in preclinical development. *Curr Opin Virol* **2014,** *8*, 45-53.
5. Jin, Z.; Du, X.; Xu, Y.; Deng, Y.; Liu, M.; Zhao, Y.; Zhang, B.; Li, X.; Zhang, L.; Duan, Y., et al, Structure of Mpro from COVID-19 virus and discovery of its inhibitors. *bioRxiv* **2020**. Li, Y., Zhang, J.; Wang, N.; Li, H.; Shi Y.; Guo, G.; Liu, K.; Zeng, H.; Zou, Q. Therapeutic drugs targeting 2019-nCoV main protease by high-throughput screening. *bioRxiv* **2020**.

6. Xiong, R.; Zhang, L.; Li, S.; Sun, Y.; Ding, M.; Wang, Y.; Zhao, Y.; Wu, Y.; Shang, W.; Jiang, X., Novel and potent inhibitors targeting DHODH, a rate-limiting enzyme in de novo pyrimidine biosynthesis, are broad-spectrum antiviral against RNA viruses including newly emerged coronavirus SARS-CoV-2. *bioRxiv* **2020**.
7. Wang, L.; Wu, Y.; Deng, Y.; Kim, B.; Pierce, L.; Krilov, G.; Lupyan, D.; Robinson, S.; Dahlgren, M. K.; Greenwood, J.; Romero, D. L.; Masse, C.; Knight, J. L.; Steinbrecher, T.; Beuming, T.; Damm, W.; Harder, E.; Sherman, W.; Brewer, M.; Wester, R.; Murcko, M.; Frye, L.; Farid, R.; Lin, T.; Mobley, D. L.; Jorgensen, W. L.; Berne, B. J.; Friesner, R. A.; Abel, R., Accurate and reliable prediction of relative ligand binding potency in prospective drug discovery by way of a modern free-energy calculation protocol and force field. *J. Am. Chem. Soc.* **2015,** *137* (7), 2695-703.
8. Steinbrecher, T. B.; Dahlgren, M.; Cappel, D.; Lin, T.; Wang, L.; Krilov, G.; Abel, R.; Friesner, R.; Sherman, W., Accurate Binding Free Energy Predictions in Fragment Optimization. *Journal of chemical information and modeling* **2015,** *55* (11), 2411-20.
9. Sun, Y.; Kollman, P. A., Determination of solvation free energy using molecular dynamics with solute Cartesian mapping: An application to the solvation of 18‐crown‐6. *J. Chem. Phys.* **1992,** *97* (7), 5108-5112.
10. Radmer, R. J.; Kollman, P. A., Free energy calculation methods: a theoretical and empirical comparison of numerical errors and a new method qualitative estimates of free energy changes. *J. Comput. Chem.* **1997,** *18* (7), 902-919.
11. Briggs, J. M.; Marrone, T. J.; McCammon, J. A., Computational science new horizons and relevance to pharmaceutical design. *Trends Cardiovas. Med.* **1996,** *6* (6), 198-203.
12. Kim, J. T.; Hamilton, A. D.; Bailey, C. M.; Domoal, R. A.; Wang, L.; Anderson, K. S.; Jorgensen, W. L., FEP-guided selection of bicyclic heterocycles in lead optimization for non-nucleoside inhibitors of HIV-1 reverse transcriptase. *J. Am. Chem. Soc.* **2006,** *128* (48), 15372-15373.
13. Jorgensen, W. L., Computer-aided discovery of anti-HIV agents. *Bioorganic & medicinal chemistry* **2016,** *24* (20), 4768-4778.
14. Abel, R.; Wang, L.; Harder, E. D.; Berne, B.; Friesner, R. A., Advancing Drug Discovery through Enhanced Free Energy Calculations. *Accounts of chemical research* **2017,** *50* (7), 1625-1632.
15. Aldeghi, M.; Heifetz, A.; Bodkin, M. J.; Knapp, S.; Biggin, P. C., Accurate calculation of the absolute free energy of binding for drug molecules. *Chemical science* **2016,** *7* (1), 207-218.
16. Aldeghi, M.; Heifetz, A.; Bodkin, M. J.; Knapp, S.; Biggin, P. C., Predictions of ligand selectivity from absolute binding free energy calculations. *Journal of the American Chemical Society* **2017,** *139* (2), 946-957.
17. Liu, X.; Li, Z.; Liu, S.; Chen, Z.; Zhao, Z.; Huang, Y.-y.; Zhang, Q.; Wang, J.; Shi, Y.; Xu, Y., et al, Therapeutic effects of dipyridamole on COVID-19 patients with coagulation dysfunction. *medRxiv* **2020**.
18. Jin, Z.; Du, X.; Xu, Y.; Deng, Y.; Liu, M.; Zhao, Y.; Zhang, B.; Li, X.; Zhang, L.; Peng, C., et al, Structure of Mpro from COVID-19 virus and discovery of its inhibitors. *bioRxiv* **2020**.
19. Halgren, T. A.; Murphy, R. B.; Friesner, R. A.; Beard, H. S.; Frye, L. L.; Pollard, W. T.; Banks, J. L., Glide: a new approach for rapid, accurate docking and scoring. 2. Enrichment factors in database screening. *Journal of medicinal chemistry* **2004,** *47* (7), 1750-1759.
20. Hess, B.; Kutzner, C.; Van Der Spoel, D.; Lindahl, E., GROMACS 4: algorithms for highly efficient, load-balanced, and scalable molecular simulation. *Journal of chemical theory computation* **2008,** *4* (3), 435-447.
21. Pronk, S.; Páll, S.; Schulz, R.; Larsson, P.; Bjelkmar, P.; Apostolov, R.; Shirts, M. R.; Smith, J. C.; Kasson, P. M.; van der Spoel, D., GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. *J Bioinformatics* **2013,** *29* (7), 845-854.
22. Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A., Development and testing of a general amber force field. *Journal of computational chemistry* **2004,** *25* (9), 1157-1174.
23. Frisch, M.; Trucks, G.; Schlegel, H.; Scuseria, G.; Robb, M.; Cheeseman, J.; Montgomery Jr, J.; Vreven, T.; Kudin, K.; Burant, J., Gaussian 03, revision C. 02; Gaussian, Inc.: Wallingford, CT, 2004. **2013**.
24. Case, D.; Babin, V.; Berryman, J.; Betz, R.; Cai, Q.; Cerutti, D.; Cheatham III, T.; Darden, T.; Duke, R.; Gohlke, H., The FF14SB force field. *Amber* **2014,** *14*, 29-31.
25. Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L., Comparison of simple potential functions for simulating liquid water. *The Journal of chemical physics* **1983,** *79* (2), 926-935.
26. Boresch, S.; Tettinger, F.; Leitgeb, M.; Karplus, M., Absolute binding free energies: a quantitative approach for their calculation. *The Journal of Physical Chemistry B* **2003,** *107* (35), 9535-9551.

27. Hummer, G.; Pratt, L. R.; García, A. E., Multistate Gaussian model for electrostatic solvation free energies. *Journal of the American Chemical Society* **1997,** *119* (36), 8523-8527.
28. Li, Z.; Huang, Y.; Wu, Y.; Chen, J.; Wu, D.; Zhan, C.-G.; Luo, H.-B., Absolute binding free energy calculation and design of a subnanomolar inhibitor of phosphodiesterase-10. *Journal of medicinal chemistry* **2019,** *62* (4), 2099-2111.
29. Van Gunsteren, W. F.; Berendsen, H. J., A leap-frog algorithm for stochastic dynamics. *J Molecular Simulation* **1988,** *1* (3), 173-185.
30. Goga, N.; Rzepiela, A.; De Vries, A.; Marrink, S.; Berendsen, H., Efficient algorithms for Langevin and DPD dynamics. *Journal of chemical theory computation* **2012,** *8* (10), 3637-3649.
31. Parrinello, M.; Rahman, A., Polymorphic transitions in single crystals: A new molecular dynamics method. *Journal of Applied physics* **1981,** *52* (12), 7182-7190.
32. Hess, B.; Bekker, H.; Berendsen, H. J.; Fraaije, J. G., LINCS: a linear constraint solver for molecular simulations. *Journal of computational chemistry* **1997,** *18* (12), 1463-1472.
33. Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G., A smooth particle mesh Ewald method. *The Journal of chemical physics* **1995,** *103* (19), 8577-8593.
34. Bennett, C. H., Efficient estimation of free energy differences from Monte Carlo data. *Journal of Computational Physics* **1976,** *22* (2), 245-268.
35. Shirts, M. R.; Chodera, J. D., Statistically optimal analysis of samples from multiple equilibrium states. *The Journal of chemical physics* **2008,** *129* (12), 124105.
36. Wang, J.; Deng, Y.; Roux, B., Absolute binding free energy calculations using molecular dynamics simulations with restraining potentials. *Biophysical journal* **2006,** *91* (8), 2798-2814.
37. Fujitani, H.; Tanida, Y.; Ito, M.; Jayachandran, G.; Snow, C. D.; Shirts, M. R.; Sorin, E. J.; Pande, V. S., Direct calculation of the binding free energies of FKBP ligands. *The Journal of chemical physics* **2005,** *123* (8), 084108.
38. Lapelosa, M.; Gallicchio, E.; Levy, R. M., Conformational transitions and convergence of absolute binding free energy calculations. *Journal of chemical theory computation* **2012,** *8* (1), 47-60.
39. Wang, M.; Cao, R.; Zhang, L.; Yang, X.; Liu, J.; Xu, M.; Shi, Z.; Hu, Z.; Zhong, W.; Xiao, G., Remdesivir and chloroquine effectively inhibit the recently emerged novel coronavirus (2019-nCoV) in vitro. *Cell research* **2020,** *30* (3), 269-271.
40. Dong, L.; Hu, S.; Gao, J. J. D. D.; Therapeutics, Discovering drugs to treat coronavirus disease 2019 (COVID-19). *Drug Discoveries & Therapeutics* **2020,** *14* (1), 58-60.

Supporting Information

Section S1. Derivation of the restrain energy distribution (RED) function.

Considering the situation where the restraint between the receptor and ligand is added at the equilibrium position, the interaction between receptor and the ligand consists of two parts including the force field and the restraint. Interaction comes from force field can be simplified as a harmonic potential since the ligand is near the equilibrium, and the restraint force is also a harmonic biasing force. Thus, the total interaction potential between receptor and ligand can be simplified as

$$E_{rec-lig} = k_{total}\,(r - r_0)^2 \qquad \text{(S1)}$$

where k_{total} is the apparent force constant, and $(r - r_0)$ is the generalized distance between the current position and the equilibrium. When adding restraints, the ΔU between adjacent λ will be sampled in order to calculate the free energy of the added restraints, and the ΔU can be represented by the following equation,

$$\Delta U_{i+1,i} = \Delta\lambda_{i+1,i} k_{res} (r - r_0)^2 \qquad \text{(S2)}$$

where i and i+1 means the adjacent ith and $(i + 1)$th window, and k_{res} is the force constant of the added restraint.

Under the above circumstances, the probability distribution $P\left(\Delta U_{i+1,i}\right)$ can be represented as

$$P\left(\Delta U_{i+1,i}\right) = \frac{exp(-\beta E_{rec-lig})\Omega(E_{rec-lig})\Delta U_{i+1,i}}{Z} \qquad \text{(S3)}$$

where Z is the partition function, $\beta = (k_B T)^{-1}$, $\Omega(E_{rec-lig})$ is the density of state. Since the generalized distance between the current position and the equilibrium $(r - r_0)$ can be considered as a three dimensional vector, the density of state $\Omega(E_{rec-lig})$ can be represented as

$$\Omega\left(E_{rec-lig}\right) = 4\pi(r - r_0)^2 \qquad \text{(S4)}.$$

Here, in Eq. (S4) we omitted $\frac{V^N}{h^{3N}N!}$ which is a constant factor and won't affect the form of the derived RED function.

By substituting Eq. (S1) (S2) and (S4) to Eq. (S3), we can get

$$P(\Delta U_{i+1,i}) = \frac{exp(-\beta k_{total}(r-r_0)^2)4\pi(r-r_0)^2\Delta\lambda_{i+1,i}k_{res}(r-r_0)^2}{Z} \text{ (S5)},$$

and by substituting $(r-r_0)^2$ term in Eq. (S5) with $\Delta U_{i+1,i}$, representing$\Delta U_{i+1,i}$by ΔU, combining the terms which are constant factors, and normalizing the distribution function, the distribution for $P(\Delta U)$ near the equilibrium has the form of Eq. (S6)

$$P(\Delta U) = a \cdot b^{\frac{3}{2}} \exp(-b\Delta U) \cdot (\Delta U)^2 \text{ (S6)},$$

where a and b are two constants, and $b^{\frac{3}{2}}$ is a factor for normalization purpose.

During the simulation, when some of the sampled points are far from the equilibrium, we need to add a Gaussian term to Eq. 6 to represent such situations, and the $P(\Delta U)$ can be represented as $P(\Delta U) =$

$$\frac{1}{\exp(c\Delta U)^{n_1}} a \cdot b^{\frac{3}{2}} \exp(-b\Delta U) \cdot (\Delta U)^2 + \frac{1}{1+\left(\frac{d}{\Delta U}\right)^{n_2}} \left(h \frac{1}{\sqrt{\pi}\mu_i} \exp\left(\frac{-(\Delta U-\mu)^2}{2\sigma}\right)\right) \text{ (S7)},$$

which is the restrain energy distribution (RED) function used in this study, where a, b, c, d, h, μ, and σ are 7 parameters to be fitted. In the RED function, $\frac{1}{\exp(c\Delta U)^{n_1}}$ and $\frac{1}{1+\left(\frac{d}{\Delta U}\right)^{n_2}}$ are two factors used to combine the two terms and keep the function to be described only by the first term (denoted as harmonic energy term hereinafter) when ΔU is small and only by the second term (denoted as Gaussian term hereinafter) when ΔU is large. The two constants n_1 and n_2 can be set to two relatively big integers, and we chose 10 and 4 for n_1 and n_2, respectively. We can also choose other values for n_1 and n_2 from 2 to 10, which do not affect the fitting results too much. The results of fitting $P(\Delta U)$ by RED function is discussed in detail in the following section (SI Section S2).

Section S2. Fitting by the RED function greatly increased the convergence of restraints steps and reduced the calculation needed in the FEP-ABFE prediction.

Although the accuracy of FEP calculation has been reported by several studies[1, 2], the calculation resources needed for a single FEP-ABFE prediction is massive. Thus, in order to rapidly discover inhibitors for clinical use against the coronavirus COVID-19, it is important to increase the convergence and decrease the calculation while keep the accuracy of FEP. Normally, more than 10 λ values will calculated during the addition of restraints, such as the works recently reported by Aldeghi *et al.* that 12 non-uniformly distributed λ values are used (0.0, 0.01, 0.025, 0.05, 0.075, 0.1, 0.15, 0.2, 0.3, 0.5, 0.75, 1.0) for the addition of restraints.[1, 2] However, with the use of the automatic restrain adding program and fitting the probability distribution of the sampled energy difference $P(\Delta U)$ function, the $P(\Delta U)$ can be fitted quite well, and the convergence for calculating restraint energy was greatly improved. As a result, the restraint energy can be accurately calculated by just one-step perturbation with λ values change directly from 0.0 to 1.0, and the calculation can be greatly accelerated. Using a HIV-1 protease-ligand complex (crystal structure PDB ID: 2QHY) as an example, the restraint was added by using one-step perturbation and 12-λ perturbation, respectively. The sampled $P(\Delta U)$'s for adding restraints are fitted by RED function, and the results are shown in Figure S1. In each picture, the yellow dots are the corresponding sampled $P(\Delta U)$, the green line is the harmonic energy term in the RED function, and the blue line is the Gaussian term in the RED function. As expected, the RED function is described only by the harmonic energy term when ΔU is small and only by the Gaussian term when ΔU is large, and the sampled $P(\Delta U)$ can be fitted well. The free energies of adding restraints are 0.912 and 0.948 kcal/mol for the 12-λ perturbation and the one-step perturbation, respectively, which means the convergence for the one-step perturbation is pretty good.

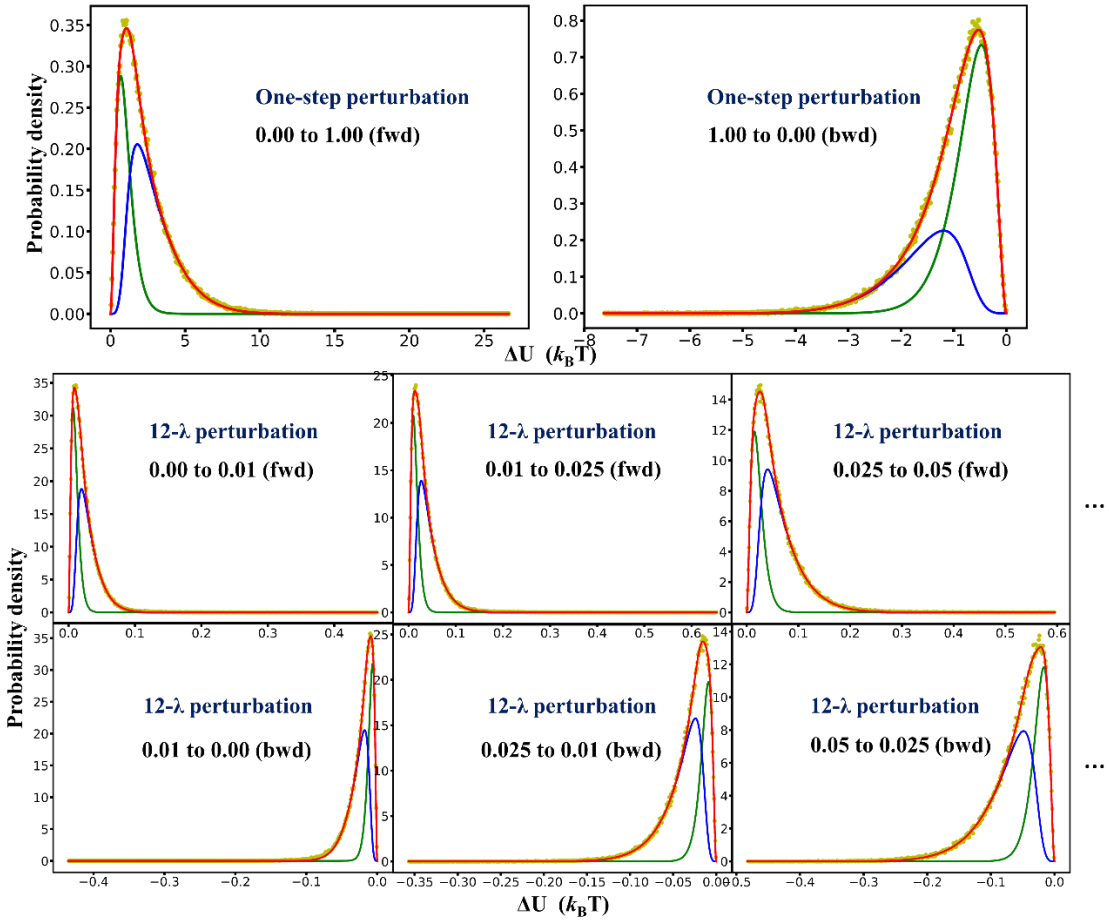

Figure S1. The sampled $P(\Delta U)$'s fitted by the RED function. The yellow dots are the corresponding sampled $P(\Delta U)$, the green line is the harmonic energy term in the RED function, and the blue line is the Gaussian term in the RED function.

In order to further justify the applicability of one-step perturbation (2 λ with values of 0.0 and 1.0, denoted as 2 λ below) in adding restraints, three different targets and 28 ligands were tested to compare the $\Delta A_{\mathrm{restr}}^{\mathrm{RL}}$ results calculated from 12 λ and 2 λ. The bond, angle, and dihedral parameters for the restraints are the same for 12 λ and 2 λ calculations. Table S1 summarized the results for the $\Delta A_{\mathrm{restr}}^{\mathrm{RL}}$, and the calculation results revealed that the single-step perturbation have good convergence with the energy difference for all the tested systems less than 0.5 kcal/mol. Thus, by using automatic restrain adding program and fitting the sampled $P(\Delta U)$ by RED function, the calculation resources needed for adding restraints can be greatly decreased while keeping the accuracy.

Table S1. The details of free energy changes for the complex in the annihilation for the traditional 42-window schemes and the 16-window schemes [a].

λ [b]	3SVG		3U5L		3U5J		4HBV	
	$\Delta A^{RL}_{42\lambda}$ [c]	$\Delta A^{RL}_{16\lambda}$ [d]	$\Delta A^{RL}_{42\lambda}$	$\Delta A^{RL}_{16\lambda}$	$\Delta A^{RL}_{42\lambda}$	$\Delta A^{RL}_{16\lambda}$	$\Delta A^{RL}_{42\lambda}$	$\Delta A^{RL}_{16\lambda}$
(0.0, 0.0, 0.0)	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
(0.01, 0.0, 0.0)	0.239	-	0.016	-	0.029	-	0.030	-
(0.025,0.0,0.0)	0.394	-	0.055	-	0.067	-	0.085	-
(0.05, 0.0, 0.0)	0.539	-	0.117	-	0.121	-	0.165	-
(0.075,0.0,0.0)	0.697	-	0.154	-	0.165	-	0.219	-
(0.10, 0.0, 0.0)	0.914	-	0.202	-	0.204	-	0.271	-
(0.15, 0.0, 0.0)	1.365	-	0.290	-	0.277	-	0.366	-
(0.20, 0.0, 0.0)	1.681	-	0.346	-	0.369	-	0.449	-
(0.30, 0.0, 0.0)	2.086	-	0.476	-	0.528	-	0.592	-
(0.50, 0.0, 0.0)	2.736	-	0.699	-	0.723	-	0.833	-
(0.75, 0.0, 0.0)	3.416	-	0.863	-	0.924	-	1.082	-
(1.0, 0.0, 0.0)	3.992	3.242	1.002	0.912	1.082	1.115	1.294	1.311
(1.0, 0.1, 0.0)	6.963	-	3.123	-	3.184	-	3.715	-
(1.0, 0.2, 0.0)	9.391	9.129	5.077	4.987	5.122	5.152	5.815	5.380
(1.0, 0.3, 0.0)	11.611	-	6.874	-	6.881	-	7.613	-
(1.0, 0.4, 0.0)	13.466	12.991	8.456	8.368	8.429	8.447	9.259	8.961
(1.0, 0.5, 0.0)	14.830	-	9.817	-	9.821	-	10.667	-
(1.0, 0.6, 0.0)	15.896	15.339	10.999	11.050	10.978	10.975	11.613	11.248
(1.0, 0.7, 0.0)	16.669	-	12.003	-	11.931	-	12.253	-
(1.0, 0.8, 0.0)	17.236	16.627	12.758	12.786	12.568	12.825	12.692	12.601
(1.0, 0.9, 0.0)	17.631	-	13.285	-	13.012	-	12.902	-
(1.0, 1.0, 0.0)	17.857	17.198	13.662	13.650	13.526	13.530	12.985	13.166
(1.0, 1.0, 0.05)	19.052	-	15.025	-	14.877	-	13.992	-
(1.0, 1.0, 0.10)	20.241	19.606	16.387	16.401	16.205	16.075	14.995	15.157
(1.0, 1.0, 0.15)	21.395	-	17.738	-	17.503	-	15.979	-
(1.0, 1.0, 0.20)	22.505	21.945	19.070	19.096	18.737	18.590	16.977	17.198
(1.0, 1.0, 0.25)	23.599	-	20.381	-	19.928	-	17.978	-
(1.0, 1.0, 0.30)	24.631	24.187	21.657	21.669	21.107	20.934	18.969	19.225
(1.0, 1.0, 0.35)	25.617	-	22.914	-	22.247	-	19.939	-
(1.0, 1.0, 0.40)	26.622	-	24.170	-	23.323	-	20.889	-
(1.0, 1.0, 0.45)	27.580	-	25.419	-	24.365	-	21.847	-
(1.0, 1.0, 0.50)	28.516	28.207	26.652	26.735	25.433	25.271	22.825	23.204
(1.0, 1.0, 0.55)	29.379	-	27.876	-	26.487	-	23.790	-
(1.0, 1.0, 0.60)	30.118	-	28.984	-	27.458	-	24.713	-
(1.0, 1.0, 0.65)	30.660	-	30.000	-	28.361	-	25.501	-
(1.0, 1.0, 0.70)	31.043	30.754	30.883	31.052	28.974	28.881	25.961	26.450
(1.0, 1.0, 0.75)	31.292	-	31.599	-	29.356	-	26.174	-
(1.0, 1.0, 0.80)	31.108	30.866	32.084	32.223	29.559	29.733	26.115	26.389
(1.0, 1.0, 0.85)	30.272	-	31.945	-	29.208	-	25.673	-
(1.0, 1.0, 0.90)	29.402	29.477	31.368	31.196	28.490	29.334	25.189	25.507
(1.0, 1.0, 0.95)	28.986	29.101	31.069	30.670	28.098	28.929	24.969	25.329
(1.0, 1.0, 1.0)	29.156	29.259	31.237	30.728	28.206	29.019	25.128	25.468

[a] All the free energy differences are calculated by BAR methods. The unit of free energy differences is kcal/mol.

[b] In the lambda arrays, the first value refers to the bonded-lambdas and the second and third value represents the value of coul-lambdas and vdw-lambdas, which define instead the lambda vectors for the decoupling of the coulombic and Lennard–Jones interactions.

[c] $\Delta A^{RL}_{42\,\lambda}$ refers to the free energy difference between the initial state of perturbation and the specific alchemical state defined by the relative lambda arrays, whose alchemical pathway is defined by 42 windows.

[d] $\Delta A^{RL}_{16\,\lambda}$ refers to the free energy difference between the initial state of perturbation and the specific alchemical state defined by the relative lambda arrays, whose alchemical pathway is defined by 16 windows.

Section S3. RED function can also correctly describe the restrain energy distribution even if the restrain is not added to the equilibrium position.

Although the automatic restrain adding program can make sure that the parameters chose for the restrain are close to the equilibrium, it is still important to have an understanding about the behavior of the probability distribution of ΔU when the parameters are not at the equilibrium. As an example, we determined the restraint parameters for 2QHY by using the original crystal structure instead by using the pre-equilibrated trajectory, and the resulted $P(\Delta U)'s$ are shown in Figure S2. As can be seen, the probability distributions for the initial several windows showed distributions with "double-peak" shapes. In theory, the "double-peak" shape distribution $P(\Delta U)$ will happen when relative position between the receptor and ligand, which we use to determine the ligand restraints addition parameters, is far from the equilibrium state, and at the same time, the harmonic potentials used to apply the restraint are not strong enough. According to FEP theory, the probability distribution near the lower tail of $P(\Delta U)$ will greatly influence the accuracy of ΔA calculation. The "double-peak" shape of the probability distribution is far from the equilibrium, and the distribution pattern near the lower tail varies from time to time during the simulation, which will result in a very unstable free energy estimation. Thus, to deal with the instability in energy calculation, in the reported studies[1, 2], the restraints are added with more than 10 λ values, and the first several λ are selected rather close to each other (0.0, 0.01, 0.025, 0.05, 0.075, 0.1). Although the purpose of the RED function is to describe the $P(\Delta U)$ when the restraints are added near the equilibrium state, the RED function can also describe the $P(\Delta U)$ well even when restraints are not added near the equilibrium. Using the forward perturbation from 0.00 to 0.01 as an example, which is shown in Figure S2, when the added restrain is relatively weak and not at the equilibrium position, the peak on the right which is described by the Gaussian term of the RED function is caused by the ligand occupying the equilibrium position, and the peak on the left is caused by the added restraints which is described by the harmonic energy term of the RED function. The RED function can describe the $P(\Delta U)$ of adding restraints quite well no

matter whether the restraints are added near the equilibrium, which further proved the correctness and rationality of the function. The RED function can be used to further optimize and give a deeper understanding of the FEP-ABFE calculation in the future studies.

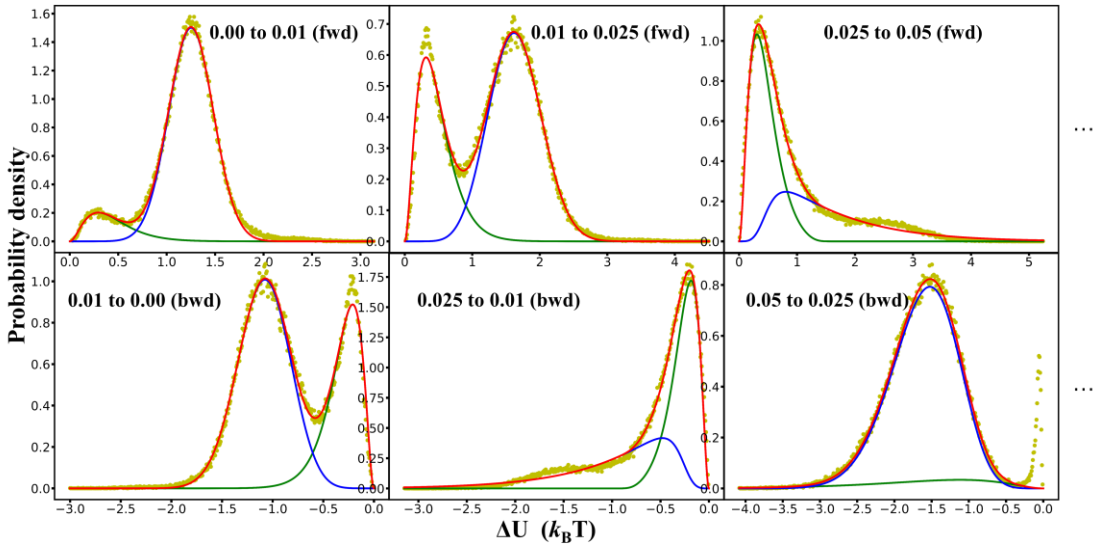

Figure S2. The sampled $P(\Delta U)$'s fitted by the RED function when the restraints are not added near the equilibrium position. The yellow dots are the corresponding sampled $P(\Delta U)$, the green line is the harmonic energy term in the RED function, and the blue line is the Gaussian term in the RED function.

Section S4. FEP calculations can be accelerated by more than 2.5 times while keeping the accuracy with appropriate selection of λ values.

In order to further accelerate the calculation, we studied the effect of decreasing the number of λ values for both charge and vdW interactions decoupling by using 4 different receptor-ligand systems. Corresponding crystal structures (PDB codes: 3SVG, 3U5L, 3U5J, 4HBV) were used in order to make sure that the binding modes between the receptor and ligands are correct. For all the 4 crystal structures, the ABFE was first calculated by using 42 λ (12 λ for restrain, 10 λ for charge, 20 λ for vdW) as reported by the recently published works[1, 2], and then the number of λ values were decreased for the decoupling of both charge and vdW interactions. As can be seen in Table S2, by using 16 λ values in total (2 λ for restrain with fitting by RED function, 5 λ for charge, and 9 λ for vdW) the energy calculation results are quite close to that calculated by using 42 λ values for all the 4 systems tested (all the energy differences are within 1 kcal/mol). Compared with 42 λ, FEP-ABFE calculation using just 16 λ values need only 38 % computation resources, and the calculation is accelerated by more than 2.5 times. With such acceleration, FEP-ABFE calculation is now possible to be applied in virtual screening to rapidly discover inhibitors against the coronavirus COVID-19.

Table S2. Comparison of ΔA_{restr}^{RL} between 12 λ and 2 λ calculations.[a]

Protein targets	PDB	$\Delta A_{12restr}^{RL}$ [b]	ΔA_{2restr}^{RL} [c]	$\Delta\Delta A_{restr}^{RL}$ [d]
HIV-1 protease	2QHY	0.912	0.948	-0.036
	4U8W	1.486	1.481	+0.006
	5UPZ	1.323	1.097	+0.226
	1AJV	0.708	0.989	-0.281
	1D4H	4.483	4.431	+0.052
	1D4I	0.897	1.127	-0.230
	1EBY	1.222	0.891	+0.331
	1EBZ	0.712	0.631	+0.080
	3A2O	0.706	0.513	+0.192
	1G2K	6.009	5.878	+0.131
Human factor Xa	1F0S	1.900	2.113	-0.214
	1FJS	1.652	1.296	+0.356
	1MQ6	1.494	1.520	-0.026
	1NFW	3.296	3.545	-0.301
	1NFX	1.706	1.651	+0.055
	2J34	2.747	2.681	+0.066
	2P16	0.725	0.703	+0.021
	2P95	0.710	0.622	+0.048
	2XC0	1.111	0.972	+0.139
	2VVV	3.114	3.545	-0.431
BRD4	3MXF	1.197	1.687	-0.490
	4MR3	1.398	1.514	-0.116
	3U5L	1.002	0.876	+0.126
	4MR4	0.761	0.716	+0.045
	3U5J	1.082	1.049	+0.034
	3SVG	1.068	1.023	+0.301
	4HBV	1.294	1.282	+0.013
	4J0R	2.839	3.009	-0.170

[a] The unit of free energy is kcal/mol.

[b] $\Delta A_{12restr}^{RL}$ is the restrain energy calculated by using 12 λ values and BAR method.

[c] ΔA_{2restr}^{RL} is the restrain energy calculated from one step perturbation with the use of 2 λ value (0.0, 1.0). The sampled energy distribution from each λ was first fitted by RED function, and then BAR method was used to calculate the energy estimates. [d] $\Delta\Delta A_{restr}^{RL}$ is the difference of $\Delta A_{12restr}^{RL}$and ΔA_{2restr}^{RL}.

Table S3. Physical chemical properties of the ligands. PDB is the PDB code the ligand belongs to. MW is the molecular weight in Daltons; Netq is the net charge; NROT is the number of rotatable bonds; HBA is the number of hydrogen bond acceptors; HBD is the number of hydrogen bond donors; cLogP is the calculated octanol/water partition coefficient (computed with XLOGP3). All the properties are gained from the PDBbind-CN Database.

PDB	MW(Da)	No Atoms	Netq(e)	NROT	HBA	HBD	cLogP
2QHY	583.643	67	0	15	5	3	3.05
4U8W	583.645	75	0	14	4	3	3.25
5UPZ	662.796	89	0	19	7	3	3.70
1AJV	574.687	75	0	12	4	2	4.42
1D4H	610.696	83	0	19	5	5	3.43
1D4I	636.284	87	0	18	5	5	3.86
1EBY	652.733	88	0	19	6	6	3.44
1EBZ	633.731	89	0	21	6	6	2.65
3A2O	568.727	80	0	15	4	4	3.97
1F0S	427.500	46	0	5	5	2	1.54
1FJS	526.500	63	0	10	4	1	4.76
1MQ6	568.860	56	0	10	3	2	4.66
1NFX	505.010	54	0	7	5	1	2.14
2J34	471.978	52	0	6	4	1	2.12
2P16	459.497	59	0	5	4	1	2.24
2P95	441.931	50	0	7	3	2	4.15
2XC0	546.543	62	0	9	5	2	1.17
3MXF	456.988	56	0	5	3	0	4.85
4MR3	326.347	42	0	4	2	2	2.67
3U5L	323.780	37	0	1	3	0	4.57
4MR4	370.399	49	0	7	2	2	2.31
3U5J	308.765	35	0	1	2	0	3.23
3SVG	261.321	38	0	4	4	1	2.21
4HBV	241.085	22	0	0	1	1	1.37
4J0R	295.332	39	0	5	3	2	3.25

Table S4. Summary of free energy calculation results (kcal/mol) among HIV-1 protease, Human factor Xa, and BRD4.

Target	PDB	ΔG_{exp}	ΔG_{cal}	$\Delta G_{cal} - \Delta G_{exp}$
HIV-1 protease	2QHY	-10.2	-8.1	+2.1
	4U8W	-15.3	-15.7	-0.4
	5UPZ	-12.5	-11.2	+1.3
	1AJV	-10.5	-11.5	-0.9
	1D4H	-13.6	-13.2	+0.4
	1D4I	-12.1	-10.7	+1.3
	1EBY	-13.2	-12.2	+1.0
	1EBZ	-12.8	-13.7	-0.8
	3A2O	-12.4	-12.6	-0.3
	1G2K	-12.2	-14.3	-2.1
Human factor Xa	1F0S	-10.6	-12.3	-1.7
	1FJS	-13.6	-15.8	-2.2
	1MQ6	-15.2	-15.3	-0.1
	1NFW	-12.2	-14.7	-2.5
	1NFX	-11.6	-11.7	-0.1
	2J34	-10.7	-11.5	-0.9
	2P16	-13.8	-15.9	-2.1
	2P95	-12.8	-15.9	-3.2
	2XC0	-9.2	-7.2	+1.9
	2VVV	-11.2	-6.5	+4.7
BRD4	3MXF	-9.6	-13.9	-4.3
	4MR3	-9.0	-12.0	-3.0
	3U5L	-8.2	-11.6	-3.4
	4MR4	-7.8	-10.2	-2.4
	3U5J	-7.4	-10.7	-3.3
	3SVG	-7.3	-9.2	-1.9
	4HBV	-6.3	-6.5	-0.2
	4J0R	-8.8	-11.9	-3.1

ΔG_{exp} and ΔG_{cal} are the experimental binding free energy and calculated binding free energy, respectively.

References for SI

1. Aldeghi, M.; Heifetz, A.; Bodkin, M. J.; Knapp, S.; Biggin, P. C., Accurate calculation of the absolute free energy of binding for drug molecules. *Chemical Science* **2016,** *7* (1), 207-218.

2. Aldeghi, M.; Heifetz, A.; Bodkin, M. J.; Knapp, S.; Biggin, P. C., Predictions of ligand selectivity from absolute binding free energy calculations. *Journal of the American Chemical Society* **2017,** *139* (2), 946-957.