Free energy perturbation–based large-scale virtual screening for effective drug discovery against COVID-19

Zhe Li1,#, Chengkun Wu2,#, Yishui Li2,#, Runduo Liu1,#, Kai Lu2, Ruibo Wang2, Jie Liu2, Chunye Gong2, Canqun Yang3, Xin Wang4,5, Chang-Guo Zhan6,7 and Hai-Bin Luo1,8

Abstract
As a theoretically rigorous and accurate method, FEP-ABFE (Free Energy Perturbation-Absolute Binding Free Energy) calculations showed great potential in drug discovery, but its practical application was difficult due to high computational cost. To rapidly discover antiviral drugs targeting SARS-CoV-2 Mpro and TMPRSS2, we performed FEP-ABFE–based virtual screening for ~12,000 protein-ligand binding systems on a new generation of Tianhe supercomputer. A task management tool was specifically developed for automating the whole process involving more than 500,000 MD tasks. In further experimental validation, 50 out of 98 tested compounds showed significant inhibitory activity towards Mpro, and one representative inhibitor, dipyridamole, showed remarkable outcomes in subsequent clinical trials. This work not only demonstrates the potential of FEP-ABFE in drug discovery but also provides an excellent starting point for further development of anti-SARS-CoV-2 drugs. Besides, ~500 TB of data generated in this work will also accelerate the further development of FEP-related methods.

Keywords
Supercomputing, SARS-CoV-2, free energy perturbation, absolute binding free energy, virtual screening

Justification for prize
The goal of this work is to develop an accurate and fast FEP-ABFE-based virtual screening approach for emergency drug discovery in the outbreak of a pandemic like COVID-19 through very large-scale parallel computation.

We
- provide a scalable platform via FEP-ABFE-based large-scale virtual screening for drug discovery to combat future pandemic;
- identified 50 active Mpro inhibitors for effective drug screening;

1School of Pharmaceutical Sciences, Sun Yat-Sen University, Guangzhou, P. R. China
2State Key Laboratory of High Performance Computing, Changsha, P. R. China
3National Supercomputer Center in Tianjin, Tianjin, P. R. China
4School of Medicine and Pharmacy, Ocean University of China, Qingdao, P. R. China
5Marine Biomedical Research Institute of Qingdao, Qingdao, P. R. China
6Molecular Modeling and Biopharmaceutical Center, College of Pharmacy, University of Kentucky, Lexington, KY, USA
7Department of Pharmaceutical Sciences, College of Pharmacy, University of Kentucky, Lexington, KY, USA
8School of Pharmaceutical Sciences, Hainan University, Haikou, P. R. China

#These authors contribute equally to this work.

Corresponding author:
Zhe Li, Sun Yat-Sen University, 132 Waihuan East Rd, Higher Education Mega Center, Guangzhou 510275, China.
Email: lizhe5@mail.sysu.edu.cn
• achieved a milestone for FEP-ABFE-based virtual screening (~12,000 ligand-receptor pairs) on a new generation of Tianhe supercomputer within six days.

**Performance attributes**

| Performance Attribute | Our submission |
|-----------------------|----------------|
| Category of achievement | Time to solution |
| Type of methods used | FEP-ABFE, Molecular dynamics (MD) |
| Results reported | Whole application including I/O with validations |
| Precision reported | Single precision |
| Measurement mechanism | Timers |
| Systems scale | Measured on 75,000 nodes of the new generation Tianhe supercomputer |

**Overview of the problem**

On-going COVID-19 pandemic caused by SARS-CoV-2 virus continues to pose detrimental impacts to the human society. Despite intensive countermeasures implemented around the world, both global incidence and mortality rates remain high with many countries facing a series of new waves of infections. Even where vaccination rates are high, the SARS-CoV-2 variants like Delta (Mlochova et al. (2021)) can significantly reduce the effectiveness of the vaccines, and people who cannot receive vaccines are even more vulnerable. Under such situation, small-molecule antiviral agents have several advantages, such as clarity in antiviral mechanism, easy mass production, and broad-spectrum antiviral activity against mutant strains. Therefore, fast development of specific antiviral agents against COVID-19 is urgently needed.

Several promising drug targets with different mechanisms for the treatment of COVID-19 have been reported. One important target that has attracted most of the attentions in anti-SARS-CoV-2 drug development is Mpro (main protease, also known as 3C-like protease), which is an important enzyme for gene replication of the virus (Adedeji and Sarafianos (2014)). Another important drug target is the cellular protease, transmembrane serine protease 2 (TMPRSS2), which is closely related to the viral cell entry (Hoffmann et al. (2020)). Therefore, both Mpro and TMPRSS2 inhibitors were predicted and discovered in this study. Potentially, their inhibitors could be used in combination as a “cocktail treatment” against COVID-19.

Generally, binding free energy prediction between a receptor and its ligand is the key for structure-based drug design because it determines the binding strength/affinity of the molecule to the target. Molecular docking is the most widely used computational method in structure-based drug design, however, its empirical scoring function-based binding pose and binding free energy predictions often suffer from low accuracy, which usually leads to a low hit-rate in subsequent experimental validations.

Statistical mechanics-based methods have the potential to achieve better predictions of free energies and increase the success rate of virtual screening. Among them, free energy perturbation-absolute binding free energy prediction (FEP-ABFE) is a representative method. It employs sampling of microscopic states to predict macroscopic properties of the target system on a rigorous theoretical basis. To note, samples corresponding to rarely occurring events is vital to the accuracy of free energy calculations. A lack of comprehensive sampling can be a major source of error. Hence, these methods are usually implemented through Monte Carlo (MC) or MD simulations to obtain sufficient sampling of the complex, ligand, and intermediate states in solution, resulting in huge computational cost in practical applications.

FEP-ABFE has been proved to be an effective method in identifying active compounds against a specific target; however, its demand for computational resources is enormous, which prevents it from being widely applied in large-scale virtual screening. Besides, external restraints between a protein and its ligand should be added to appropriately estimate the free energy change, and the difficulty in automating the restraint addition process is another setback for the large-scale applications of the FEP-ABFE method. Based on the recently derived restraint energy distribution (RED) function, the FEP-ABFE calculation can be performed automatically following the double decoupling method (Li et al. (2020)) (as shown in Figure 1), with 1 preliminary equilibrate simulation and 41 λ simulations (21 for receptor–ligand complex and 20 for ligand in water) to be performed for each FEP-ABFE calculation.

Herein, during the initial outbreak of SARS-CoV-2, in order to rapidly identify potential Mpro inhibitors, FDA-approved drug library containing about 2500 drugs were virtually screened. With all the 2500 compounds docked to Mpro followed by 100 FEP-ABFE calculations, 25 compounds were selected for further experimental validation (Li et al. (2020)). Among them, 16 compounds were proved to be active Mpro inhibitors with the half maximal inhibitory concentration (IC50) values <100 μM (Table 1). The most potent one, dipyridamole (IC50 = 0.6 μM), has shown remarkable therapeutic effects in following clinical studies for treatment of patients with COVID-19 (Liu et al. (2020)).

Further, to accelerate development of broad-spectrum antiviral drugs against COVID-19, new inhibitors of Mpro and TMPRSS2 were discovered through FEP-ABFE–based large-scale virtual screening of two larger libraries of compounds performed on the new generation Tianhe supercomputer. First,
we virtually screened more than 3.6 million compounds from commercially available databases by docking methods, followed by the FEP-ABFE calculations for about 12,000 compounds to achieve FEP-based binding free energy results. Second, 73 and 66 compounds selected from the SPECS and ChemDiv libraries were subjected to experimental assays against M\textsubscript{pro} and TMPRSS2, respectively. Among them, 34 compounds were identified as active M\textsubscript{pro} inhibitors and 16 compounds were proven to be potential TMPRSS2 inhibitors under the threshold of 100 µM with the most potent one at IC\textsubscript{50} of 4.9 µM. Besides, all the 50 discovered M\textsubscript{pro} inhibitors (16 from FDA-approved drugs and 34 from commercially available libraries, hit rate >51%) and 16 TMPRSS2 inhibitors (hit rate of 24%) have different scaffolds, which was done deliberately when selecting the compounds for bioassay, in order to provide as much starting points for the further drug discovery as possible.

Figure 1. The alchemical pathways of the FEP-ABFE calculation and the curve that the free energy changes with simulation windows. A ligand with carbon atoms in green indicates that it still has Coulomb interaction and vdW (van der Waals) interaction with the surrounding environment. The gray ligand and the transparent ligand indicate the end points of the disappearance of the Coulomb interaction and the vdW interaction, respectively. The attached red spring represents the harmonic potential restraint that restrains the relative position of the ligand with respect to the target.

Table 1. The hit compounds against M\textsubscript{pro} validated by in vitro activity assays.

| Database | Number of tested compounds | >50% inhibition at 100 µM | >33% inhibition at 100 µM |
|----------|----------------------------|--------------------------|--------------------------|
| SPECS    | 38                         | 18                       | 24                       |
| ChemDiv  | 35                         | 16                       | 19                       |
| FDA      | 25                         | 16                       | 20                       |
| Total    | 98                         | 50 (51%)                 | 63 (64%)                 |
Current state of the art

Binding between a protein and a ligand directly determines the pharmacological effects of a drug or drug candidate. Thus, how to efficiently and accurately predict the binding affinities between drugs and their targets, which is challenging due to the complexity of the system, is considered as one of the most important issues in rational drug design as well as in computational chemistry. There are various methods in predicting the binding affinity, including the very fast and approximate empirical scoring functions (e.g., docking), the relatively more sophisticated methods with moderate accuracy and performance (e.g., Molecular Mechanics/Poisson-Boltzmann/surface area, MM-PBSA; Molecular Mechanics/Poisson-Boltzmann surface area, MM-GBSA (Wang et al. (2019))), and the theoretically rigorous and accurate but very computational expensive free energy methods (e.g., FEP) (Pohorille et al. (2010)). Applications of these methods will be discussed in this section.

Docking approaches for virtual screening

Molecular docking programs predict binding poses, and binding affinities between small molecules and targets could be predicted based on scoring functions. For the prediction of binding poses, various docking methods have achieved considerable results in different target protein systems (Kontoyianni et al. (2004); Oum et al. (2020); Zhang et al. (2021)). However, in terms of binding affinity prediction, the empirical scoring functions are very approximate and cannot reliably distinguish active molecules. Despite scoring function-based docking methods are not accurate enough, their fast calculation speed enables them to screen billions of compounds within an acceptable period of time. In 2020, Gorgulla et al. developed a docking-based high-throughput virtual screening platform VirtualFlow, which was used to screen more than 1.4 billion small molecules (Gorgulla et al. (2020)). In that study, 590 compounds were tested by experimental validations and 10 potential compounds (IC50 <60 μM) were identified (hit rate <2%). In view of their low hit rate and accuracy, the efficiency of scoring function-based virtual screening still needs to be improved.

Methods with moderate speed and accuracy

Molecular docking can only predict binding affinities based on static binding conformations between biomolecules and small molecules. With MD simulations, calculating energy average along the MD trajectories can bring us more information and improve the prediction accuracy. The most commonly used methods include MM-GBSA or MM-PBSA (Wang et al. (2019); Xu et al. (2013); Genheden and Ryde (2015); Wang et al. (2016)). Kuhn (Kuhn et al. (2005)) et al. used the MM-PBSA method to calculate the binding affinities between eight protein systems and their corresponding small-molecule ligands, which showed that the MM-PBSA method is more robust than the traditional scoring function-based docking approaches. However, these methods are dependent on many empirical parameters, and the accuracy is still not comparable to that of the more theoretically rigorous methods.

Theoretically rigorous and accurate FEP-ABFE methods

In the past five years, the FEP-ABFE methods began to show their potential in drug design and discovery research. Aldeghi (Aldeghi et al. (2016)) et al. used FEP-ABFE to identify 11 drug-like molecules with diverse scaffold binding to BRD4 in 2016 (with a root mean square error of 0.6 kcal/mol), and the optimal binding conformation was also successfully predicted for different ligands by the FEP-ABFE method. Our group (Li et al. (2019)) recently added a post-processing step based on Gaussian function fitting to the FEP-ABFE calculation, and successfully optimized the chemical structure of a series of PDE10 inhibitors, leading to a subnanomolar inhibitor with a nearly 2000-fold increase in activity compared to the starting compound. However, to date, there has been no study that has increased the scale of FEP-ABFE calculations to a level of several hundred small molecules or more, which was restricted due to the massive demand of computing power. Therefore, we need to implement the feasibility of applying efficient FEP-ABFE calculations towards large-scale virtual screening empowered by modern supercomputing infrastructures.

Innovations realized

Computational challenges

Our implementation of FEP-ABFE is based on the GROMACS package (Prönk et al. (2013); Páll et al. (2014); Abraham et al. (2015)). A FEP-ABFE calculation of a ligand and a target protein involves 1 equilibrium simulations and 41 value simulations (21 for receptor–ligand complex and 20 for ligand in water). On average, calculations at each value could take 24 h on a typical Intel Xeon CPU. Herein, we aim to complete a total number of 0.5 million MD tasks within an acceptable range of time, preferably less than a week, to respond quickly to the need of repurposing an approved drug for emergency use or for further development.

The number of jobs (about 0.5 million) would be too large for the commonly used job management systems (such as slurm or PBS) to efficiently control. To finish all the jobs within one week on 75,000 nodes in the new generation Tianhe system, how to correctly and efficiently submit, queue, schedule, control, and monitor all the 500,000 jobs is
very important and is also a big challenge. A new customized job management system needs to be developed for those ultra-large-scale calculations.

**Workflow**

In this work, based on the binding conformations generated by docking program Glide (Friesner et al. (2004); Halgren et al. (2004)) in Schrödinger (2015), we implemented an integral workflow of the FEP-ABFE-based large-scale virtual screening on the new generation of Tianhe system (as shown in Figure 2). First, propka (Olsson et al. (2011); Søndergaard et al. (2011)) program was used to calculate the protonation state of all the residues in the protein. Then, we used an in-house script—AutoMD (Li et al. (2013)) to automatically generate the topology and coordinate files required for each system simulation. The detailed steps involved are as follows: The partial atom charges for the small molecules were generated by Austin Model 1-Bond Charge Correction (AM1-BCC) (Jakalian et al. (2000); Jakalian et al. (2002)). The restrained electrostatic potential (RESP) each small molecule was generated by the Antechamber (Wang et al. (2006)) module of AmberTools 20. Then we used generalized Amber force field (GAFF) (Wang et al. (2004)) and Amber ff14SB force field (Maier et al. (2015)) to parameterize the small molecules and proteins, respectively. After that, each complex was solved in a dodecahedron TIP3P (Jorgensen et al. (1983)) water box with a minimum distance of 12 Å between the solute and the box boundary. Counter ions (Na+ or Cl−) were added to the system to keep the electric neutrality of the system. We used Acype (Da Silva and Vranken (2012)) to generate the GROMACS topologies and coordinates.

In the subsequent simulation steps, before performing FEP simulation of the complex, we needed to perform a preliminary simulation of the complex, and used our own in-house program to generate six degrees of freedom restraints for keeping the relative position between ligands and the receptor according to this preliminary trajectory, which are required in FEP-ABFE calculations. Then, we carried out 21 and 20 FEP simulation windows on the decoupling process of complex and ligand, respectively. After the FEP simulation, each window generates a corresponding XVG file for subsequent free energy calculation, recording the internal energy information of the conformation in the simulation process under different force fields.

Finally, in the post-processing step, we used the energy extraction module in the alchemlyb (Klimovich et al. (2015)) library to extract the energy information in the XVG file of each window and used our in-house free energy analysis code implemented based on the bar calculation module of the pymbar (Shirts and Chodera (2008)) library to calculate ABFE values for each ligand–receptor pair.

Given required input data and configuration parameters like the number of compute nodes and processes for each sub-task, the workflow can be easily initiated with a single hit and automatically perform FEP-ABFE-based virtual screening towards different compound databases.

**High-throughput optimization of large-scale FEP-ABFE calculations on the Tianhe system**

This work involves over 500,000 FEP-ABFE related MD jobs. Each MD job requires a significant number of
computing resources to finish within a reasonable amount of time, which can be executed \textit{via} parallel processes with multi-threads. We designed a hybrid high-throughput computing (HTC) task management system in order to optimize the overall time to solution of all FEP-ABFE calculations, as shown in Figure 3. The systems perform dynamic job allocation by continuously fetching idle nodes from a constantly updated list of available compute nodes to perform unfinished jobs. Each job is characterized with its resource requirements. Generally speaking, each job can run on multiple nodes. However, the number of jobs goes far beyond the number of available compute nodes. Therefore, in this large-scale application, each job runs on 16 threads on a single node. If a job/node failure was detected, then the job would be allocated another idle node. A job would be terminated and removed if it could not be finished after 3 trials, where the problem is most likely to be caused by the physics model rather than a hardware problem. Each node is released to the idle list once the assigned job is completed successfully, which allows execution of remaining jobs. If the number of idle nodes exceeds the requirement of remaining jobs, extra nodes will be released. The dynamically maintained queues and lists of jobs and nodes form an execution loop until all jobs are processed.

When the management system checked that all simulation tasks were completed, the system should check whether each ligand sample includes at least 35 completed 4-ns production MD simulations. If a ligand sample meets the above-mentioned requirement, a free energy calculation script would be executed immediately.

\textbf{Computational results were validated by in vitro activity experiments}

Considering the initial goal of drug discovery and repurposing, experimental validations were performed after the FEP-ABFE-based virtual screening. As a result, among FDA-approved drugs, 16 active M\text{pro} inhibitors (Figure 4) were identified with IC\textsubscript{50} < 100 μM by in vitro activity assays (Li et al. (2020)). In the subsequent proof-of-concept clinical trials, we studied the therapeutic effects of the most potent M\text{pro} inhibitor dipyridamole in COVID-19 patients. The results revealed that dipyridamole significantly increased the survival rate of severely ill patients with COVID-19 (Liu et al. (2020)). For the further much larger-scale of FEP-ABFE based virtual screening (with about 12,000 FEP-ABFE calculations performed) targeting M\text{pro} and TMPRSS2, experimental validations also showed encouraging outcomes. Currently, most of the known M\text{pro} and TMPRSS2 inhibitors are covalent inhibitors. From the commercially available databases, 34 M\text{pro} non-covalent inhibitors (under the threshold of IC\textsubscript{50} < 100 μM) with novel scaffolds were discovered, which provided excellent starting points for further development of antiviral M\text{pro} non-covalent inhibitors.

Overall, among all the 98 compounds (25 from FDA-approved drugs and 73 from SPECS and ChemDiv) subjected to experimental tests against M\text{pro}, 50 showed significant inhibitory activity with IC\textsubscript{50} values better than 100 μM (hit rate was 51%). Using a looser threshold, 63 compounds exhibited more than 33% inhibition at concentration 100 μM and the relevant hit rate was 64% (details are given in Table 1), indicating most of the virtual screened compounds can bind with the target. Subsequent

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{fig3.png}
\caption{A hybrid HTC (high-throughput computing) task management system.}
\end{figure}
structure optimizations for the M\textsuperscript{pro} inhibitors as well as the experimental validations for non-covalent TMPRSS2 inhibitors are also being performed.

Similarly, 16 out of 66 selected compounds showed more than 50% inhibition of TMPRSS2 at a concentration of 100 \( \mu \text{M} \) (24% hit rate) after validation of \textit{in vitro} activity assays (Table 2). The screening success rate for TMPRSS2 inhibitors was lower than that of M\textsuperscript{pro}. The reason could be that the crystal structure of the TMPRSS2 protein contained a covalent inhibitor, which was not suitable for screening non-covalent inhibitors due to the conformation difference. Once the initial binding conformation cannot be accurately predicted by the molecular docking program, the accuracy of the FEP calculation can be significantly influenced.

### Performance measurement

#### Systems used

The large-scale virtual screening tasks of two targets were performed on the new generation Tianhe supercomputer. 16 CPU cores and 64 GB main memory are used on each compute node. The total number of sub-tasks involved was about 500,000. The execution time of different sub-tasks varies from a few hours to two days. Therefore, each compute node was allocated exclusively to one single sub-task.

Initial structures and input files were prepared following the steps described in section 5.2. The half-million-scale simulation tasks were assigned to 75,000 compute nodes. The final part of this work was free energy calculation and trajectory analysis. We used 1000 nodes for free energy calculation via \texttt{pymbar}. Libraries and applications used are provided in Table 3.

#### Application benchmarked

For M\textsuperscript{pro}, FEP-ABFE were performed for a total of 6270 small compounds (100 FDA-approved drugs, 3143 compounds in ChemDiv, and 3027 compounds in SPECS). For TMPRSS2, 5829 small compounds (3004 compounds in ChemDiv and 2825 compounds in SPECS) were subjected to FEP-ABFE calculations.

For each compound, we set 20 \( \lambda \) values for ligand system and 21 \( \lambda \) values for complex system. Each \( \lambda \) value corresponds to an independent task and all these tasks can be performed in parallel. We set up a single process with

![Figure 4. (a). Representative M\textsuperscript{pro} inhibitors obtained via FEP-ABFE-based virtual screening against the FDA-approved drug database. (b). The predicted binding modes of dipyridamole-M\textsuperscript{pro} (left) and candesartan-M\textsuperscript{pro} (right). The yellow and cyan sticks represent the ligands and key residues, respectively. The grey cartoons represent the protein structures.](image-url)

### Table 2. The hit compounds against TMPRSS2 validated by \textit{in vitro} activity assays.

| Database | Number of tested compounds | >50% inhibition at 100 \( \mu \text{M} \) | >33% inhibition at 100 \( \mu \text{M} \) |
|----------|-----------------------------|-----------------------------|-----------------------------|
| SPECS    | 35                          | 9                           | 24                          |
| ChemDiv  | 31                          | 7                           | 20                          |
| Total    | 66                          | 16 (24%)                    | 44 (67%)                    |

### Table 3. The information of the applications and the respective libraries used in the FEP-ABFE calculations.

| Software | Version |
|----------|---------|
| Gromacs  | 2019.5  |
| GCC      | 9.3.0   |
| Pymbar   | 3.0.5   |
| Python   | 3.7.10  |


\[\text{Li et al.} 51\]

Li et al.
16 threads for each task. A series of restraint parameters are needed when performing complex-system simulations, which can be extracted from a short pre-equilibrate trajectory of the complex system. Thus, a pre-equilibrate task with `free energy` and `calc_lambda_state` options off should be finished before a complex-system simulation.

**Large-scale jobs run on the Tianhe supercomputer system**

After finished the preparation work, we carried out simulations on the new generation Tianhe system. For both Mpro and TMPRSS2, a task list consisted of all ligand–receptor pairs’ tasks that each pair corresponded to 20 ligand tasks, one pre-equilibrate task and 21 complex tasks was generated. Meanwhile, a usable node partition was obtained, and the node map was applied to unsubmitted tasks. The pre-equilibrate tasks were mapped to all idle nodes firstly, followed by ligand system. Complex-system tasks were performed after all the pre-equilibrate tasks finished.

There were four simulation stages in every single ligand and complex task. First, using the steepest descent to minimize energy with 5000 steps and then heated system in an NVT ensemble for 5000 steps. The system was subsequently equilibrated in an NPT ensemble for 5000 steps and followed by a 4-ns with 2,000,000 steps production MD simulation. The time interval, dt, was set as 0.002 ps. As for the pre-equilibrate task, we set energy minimization, NVT ensemble, and NPT ensemble simulations to 10,000 steps, 10,000 steps, and 25,000 steps, respectively. We used our high-throughput screening management system implemented to administrate these simulation tasks. All tasks were listed in order of the execution sequence in a file in order of the execution sequence, and a file containing a list of idle nodes was generated as well. The management system is in charge of mapping tasks and nodes, monitoring running jobs and nodes, and submitting queued tasks when nodes become available. We carried out the final large-scale run on the new generation Tianhe supercomputing system.

**Post-processing**

The free energy between each pair of ligand and complex was calculated by pymbar (Shirts and Chodera (2008)), a python package that implements BAR and MBAR algorithms. Once the management system has checked and found that a ligand–receptor pair has all the 41 $\lambda$ tasks finished (either finished normally or quitted abnormally), among which at least 35 $\lambda$ tasks completed 4-ns production MD simulations, free energy calculation program would be ran automatically.

**Performance results**

**Time to solution for the FEP-ABFE pipeline**

To perform all the simulations with high efficiency, we conducted several performance tests on the same ligand to find the most suitable number of processes and threads. The results are shown in Table 4.

The performance using 1 process (3.5 ns/d) and 4 processes (0.68 ns/d) within a single node indicate that the increasing number of processes brings more computational cost. Although the test with 3 nodes had less computation time, with the same computing resources, computing a task with a single process (16 threads) in 1 node performs more efficiently. Taking into account both the computation efficiency and manageability of the tasks, we ultimately chose the first setting, only use one node, one process, and 16 threads for a single task.

Pre-equilibrate tasks must be finished before the $\lambda$ simulations of the complex system can be started. Therefore, in the beginning, we used 12,000 compute nodes to run pre-equilibrate tasks and remaining 63,000 nodes to run ligand tasks simultaneously (Figures 5 and 6). Following this strategy, after 27.4 h, 0.25 million tasks can be completed and complex tasks would be submitted. 75,000 compute nodes were applied to perform the remaining 0.25 million complex tasks for 114.5 h.

**Table 4. MD simulation performance with different number of processes and threads.**

| #Node | #Processes | Threads per process | Performance |
|-------|------------|---------------------|-------------|
| 1     | 1          | 16                  | 3.5 ns/d    |
| 1     | 4          | 4                   | 0.68 ns/d   |
| 3     | 9          | 5                   | 7.6 ns/d    |

![Figure 5. Number of nodes and the runtime of each component in MD simulations.](image)
**Free energy calculations from trajectories generated via MD simulations**

We obtained .xvg files as the result of each simulation, and with the aid of pymbar, a python library used for extracting and analyzing energy data, we calculated binding affinity between each ligand and the respective receptor. In this step, 1000 compute nodes were provided for these calculation tasks. When all simulation tasks for a complex/ligand system finished (either completed normally or failed abnormally), the free energy calculation task would be carried out if at least 35 tasks have completed the whole 4 ns simulation. All about 12,000 calculation tasks were completed in 28 min.

**Comparison of different targets and compound databases**

Mpro systems have about 22,000 atoms, about 7000 more than that of TMPRSS2. As a result, the simulation time of Mpro lasted significantly longer than that of TMPRSS2 (Table 5). Figure 6 illustrates the aggregate time and proportion of the different parts of the respective compound library. Pre-equilibrate and complex simulation take at least 91.3% of the total time. In view of that the number of available nodes is much more than the number of pre-equilibrate tasks, we conduct ligand tasks at the same time to overlap ligand computation cost.

In conclusion, we completed FEP simulations of about 12,000 ligands with Mpro and TMPRSS2 proteins within six days. After experimental validation, 50 hits among 98 selected compounds exhibited significant inhibitory potency towards Mpro, and one drug dipyridamole showed remarkable outcomes in subsequent clinical trials. Due to the limitations of existing docking techniques for binding conformation predictions and the fact that inaccurate initial complex structures can lead to non-convergent FEP calculations, the other 48 selected compounds had false-positive predictions. The whole FEP-ABFE calculations consist of totally about 2.1 ms MD simulations, involving a total of more than 500,000 MD simulation tasks. On average, we achieved 350 μs MD runs per day with 75,000 computing nodes on the new generation Tianhe supercomputer system.

**Implications**

In recent decades, with the continued development of high-performance computing, computational methods in drug design are experiencing a renaissance. FEP is recognized as a reliable method for binding free energy calculations with satisfactory accuracy, which is vital for the virtual screening to identify new hits or leads. However, FEP-ABFE calculations are extremely expensive and time-consuming. Previously, it was not feasible to predict ABFE for each ligand binding with the target at a large-scale without the support of massively parallel supercomputers. Therefore,

---

**Table 5. The number of jobs and average MD simulation time of drug libraries.**

| Target  | Ligand DB | Pre-equilibrate | | Ligand | Complex |
|---------|-----------|----------------|---|--------|--------|
|         |           | #Jobs | Time   | #Jobs | Time   | #Jobs | Time   |
| Mpro    | FDA       | 100   | 27.2 h | 2000 | 4.7 h  | 2100 | 29.7 h |
|         | ChemDiv   | 3143  | 27.1 h | 62,860| 4.8 h  | 66,003| 30.3 h |
|         | SPECS     | 3027  | 20.7 h | 60,540| 4.3 h  | 63,567| 30.6 h |
| TMPRSS2 | ChemDiv   | 3004  | 20.7 h | 60,060| 4.1 h  | 63,084| 22.8 h |
|         | SPECS     | 2825  | 20.6 h | 56,500| 4.0 h  | 59,325| 21.5 h |
| Total   |           | 12,099|        | 241,960|       | 254,079|       |
|         |           | 508,138|      |
there was no report about the application of FEP-ABFE for virtual screening.

To fight against COVID-19 and other potential pandemic in the future, we combined the frontiers of drug discovery and computational chemistry with high-performance computing, and several achievements were made.

**Large-scale FEP-ABFE calculations with up to several milliseconds of MD simulations**

In this work, after molecular docking-based prescreening of more than 3.6 million compounds, about 12,000 FEP-ABFE calculations were performed to achieve a more accurate prediction of the drug-target binding affinities. For each FEP-ABFE calculation, 42 MD simulations jobs with a simulation time of 4 ns were performed to get the free energy estimation, resulting in a total of more than 500,000 MD trajectories and a cumulative simulation time over 2 milliseconds.

**First application of the FEP-ABFE–based large-scale virtual screening for drug discovery**

There are various methods in predicting the binding affinity between targets and drugs. Due to the computational cost, the fast but approximate empirical scoring functions implemented in different docking methods are most widely used. But hit rates are limited in finding active compounds by using such methods. If more computational resources are available, there are also reports using more complicated methods with improved accuracies such as MM-PBSA (Wang et al. (2019)), MM-GBSA (Genheden and Ryde (2015)), and LIE (Åqvist et al. (1994); Hansson et al. (1998)), but there are also a lot of approximations introduced in these methods (e.g., the entropy changes are omitted and use the implicit solvent model for predicting solvation energy). FEP-ABFE is theoretically rigorous and accurate (Aldeghi et al. (2016); Aldeghi et al. (2017); Boyce et al. (2009); Li et al. (2019); Li et al. (2020); Pohorille et al. (2010)), which is considered to be one of the most promising next-generation drug design methods. However, FEP-ABFE is also extremely expensive and time-consuming, which refrains it from being applied to large-scale virtual screening in previous studies. In this work, with the new generation Tianhe supercomputing system, we applied the FEP-ABFE–based large-scale virtual screening for the first time, which shows encouraging outcomes and suggests that the FEP-ABFE–based virtual screening is a promising approach to perform hit identification and drug repurposing.

**Experimental validations and clinical studies show the effectiveness of supercomputing in drug discovery**

Our goal was to find $M^{\text{pro}}$ and TMPRSS2 inhibitors with the assistance of the new generation Tianhe system. Thus, experimental validations are of vital importance, and compounds selected from virtual screening were validated by experiment tests in this work. For FDA-approved drug database, among the 25 compounds selected by FEP-ABFE–based virtual screening, 16 were proved to be active $M^{\text{pro}}$ inhibitors with IC$_{50}$ <100 $\mu$M (Li et al. (2020)). The most potent one is dipyridamole (IC$_{50}$ = 0.6 $\mu$M), which showed remarkable therapeutic effects in clinical studies for treatment of patients with COVID-19 (Liu et al. (2020)). From the commercially available databases SPECS and ChemDiv, 73 and 66 compounds were subjected to the experimental validations against $M^{\text{pro}}$ and TMPRSS2, respectively. As a result, under the threshold of IC$_{50}$ lower than 100 $\mu$M, 34 and 16 active inhibitors were identified for $M^{\text{pro}}$ (hit rate of 51%) and TMPRSS2 (hit rate of 26%), respectively.

More experiments are being performed for both $M^{\text{pro}}$ and TMPRSS2, which will hopefully promote the further development of anti-SARS-CoV-2 drugs.

**A complete computational platform for emergency drug discovery against further possible pandemic**

In this work, the FEP-ABFE-based large-scale virtual screening was proved to be a promising approach in hit discovery and drug repurposing. With the tools and innovations implemented, this method can be easily applied when needed, for example, the possible outbreak of pandemic caused by other viruses in the future. The ultra-high-throughput task management can also be applied to other complicated simulations with up to millions of independent simulation tasks.

**A valuable repository of MD simulation data for further development of machine learning based FEP-ABFE method**

For the statistical mechanism-based methods like FEP, sufficient MD/MC sampling is essential for the convergence of the computational data, which will consume a huge amount of computer resources. To make FEP a widely used method in drug discovery, reducing the computer resources for the FEP calculations is a key factor. Machine learning is an effective tool that can be used to learn and predict the convergence behavior and thus vastly accelerate the calculations. The 500,000 MD simulations results can be a precious training set for learning an AI model for FEP-
ABFE, which will also promote the development of the FEP method.

Conclusion

Rapid antiviral drug discovery plays important role in fighting against the outbreak of pandemic. To screen emergency antiviral agents against COVID-19, we implemented FEP-based virtual screening for ∼12,000 ligands on the Tianhe system. 500,000 simulation tasks were performed on 75,000 compute nodes within 142h. For Mpro and TMPRSS2, 63 and 44 active inhibitors were discovered, respectively, with the inhibition rate of more than 33% at 100 μM. Meanwhile, about 500 TB of data was obtained, which could be used to enhance molecular simulations by machine learning in further studies. Based on the results and performance of the FEP-ABFE method, the relating algorithms can further be optimized and new alchemical transformation pathways can be designed to accelerate the virtual screening process.

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: We cordially acknowledge National Key R&D Program of China (2017YFB0202600), National Natural Science Foundation of China (81903542, 21877134, 22077143, and U1811462), Fundamental Research Funds for Hainan University (KYQD (ZR)-21,031), Science Foundation of Guangzhou City (202102021151, 201904020023), Guangdong Province Higher Vocational Colleges & Schools Pearl River Scholar Funded Scheme (2016), the National Science Foundation (NSF, grant CHE-1111761), the Taishan Scholars Program (tsqn201909170), the Innovative Leader of Qingdao Program (19-3-2–26-zhc), the special scientific research fund for COVID-19 from the Pilot National Laboratory for Marine Science and Technology (QNLM202001), and open fund from the State Key Laboratory of High Performance Computing (201901–11).

ORCID iDs

Zhe Li https://orcid.org/0000-0002-9357-6839
Chengkun Wu https://orcid.org/0000-0002-9688-5311
Runduo Liu https://orcid.org/0000-0002-8446-0404

References

Abraham MJ, Murtola T, Schulz R, et al. (2015) Gromacs: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. SoftwareX 1: 19–25.

Adedeji AO and Sarafianos SG (2014) Antiviral drugs specific for coronaviruses in preclinical development. Current Opinion in Virology 8: 45–53.

Aldeghi M, Heifetz A, Bodkin MJ, et al. (2016) Accurate calculation of the absolute free energy of binding for drug molecules. Chemical Science 7(1): 207–218.

Aldeghi M, Heifetz A, Bodkin MJ, et al. (2017) Predictions of ligand selectivity from absolute binding free energy calculations. Journal of the American Chemical Society 139(2): 946–957.

Åqvist J, Medina C and Samuelsson JE (1994) A new method for predicting binding affinity in computer-aided drug design. Protein Engineering, Design and Selection 7(3): 385–391.

Boyce SE, Mobley DL, Rocklin GJ, et al. (2009) Predicting ligand binding affinity with alchemical free energy methods in a polar model binding site. Journal of Molecular Biology 394(4): 747–763.

Da Silva AWS and Vranken WF (2012) Acyppe-antechamber python parser interface. BMC Research Notes 5(1): 1–8.

Friesner RA, Banks JL, Murphy RB, et al. (2004) Glide: a new approach for rapid, accurate docking and scoring. 1. method and assessment of docking accuracy. Journal of Medicinal Chemistry 47(7): 1739–1749.

Genheden S and Ryde U (2015) The mm/pbsa and mm/gbsa methods to estimate ligand-binding affinities. Expert Opinion on Drug Discovery 10(5): 449–461.

Gorgulla C, Boeszoemenyi A, Wang ZF, et al. (2020) An open-source drug discovery platform enables ultra-large virtual screens. Nature 580(7805): 663–668.

Halgren TA, Murphy RB, Friesner RA, et al. (2004) Glide: a new approach for rapid, accurate docking and scoring. 2. enrichment factors in database screening. Journal of Medicinal Chemistry 47(7): 1750–1759.

Hansson T, Marelius J and Åqvist J (1998) Ligand binding affinity prediction by linear interaction energy methods. Journal of Computer-Aided Molecular Design 12(1): 27–35.

Hoffmann M, Kleine-Weber H, Schroeder S, et al. (2020) Sars-cov-2 cell entry depends on ace2 and tmprss2 and is blocked of high-quality atomic charges. am1-bcc model: I. parametrization and validation. Journal of Computational Chemistry 23(16): 1623–1641.

Jorgensen WL, Chandrasekhar J, Madura JD, et al. (1983) Comparison of simple potential functions for simulating liquid water. The Journal of Chemical Physics 79(2): 926–935.

Klimovich PV, Shirts MR and Mobley DL (2015) Guidelines for the analysis of free energy calculations. Journal of Computer-Aided Molecular Design 29(5): 397–411.
Kontoyianni M, McClellan LM and Sokol GS (2004) Evaluation of docking performance: comparative data on docking algorithms. *Journal of Medicinal Chemistry* 47(3): 558–565.

Kuhn B, Gerber P, Schulz-Gasch T, et al. (2005) Validation and use of the mm-pbsa approach for drug discovery. *Journal of Medicinal Chemistry* 48(12): 4040–4048.

Li Z, Cai YH, Cheng YK, et al. (2013) Identification of novel phosphodiesterase-4d inhibitors prescreened by molecular dynamics-augmented modeling and validated by bioassay. *Journal of Chemical Information and Modeling* 53(4): 972–981.

Li Z, Huang Y, Wu Y, et al. (2019) Absolute binding free energy calculation and design of a subnanomolar inhibitor of phosphodiesterase-10. *Journal of Medicinal Chemistry* 62(4): 2099–2111.

Li Z, Li X, Huang YY, et al. (2020) Identify potent sars-cov-2 main protease inhibitors via accelerated free energy perturbation-based virtual screening of existing drugs. *Proceedings of the National Academy of Sciences* 117(44): 27381–27387.

Liu X, Li Z, Liu S, et al. (2020) Potential therapeutic effects of dipyrindamole in the severely ill patients with Covid-19. *Acta Pharmaceutica Sinica B* 10(7): 1205–1215.

Maier JA, Martinez C, Kasavajhala K, et al. (2015) ff14sb: improving the accuracy of protein side chain and backbone parameters from ff99sb. *Journal of Chemical Theory and Computation* 11(8): 3696–3713.

Mlechina P, Kemp S, Dhar MS, et al. (2021) Sars-cov-2 b. 1.617. 2 delta variant replication and immune evasion. *Nature* 599: 1–8.

Olsson MH, Søndergaard CR, Rostkowski M, et al. (2011) Propka3: consistent treatment of internal and surface residues in empirical p k a predictions. *Journal of Chemical Theory and Computation* 7(2): 525–537.

Oum YH, Kell SA, Yoon Y, et al. (2020) Discovery of novel aminopiperidinyl amide exer4 modulators through virtual screening and rational drug design. *European Journal of Medicinal Chemistry* 201: 112479.

Pall S, Abraham MJ, Kutzner C, et al. (2014) Tackling exascale software challenges in molecular dynamics simulations with gromacs. In: *International conference on exascale applications and software*. Springer, pp. 3–27.

Pohorille A, Jarzynski C and Chipot C (2010) Good practices in free-energy calculations. *Journal of Palliative Care* 114(32): 10235–10253.

PrOnek S, Páll S, Schulz R, et al. (2013) Gromacs 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. *Bioinformatics* 29(7): 845–854.

Shirts MR and Chodera JD (2008) Statistically optimal analysis of samples from multiple equilibrium states. *The Journal of Chemical Physics* 129(12): 124105.

Søndergaard CR, Olsson MH, Rostkowski M, et al. (2011) Improved treatment of ligands and coupling effects in empirical calculation and rationalization of p k a values. *Journal of Chemical Theory and Computation* 7(7): 2284–2295.

Wang C, Nguyen PH, Pham K, et al. (2016) Calculating protein–ligand binding affinities with mmpbsa: Method and error analysis. *Journal of Computational Chemistry* 37(27): 2436–2446.

Wang E, Sun H, Wang J, et al. (2019) End-point binding free energy calculation with mm/pbsa and mm/gbsa: strategies and applications in drug design. *Chemical Reviews* 119(16): 9478–9508.

Wang J, Wang W, Kollman PA, et al. (2006) Automatic atom type and bond type perception in molecular mechanical calculations. *Journal of Molecular Graphics and Modelling* 25(2): 247–260.

Wang J, Wolf RM, Caldwell JW, et al. (2004) Development and testing of a general amber force field. *Journal of Computational Chemistry* 25(9): 1157–1174.

Xu L, Sun H, Li Y, et al. (2013) Assessing the performance of mm/pbsa and mm/gbsa methods. 3. the impact of force fields and ligand charge models. *Journal of Palliative Care* 117(28): 8408–8421.

Zhang Y, Xu L, Zhang Y, et al. (2021) Discovery of novel mif inhibitors that attenuate microglial inflammatory activation by structures-based virtual screening and in vitro bioassays. *Acta Pharmacologica Sinica* 43: 1–13.

**Author biographies**

Zhe Li is currently an associate professor in the School of Pharmaceutical Sciences at Sun Yat-Sen University (SYSU). He received his PhD degree in medicinal chemistry from SYSU in 2017. His research in medicinal chemistry focuses on the development of drug design methods and the applications of these methods in the discovery of novel inhibitors against various kinds of targets.

Cheng-Kun Wu is currently an associate professor in State Key Laboratory of High-Performance Computing, College of Computer, National University of Defense Technology. His research focuses on High-Performance Computing, Computational Systems Biology, and Biomedical Data Mining.

Yishui Li is currently a PhD student in the College of Computer, National University of Defense Technology, China. Her research focus on Biological Information and High-Performance Computing.

Runduo Liu is currently pursuing his master’s degree in the School of Pharmaceutical Sciences at Sun Yat-Sen University. His research focuses on the development of free energy calculation methods and the applications of these methods in drug design.

Kai Lu is currently the Dean of the College of Computer, National University of Defense Technology. His research focuses on Operating System, Computer Architecture, High-Performance Computing, and Quantum computing.
Ruibo Wang received his BS and PhD degrees in computer science from the National University of Defense Technology in 2003 and 2011, respectively. His research interests include high-performance computing and operating system.

Jie Liu is currently a professor in Science and Technology on Parallel and Distributed Processing Laboratory, College of Computer, National University of Defense Technology. His researches focus on Computer Architecture, Parallel and Distributed Computing, Machine Learning and Scientific and Engineering Computing.

Chunye Gong, PhD, associate professor in Science and Technology on Parallel and Distributed Processing Laboratory, College of Computer, National University of Defense Technology. His researches focus on Parallel and Distributed Computing and Scientific and Engineering Computing.

Canqun Yang is currently the Director of the National SuperComputer Center in Tianjin. His researches focus on Computer Architecture, High-Performance Computing, and High Precision Numerical Simulation.

Xin Wang Dr Wang mainly studies the interaction between viral infection and the host, and develops antiviral drugs. Using molecular biology and virology, his team focuses on the mining and verification of new drug targets, the regulation function of sugar substances on immunity, and the feasibility of non-coding RNA as a drug target.

Chang-Guo Zhan Dr Zhan’s main research interest is drug design, discovery, and development through integrated computational-experimental studies. Drugs designed and discovered in Dr Zhan’s lab are either small molecules (such as inhibitors of enzymes or agonists/antagonists of receptor proteins or DNA-binding molecules etc.) or engineered proteins (mutants with significantly improved biological functions and/or increased circulation time in the body).

Hai-Bin Luo is currently the Dean of the School of Pharmaceutical Sciences, Hainan University. He received his Bachelor’s degree and PhD degree from Xiamen University in 1999 and from Hong Kong Baptist University in 2005, respectively. At the end of 2006, he joined Sun Yat-Sen University as an Associate Professor and was promoted to a Full Professor in 2011. In 2021, he joined Hainan University. His research currently encompasses the rational design and discovery of novel phosphodiesterase inhibitors for the treatment of several diseases, such as COVID-19, pulmonary fibrosis, and pulmonary hypertension.