Article

Absolute binding free energy calculation and design of a subnanomolar inhibitor of phosphodiesterase-10

Zhe Li, Yiyou Huang, Yinuo Wu, Jing-Yi Chen, Deyan Wu, Chang-Guo Zhan, and Hai-Bin Luo

J. Med. Chem., **Just Accepted Manuscript** • DOI: 10.1021/acs.jmedchem.8b01763 • Publication Date (Web): 28 Jan 2019

Absolute binding free energy calculation and design of a subnanomolar inhibitor of phosphodiesterase-10

Zhe Li[a, b, #], Yiyou Huang[a, #], Yinuo Wu[a], Jingyi Chen[a], Deyan Wu[a], Chang-Guo Zhan[b,*], and Hai-Bin Luo[a,*]

[a] School of Pharmaceutical Sciences, Sun Yat-Sen University, Guangzhou 510006, P.R. China

[b] Department of Pharmaceutical Sciences, College of Pharmacy, University of Kentucky, 789 South Limestone Street, Lexington, KY, 40536

Abstract

Accurate prediction of absolute protein-ligand binding free energy could considerably enhance the success rate of structure-based drug design, but is extremely challenging and time-consuming. Free energy perturbation (FEP) has been proven reliable, but limited to prediction of relative binding free energies of similar ligands (with only minor structural differences) in binding with a same drug target in practical drug design applications. Herein, a Gaussian algorithm enhanced FEP (GA-FEP) protocol has been developed to enhance the FEP simulation performance, enabling to efficiently carry out the FEP simulations on vanishing the whole ligand and, thus, predict the absolute binding free energies (ABFE). Using the GA-FEP protocol, the FEP simulations for the ABFE calculation (denoted as GA-FEP/ABFE) can achieve a satisfactory accuracy for both structurally similar and diverse ligands in a dataset of more than 100 receptor-ligand systems. Further, our GA-FEP/ABFE-guided lead optimization against phosphodiesterase-10 led to discovery of a subnanomolar inhibitor (IC_{50}=0.87 nM, ~2,000-fold improvement in potency) with cocrystal confirmation.

Introduction

Receptor-ligand binding energy prediction is a primary objective of structure-based drug design, and there have been numerous efforts made to improve the accuracy of binding free energy prediction.[1-7] Progress has been made on the applications of many computational methods, such as various scoring functions, molecular mechanics energies combined with the Poisson-Boltzmann or generalized Born and surface area continuum solvation (MM-PBSA or MM/GBSA), and statistical mechanics based methods for drug design and discovery.[8-12] Although computational methods have contributed considerably to the hit discovery and lead optimization, there still exist many problems associated with the computational accuracy and efficiency of these drug design methods. [13] It is highly desirable to develop the receptor-ligand binding affinity prediction methods for more reliable and/or efficient prediction of the binding free energies in structure-based drug design.

Interestingly, the free energy perturbation (FEP) method has demonstrated the promise of reliably predicting the relative protein-ligand binding free energies.[14-20] The FEP simulations were usually carried out to calculate the protein-ligand binding free energy change when a ligand changes from a reference structure to another one, thus predicting the relative binding free energy (RBFE). Recently, Wang *et al*. reported a FEP protocol to calculate the RBFE for a series of ligands with the same scaffolds and tested the accuracy for >10 targets.[14] Their FEP protocol gave rather accurate statistical results with a conventional correlation coefficient (R) of 0.81 between the FEP-predicted binding free energies and the corresponding experimental data. In a recently published report, they further improved the RBFE calculation results by using different enhanced sampling methods.[21] Although the RBFE calculations have demonstrated satisfactory accuracy, their actual applications to drug design are still limited, because these FEP protocols are mainly focused on simulating minor structural changes of the

ligands.[22] To deal with ligands with a variety of diverse structures, one has to calculate the absolute binding free energy (ABFE) for each ligand without use of a reference ligand structure. The practical applications of the FEP method in drug design would be extended greatly, if the FEP simulations could be used for the ABFE prediction. However, compared to the traditional FEP simulations for the RBFE calculation, the FEP simulations for the ABFE calculation would be much more computationally intensive and challenging.

In the present study, we evaluated the FEP-based ABFE (FEP-ABFE) calculations on a variety of receptor-ligand binding systems by using a new, efficient algorithm called Gaussian algorithm enhanced free energy perturbation (GA-FEP) protocol to simulate annihilation of the whole ligand structure with improved integration accuracy of the simulation. The receptor-ligand binding systems tested in our GA-FEP/ABFE calculations include seven drug targets with >100 ligands. Four out of the seven targets have ligands with similar scaffolds, and the rest three targets have structurally diverse ligands. The GA-FEP/ABFE calculations have demonstarted the reasonable accuracy for all kinds of ligands tested, with or without similar scaffolds. For ligands with similar scaffolds, the accuracy of the GA-FEP/ABFE calculations is comparable to that of the corresponding RBFE calculations. Further, the GA-FEP/ABFE calculations were performed to design new inhibitors against phosphodiesterase-10 (PDE10). PDE10 is recognized as a promising drug target for treatment of colon cancer[23] and central nervous system (CNS) disorders such as schizophrenia and Huntington's disease[24] with several inhibitors including MP-10 and OMS-824 in clinical trials.[25] However, there is still no PDE10 inhibitor approved for clinical use. It is still highly desirable to design new, more potent PDE10 inhibitors to accelerate the drug development targeting PDE10. Herein, starting from a hit H1[26] (**LHB-1**) with an IC_{50} of 1.8 μM, two series of new PDE10 inhibitors were designed, synthesized, and assayed for their

inhibitory activity. As a result, the GA-FEP/ABFE guided lead optimization on **LHB-1** led to a ~2,000-fold improvement in the inhibitory potency against PDE10. In addition, their predicted binding patterns were further verified by determining a co-crystal structure of the PDE10-inhibitor complex, and the predicted binding free energies correlate well with the experimental activity data. To the best of our knowledge, this is the first FEP protocol of systematic ABFE predictions for a large number of receptor-ligand systems, demonstrating the remarkable accuracy and efficiency of this method. The validated GA-FEP/ABFE method may be valuable in a variety of structure-based ligand design efforts for rational design of novel drugs and chemical probes.

Results and Discussion

Fitting the probability distribution by Gaussian functions can greatly improve the convergence and accuracy. Using the GA-FEP protocol (see Methods section for the details), we first carried out the FEP-ABFE calculations on seven protein targets binding with more than 100 ligands to examine the GA-FEP protocol. Among the seven targets, the crystal structures of CDK2,[27] JNK1,[28] Thrombin, TYK2,[29] and their relevant ligands with similar scaffolds were selected from Wang and Abel's work (RBFE method)[14] in order to compare the accuracy of the GA-FEP/ABFE method with that of the RBFE method. All of the 63 ligands for the four targets used in Wang and Abel's work were predicted by our GA-FEP/ABFE method, except for ligand **1d** against Thrombin since it contains an iodine atom without the 6-31G* basis set available for the restricted electrostatic potential (RESP) calculations at the HF/6-31G* level (a level used for all of the ligands in this study). The structurally diverse ligands for the other three targets, *i.e.* cAMP-specific phosphodiesterase-4 (PDE4), cGMP-specific PDE5,[30] and cGMP-specific PDE9,[31] were selected from BindingDB[32] and other

publications[31, 33] (see Supporting Information Section S1 for the detailed structures). Figure 1 shows the thermodynamic cycle designed to calculate the ABFE between a receptor and its ligand.

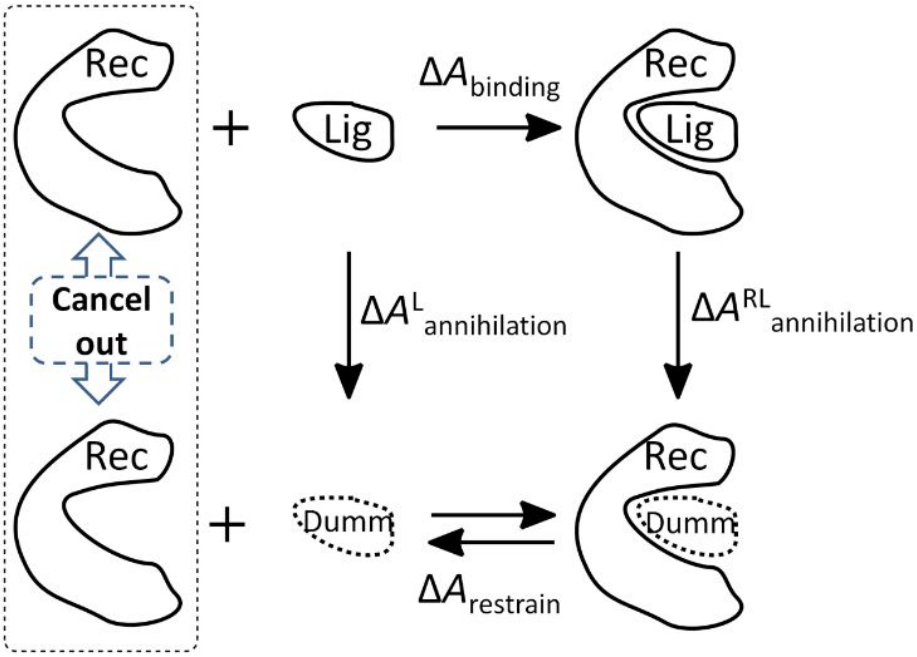


Figure 1. Thermodynamic cycle designed to calculate the absolute binding free energy between a receptor and its ligand.

As well known, usual FEP simulations with large perturbations can often have convergence problems. In particular, the negative $-\Delta U$ tail in $P(\Delta U)$, which is always poorly sampled, could significantly influence the calculation results. Considering the Gaussian-like distribution of the sampled probability $P(\Delta U)$ as well as the well sampled part around the peak of $P(\Delta U)$, five Gaussian functions were used to fit the probability distribution $P(\Delta U)$. As a result, the well-sampled part around the peak of $P(\Delta U)$ can be fully considered, and the negative $-\Delta U$ tail could be refined through extrapolation. As shown in Figure 2, by using five Gaussian functions, the distribution of ΔU can be fitted very well.

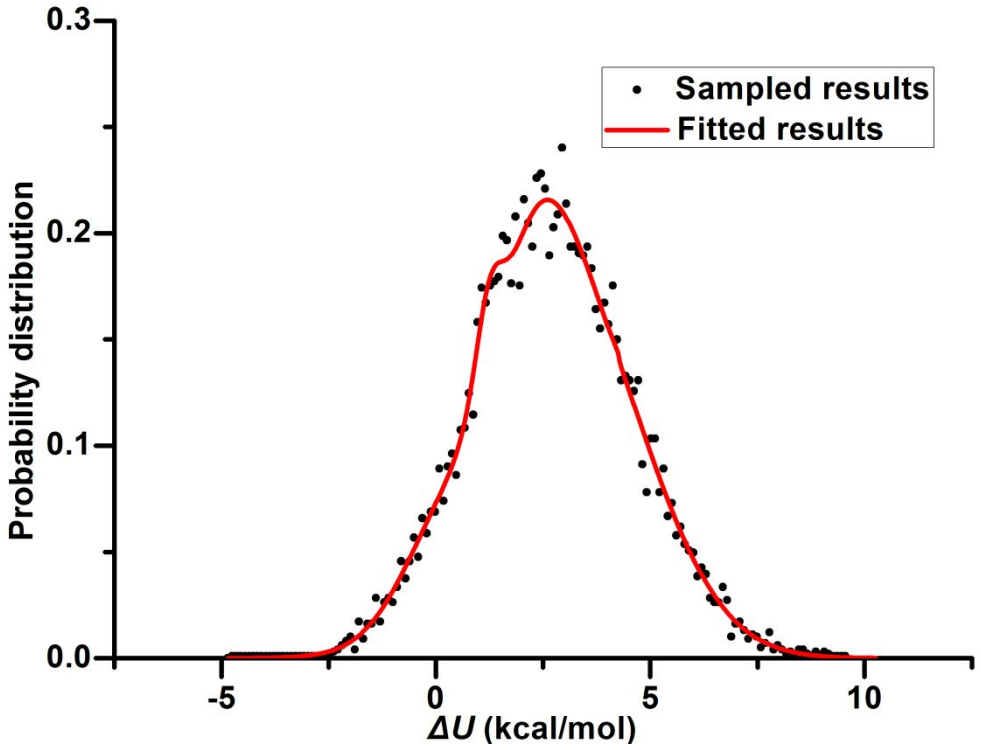


Figure 2. Fitting a sampled probability distribution by five Gaussian functions. Sampling information can be fully considered, and the poorly sampled part can be refined by extrapolation.

By using the strategy of fitting the probability distribution by Gaussian functions, along with the Bennett Acceptance Ratio (BAR) method, the convergence of FEP calculation was greatly increased. In the GA-FEP/ABFE calculations, the largest perturbations mainly come from the annihilation of electrostatic interactions that may contribute to more than 90% of the total interaction energies for most of the receptor-ligand binding systems. Summarized in Table S1 (Supporting Information) are the computational data obtained for an example, *i.e.* **lig_16** binding with CDK2. The total interaction energies were about 270 kcal/mol for both the complex and ligand systems, and the electrostatic interaction energies of these two systems were more than 250 kcal/mol, about 95% of the total interaction energies (see Supporting Information Section S2 for the details). The standard devation (calculated based on the forward energy, backward energy, and Bennett Acceptance Ratio (BAR) energy) of the electrostatic interaction calculation was only 0.9 kcal/mol, indicating good convergence

of the electrostatic interaction calculation. For all the >100 ligands that were used to test the GA-FEP/ABFE method, as shown in Figure 3a, the standard deviations of the electrostatic interaction calculations were reduced by more than 50% after fitting probability distributions by Gaussian functions, with almost all of them within 2 kcal/mol. The effects of fitting probability distributions to both electrostatic interaction and total interaction can be seen in Figure 3b. Without fitting the probability distribution, the standard deviations of both the electrostatic energy and total energy spread much wider, making the results hardly to be accurate. While after fitting the probability distributions, for more than 98% of the electrostatic energy results and more than 95% of the total energy results, the standard deviations can be lower than 2 kcal/mol.

We further examined the convergence of the ABFE calculations (containing 10 alchemical states, or lambda windows, with 4 ns simulations for each window) on representative CDK2 inhibitors, including **30**, **28**, and **1oiy**, with increasing the number of lambda windows (with doubled and quadrupled numbers of lambda windows) and extending the simulation time for each window (from 4 ns to 20 ns). It turned out that further increasing the number of windows and/or further extending the simulation time for each window did not affect the calculated ABFE results too much (see detailed data in Supporting Information Section S3 and Supporting Dataset S1).

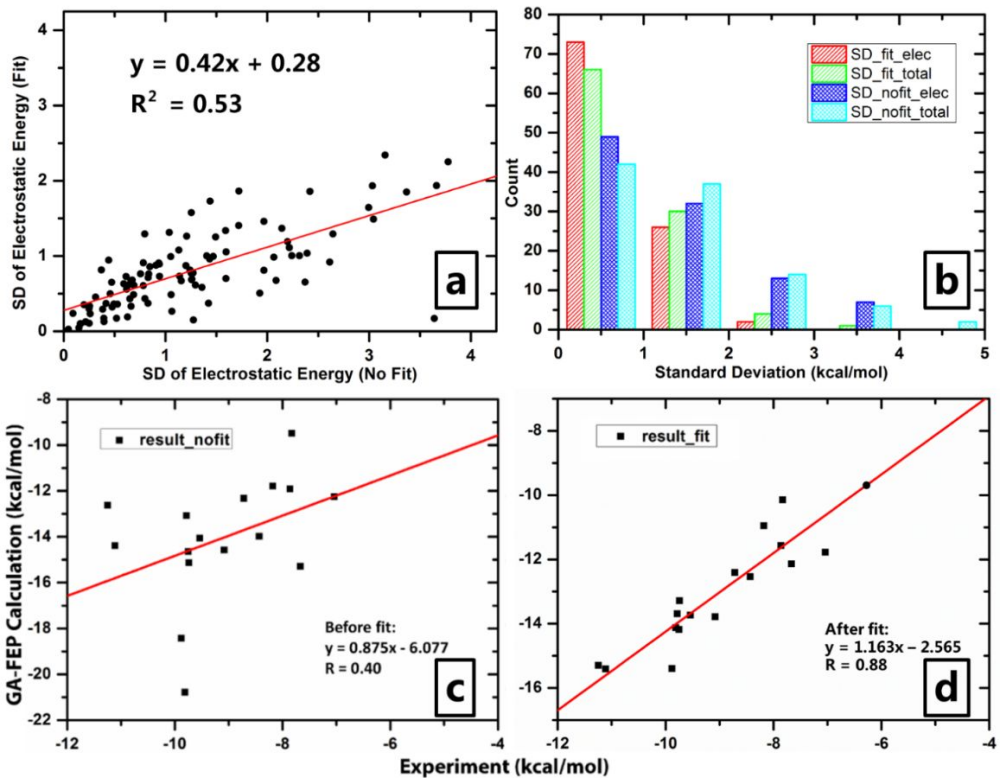


Figure 3. Fitting probability distribution can significantly reduce the standard deviations and improve the accuracy of the energy calculation results. (a) Standard deviations of electrostatic energy calculations with and without fitting probability distributions, denoted by Fit and No Fit, respectively. (b) Distributions of standard deviations. Red and green bars represent electrostatic energy and total energy after fitting probability distributions, respectively. Blue and cyan bars represent electrostatic energy and total energy without fitting probability distributions, respectively. (c) The result of CDK2 without fitting by Gaussian functions. (d) The result of CDK2 after fitting by 5 Gaussian functions.

The improvement in the accuracy is also notable. To show the effect of fitting the energy distribution to the final results, the binding free energies for the 16 ligands of CDK2 were also calculated by only the BAR method without fitting using the Gaussian functions (denoted as nofit-BAR below). As shown in Figure 3c-d, when the nofit-BAR method was used to calculate the binding free energies, the correlation coefficient (R) between the predicted and experimental values was only 0.39. In contrast, with the probability distribution fitting before the BAR calculations, the correlation coefficient improved to 0.88. Thus, fitting the energy distribution can considerably improve both the

convergence and accuracy. Below, we will only discuss the FEP-ABRE results calculated by using the standard GA-FEP protocol with the use of five Gaussian functions.

For ligands with similar scaffolds, the overall accuracy of our GA-FEP/ABFE method is comparable to that of the RBFE method. Herein, the experimental binding affinity (binding free energy) values (ΔG_{exp}) were derived from available experimental IC_{50} values by using equation $\Delta G_{exp} \approx RT\ln(IC_{50})$, whereas the predicted binding affinity values (ΔG_{pred}) were calculated by the GA-FEP/ABFE method. Summarized in Table 1 and depicted in Figure 4a are the statistical results of the four retrospective drug targets and the comparison between the GA-FEP/ABFE and RBFE results. Their individual linear regression results for the four targets between the ΔG_{exp} and ΔG_{pred} values are given in Supporting Information Section S4. Due to the intrinsically insufficient simulation on the hydration of the receptor binding site, the predicted binding free energies for a same receptor tend to systematically shift to the same direction compared to the corresponding experimental values, resulting in a non-zero intercept of the fitted equation between the experimental and calculated results. Thus, to put all of the results together and show the overall performance of the GA-FEP/ABFE method, the computed binding energies for each target were shifted back to make the intercept of the fitted equation close to zero. As reported in Abel and Wang's study, the overall correlation coefficient (R) for all of the eight targets used in their study was 0.81.[14] Four out of the eight drug targets were selected as the test set in this study, and the R value for the four targets was 0.73. Our GA-FEP/ABFE method, which uses the enhanced sampling method as mentioned above, has the accuracy with R = 0.79 that is comparable to that (with R = 0.73) of the RBFE method when the same test set of 63 complexes were used. The R value is also comparable to that (R = 0.81) for the overall accuracy of the RBFE method. It should be noted that the RBFE is still preferable when there is a known reference ligand and when

the structural changes from the reference compound to new compounds are minor due to the relatively smaller perturbation and better convergence. However, our GA-FEP/ABFE method will greatly extend the applications of the FEP method in drug discovery because the method does not need a reference compound for the ABFE calculation.

Table 1. Comparison of the statistical results between absolute (GA-FEP/ABFE) and relative (FEP/REST) binding free energy calculations. RBFE (relative binding free energy) are the FEP/REST (free energy perturbation/replica exchange with solute tempering) calculation results from Wang and Abel's work,[14] and the ABFE (absolute binding free energy) data are calculated by using our GA-FEP/ABFE method described in the current study. RMSE refers to the root mean square error. More details including the calculated ABFE of each protein-ligand complex, unsigned error *etc*. can be found in Supporting Dataset S1.

Target		CDK2	JNK1	TYK2	Thrombin	Overall[a]	PDE4	PDE5	PDE9
No. of compounds		16	21	16	10	63	20	11	7
Crystal structure		1H1Q	2GMX	4GIH	2ZFF		1XOQ	2H42	4GH6
RBFE									
	***R*[14]**	0.48	0.85	0.89	0.71	0.73	–	–	–
	Slope[14]	0.27	1.78	0.80	1.03	0.88	–	–	–
	RMSE[14]	1.11	1.00	0.93	0.93	–	–	–	–
ABFE									
	R	0.88	0.69	0.72	0.70	0.79	0.74	0.69	0.94
	Slope	1.16	0.86	1.08	1.00	1.07	0.99	0.91	0.71
	RMSE	0.73	0.76	1.36	0.62	–	1.23	1.54	0.65
MM/PBSA									
	R	0.00	0.70	0.32	0.14	–	0.17	0.14	0.20
	Slope	-0.03	3.06	0.62	0.91	–	-1.95	0.69	0.53
	RMSE	2.47	2.71	2.37	3.87	–	13.11	8.24	6.28
MM/GBSA									
	R	0.10	0.60	0.33	0.30	–	0.40	0.32	0.69
	Slope	-0.22	2.14	0.62	1.29	–	1.45	1.44	1.91
	RMSE	4.92	2.44	2.31	2.40	–	4.25	6.41	4.94

[a] Including CDK2, JNK1, TYK2, and Thrombin in order to compare ABFE with RBFE.

The GA-FEP/ABFE method also shows a remarkable accuracy for structurally diverse ligands. The statistical results associated with targets PDE4, PDE5, and PDE9, as shown in Table 1 and Figure 4b-d, reveal that the GA-FEP/ABFE method can also produce accurate results for these three targets with structurally diverse ligands. The statistical regression coefficient *R* values for PDE4, PDE5, and PDE9 are 0.74, 0.69, and 0.94, respectively, suggesting that the GA-FEP/ABFE method has a remarkable accuracy for predicting the binding affinities (ΔG_{pred}) compared to the corresponding experimental binding data (ΔG_{exp}). Furthermore, for these three targets in this study, the three slopes (0.997, 0.913, and 0.712) of the linear regression results between ΔG_{pred} and ΔG_{exp} are all reasonably close to 1 (the theoretically ideal value).

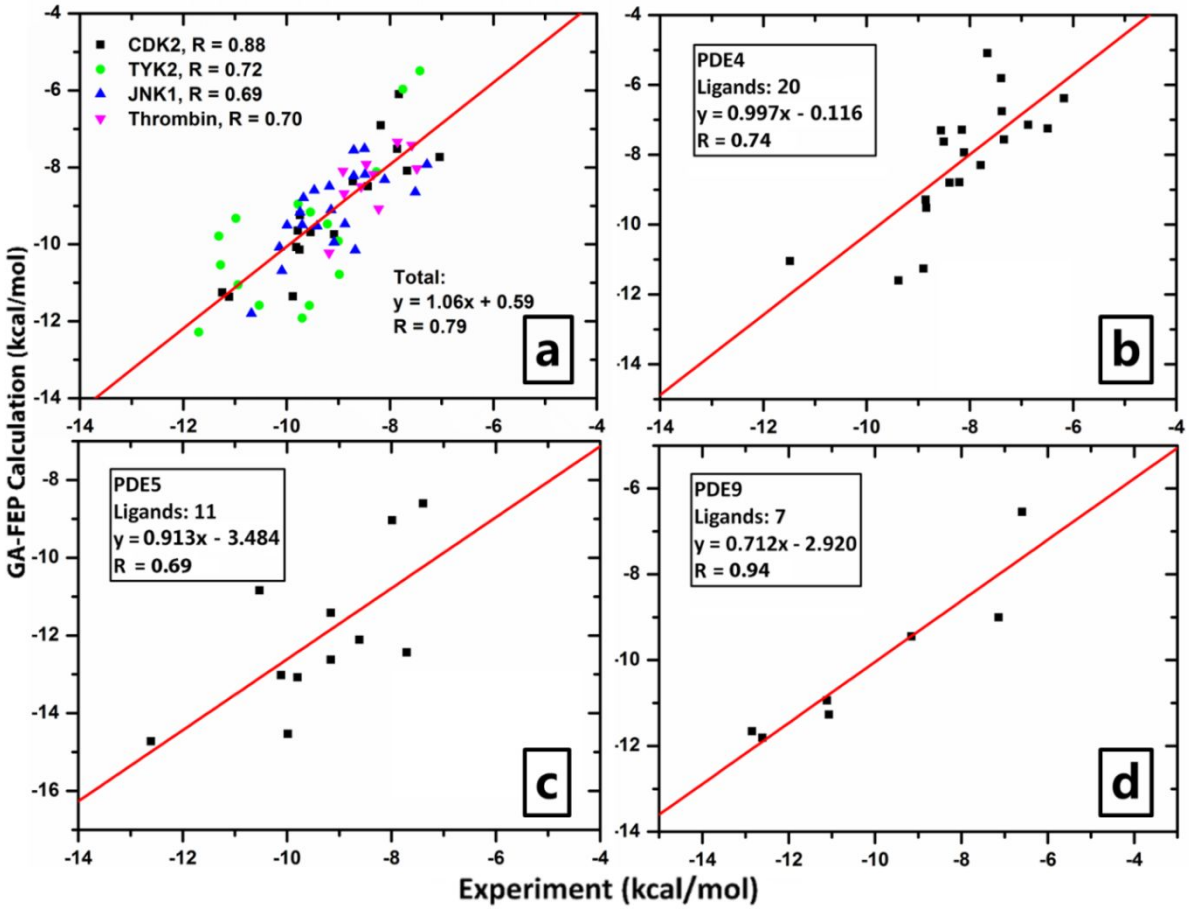


Figure 4. The correlations between the GA-FEP predicted and experimental affinities. (a) Results for the structurally similar ligands. (b), (c), and (d) Results for the structurally diverse ligands.

Residual distribution of the GA-FEP/ABFE method. The residual distribution of the GA-FEP/ABFE method between the calculated and experiment results for each drug target is shown in Figure 5. For CDK2, JNK1, TYK2, and Thrombin, the ligands of which have the same scaffolds, the residuals are all within 2 kcal/mol from their medians. For PDE4, PDE5, and PDE9 with structurally diverse ligands, the residual distributions spread a little bit wider than those of the above four targets and the errors reach 2-3 kcal/mol from their medians. For the targets with structurally diverse ligands, the accuracy of this method is relatively lower than those with structurally similar ligands, but the correlation between ΔG_{pred} and ΔG_{exp} is still reasonable. To find out the residual distribution property of the results from our GA-FEP/ABFE calculations, we put all of the >100 results together, and as described in Supporting Information Section S5, the result is a Gaussian-like distribution. All the residual distribution data, RMSE, error (Dif_cal-exp), *etc*. can be found in Supporting Dataset S1.

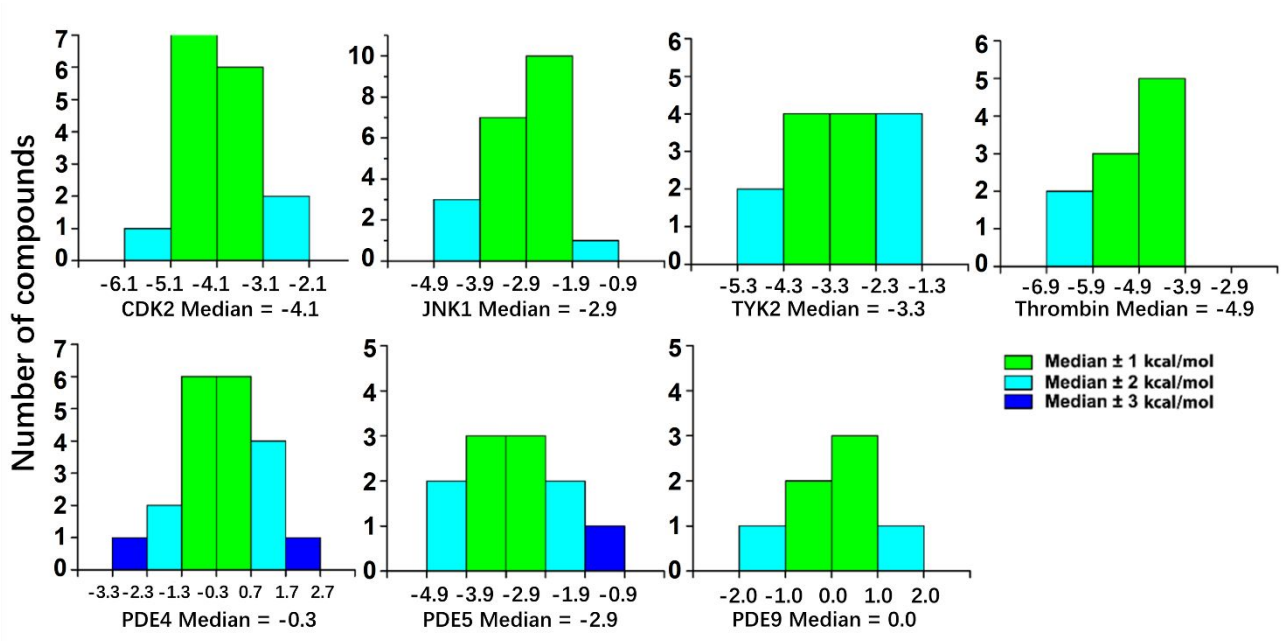


Figure 5. Residual distributions of the GA-FEP/ABFE method against each drug target. For the four targets in the first row, they have structurally similar ligands, and for the three targets in the second

row, they have structurally diverse ligands. Different colors stand for how far the values are from their median in kcal/mol.

Comparison of the GA-FEP/ABFE results with the corresponding MM/PBSA and MM/GBSA results. The MM/PBSA and MM/GBSA methods have been widely used for binding free energy prediction in drug design. In comparison with traditional FEP calculations that would be extremely difficult to predict the binding free energies for structurally diverse ligands, the MM/PBSA and MM/GBSA calculations are practically available for predicting the binding free energies as shown in reported studies.[28, 31] However, our GA-FEP protocol made the FEP calculations possible for the ABFE predictions with acceptable speed (under the GPU acceleration). Thus, to know how much improvement in the accuracy the GA-FEP/ABFE method can make, the MM/PBSA and MM/GBSA methods were also used to predict all of the binding energies of the same (>100) ligands for comparison. In order to calculate the MM/PBSA and MM/GBSA binding energies, 4 ns MD simulations were first applied to each of the receptor-ligand complexes. For each complex, 100 snapshots were extracted from the last 1 ns trajectory with an interval of 10 ps. Then, the MM/PBSA and MM/GBSA calculations were carried out on these snapshot structures. The detailed results from the MM/PBSA and MM/GBSA calculations are summarized in Table 1 and given in Supporting Dataset S1. As seen in Table 1, the correlation coefficnient (R) values associated with targets CDK2, JNK1, TYK2, Thrombin, PDE4, PDE5, and PDE9 are 0.10, 0.60, 0.33, 0.30, 0.40, 0.32, and 0.69, respectively, for the MM/GBSA method, and 0.00, 0.70, 0.32, 0.14, 0.17, 0.14, and 0.20, respectively, for the MM/PBSA method. In comparison, for the GA-FEP/ABFE method, the corresponding R values are 0.88, 0.69, 0.72, 0.70, 0.74, 0.69, and 0.94, respectively. The overall accuracy has been improved significantly. As seen from Table 1, the accuracies of the MM/PBSA and MM/GBSA methods are significantly lower than those of the two types of FEP methods for all of the targets, except JNK1. The

MM/PBSA and MM/GBSA calculations were reasonable for JNK1, but not satisfactory for most of the targets. So, only the FEP calculations produced consistently reasonable binding free energies.

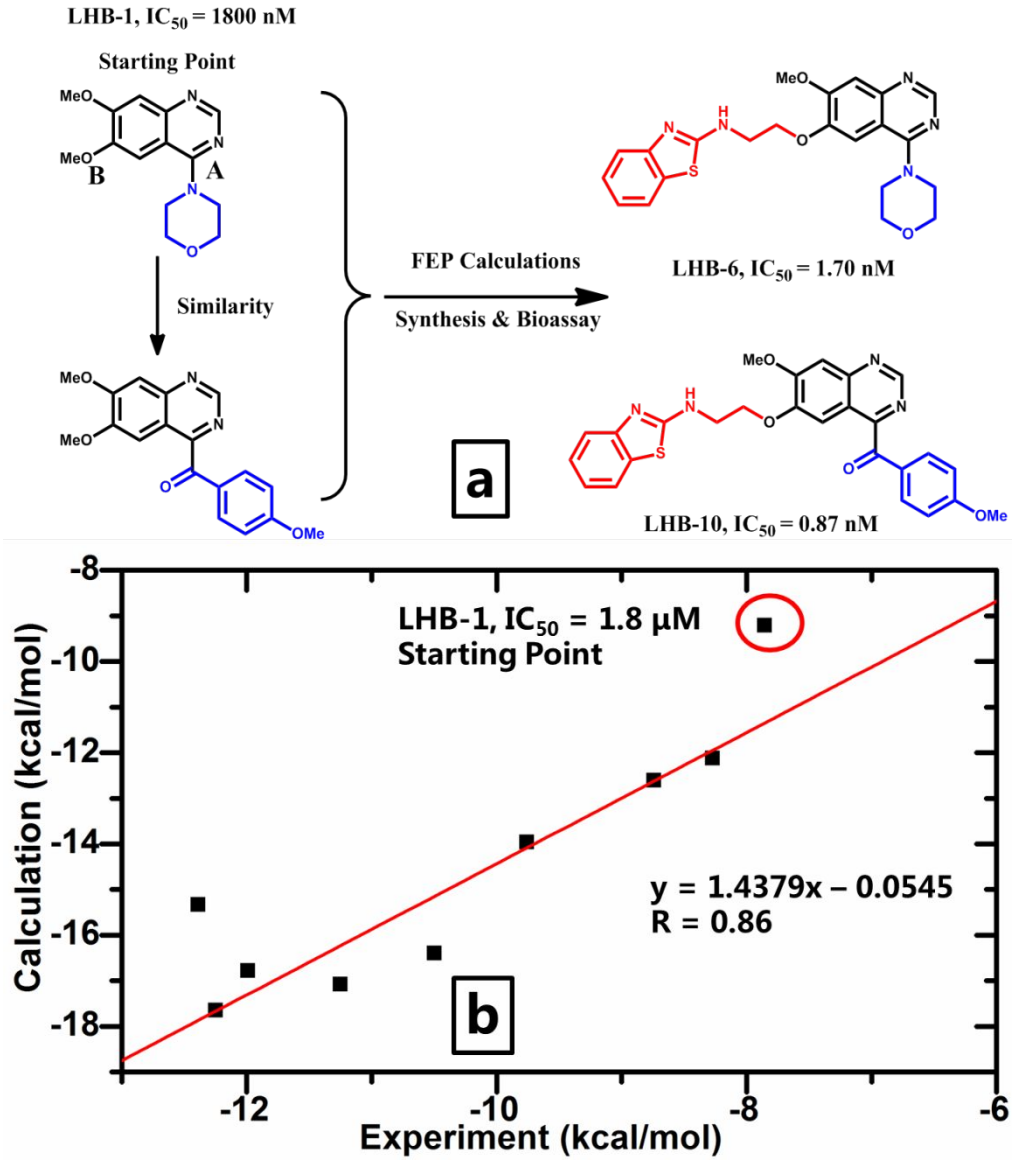


Figure 6. Lead optimization strategy starting from the hit **LHB-1** with a moderate IC_{50} of 1.8 μM. (a) Lead optimization strategy. (b) The predicted results of the designed molecules showed remarkable correlation with the experimental results with a regression coefficient of 0.86.

The GA-FEP/ABFE method has led to discovery of highly potent inhibitors of PDE10. For a practical drug discovery effort, and also for further external validation of the GA-FEP method, the GA-FEP/ABFE calculations were performed to rationally design new, more potent PDE10 inhibitors

starting from a known PDE10 inhibitor for lead optimization. Compound **LHB-1 (**Figure 6a), which has a reasonable molecular weight of 275 and a moderate IC_{50} of 1.8 μM, was taken as a starting fragment for the purpose of lead optimization. Since **LHB-1** is a very small molecule, we first identified the core structure that should be kept and the side chain that should be optimized. According to the docking structure, the 6,7-dimethoxyquinazolin group (black colored part of the structure on Figure 6) forms a hydrogen bond and π-π interaction with the conserved amino acids Q716 and F719, respectively. Hence, the 6,7-dimethoxyquinazolin structure was chosen as the core structure. Based on the core structure, side chains on two positions (positions A and B denoted on Figure 6) can be easily modified due to their orientations pointing to outside of the pocket and, thus, there are sufficient rooms for their structural modifications. We first tried to optimize the side chain on position A. Through the GA-FEP/ABFE calculation, we found that the calculated binding energy was significantly increased by ~3 kcal/mol when the side chain was benzoyl group such as **LHB-2** and **LHB-3**, and their actual activities increased to 890 and 403 nM, respectively. We further tried to optimize the side chain on position B. Based on the protein structure, along the orientation of position B we found Y683 could be a potential proton donor and there also exists a hydrophobic pocket. According to the GA-FEP/ABFE calculation, when connecting a benzo[d]imidazole or a benzo[d]thiazole group to B position through a 3-atoms flexible chain, like the 7 ligands from **LHB-4** to **LHB-10**, the binding free energy may be further increased by 3-5 kcal/mol. The substituents on position B of these ligands could form a hydrogen bond and hydrophobic interaction with Y683 and the hydrophobic pocket, respectively, and the IC_{50} of the most potent ligand **LHB-10** reached to 0.87 nM (the actual experimental value). The designed structures resulted in a total of 9 ligands that can be classified into two series with slightly different scaffolds. Their structural details, activity data, and the calculated ABFE results are given in

Supporting Information Section S6.

PAINS screening of the newly designed compounds was carried out *via* an online program PAINS-Remover (http://cbligand.org/PAINS)[34] in order to prevent false positive results, and all the designed compounds passed this screening. As a result, most of the synthesized compounds showed improved inhibitory activity against PDE10 compared to **LHB-1**.

Among all the designed molecules, four of them had IC_{50} lower than 10 nM, and the most potent compound **LHB-10** showed an IC_{50} of 0.87 nM against PDE10, with a ~2000-fold improved binding affinity compared to **LHB-1**. Another potent inhibitor is **LHB-6** with an IC_{50} of 1.7 nM against PDE10. Depicted in Figure 7a-b are the predicted binding structures of PDE10 with **LHB-10** and **LHB-6**, showing some common favorable interactions of the two inhibitors with residues Q716, F719, and Y683. As described in the above ligand design procedure, the interactions with Q716 and F719 are common for all of the designed compounds starting from **LHB-1**, compounds **LHB-6** and **LHB-10** have the unique hydrogen bonding interaction with the side chain of Y683 and more favorable hydrophobic interactions with surrounding amino acids.

Further, the computationally predicted PDE10-inhibitor binding mode for these new, potent inhibitors was confirmed by our X-ray structural analysis. In particular, the X-ray crystal structure (Figure 7c) of PDE10 in complex with **LHB-6** was determined (PDB ID: 5ZNL, Supporting Information Section S7) to verify their predicted binding mode derived from our GA-FEP/ABFE calculations. In Figure 7d, the binding mode in the crystal structure is indeed consistent with the predicted binding pattern. Meanwhile, the ΔG_{pred} values for the nine compounds and **LHB-1** showed statistically linear correlation with the corresponding experimental ΔG_{exp} values as reflected by the regression coefficient *R* of 0.86 (Figure 6b). Apparently, the ABFE calculations using the GA-FEP

protocol showed a remarkable accuracy and can be used for practical structure-based drug design.

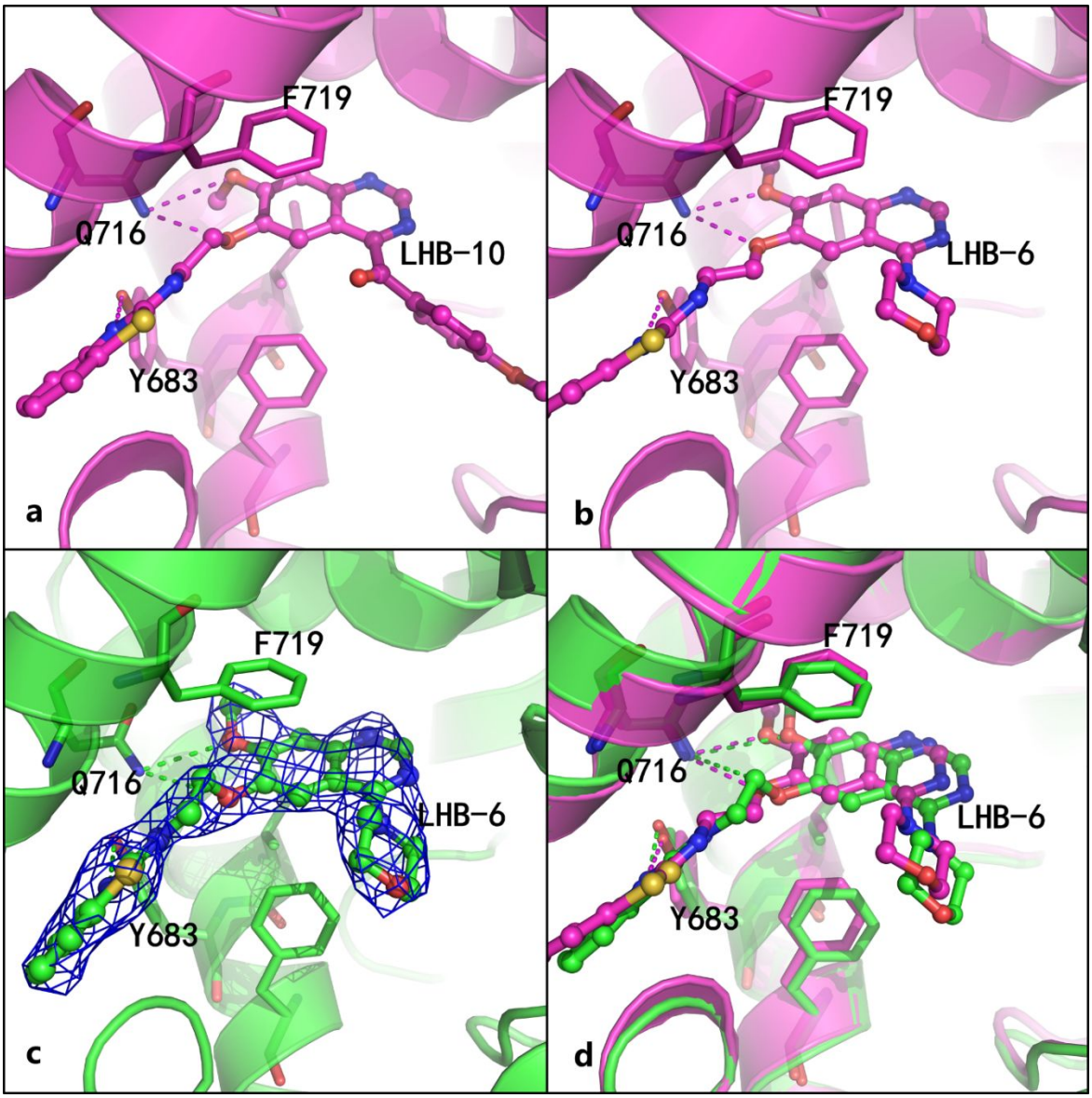


Figure 7. The predicted binding mode was confirmed by crystal structure. (a) Predicted binding mode between PDE10 (PDB ID: 2O8H) and the most potent inhibitor **LHB-10**. (b) Predicted binding mode between PDE10 (PDB ID: 2O8H) and another potent inhibitor **LHB-6**. (c) Determined crystal structure of PDE10 with **LHB-6** (PDB ID: 5ZNL). The 2Fo–Fc electron density is contoured in dark blue at 1.0 δ. (d) The superimposed structure between predicted binding mode of **LHB-6** and the crystal structure. All the predicted structures are depicted in magenta, and the crystal structure is depicted in green.

Conclusion

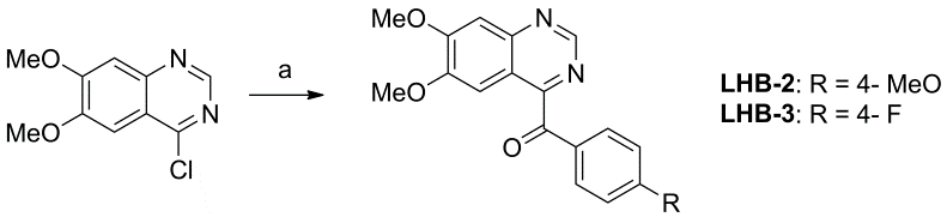
In this study, we reported a relaible and efficient GA-FEP protocol for the FEP-ABFE calculations on protein targets binding with ligands. An extensive test including more than 100 ligands and 7 targets has demonstrated that fitting probability distribution by Gaussian functions can efficiently increase the convergence and accuracy of the computational results. The computational results were able to reach the satisfactory accuracy for all kinds of ligands tested, with or without similar scaffolds, which may considerably extend the pratical applications of the FEP approach.

Further, using the GA-FEP protocol, the FEP-ABFE calculation-based lead optimization targeting PDE10 resulted in several new, highly potent PDE10 inhibitors with sub-nanomolar IC_{50}, illustrating the value of the computational protocol in practical structure-based drug design. In addition, the GA-FEP calculations are computationally effieicnt, *e.g.* one can complete the calculations on 2-3 ligands per day for their absolute binding free energies by using 8 Nvidia Geforce GTX580 GPUs. Thus, the GA-FEP protocol described in this report is promising for the ABFE prediction and structure-based drug design.

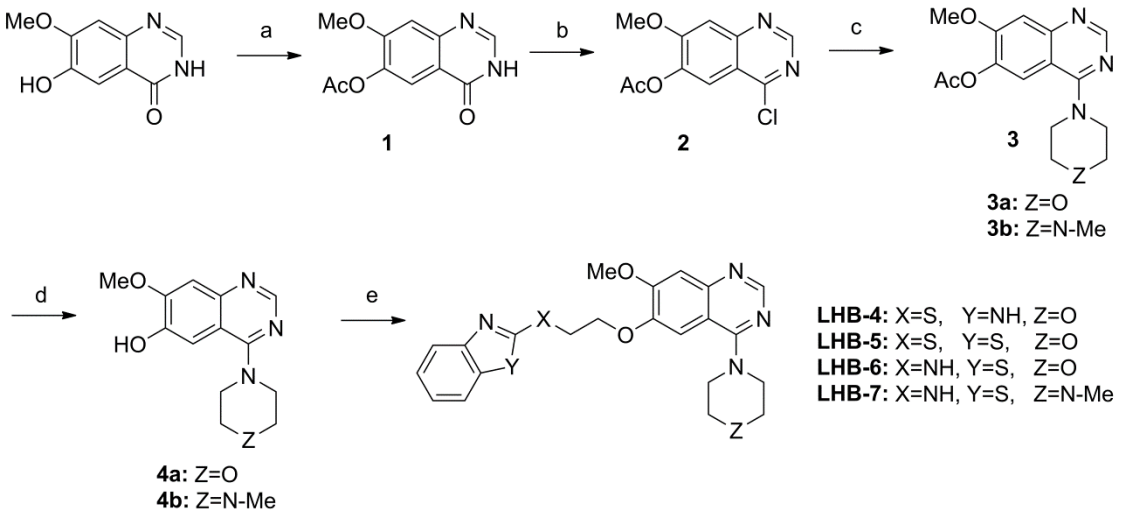
EXPERIMENTAL SECTION

Chemistry: General Methods. The designed compounds were synthesized through the routes outlined in Schemes 1-3. ^{1}H NMR and ^{13}C NMR spectra were recorded on a BrukerBioSpin GmbH spectrometer at 400.1 and 100.6 MHz, respectively. Coupling constants are given in Hz. MS spectra were recorded on an Agilent LC-MS 6120 instrument with an ESI and APCI mass selective detector. The high-resolution mass spectrum was run on a Shimadzu LCMS-IT-TOF. Flash column chromatography was performed using silica gel (200–300 mesh) purchased from Qingdao Haiyang

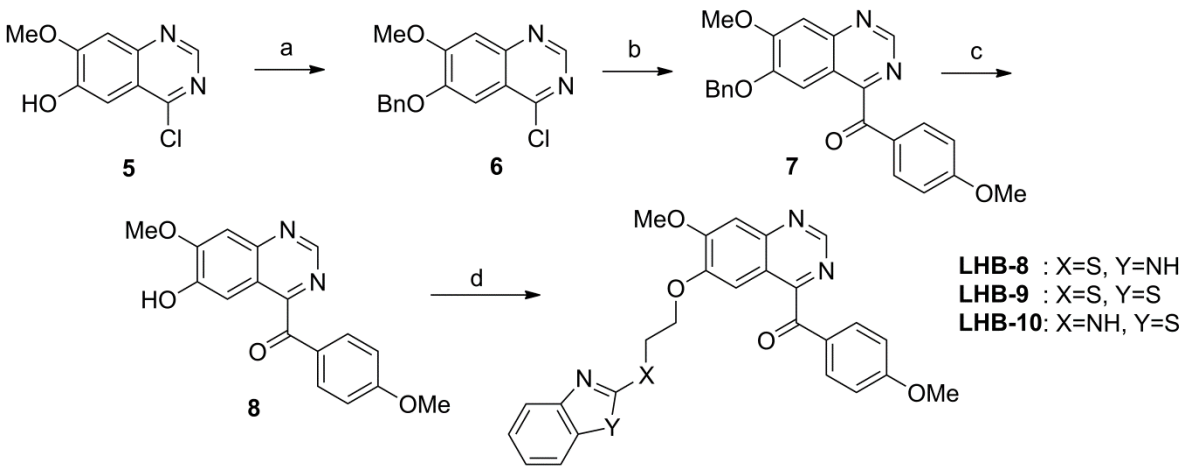
Chemical Co. Ltd. Thin-layer chromatography was performed on precoated silica gel F-254 plates (0.25 mm, Qingdao Haiyang Chemical Co. Ltd) and was visualized with UV light. All the starting materials and reagents were purchased from commercial suppliers and used directly without further purification. The purity of these compounds was determined by reverse-phase high performance liquid chromatography (HPLC) analysis and confirmed to be better than 95%. HPLC instruments: Shimadzu LC-20AT (SPD-20A UV/vis detector, UV detection at 254 nm; column Hypersil BDS C18, 5.0 μm, 4.6 × 150 mm (Elite); Elution, MeOH in water (70%-90%, v/v); T = 25 ℃; and flow rate = 1.0 mL/min). All the compounds are synthesized following Schemes 1-3. Compounds **1** to **8** are intermediate compounds, and compounds **LHB-2** to **LHB-10** are the designed target compounds starting from the hit **LHB-1** (which was synthesized according to method described previously[26]).



Scheme 1. The synthesis route of compounds LHB-2 and LHB-3. Reagents and conditions: (a) Benzaldehyde, 1, 3-dimethylimidazolium iodide, NaH, THF, reflux.



Scheme 2. The synthesis route of compounds LHB-4, LHB-5, LHB-6, and LHB-7. Reagents and conditions: (a) Acetic anhydride, pyridine, 100 °C; (b) $SOCl_2$, 80 °C; (c) Morpholine or 1-methylpiperazine, DMF, 80°C; (d) Ammonia, methanol, reflux; (e) The corresponding bromide, K_2CO_3, CH_3CN, reflux.



Scheme 3. The synthesis route of compounds LHB-8, LHB-9, and LHB-10. Reagents and conditions: (a) Benzyl bromide, acetone, K_2CO_3, 80°C; (b) 4-methyoxylbenzaldehyde, 1,3-dimethylimidazolium iodide, NaH, THF, reflux; (c) Boron tribromide, DCM, -78°C; (d) The corresponding bromide, K_2CO_3, CH_3CN, reflux.

Compound 1: 7-methoxy-4-oxo-3,4-dihydroquinazolin-6-yl acetate

Pyridine (4.0 mL) was added dropwise to the solution of 6-hydroxy-7-methoxyquinazolin-4(3H)-one (1.92 g, 10 mmol) in acetic anhydrate (20 mL). The reaction mixture was heated at 100 °C for 2 h and then cooled to room temperature. After the mixture was poured into ice water, a white solid was precipitated. The precipitate was collected, washed and dried to give compound **1** (2.32 g, Yield: 99%) as a white solid. ^{1}H NMR (400 MHz, DMSO-d_6) δ 8.09 (s, 1H), 7.76 (s, 1H), 7.28 (s, 1H), 3.92 (s, 3H), 2.30 (s, 3H).

Compound 2: 4-chloro-7-methoxyquinazolin-6-yl acetate

Compound **1** (2.34 g, 10 mmol) was dissolved in $SOCl_2$ (20 mL). The mixture was stirred at 80

°C for 2.5 h and then concentrated under vacuum, providing compound **2** (2.22 g, Yield: 88%) which could be used in the next step without further purification. ^{1}H NMR (400 MHz, DMSO-d_6) δ 9.02 (s, 1H), 8.02 (s, 1H), 7.65 (s, 1H), 4.03 (s, 4H), 2.36 (s, 3H).

Compound 4a: 7-methoxy-4-morpholinoquinazolin-6-ol

A solution of compound **2** (2.52 g, 10 mmol) and morpholine (1.04 g, 12 mmol) in DMF (20 mL) was stirred at 80 °C for 6 h. The mixture was then poured onto the ice water and a white solid was precipitated, which was collected and washed with ice water to afford compound **3a**. Compound **3a** was dissolved in methanol (20 mL). Ammonia (1.6 mL) was added to the mixture and the mixture was then stirred under reflux for 2 h. The solvents were evaporated under vacuum. The crude product was recrystallized using methanol to afford compound **4a** (1.87 g, Yield: 72%) as a white solid. ^{1}H NMR (400 MHz, DMSO-d_6) δ 8.52 (s, 1H), 7.24 (s, 1H), 7.21 (s, 1H), 3.93 (s, 3H), 3.83 – 3.75 (m, 4H), 3.54 – 3.47 (m, 4H).

Compound 4b: 7-methoxy-4-(4-methylpiperazin-1-yl)quinazolin-6-ol

Compound **4b** (1.54 g, Yield: 56%) as a white solid was synthesized using the same procedure of preparing compound **4a** but with 1-methylpiperazine as the starting material. ^{1}H NMR (400 MHz, DMSO-d_6) δ 9.93 (s, 1H), 8.49 (s, 1H), 7.23 (s, 1H), 7.19 (s, 1H), 3.93 (s, 3H), 3.52 (d, J = 4.8 Hz, 4H), 2.51 (dd, J = 3.6, 1.8 Hz, 4H), 2.26 (s, 3H).

Compound 6: 6-(benzyloxy)-4-chloro-7-methoxyquinazoline

Ammonia (25%, 2.2 mL, 25 mmol) was added to the solution of compound **2** (2.52 g, 10 mmol) in methanol (50 mL). The mixture was stirred at room temperature for 2 h. And then the solvent was evaporated under vacuum to give the crude product of compound **5**. The mixture of compound **5**, benzyl bromide (1.71 g, 15 mmol) and potassium carbonate (2.76 g, 20 mmol) were suspended in the

acetone and then stirred at 80 °C for 2 h. Water was added and the mixture was extracted with dichloromethane three times. The combined organic extract was washed with brine, dried over Na_2SO_4 and concentrated. The residue was purified on a silica gel column to give compound **6** (2.48 g, Yield: 83%) as a white solid. ^{1}H NMR (400 MHz, $CDCl_3$) δ 8.85 (s, 1H), 7.51 (d, J = 7.3 Hz, 2H), 7.47 (s, 1H), 7.42 (t, J = 7.3 Hz, 2H), 7.36 (t, J = 7.2 Hz, 1H), 7.34 (s, 1H), 5.31 (s, 2H), 4.06 (s, 3H).

Compound 7: (6-(benzyloxy)-7-methoxyquinazolin-4-yl)(4-methoxyphenyl)methanone

A mixture of compound **6** (2.41 g, 8.0 mmol), 1, 3-dimethylimidazolium iodide (90 mg, 0.4 mmol) and 4-methoxylbenzaldehyde (164 mg, 9.6 mmol) was dissolved in THF (80 mL). NaH (230 mg, 9.6 mmol) was added to the mixture and then stirred under reflux overnight with Ar protection. The mixture was cooled to room temperature and poured into ice water. The aqueous layer was extracted with ethyl acetate three times. The combined organic layers were washed with brine, dried over Na_2SO_4, and concentrated. The residue was purified on the silica column to give compound **7** (2.15 g, Yield: 68%) as an off-white solid. ^{1}H NMR (400 MHz, Acetone-d_6) δ 9.11 (s, 1H), 7.90 (d, J = 8.7 Hz, 2H), 7.46 (d, J = 8.8 Hz, 2H), 7.45(s, 1H), 7.40 (s, 1H), 7.37 – 7.27 (m, 3H), 7.06 (d, J = 8.7 Hz, 2H), 5.19 (s, 3H), 4.08 (s, 4H), 3.91 (s, 4H).

Compound 8: (6-hydroxy-7-methoxyquinazolin-4-yl)(4-methoxyphenyl)methanone

Compound **7** (2.00 g, 5.0 mmol) was dissolved in dichloromethane and the mixture was cooled to -78^{o}C. BBr_3 (1.0 M, 6.0 mmol, 6.0 mL) in dichloromethane was added dropwise to the mixture. The reaction was stirred at -78^{o}C for 1 h and then poured into ice water. A precipitate was formed and filtered, giving compound **8** (1.00 g, Yield: 65%) as a white solid. ^{1}H NMR (400 MHz, $CDCl_3$) δ 10.49 (s, 1H), 9.12 (s, 1H), 7.83 (d, J = 9.0 Hz, 2H), 7.46 (s, 1H), 7.09 (d, J = 9.0 Hz, 2H), 7.08 (s, 1H), 4.02 (s, 3H), 3.86 (s, 3H).

Compound LHB-2: (6,7-dimethoxyquinazolin-4-yl)(4-methoxyphenyl)methanone

A mixture of 4-chloro-6,7-dimethoxyquinazoline (224 mg, 1.0 mmol), 1, 3-dimethylimidazolium iodide (12 mg, 0.05mmol) and 4-methoxylbenzaldehyde (164 mg, 1.2 mmol) was dissolved in THF (10 mL). NaH (29 mg, 1.2 mmol) was added to the mixture and then stirred under reflux overnight with Ar protection. The mixture was cooled to room temperature and poured into ice water. The aqueous layer was extracted with ethyl acetate three times. The combined organic layers were washed with brine, dried over Na_2SO_4, and concentrated. The residue was purified on the silica column (dichloromethane: methanol = 80:1) to give compound **LHB-2** (146 mg, Yield: 45%) as a white solid. ^{1}H NMR (400 MHz, $CDCl_3$) δ 9.23 (s, 1H), 7.57 (s, 1H), 7.47 - 7.36 (m, 3H); 7.35 (s, 1H), 7.20 (d, *J* = 9.0 Hz, 1H), 4.10 (s, 3H), 3.96 (s, 3H), 3.87 (s, 3H).^{13}C NMR (101 MHz, $CDCl_3$) δ 193.60, 159.83, 159.69, 156.67, 152.45, 151.43, 149.93, 136.82, 129.68, 124.00, 121.04, 118.28, 114.25, 106.92, 102.39, 56.58, 56.34, 55.55. LC-MS (ESI) m/z [M]+ 325.2.

Compound LHB-3: (6,7-dimethoxyquinazolin-4-yl)(4-fluorophenyl)methanone

Compound **LHB-3** (181 mg, Yield: 58%) as a white solid was synthesized using the same procedure of preparing compound **LHB-2** but with 4-fluorobenzaldehyde as the starting material; ^{1}H NMR (400 MHz, $CDCl_3$) δ 9.23 (s, 1H), 8.09 – 7.98 (m, 2H), 7.42 (s, 1H), 7.40 (s, 1H), 7.23 – 7.15 (m, 2H), 4.10 (s, 3H), 3.99 (s, 3H); ^{13}C NMR (101 MHz, $CDCl_3$) δ 192.14, 158.99, 156.73, 152.36, 151.55, 150.13, 133.73, 133.63, 131.98, 118.36, 116.05, 115.83, 106.95, 102.36, 56.59, 56.36. HRMS (ESI-TOF) m/z [M + H]+ calcd for C17 H13 N2 O3 F 313.0983, found 313.0978.

Compound LHB-4: 4-(6-(2-((1H-benzo[d]imidazol-2-yl)thio)ethoxy)-7-methoxyquinazolin-4-yl)morpholine

The mixture of 2-((2-bromoethyl)thio)-1H-benzo[d]imidazole (0.308 g, 1.2 mmol), potassium

carbonate (0.207 g, 1.5 mmol), and compound **4a** (0.261 g, 1.0 mmol), were suspended in acetonitrile (30 mL). The mixture was then stirred under reflux for 16 h. The solid was filtered and the filtrate was concentrated under vacuum. The resulting residue was purified gel column (dichloromethane: methanol = 80:1) to give compound **LHB-4** (189 mg, Yield: 43%) as a white solid. ^{1}H NMR (400 MHz, $CDCl_3$) δ 8.69 (s, 1H), 7.55 (s, 2H), 7.34 (s, 1H), 7.24 (dd, *J* = 5.5, 3.3 Hz, 2H), 7.18 (s, 1H), 4.50 (t, *J* = 5.3 Hz, 2H), 4.12 (dd, *J* = 14.2, 7.1 Hz, 2H), 4.05 (s, 3H), 3.90 – 3.80 (m, 4H), 3.70 – 3.60 (m, 4H); ^{13}C NMR (101 MHz, DMSO-d_6) δ 163.40, 154.84, 152.96, 150.17, 149.15, 147.40, 135.99, 130.12, 122.12, 121.62, 117.79, 110.81, 107.91, 105.21, 100.00, 67.88, 66.32, 56.36, 50.17, 30.61; HRMS (ESI-TOF) m/z [M + H]+ calcd for C22 H23 N5 O3 S 438.1594, found 438.1605.

Compound LHB-5: 4-(6-(2-(benzo[d]thiazol-2-ylthio)ethoxy)-7-methoxyquinazolin-4-yl)morpholine

Compound **LHB-5** (277 mg, Yield: 61%) as a pale-yellow solid was synthesized using the same procedure of preparing compound **LHB-4** but with 2-((2-bromoethyl)thio)benzo[d]thiazole as the starting material; ^{1}H NMR (400 MHz, $CDCl_3$) δ 8.68 (s, 1H), 7.85 (d, *J* = 8.0 Hz, 1H), 7.77 (d, *J* = 7.7 Hz, 1H), 7.50- 7.38 (m, 1H), 7.37 – 7.29 (m, 1H), 7.27 (s, 1H), 7.22 (s, 1H), 4.55 (t, *J* = 6.5 Hz, 2H), 3.97 (s, 3H), 3.87 (t, *J* = 6.5 Hz, 2H), 3.84 - 3.73 (m, 4H), 3.73- 3.50 (m, 4H); ^{13}C NMR (101 MHz, $CDCl_3$) δ 165.59, 163.79, 155.08, 153.24, 153.00, 149.60, 147.37, 135.39, 126.16, 124.52, 121.52, 121.09, 111.23, 108.00, 105.84, 68.07, 66.60, 56.15, 50.24, 31.93. HRMS (ESI-TOF) m/z [M + H]+ calcd for C22 H22 N4 O3 S2 455.1206, found 455.1214.

Compound LHB-6: N-(2-((7-methoxy-4-morpholinoquinazolin-6-yl)oxy)ethyl)benzo[d]thiazol-2-amine

Compound **LHB-6** (166 mg, Yield: 38%) as a white solid was synthesized using the same

procedure of preparing compound **LHB-4** but with N-(2-bromoethyl)benzo[d]thiazol-2-amine as the starting material. ^{1}H NMR (400 MHz, $CDCl_3$) δ 8.67 (s, 1H), 7.58 (t, *J* = 7.5 Hz, 2H), 7.31 (t, *J* = 7.9 Hz, 1H), 7.27 (s, 1H), 7.15 (s, 1H), 7.11 (t, *J* = 7.3 Hz, 1H), 6.02 (br, 1H), 4.35 (t, *J* = 4.9 Hz, 2H), 4.02 (t, *J* = 4.9 Hz, 2H), 3.99 (s, 3H), 3.84 (m, 4H), 3.69 – 3.58 (m, 4H); ^{13}C NMR (101 MHz, $CDCl_3$) δ 166.63, 163.81, 154.85, 153.26, 152.28, 149.44, 147.50, 130.52, 126.08, 121.98, 120.86, 119.07, 111.26, 107.90, 105.54, 68.07, 66.66, 56.16, 50.22, 44.10. HRMS (ESI-TOF) m/z [M + H]+ calcd for C22 H23 N5 O3 S 438.1594, found 438.1606.

Compound LHB-7: N-(2-((7-methoxy-4-(4-methylpiperazin-1-yl)quinazolin-6-yl)oxy)ethyl)benzo[d]thiazol-2-amine

Compound **LHB-7** (234 mg, Yield: 52 %) as a white solid was synthesized using the same procedure of preparing compound **LHB-4** but with N-(2-bromoethyl)benzo[d]thiazol-2-amine and compound **4b** as the starting material. ^{1}H NMR (400 MHz, $CDCl_3$) δ 8.65 (s, 1H), 7.58 (t, *J* = 7.7 Hz, 2H), 7.31 (t, *J* = 7.7 Hz, 1H), 7.25 (s, 1H), 7.17 (s, 1H), 7.11 (t, *J* = 7.6 Hz, 1H), 5.89 (s, 1H), 4.35 (t, *J* = 4.9 Hz, 2H), 4.03 – 3.99 (m, 5H), 3.69 – 3.63 (m, 4H), 2.60 – 2.55 (m, 4H), 2.35 (s, 3H); ^{13}C NMR (126 MHz, $CDCl_3$) δ 166.60, 163.75, 154.80, 153.28, 152.31, 149.43, 147.31, 130.57, 126.04, 121.95, 120.84, 119.11, 111.26, 107.86, 105.98, 68.14, 56.12, 54.72, 49.48, 46.00, 44.11; HRMS (ESI-TOF) m/z [M + H]+ calcd for C23 H26 N6 O2 S F 451.1911, found 451.1910.

Compound LHB-8: (6-(2-((1H-benzo[d]imidazol-2-yl)thio)ethoxy)-7-methoxyquinazolin-4-yl)(4-methoxyphenyl)methanone

Compound **LHB-8** (41 mg, Yield: 43%) as a white solid was synthesized using the same procedure of preparing compound **LHB-4** but with compound **8** as the starting material. ^{1}H NMR (400 MHz, $CDCl_3$) δ 9.27 (s, 1H), 8.05 – 7.93 (m, 2H), 7.56 (dd, *J* = 5.3, 2.9 Hz, 2H), 7.48 (s, 1H), 7.45 (s,

1H), 7.27 – 7.21 (m, 2H), 7.06 – 6.94 (m, 2H), 4.49 (t, *J* = 5.4 Hz, 2H), 4.12 (s, 3H), 3.91 (s, 3H), 3.73 – 3.57 (t, *J* = 5.4 Hz, 2H); ^{13}C NMR (101 MHz, $CDCl_3$) δ 191.98, 164.71, 160.66, 156.14, 152.83, 149.89, 149.73, 149.59, 133.37, 128.29, 122.58, 118.19, 114.07, 107.51, 104.06, 69.13, 56.86, 55.68, 32.59. HRMS (ESI-TOF) m/z [M + H]+ calcd for C26 H22 N4 O4 S 487.1435, found 487.1442.

Compound LHB-9: (6-(2-(benzo[d]thiazol-2-ylthio)ethoxy)-7-methoxyquinazolin-4-yl)(4-methoxyphenyl)methanone

Compound **LHB-9** (41 mg, Yield: 41%) as a pale-yellow solid was synthesized using the same procedure of preparing compound **LHB-4** but with compound **8** and 2-((2-bromoethyl)thio)benzo[d]thiazole as the starting material. ^{1}H NMR (400 MHz, $CDCl_3$) δ 9.21 (s, 1H), 7.99 – 7.89 (m, 2H), 7.87 – 7.80 (m, 1H), 7.78 – 7.70 (m, 1H), 7.45 – 7.39 (m, 1H), 7.38 (s, 1H), 7.37 (s, 1H), 7.33 – 7.27 (m, 1H), 7.01 – 6.89 (m, 2H), 4.49 (t, *J* = 6.4 Hz, 2H), 4.00 (s, 3H), 3.87 (s, 3H), 3.84 (t, *J* = 6.4 Hz, 2H); ^{13}C NMR (101 MHz, $CDCl_3$) δ 198.77, 192.04, 168.17, 164.56, 159.28, 156.77, 153.04, 152.61, 149.93, 135.43, 133.28, 128.48, 126.04, 124.37, 121.69, 120.99, 118.15, 113.99, 107.12, 104.00, 67.77, 56.40, 55.61, 31.71; HRMS (ESI-TOF) m/z [M + H]+ calcd for C26 H21 N3 O4 S2 504.1046, found 504.1038.

Compound LHB-10: (6-(2-(benzo[d]thiazol-2-ylamino)ethoxy)-7-methoxyquinazolin-4-yl)(4-methoxyphenyl)methanone

Compound **LHB-10** (44 mg, Yield: 45%) as a pale-yellow solid was synthesized using the same procedure of preparing compound **LHB-4** but with compound **8** and N-(2-bromoethyl)benzo[d]thiazol-2-amine as the starting material. ^{1}H NMR (400 MHz, $CDCl_3$) δ 9.21 (s, 1H), 7.94 (d, *J* = 8.8 Hz, 2H), 7.58 – 7.50 (m, 2H), 7.39 (s, 1H), 7.35 (s, 1H), 7.32 – 7.22 (m, 1H), 7.08 (t, *J* = 7.6 Hz, 1H), 6.95 (d, *J* = 8.8 Hz, 2H), 4.31 (t, *J* = 4.9 Hz, 2H), 4.04 (s, 3H), 3.96 (t, *J* = 4.8 Hz,

2H), 3.87 (s, 3H); ^{13}C NMR (101 MHz, $CDCl_3$) δ 191.98, 166.57, 164.65, 160.60, 156.69, 152.71, 152.32, 150.20, 149.92, 133.30, 130.62, 128.43, 125.99, 121.92, 120.80, 119.22, 118.13, 114.05, 107.18, 104.27, 67.84, 56.40, 55.64, 43.93; HRMS (ESI-TOF) m/z [M + H]+ calcd for C26 H22 N4 O4 S 487.1435, found 487.1444.

General strategy of the GA-FEP/ABFE method. To calculate ABFE of a ligand with a receptor, the whole ligand molecule must be annihilated during the FEP calculation and, thus, the perturbation is expected to be generally greater than that for the usual RBFE calculation. The greater perturbation could lead to a problem for the convergence of the calculated ABFE, making the usual FEP procedure inappropriate for the practical ABFE calculations. The numerical integration accuracy, which is directly associated with sampled probability distribution, must be improved for the FEP method to be suitable for the ABFE calculations. In general, there are several computational algorithms developed to improve numerical integration accuracy in both mathematics and physics. The commonly used algorithms for numerical analysis, such as the Shanks transformation and Richardson extrapolation, are based on fitting the hypothesized property of a sequence by some specific functions, and could greatly increase the rate of convergence of a sequence.[35, 36] Hummer *et al*. also used a multistate Gaussian model to fit the probability distribution of electrostatic interactions, and successfully increased the accuracy of electrostatic solvation free energy prediction.[37] Inspired by these works, and considering the Gaussian-like distribution property of the sampling results $P(\Delta U)$ during the FEP calculations, probability distribution of ΔU, *i.e.* $P(\Delta U)$, was modelled as a superposition of a series of Gaussian functions, and the poorly sampled part of $P(\Delta U)$ could be refined by extrapolation. This method is also referred to as Gaussian algorithm enhanced free energy perturbation (GA-FEP) protocol. Based on the modeled probability distribution, Bennett acceptance ratio (BAR) method was used to

further increase the accuracy of calculations. Seven types of drug targets and over 100 ligands were selected to test the GA-FEP method. After fitting probability distributions by Gaussian functions, the convergence of the FEP calculations was greatly improved, and the result also showed good accuracy for all the tested ligands, with or without similar scaffolds. For ligands with similar scaffolds, the ABFE results could even reach to the similar accuracy to the RBFE results.

The GA-FEP method. In order to calculate the binding free energy of a protein-ligand system by using alchemical transformation, a thermodynamic cycle was designed as Figure 1, where Rec represents the receptor, Lig represents the ligand, and Dumm represents the dummy ligand after annihilation ("annihilation" means all atoms of the ligand vanish and become dummy atoms with zero charge and vdW interactions with other atoms, and surrounding solvent or receptor cannot "feel" the existence of the ligand just like the ligand is "annihilated"). In order to evaluate $\Delta A_{\mathrm{binding}}$, a series of non-physical states that connect the initial reference state (all atoms in this state were described by the force field) and the final target state (ligand atoms were annihilated) were added. According to the thermodynamic cycle, $\Delta A_{\mathrm{binding}}$ can be calculated as

$$\Delta A_{\mathrm{binding}} = \Delta A^{\mathrm{L}}_{\mathrm{annihilation}} - \Delta A^{\mathrm{RL}}_{\mathrm{annihilation}} \quad (1)$$

As a commonly used practice in most FEP calculations, $\Delta A_{\mathrm{restrain}}$ in the lower horizontal transformation, which represents the changes of translational and rotational entropies during the binding process, was omitted due to the existing difficulties in calculating entropies. The final result was not affected too much after $\Delta A_{\mathrm{restrain}}$ was omitted due to the error cancellation, $\Delta A^{\mathrm{RL}}_{\mathrm{annihilation}}$ was calculated by annihilating the ligand in the RL complex, and $\Delta A^{\mathrm{L}}_{\mathrm{annihilation}}$ was calculated by just annihilating ligand in water, because the energy differences between receptors may roughly be cancelled out. For both the $\Delta A^{\mathrm{L}}_{\mathrm{annihilation}}$ and $\Delta A^{\mathrm{RL}}_{\mathrm{annihilation}}$ calculations, the annihilations of

electrostatic and vdW interactions were decoupled. A set of five alchemical states were first used to annihilate electrostatic interactions followed by another set of five states to annihilate the vdW interactions. Based on the FEP theory, when the mass of the system does not change during the whole alchemical transformation, the final $\Delta A^{\mathrm{L}}_{\mathrm{annihilation}}$ and $\Delta A^{\mathrm{RL}}_{\mathrm{annihilation}}$ could be calculated by the following equations (2) and (3).[38]

$$\Delta A = -\frac{1}{\beta}\sum_{i=1}^{N-1} \ln\langle \exp[-\beta\, \Delta U_{i,i+1}] \rangle_i \,, \qquad (2)$$

$$\Delta A = -\frac{1}{\beta}\sum_{i=1}^{N-1} \ln \int \exp[-\beta\, \Delta U_{i,i+1}] P_i[\Delta U_{i,i+1}]\, \mathrm{d}\Delta U_{i,i+1}, \qquad (3)$$

where *N* represents the total number of states in the alchemical transformation and *i* refers to the *i*th state. Equation (3) is the integral form of equation (2). According to these equations, the free energy difference ΔA between two neighboring states (*i* and *i+1*) may be calculated based on the trajectory of the *i*th state.

The above GA-FEP method was implemented in a local version of the AMBER 12 program,[39-41] and the implementation enabled to perform the FEP simulations under the GPU acceleration. In the input preparation step, the force field parameters for the initial reference state (such as Lig or Rec-Lig state in Figure 1) were generated first. With the Rec-Lig state as an example (we used the original version of GAFF for the ligand in the reference state): the partial atomic charge of the ligand was calculated by using the Gaussian 03 program[42] at the HF/6-31G* level. The Antechamber module of the AMBER 12 program[43] was then used to obtain the RESP charges and assign the general AMBER force field (GAFF), and the Amber03 force field was used for the proteins. According to the net charge of the receptor-ligand complex, counter ions (either Na^+ or Cl^-) were used to neutralize the system.

Truncated octahedral box of simple point charge/flexible water (SPC/Fw) [44] molecules was then added with the surface of the box at least 5 Å away from any atom in the receptor-ligand complex, and the radius of the boxes are at least 12 Å. For the Lig state, the preparation method was very similar to that mentioned above, except that the size of the truncated octahedral water box was set to 8 Å.

The overall alchemical transformation of either Rec-Lig system or Lig system contains 10 alchemical states. Force field parameters of all the non-physical intermediate states were generated based on the reference state. To annihilate the charge interaction, the partial atomic charge of each atom in the ligand was linearly changed to zero in the first five states with the vdW parameters unchanged. In the next five states, the vdW interaction was gradually annihilated. Considering the vdW interaction is calculated by using the Lennard-Jones potential,

$$V_{LJ} = 4\varepsilon_{ij}\left(\frac{r_{ij}}{r}\right)^{12} - 2\left(\frac{r_{ij}}{r}\right)^{6}, \quad (4)$$

to prevent the 'origin singularity' effect, the ε of every atom decreases as a $(1\text{-}0.2\lambda_i)^6$ manner, where λ_i represents the λ value of the ith state in the annihilation of vdW interaction.[45-47] After all the input files were prepared, the MD simulations were performed. Since the trajectories could become very unstable when the SHAKE constrains[48, 49] were applied during the annihilation steps, the SHAKE algorithm was not used in the MD simulations, and the time step was set to 1 fs. Along with the periodic boundary condition (PBC), the particle mesh Ewald (PME) method[50] was used to calculate the long-range electrostatic interactions. During the MD simulations, weak restrains were applied to all the ligands in the Lig systems with a force constant of 10 kcal/(mol·Å^2). Unless explicitly mentioned otherwise, 4 ns MD simulations were performed for each alchemical state, and the snapshots from the last 2 ns were saved into production MD trajectory files with an interval of 100 fs for post-simulation processing.

Energy calculations were carried out in the post-simulation processing steps. Based on the trajectories of the last 2 ns and the force field parameter files of the forward/backward step, the forward/backward energy difference of each snapshot was calculated. To improve the accuracy of energy calculation, some enhanced sampling methods were used here instead of directly using equation (2) or (3). In the FEP calculations, as the probability distribution of ΔU, *i.e.* $P(\Delta U)$, is weighted by $\exp(-\beta\,\Delta U_{i+1,i})$ as seen in equation (3), the negative $-\Delta U$ tail could remarkably influence the energy calculation, and the main part of ΔU, however, is less important to the final result.[51] To fully use the knowledge of the whole distribution $P(\Delta U)$, and improve the accuracy of the sampling of the energy near the negative $-\Delta U$ tail, the overall distribution was fitted by 5 Gaussian functions, as shown in equation (5) with a total of 15 different parameters to be fitted.

$$P(\Delta U) = \sum_{i=1}^{5} c_i \exp\left[-\frac{(\Delta U - \mu_i)^2}{2{\sigma_i}^2}\right] \qquad (5)$$

Least-square method was used to optimize all these 15 parameters to get the fitted function, and the poorly sampled negative $-\Delta U$ tail was refined by the extrapolation. The details of the fitting procedure are given in Supporting Information Section S8

Based on the fitted distribution functions, the results were further improved by the Bennett acceptance ratio (BAR) method[52, 53] which combines the forward and backward transformations. According to the BAR method, equations (6) and (7) were used to calculate energy differences between adjacent alchemical states:

$$\exp\left[-\beta\,\Delta A_{i,i+1}\right] = \frac{\langle\left[1+\exp\left[\beta\left(\Delta U_{i,i+1} - C\right)\right]\right]^{-1}\rangle_i}{\langle\left[1+\exp\left[-\beta\left(\Delta U_{i,i+1} - C\right)\right]\right]^{-1}\rangle_{i+1}}\exp(-\beta C), \qquad (6)$$

$$C = \Delta A_{i,i+1} + \frac{1}{\beta}\ln\frac{n_i}{n_{i+1}}. \quad (7)$$

The BAR method will insert an additional intermediate state between the *i*th and (*i*+*1*)th state. The constant *C* determines the position of the additional intermediate state, and n_i and n_{i+1} are sample sizes collected from states i and $i+1$. Combining the information from both forward and backward calculations, the accuracy of the calculation can be improved further. Except for the final state, the energy differences for all of the other states were calculated by using the BAR method. In the final state, all the surrounding atoms cannot feel the existence of the dummy ligand and, thus, can occupy the original position of the ligand. When doing backward calculation based on the final state, the ligand atoms will reappear, and can clash against the surrounding atoms, which makes inaccurate prediction for the backward energy. This technical problem is also known as the end-point catastrophes[54]. Thus, the energy of the final state was just calculated by the forward energy calculation based on the trajectory of the second last state. Then, based on the sum of all the energies of these states, we can obtain $\Delta A^{\mathrm{L}}_{\mathrm{annihilation}}$ and $\Delta A^{\mathrm{RL}}_{\mathrm{annihilation}}$, and the final binding free energy can be calculated by using equation (1).

Due to very small vdW interactions between a ligand and surrounding atoms in the second last state, water molecules cannot fully occupy the binding site of the receptor, thus the hydration energy of the receptor might not be sufficiently accounted for. According to the chemical environment of the binding site, the energy calculation results of a same receptor may systematically shift to a same direction due to the insufficient simulation of the hydration energy of the receptor binding site, and the systematical shift may reflect the magnitude of the hydration energy of the receptor binding site.

A new intermediate state generation method for the GA-FEP calculation. In common FEP practices, there are mainly two approaches for generating alchemical states. One is the single-topology

paradigm[55] in which a common topology is used for both reference and target state. According to the single-topology paradigm, after the force field parameters for the initial reference state and the final target state are prepared, the force field parameters for all the intermediate non-physical states are expressed as linear combinations of those for the reference and target states. Another approach is known as the dual-topology paradigm[56], in which the topologies of reference and target states coexist during the alchemical transformation, and the potential energy for each intermediate state is a linear combination of those of the reference and target states.

The intermediate state parameter generation method used here is different from the commonly used two approaches described above. In this study, intermediate state parameters were not described as combinations of the reference and target states. Instead, the force field parameters of each intermediate state were generated individually, and each intermediate corresponds to a single force field parameter file. Thus, reading two force field parameter files in one MD simulation, which is a common practice in the FEP calculation, is no longer needed by using our new protocol, and the MD simulations on all the intermediate states became the usual MD simulations. By using this method, the force field parameters of all the intermediate states can be customized and, more importantly, the speed advantage of the AMBER under the GPU acceleration can be fully exerted.

Eight Nvidia Geforce GTX580 GPUs were used in this study. For the seven targets used in this study, this GA-FEP method can predict the binding free energies for 2-3 ligands per day, which is an acceptable speed to predict the ABFE for the purpose of practical structure-based drug design.

FEP-based lead optimization targeting PDE10. Phosphodiesterase-10 (PDE10) is known as a promising drug target for treatment of colon cancer[23] and central nervous system disorders such as schizophrenia and Huntington's disease.[24] Despite extensive efforts in development of PDE10

inhibitors, there is still no PDE10 inhibitor approved for the desirable clinical use.[25] Structurally diverse and potent PDE10 inhibitors are urgently needed in order to accelerate the development of PDE10 inhibitors. In order to discover structurally diverse and highly potent PDE10 inhibitors, as well as externally validate the accuracy and efficiency of our GA-FEP protocol, lead optimization starting from a small molecule **LHB-1** with IC_{50} = 1.8 μM (Figure 6a) against PDE10 were carried out by FEP-based rational design, chemical synthesis, and bioassay validation. Starting from the hit **LHB-1**, two series of molecules with different kinds of scaffolds were designed in light of the computational modeling. All of the designed molecules were docked into the binding site of PDE10 (PDB ID: 2O8H), and their ABFE values were then predicted by our GA-FEP protocol. Based on their ABFE results, nine molecules were selected for further chemical synthesis and bioassay studies. Details about the designed molecules are provided in Supporting Information Section S6.

To validate the predicted binding mode, an X-ray cocrystal structure of PDE10 with bound **LHB-6** was determined. The experimental procedures of the subcloning and protein expression of human PDE10A2 (catalytic domain, residues 449-770) were similar to those described previously,[57] and the crystals of PDE10A2 were grown by hanging drop. The unliganded PDE10A2 (10 mg/ml in a buffer of 50 mM NaCl, 20 mM Tris-HCl (pH 7.5), 1 mM β-mercaptoethanol and 1 mM EDTA) was vapor-diffused against the well buffer of 0.1 M Hepes (pH 7.5), 0.2 M $MgCl_2$, 18% PEG3350 and 50 mM 2-mercaptoethanol. The complex of PDE10A2 with **LHB-6** was prepared by soaking the unliganded crystals in 20 mM **LHB-6** in a buffer of 16% PEG 8000, 0.1 M Hepes (pH 7.5), 0.1 M $MgCl_2$ and 60 mM 2-ME at 4°C for 24 h. The crystallization buffer plus 20% ethylene glycolwas used as the cryosolvent. Diffraction data were collected at 100 K on an in-house Oxford Diffraction Xcalibur Nova diffractometer. The data were processed using the program *CrysAlis Pro* and the structure was solved

and refined by using the CCP4.[58] The coordinates and structure factors have been deposited in the Protein Data Bank with PDB ID **5ZNL**.

ACKNOWLEDGEMENTS

This work was supported by the National Key R&D Program of China (2017YFB0202600), Natural Science Foundation of China (21572279, 21877134, 81522041, and 21402243), Science Foundation of Guangdong Province (2018A030313215 and 2014A020210009), Fundamental Research Funds for the Central Universities (Sun Yat-Sen University, 17ykjc03 and 18ykpy23), Guangdong Province Higher Vocational Colleges & Schools Pearl River Scholar Funded Scheme (2016), and the National Science Foundation (NSF, grant CHE-1111761). We cordially thank Prof. H. Ke from Department of Biochemistry and Biophysics at the University of North Carolina, Chapel Hill, for his help with PDEs.

AUTHOR INFORMATION

Corresponding Authors

* Tel: +1-859-323-3943 or +86-20-39943031. Fax: +1-859-257-7585 or +86-20-39943000. E-mail: zhan@uky.edu (C.-G. Zhan) or luohb77@mail.sysu.edu.cn (H.-B. Luo).

Author Contributions

# These authors contributed equally to this study.

DECLARATION OF INTERESTS

The authors declare no competing financial interest.

Supporting Information Availability

PDE4, PDE5, and PDE9 inhibitors used in this study; an example to show how the GA-FEP

calculations work; use of 10 lambda windows reaching to similar accuracy with that of 20 or 39 lambda windows; comparison between the ABFE and RBFE for the same data set; designed PDE10 inhibitors; diffraction data and structure refinement statistics for PDE10A-LHB-6 structure; the calculation results of all ligands; ^{1}H NMR, ^{13}C NMR, High-resolution mass spectra (HRMS) and purity data for representative compounds; HPLC spectrum for the purity of representative target compounds; and Molecular Formula Strings. This material is available free of charge *via* the Internet at http://pubs.acs.org.

PDB ID Code

The atomic coordinates and structure factors have been deposited into the RCSB Protein Data Bank with accession number 5ZNL. Authors will release the atomic coordinates and experimental data upon article publication.

Abbreviations Used

FEP: Free energy perturbation; GA-FEP: Gaussian algorithm enhanced free energy perturbation; ABFE: Absolute binding free energy; RBFE: Relative binding free energy; MM-PBSA: Molecular mechanics/Poisson-Boltzmann and surface area continuum solvation; MM-GBSA: Molecular mechanics/generalized Born and surface area continuum solvation; RESP: Restricted electrostatic potential; GAFF: Generalized Amber force field; BAR: Bennett acceptance ratio; REST: replica exchange with solute tempering; JNK: c-Jun N-terminal kinase; TYK: Tyrosine kinase

REFERENCES

1. Wong, C.F.; McCammon, J.A. Dynamics and Design of Enzymes and Inhibitors. *J. Am. Chem. Soc.* **1986**, *108*, 3830-3832.

2. McCammon, J.A. Computer-aided Molecular Design. *Science* **1987**, *238*, 486-491.

3. Gilson, M.K.; Given, J.A.; Bush, B.L.; McCammon, J.A. The Statistical-Thermodynamic Basis for Computation of Binding Affinities: A Critical Review. *Biophys. J.* **1997**, *72*, 1047-1069.

4. Wereszczynski, J.; McCammon, J.A. Statistical Mechanics and Molecular Dynamics in Evaluating Thermodynamic Properties of Biomolecular Recognition. *Q. Rev. Biophys.* **2012**, *45*, 1-25.

5. Deng, Y.; Roux, B. Computations of Standard Binding Free Energies with Molecular Dynamics Simulations. *J. Phys. Chem. B* **2009**, *113*, 2234-2246.

6. Christ, C.D.; Mark, A.E.; van Gunsteren, W.F. Basic Ingredients of Free Energy Calculations: A Review. *J. Comput. Chem.* **2010**, *31*, 1569-1582.

7. Chodera, J.D.; Mobley, D.L.; Shirts, M.R.; Dixon, R.W.; Branson, K.; Pande, V.S. Alchemical Free Energy Methods for Drug Discovery: Progress and Challenges. *Curr. Opin. Struct. Biol.* **2011**, *21*, 150-160.

8. Kolář, M.H.; Hobza, P. Computer Modeling of Halogen Bonds and Other σ-hole Interactions. *Chem. Rev.* **2016**, *116*, 5155-5187.

9. Jorgensen, W.L. Efficient Drug Lead Discovery and Optimization. *Acc. Chem. Res.* **2009**, *42*, 724-733.

10. Chodera, J.D.; Mobley, D.L.; Shirts, M.R.; Dixon, R.W.; Branson, K.; Pande, V.S. Alchemical Free Energy Methods for Drug Discovery: Progress and Challenges. *Curr. Opin. Struct. Biol.* **2011**, *21*, 150-160.

11. Mauser, H.; Guba, W. Recent Developments in de novo Design and Scaffold Hopping. *Curr. Opin. Drug Discov.* **2008**, *11*, 365-374.

12. Schneider, G. Virtual Screening: An Endless Staircase? *Nat. Rev. Drug Discov.* **2010**, *9*, 273-276.

13. Cournia, Z.; Allen, B.; Sherman, W. Relative Binding Free Energy Calculations in Drug Discovery: Recent Advances and Practical Considerations. *J. Chem. Inf. Model.* **2017***, 57*, 2911-2937.

14. 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*, 2695-2703.

15. 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. *J. Chem. Inf. Model.* **2015**, *55*, 2411-2420.

16. 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*, 5108-5112.

17. 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*, 902-919.

18. Briggs, J.M.; Marrone, T.J.; McCammon, J.A. Computational Science New Horizons and Relevance to Pharmaceutical Design. *Trends Cardiovas. Med.* **1996**, *6*, 198-203.

19. 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*, 15372-15373.

20. Jorgensen, W.L. Computer-aided Discovery of Anti-HIV Agents. *Bioorg. Med. Chem.* **2016**, *24*, 4768-4778.

21. Abel, R.; Wang, L.; Harder, E.D.; Berne, B.; Friesner, R.A. Advancing Drug Discovery Through Enhanced Free Energy Calculations. *Acc. Chem. Res.* **2017**, *50*, 1625-1632.

22. Mobley, D.L.; Dill, K.A. Binding of Small-Molecule Ligands to Proteins: "What You See" Is Not Always "What You Get". *Structure* **2009**, *17*, 489-498.

23. Li, N.; Lee, K.; Xi, Y.; Zhu, B.; Gary, B.D.; Ramírez-Alcántara, V.; Gurpinar, E.; Canzoneri, J.C.; Fajardo, A.; Sigler, S. Phosphodiesterase 10A: A Novel Target for Selective Inhibition of Colon Tumor Cell Growth and β-catenin-dependent TCF Transcriptional Activity. *Oncogene* **2015**, *34*, 1499-1509.

24. Hebb, A.L.; Robertson, H.A. Role of Phosphodiesterases in Neurological and Psychiatric Disease. *Curr. Opin. Pharmacol.* **2007**, *7*, 86-92.

25. Chappie, T.A.; Helal, C.J.; Hou, X. Current Landscape of Phosphodiesterase 10A (PDE10A) Inhibition. *J. Med. Chem.* **2012**, *55*, 7299-7331.

26. Sawasdee, P.; Wacharasindhu, S.; Uetrongchit, Y. Quinazoline Derivatives with Anti-cholinesterase Activity. *Reports of Research Assisted by the Asahi Glass Foundation*. **2014**, 1-13.

27. Davies, T.G.; Bentley, J.; Arris, C.E.; Boyle, F.T.; Curtin, N.J.; Endicott, J.A.; Gibson, A.E.; Golding, B.T.; Griffin, R.J.; Hardcastle, I.R.; Jewsbury, P.; Johnson, L.N.; Mesguiche, V.; Newell, D.R.; Noble, M.E.; Tucker, J.A.; Wang, L.; Whitfield, H.J. Structure-Based Design of a Potent purine-based Cyclin-dependent Kinase Inhibitor. *Nat. Struct. Biol.* **2002**, *9*, 745-749.

28. Szczepankiewicz, B.G.; Kosogof, C.; Nelson, L.T.; Liu, G.; Liu, B.; Zhao, H.; Serby, M.D.; Xin, Z.; Liu, M.; Gum, R.J.; Haasch, D.L.; Wang, S.; Clampit, J.E.; Johnson, E.F.; Lubben, T.H.; Stashko, M.A.; Olejniczak, E.T.; Sun, C.; Dorwin, S.A.; Haskins, K.; Abad-Zapatero, C.; Fry, E.H.; Hutchins, C.W.; Sham, H.L.; Rondinone, C.M.; Trevillyan, J.M. Aminopyridine-based c-Jun N-terminal Kinase Inhibitors with Cellular Activity and Minimal Cross-

kinase Activity. *J. Med. Chem.* **2006**, *49*, 3563-3580.

29. Liang, J.; Tsui, V.; Van Abbema, A.; Bao, L.; Barrett, K.; Beresini, M.; Berezhkovskiy, L.; Blair, W.S.; Chang, C.; Driscoll, J.; Eigenbrot, C.; Ghilardi, N.; Gibbons, P.; Halladay, J.; Johnson, A.; Kohli, P.B.; Lai, Y.; Liimatta, M.; Mantik, P.; Menghrajani, K.; Murray, J.; Sambrone, A.; Xiao, Y.; Shia, S.; Shin, Y.; Smith, J.; Sohn, S.; Stanley, M.; Ultsch, M.; Zhang, B.; Wu, L.C.; Magnuson, S. Lead Identification of Novel and Selective TYK2 Inhibitors. *Eur. J. Med. Chem.* **2013**, *67*, 175-187.

30. Huang, Y.Y.; Li, Z.; Cai, Y.H.; Feng, L.J.; Wu, Y.; Li, X.; Luo, H.-B. The Molecular Basis for the Selectivity of Tadalafil toward Phosphodiesterase 5 and 6: A Modeling Study. *J. Chem. Inf. Model.* **2013**, *53*, 3044-3053.

31. A) Meng, F.; Hou, J.; Shao, Y.X.; Wu, P.Y.; Huang, M.; Zhu, X.; Cai, Y.; Li, Z.; Xu, J.; Liu, P.; Luo, H.-B.; Wan, Y.; Ke, H. Structure-based Discovery of Highly Selective Phosphodiesterase-9A Inhibitors and Implications for Inhibitor Design. *J. Med. Chem.* **2012**, *55*, 8549-8558. B) Li, Z.; Lu, X.; Feng, L.J.; Gu, Y.; Li, X.; Wu, Y.; Luo, H.-B. Molecular Dynamics-Based Discovery of Novel Phosphodiesterase-9A Inhibitors with Non-pyrazolopyrimidinone Scaffolds. *Mol. Biosyst.* **2015**, *11*, 115-125.

32. Gilson, M.K.; Liu, T.; Baitaluk, M.; Nicola, G.; Hwang, L.; Chong, J. BindingDB in 2015: A Public Database for Medicinal Chemistry, Computational Chemistry and Systems Pharmacology. *Nucleic Acids Res.* **2016**, *44*, D1045-1053.

33. Shao, Y.X.; Huang, M.; Cui, W.; Feng, L.J.; Wu, Y.; Cai, Y.; Li, Z.; Zhu, X.; Liu, P.; Wan, Y.; Ke, H.; Luo, H.-B. Discovery of a Phosphodiesterase 9A Inhibitor as a Potential Hypoglycemic Agent. *J. Med. Chem.* **2014**, *57*, 10304-10313.

34. Baell, J.B.; Holloway, G.A. New Substructure Filters for Removal of Pan Assay Interference Compounds (PAINS) from Screening Libraries and for Their Exclusion in Bioassays. *J. Med. Chem.* **2010**, *53*, 2719-2740.

35. Shanks, D. Non−linear Transformations of Divergent and Slowly Convergent Sequences. *Stud. Appl. Math.* **1955**, *34*, 1-42.

36. Richardson, L.F. The Approximate Arithmetical Solution by Finite Differences of Physical Problems Involving Differential Equations, with an Application to the Stresses in a Masonry Dam. *Phil. Trans. R. Soc. Lond. A* **1911**, *210*, 307-357.

37. Hummer, G.; Pratt, L.R.; García, A.E. Multistate Gaussian Model for Electrostatic Solvation Free Energies. *J. Am. Chem. Soc.* **1997**, *119*, 8523-8527.

38. Zwanzig, R.W. High‐Temperature Equation of State by a Perturbation Method. I. Nonpolar Gases. *J. Chem. Phys.* **1954**, *22*, 1420-1426.

39. Salomon-Ferrer, R.; Gotz, A.W.; Poole, D.; Le Grand, S.; Walker, R.C. Routine Microsecond Molecular Dynamics Simulations with AMBER on GPUs. 2. Explicit Solvent Particle Mesh Ewald. *J. Chem. Theory Comput.* **2013**, *9*, 3878-3888.

40. Gotz, A.W.; Williamson, M.J.; Xu, D.; Poole, D.; Le Grand, S.; Walker, R.C. Routine Microsecond Molecular Dynamics Simulations with AMBER on GPUs. 1. Generalized Born. *J. Chem. Theory Comput.* **2012**, *8*, 1542-1555.

41. Le Grand, S.; Götz, A.W.; Walker, R.C. SPFP: Speed Without Compromise—A Mixed Precision Model for GPU Accelerated Molecular Dynamics Simulations. *Comput. Phys. Commun.* **2013**, *184*, 374-380.

42. Frisch, M.J.; Trucks, G.W.; Schlegel, H.B.; Scuseria, G.E.; Robb, M.A.; Cheeseman, J.R.; Montgomery, J.A.; Vreven, T.; Kudin, K.N.; Burant, J.C.; Millam, J.M.; Iyengar, S.S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, G.A.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J.E.; Hratchian, H.P.; Cross, J.B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R.E.; Yazyev, O.; Austin, A.J.; Cammi, R.; Pomelli, C.; Ochterski, J.W.; Ayala, P.Y.; Morokuma, K.; Voth, G.A.; Salvador, P.; Dannenberg, J.J.; Zakrzewski, V.G.; Dapprich, S.; Daniels, A.D.; Strain, M.C.; Farkas, O.; Malick, D.K.; Rabuck, A.D.; Raghavachari, K.; Foresman, J.B.; Ortiz, J.V.; Cui, Q.; Baboul, A.G.; Clifford, S.; Cioslowski, J.; Stefanov, B.B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R.L.; Fox, D.J.; Keith, T.; Al-Laham, M.A.; Peng, C.Y.; Nanayakkara, A.; Challacombe, M.; Gill, P.M.W.; Johnson, B.; Chen, W.; Wong, M.W.; Gonzalez, C.; Pople, J.A. Gaussian 03, Revision D. 01. *Gaussian, Inc., Wallingford CT*. **2004**.

43. Wang, J.; Wang, W.; Kollman, P.A.; Case, D.A. Automatic Atom Type and Bond Type Perception in Molecular Mechanical Calculations. *J. Mol. Graph. Model.* **2006**, *25*, 247-260.

44. Wu, Y.; Tepper, H.L.; Voth, G.A. Flexible Simple Point-charge Water Model with Improved Liquid-state Properties. *J. Chem. Phys.* **2006**, *124*, 024503.

45. Steinbrecher, T.; Mobley, D.L.; Case, D.A. Nonlinear Scaling Schemes for Lennard-Jones Interactions in Free Energy Calculations. *J. Chem. Phys.* **2007**, *127*, 214108.

46. Steinbrecher, T.; Hrenn, A.; Dormann, K.L.; Merfort, I.; Labahn, A. Bornyl (3, 4, 5-trihydroxy)-cinnamate An Optimized Human Neutrophil Elastase Inhibitor Designed by Free Energy Calculations. *Bioorg. Med. Chem.* **2008**, *16*, 2385-2390.

47. Simonson, T. Free Energy of Particle Insertion: An Exact Analysis of the Origin Singularity for Simple Liquids. *Mol. Phys.* **1993**, *80*, 441-447.

48. Ryckaert, J.-P.; Ciccotti, G.; Berendsen, H.J.C. Numerical Integration of the Cartesian Equations of Motion of a

System with Constraints: Molecular Dynamics of N-alkanes. *J. Comput. Phys.* **1977**, *23*, 327-341.

49. Miyamoto, S.; Kollman, P.A. SETTLE: An Analytical Version of the SHAKE and RATTLE Algorithm for Rigid Water Models. *J. Comput. Chem.* **1992**, *13*, 952-962.

50. Darden, T.; York, D.; Pedersen, L. Particle Mesh Ewald: An N· log (N) Method for Ewald Sums in Large Systems. *J. Chem. Phys.* **1993**, *98*, 10089-10092.

51. Chipot, C.; Pohorille, A. Calculating Free Energy Differences Using Perturbation Theory. In *Free Energy Calculations;* Chipot, C.; Pohorille, A. Eds.; Springer: Berlin, Heidelberg, 2007; pp 37-40.

52. Bennett, C.H. Efficient Estimation of Free Energy Differences from Monte Carlo Data. *J. Comput. Phys.* **1976**, *22*, 245-268.

53. Shirts, M.R.; Chodera, J.D. Statistically Optimal Analysis of Samples from Multiple Equilibrium States. *J. Chem. Phys.* **2008**, *129*, 124105.

54. Pitera, J.W.; van Gunsteren, W.F. A Comparison of Non-bonded Scaling Approaches for Free Energy Calculations. *Mol. Simul.* **2002**, *28*, 45-65.

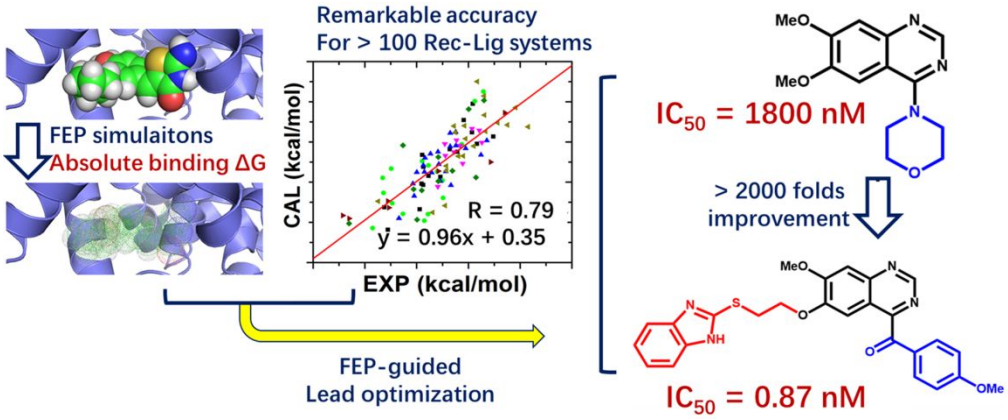
55. Jorgensen, W.L.; Ravimohan, C. Monte Carlo Simulation of Differences in Free energies of Hydration. *J. Chem. Phys.* **1985**, *83*, 3050-3054.

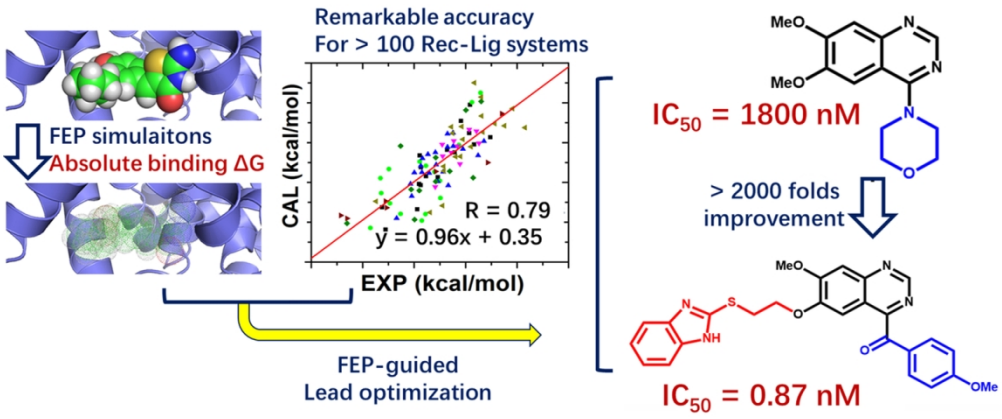
56. Gao, J.; Kuczera, K.; Tidor, B.; Karplus, M. Hidden Thermodynamics of Mutant Proteins: A Molecular Dynamics Analysis. *Science* **1989**, *244*, 1069-1072.

57. Wang, H.; Liu, Y.; Hou, J.; Zheng, M.; Robinson, H.; Ke, H. Structural Insight Into Substrate Specificity of Phosphodiesterase 10. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 5782-5787.

58. Winn, M.D.; Ballard, C.C.; Cowtan, K.D.; Dodson, E.J.; Emsley, P.; Evans, P.R.; Keegan, R.M.; Krissinel, E.B.; Leslie, A.G.; McCoy, A. Overview of the CCP4 Suite and Current Developments. *Acta Crystallogr. D* **2011**, *67*, 235-242.

Table of Contents graphic.





TOC graph

132x54mm (300 x 300 DPI)