

A Cloud Computing Platform for Scalable Relative and Absolute Binding Free Energy Prediction: new opportunities and challenges for drug discovery

Zhixiong Lin^{‡1}, Junjie Zou^{‡1}, Chunwang Peng¹, Shuai Liu^{1,2}, Zhipeng Li¹, Xiao Wan¹,
Dong Fang¹, Jian Yin¹, Gianpaolo Gobbo², Yongpan Chen¹, Jian Ma¹, Shuhao Wen^{1,2},
Peiyu Zhang^{*1}, Mingjun Yang^{*1}

1. Shenzhen Jingtai Technology Co., Ltd. (XtalPi Inc.), Floor 4, No. 9, Hualian Industrial Zone, Dalang Street, Longhua District, Shenzhen 518100, China
2. XtalPi Inc., 245 Main Street, Cambridge, Massachusetts 02142, United States

[‡]These authors contributed equally to this work

Corresponding Author

*Peiyu Zhang: peiyu.zhang@xtalpi.com

*Mingjun Yang: mingjun.yang@xtalpi.com

KEYWORDS: free energy perturbation, drug selectivity, R-group substitutions, core hopping, scaffold hopping, macrocycle design

ABSTRACT

Free energy perturbation (FEP) has become widely used in drug discovery programs for binding affinity prediction between candidate compounds and their biological targets. Simultaneously limitations of FEP applications also exist, including but not limited to, the high cost, long waiting time, limited scalability and application scenarios. To overcome these problems, we have developed a scalable cloud computing platform (XFEP) for both relative and absolute free energy predictions with refined simulation protocols. XFEP enables large-scale FEP calculations in a more efficient, scalable and affordable way, e.g. the evaluation of 5,000 compounds can be performed in one week using 50-100 GPUs with a computing cost approximately corresponding to the cost for one new compound synthesis. Together with artificial intelligence (AI) techniques for goal-directed molecule generation and evaluation, new opportunities can be explored for FEP applications in the drug discovery stages of hit identification, hit-to-lead, and lead optimization with R-group substitutions, scaffold hopping, and completely different molecule evaluation. We anticipate scalable FEP applications will become widely used in more drug discovery projects to speed up the drug discovery process from hit identification to pre-clinical candidate compound nomination.

Introduction

Drug discovery is a challenging task with multiple parameter to be optimized to pre-defined criteria for various pharmaceutical properties¹⁻². For drugs targeting specific biological molecules, sufficient binding affinity between the drug and its target is the basis for developing potency effect of the drug. Therefore, binding affinity is one of the critical optimization goals during the drug discovery stages of hit identification, lead generation (hit to lead) and optimization. For hit identification, a set of compounds are identified with confirmed activity to the biological target usually using high throughput screening or virtual screening followed by experimental validation from a library of diverse compounds. During the hit-to-lead phase, the structure activity relationship (SAR) is explored with various analogous compounds from two or three selected hit series, where scaffold hopping involved in the structural change of molecules is a common design strategy at this stage. For lead optimization, the SAR is further exploited with chemical group substitutions and modifications. During these consecutive phases, the change of molecules is gradually narrowed down in structural space to achieve sufficient binding affinity. Besides binding affinity, other important pharmaceutical properties are also optimized to reach pre-defined criteria. These properties include but are not limited to target selectivity, membrane permeability, toxicity, metabolism and pharmacokinetics, and solubility etc.. The whole process of drug discovery involves many rounds of iterative design (or hypothesis), make, test and analysis (DMTA) to tailor pharmaceutical profiles to pre-defined criteria. It takes on average 3-5 years and \$700 million to bring a satisfactory compound from the discovery

stage to the even higher-risk preclinical and clinical development phases³⁻⁵. Thus, state-of-art methods, including both experimental and computational approaches that can speed up the project progress and discover high-quality compounds, are continuously explored and widely adopted in drug discovery researches.

Many in-silico methods, including both rule-based physical models and data-based machine learning (ML) or artificial intelligence (AI) models, have been adopted in drug discovery projects for binding affinity prediction with continuously improved performance.⁶⁻¹⁶ Among these computational methods, free energy perturbation (FEP) has attracted increasing attention for binding affinity predictions between candidate compounds and their biological target due to its reliable performance in accuracy and efficiency. FEP employs a series of well-defined alchemical states to change the system from one real ligand to another one (the relative binding free energy method, RBFE) or from the target-ligand complex to the separated target and ligand state (the absolute binding free energy method, ABFE).¹⁷⁻⁴¹ Besides binding affinity, the time-dependent information can also be derived from FEP simulations, including the interactions and conformational dynamics of the ligand and its target. With this available information, FEP can be used to optimize and validate binding pose, predict binding affinity of different ligands, interpret binding affinity cliff, and construct virtual SAR, etc. In recent years, many studies have been published from both academic and industrial communities to further optimize the FEP protocol and methodology for more efficient and accurate predictions^{20-27, 32, 35-40, 42}, identify and extend the domain of applicability to more challenging target-ligand systems and scenarios^{35, 43-45}, and build open toolkits

to reduce the access barrier for FEP applications^{26, 37-39, 46}. Most of these studies focus on the use of RBFE for R-group substitution and core hopping in corresponding drug discovery scenarios of lead optimization and hit-to-lead stages. The use of ABFE is relatively limited due to accuracy issues arising from significant structural changes between terminal states of FEP thermodynamic cycle^{32, 36, 38}. All these efforts have made FEP more widely applied in drug discovery research. However, limitations in scalable FEP applications in drug discovery programs still exist due to issues in the scalability, affordability, efficiency, and applicability of scenarios³³.

To overcome these limitations, we have developed a cloud computing platform (XFEP) for large-scale FEP simulations, for which the RBFE and ABFE protocols are further optimized with system specifically refined force fields and enhanced sampling. Together with AI models and wet-lab experiments, we have built an efficient workflow for goal-directed design idea generation from AI, quick evaluation and prioritization of hypothesized ideas using verified FEP protocol for the given target system, wet-lab synthesis and analysis. This makes FEP applications more scalable, affordable and efficient in scenarios of hit identification, hit to lead and lead optimization in drug discovery projects. In this study, we present the performance of our XFEP platform in RBFE and ABFE calculations corresponding to various application scenarios at the three phases of drug discovery, based on which the new opportunities and challenges of the scalable FEP applications are discussed.

The remainder of the paper is organized as follows. In the method section, the protocols of RBFE, ABFE, system specific force field refinement and XFEP platform

introduction are addressed. The performance of the RBFE and ABFE protocols is shown using representative examples for R-group substitutions, scaffold hopping, macrocycle design, protein residue mutations and hit identification in the results section, respectively. The new opportunities and remaining challenges are then discussed in the discussion and conclusion sections. We anticipate more impactful applications of our scalable XFEP platform in academic studies and industrial projects.

Method

1. RBFE

The foundations for estimating free energy differences were developed many decades ago by Kirkwood in 1935¹⁸. In this seminal work, Kirkwood introduced the notion of coupling parameter, which he used to calculate the free energy difference between two well defined thermodynamic states. This laid the groundwork of the perturbation theory and thermodynamic integration (TI) methods⁴⁷⁻⁴⁸, both of which rely upon a coupling parameter to describe chemical changes between two states. In 1954, Zwanzig proposed the FEP method¹⁷. This method relates the free energy difference between an initial (reference) and a final (target) state of a system to the thermodynamic average of a function of their energy difference evaluated in the ensemble of the initial state.

The XFEP workflow utilizes the AMBER software package³⁴ for free energy calculations. Prepared protein structures and docked ligand structures serve as the inputs. Common atoms between reference and target compounds are identified using structure and maximum common substructure criteria. The input ligands and complexes structures are first equilibrated in solution, and then the morphing is performed by superimposing the target molecule onto the reference molecule. This step is followed by free energy simulations using the AMBER software³⁴. We refer the reader to a recently published review paper for details about the AMBER free energy simulation protocol³⁴. In order to speed up the calculations, the hydrogen mass transfer (HMR) method as well as concerted λ scheme²⁷ are enabled. HMR allows the use of a larger timestep, which can speed up the simulations by a factor of two. The concerted λ

scheme allows to use fewer lambda windows compared to the more widely used stepwise scheme which can also potentially speed up the simulations by another factor of two²⁷. Finally, the free energy is evaluated using TI and the multistate Bennett acceptance ratio methods⁴⁸⁻⁴⁹.

2. ABFE

The calculation of absolute binding affinities is a special case of relative free energy calculations, where the target ligand is a dummy ligand. The ABFE approach incorporates a considerably different perturbation from the RBE method and carries its own set of difficulties and special considerations. For example, due to the larger size of the perturbation, ABFE is much more difficult to converge than RBE. Also, because the target ligand is treated as a dummy ligand, that is the whole ligand is “disappeared” from the system, specific treatment for restraining needs to be applied. Moreover, the rearrangement of the binding site upon unbinding and the associated water diffusion can be very slow processes. As a consequence, ABFE calculations generally demand longer simulation times than RBE to attain comparable accuracy, if these movements can ever be sampled properly.

In the XFEP implementation, Boresch restraining²⁸ is applied. Six restraints, including one bond, two angles and three dihedrals, are switched on first, followed by switching off the partial charges on all atoms. Then the van der Waals (vdW) interactions between the ligands and the environment are turned off keeping the restraints on. At the end, an analytical correction term is applied to account for the effect²⁸ of the restraints.

Additionally, for both RBE and ABFE, replica exchange with solute tempering

(REST2)⁵⁰ is implemented to enhance the conformation sampling of subsystems of interest in XFEP.

3. System specific force field refinement protocol

Force field (FF) is one of the critical factors that govern the accuracy of FEP predictions. The initial FF parameters for the target specific ligands are derived from an advanced general force field developed internally for drug-like molecules (XFF), the parametrization philosophy of which is consistent with AMBER macromolecular force fields⁵¹⁻⁵². The ligands were firstly fragmented to construct smaller model compounds for QM calculation at B3LYP/6-31G**//RI-MP2/cc-pVTZ level using the PSI4 package⁵³. The model compounds were constructed by reserving the chemical environment of individual flexible torsions for FF validation and further refinement. Atomic charges are derived with RESP using multiple low energy conformations for the model compounds⁵⁴. vdW parameters are transferred from the GAFF1.8.1 force field⁵². Using QM energy profiles derived from model compounds, the torsion parameters are refitted and then transferred to full ligand compounds for FEP simulations. Table 1 shows the comparison of MM and QM torsion profile agreement before and after refitting for ligands of the eight biological targets discussed below for the RBFE benchmark.

4. Cloud computing platform for FEP simulations

To get access to scalable computing resources⁵⁵⁻⁵⁶, we have developed the XFEP cloud computing platform for FEP based simulations. The platform enables simulation tasks of RBFE, ABFE and the related relative hydration free energy (RHFE) and absolute

hydration free energy (AHFE). Therefore, XFEP supports studies of binding of small-molecule ligands, peptides or mutated proteins to their biological targets. XFEP provides multiple options via a GUI to combine perturbation pair design for RBFE and RHFE, force field parameter derivation from existing FF repository and for further refinement, enhanced sampling choices for selected subsystems of interest, correction schemes for charge change, decoupled as well as concerted schemes for electrostatic and vdW interactions, simulation settings for different tasks and for both efficient and full protocol FEP (Figure 1). A number of 5,000 FEP pairs can be completed reliably within one week using the efficient protocol of XFEP on 50~100 GPU cards.

Results

1. RBFE performance benchmark with R-group substitutions

We validated our XFEP workflow and in house force field parameters (XFF with system specific refinement) against eight datasets from Wang et al.²⁰: Bace, CDK2, Jnk1, MCL1, p38, PTP1B, Thrombin, and Tyk2. In order to be able to compare with FEP+ results directly, we used exactly the same pairs as were used in Roos et al.⁵⁷. The protein and ligands poses were taken directly from Wang et al.²⁰. All simulations were carried out for five repetitions starting from independently randomized velocities. The simulation time for each lambda window was 2 ns. The resulting $\Delta\Delta G$ values were calculated from the averages of the five runs. The correlation coefficient R^2 , the root-mean-square error (RMSE), the mean unsigned error (MUE) and the Kendall's rank coefficient τ were calculated and compared with the raw $\Delta\Delta G$ results from FEP+ without cycle closure correction added, see Table 2 and Figure 2.

In terms of accuracy of the FEP results for these eight targets, our XFEP workflow combined with our in-house force field parameters outperforms FEP+ in combination with op1s2.1 in terms of both RMSE and R^2 ⁵⁸. Also, our results show comparable R^2 as op1s3e. This indicates that in terms of correlation with experimental data, our XFEP workflow in combination with our in-house force field parameters show as good performance as FEP+ in combination with the most up-to-date OPLS force field⁵⁹. In terms of RMSE, our workflow shows slightly worse performance than op1s3e, with 0.15 kcal/mol larger RMSE. This is likely due to some outliers and relatively high RMSE in the MCL1 system. In depth studies of these outliers will be one of our main priorities

for next-round refinement of XFEP.

The sampling time per perturbation is also of interest as computational resources in drug design projects are limited. In opls2.1 simulations, a simulation time of 60 ns per perturbation was used²⁰, while in opls3e simulations, a simulation time of 300 ns per perturbation was used⁵⁷. The standard XFEP workflow uses 22 lambda windows and 2 ns per lambda window. Hence, the cumulative simulation time is 220 ns per perturbation (accounting for the five repetition per perturbation) in the full-protocol FEP. Therefore, the current simulation time per perturbation in our workflow is comparable to that of FEP+. However, in an effort to improve computational efficiency, we have also tried to reduce the number of lambda windows to 11, using a concerted λ scheme²⁷, and the simulation time to 1 ns per lambda window. We have verified that using this procedure we obtain results that are very consistent with those obtained with the standard workflow described above (data not shown). This optimized setup allows to reduce the simulation cost by a factor of 4 in the efficient-protocol FEP.

2. RBFE performance benchmark with scaffold hopping

Besides R-group perturbations, FEP calculations have also been employed to calculate the free energy changes caused by altering the bonding topology of compounds, which is usually referred to as core (scaffold)-hopping in the field of FEP calculations. Modeling the changes of bonding topology using FEP is technically challenging but can significantly broaden the application of FEP calculations in drug design. Successful estimation of the effect caused by altering the bonding topology of compounds requires accurate modeling of both the enthalpy and entropy change upon binding. These effects

can be well captured with MD-based methods like FEP calculations. Here we demonstrate how XFEP can be applied to several different scenarios of drug design, in which the bonding topologies of compounds are altered to different extents (Figures 3-6).

Minor modification of the bonding topologies such as ring opening/closure, ring and chain contraction/expansion are often used by medicinal chemists during rational design and optimization of compounds. Two transformations involving ring closure and chain expansion⁶⁰⁻⁶¹ using XFEP are studied here and the results are shown in Figure 3.

Macrocyclization of linear compounds is a common strategy employed by medicinal chemists, which enhances the binding affinity between compounds and protein mainly by minimizing the entropy loss upon binding. Macrocyclized compounds can be further optimized by rigidification and relieving of strain, which is usually achieved by changing the ring size of the compounds and by modifying the chemistry of the ring scaffolds (Figure 4).

Besides minor changes on the bonding topologies of compounds, XFEP also offers solutions for transformations between cores with highly different bonding topologies. This allows for the evaluation of molecules constructed by using fragment-based design, in which ligands with different sizes and shapes of cores are proposed (Figure 5).

During the design of bi-valent, bi-substrate and PROTAC compounds, the length, shape and rigidity of linkers can greatly affect the potency of the compounds. Choosing the optimal linkers usually requires careful balance in the gain and loss of enthalpy and entropy upon binding. The capability of XFEP to model with high accuracy both the

enthalpy and entropy change upon binding opens the door for a more reliable design of linkers (Figure 6).

3. ABFE performance benchmark for molecules with different scaffolds

Apart from the conventional R-group substitutions and core hopping discussed above, more dramatic structural changes may occur in the hit-to-lead stage of the drug design cycle. Sometimes, the molecules considered at this stage may not share any common part, which makes it difficult for R group substitution FEP or even core hopping to rank them. What's more, in the hit finding stage, where virtual ligand screening is performed, one usually needs to rank dissimilar compounds. In this case, structure-based tools such as docking involve significant approximations that limit their accuracy. In all these instances ABFE calculations may prove very useful. Since ABFE does not rely on any reference compound, it can be used to compare binding affinities of very different molecules.

We have implemented ABFE into our XFEP workflow, and tested the implementation using four of eight targets shown in Table 1: CDK2, Jnk1, p38 and Tyk2. These four targets were selected because all these ligands are neutral compounds. ABFE was carried out using the XFEP workflow in combination with in-house force field parameters (XFF with system specific refinement). The simulation setup is as follows. 11 lambda windows were used for the decharge step to set all the partial charges of the ligand to zero, and 16 lambda windows were used for the vdW step to decouple the vdW interactions between the ligand and the environment. Simulation time was 2 ns for each lambda window. The results are shown in Figure 7. The performance of ABFE

is comparable with that of RBFE (Figure 1). Additionally, we further tested ABFE on the BRD4 system. To this purpose, we chose 18 compounds with 15 different scaffolds, see Figure 8. For this system, the GAFF2 force field in AMBER20⁵² was used. The correlation between predicted ABFE and experimental data is shown in Figure 8. ABFE results correlate with experimental data very well, with a correlation coefficient R^2 of 0.78.

However, we did also notice a systematic shift of the ABFE results compared with experimental results for both test systems. The range of ABFE results (from -15 kcal/mol to -8 kcal/mol for BRD4 system) is almost twice as big as the range of experimental data (from -10 kcal/mol to -6 kcal/mol for BRD4 system). This systematic shift may be due to a deficiency of the force field parameters, or to inaccurate accounting of the effect of binding site changes upon unbinding due to the limited simulation time.

4. RBFE performance benchmark for protein mutations

FEP calculations have been widely used for studying the effect of protein mutations on various thermodynamic quantities. Such calculations have been successfully used to predict drug resistance, design selective compounds, optimize binding affinities and thermo-stabilities of peptide inhibitors⁶²⁻⁶⁴. Typically, the FEP calculations for studying protein mutations are carried out by perturbing the side chains of residues from one to another, which is similar in spirit to the R-group perturbations described above.

We carried out FEP calculations to calculate the changes of binding affinity between the *Streptomyces griseus* proteinases B (SGPB) and the turkey ovomucoid third domain

(OMTKY3) caused by single mutations⁶⁵. This protein-protein complex has been used to validate the accuracy of FEP calculations in several previous studies^{64, 66}. FEP calculations were conducted on five selected mutations, which display changes in binding affinity ranging from 3.0 to 8.5 kcal/mol. The calculated and experimental values have an MUE, RMSE, Kendall's τ and R^2 of 0.55 kcal/mol, 0.76 kcal/mol, 0.95 and 0.97 respectively (Table 3).

FEP calculations were also conducted to calculate the change of binding affinity between the Abl kinase and the drug compound Axitinib caused by single mutations⁶⁷. Such calculations can be used to predict mutations that lead to drug resistance due to weakened binding (Table 4).

Selectivity calculations were conducted on the compound CEP-701 and Mitogen-activated protein kinase kinase kinase kinase (MAP4K) family. The relative binding free energy difference of MAP4K1/HPK1 and other isoforms were computed by mutating the residues which form direct contact with the bound compound. An example is shown in the following table (Table 5).

Discussions

1. A statistical view of prediction performance for RBFE

Based on data from 333 perturbation pairs of the eight systems studied above, the FEP performance was also examined through a statistical analysis of the distribution and range of the errors of the predicted values with respect to the experimental data. For individual pairs, the predicted $\Delta\Delta G$ result is used to assess if the target molecule is more favorable in binding affinity than the reference compound. At this level the prediction accuracy is mainly dictated by the force field, the physical rigorousness of the FEP protocol, the accurate modeling of systems, and the sampling of important conformational states etc. However, under most circumstances, multiple molecules are evaluated together for priority ranking. In this case, the performance of the predictions is usually characterized by the correlation coefficient (R^2) and the Kendall's rank correlation coefficient (τ) between the experimental and predicted values. We notice that, as hypothesized in other studies^{44, 68}, the distribution of the prediction error ($\Delta\Delta G_{\text{exp}} - \Delta\Delta G_{\text{FEP}}$) is well represented by a Gaussian distribution $N(\mu, \sigma)$. We examined this hypothesis by fitting a Gaussian model to the 333 pairs from different force fields, indicating a well Gaussian distribution (Figure 9). Interestingly, the distribution of experimental values ($\Delta\Delta G_{\text{exp}}$) also generally follows a Gaussian distribution with a standard deviation of the mean value $\sigma_{\text{exp}} = 1.14$ kcal/mol. Based on these observations, we carried out a model study of the R^2 and τ dependence on the variation range of the experimental values, and the variation of the predicted errors with respect to the experimental values. Figure 10 shows the dependence of the average R^2 and τ to the

two parameters. A dramatic change can be observed along with the range of experimental values and the predicted errors. For typical values of predicted errors at $RMSE = 1.00$ kcal/mol, the average R^2 and τ are at 0.56 and 0.54, which is close to our observations from studies with different force fields (Table 2). This suggests the interpretation and expectation of prediction performance using R^2 and τ should consider the range and accuracy of experimental values included in the dataset. For typical inter-lab measurements, the average experimental variations are about 0.6 kcal/mol while for intra-lab measurements the typical variations are of ⁶⁹⁶² These average variations correspond respectively to an upper limit of the predicted average $R^2=0.78$ and 0.97, average $\tau= 0.69$ and 0.89 at $RMSE=0.6$ kcal/mol and 0.2 kcal/mol at an experimental range of $\sigma_{exp} = 1.14$ kcal/mol. Similar considerations can also be applied to the ABFE cases if a similar distribution of errors of the predictions is observed. In summary, the statistical analysis of the agreement between predicted and experimental values gives a measure of how much the FEP performance can be improved.

2. Scalable and efficient FEP prediction with cloud computing platforms

The high cost and long waiting time for FEP calculations are two other main limitations for the application of FEP to real drug discovery projects that require efficient DMTA iteration and evaluation of many different molecules for priority ranking and hypothesis testing. Using scalable resources on cloud computing platforms, the typical number of instances of FEP calculations for a project can be expanded from tens or hundreds to thousands or even more, potentially allowing for a two-fold increase in the number of evaluated molecules. For example, for a medium-sized system (50,000 atoms), with

sufficient GPU cards available in spot-instance mode on a cloud platform, e.g. Amazon Web Services, all states for one perturbation pair can be run in parallel and completed within a walltime of half an hour for 5 ns simulation length per window. Furthermore, with a thoroughly validated efficient FEP protocol for a given system, it is possible to further reduce the GPU cost of an additional factor of 4-8 as discussed in the methods section. Within this framework and under the assumption of enough GPUs available in spot-instance mode, the cost of FEP calculations for one molecule can vary from less than \$1 to \$10 depending on which FEP protocol (efficient or full) is adopted.

3. New opportunities for FEP applications in drug discovery projects

With persistent efforts in FEP methodology and protocol improvement from both the academic and industrial communities, reasonably good prediction performance has been achieved in predicting both relative and absolute binding affinities for many biological systems. Together with the scalable cloud computing resources available, FEP calculations can be applied to thousands of molecules in one week with an average GPU cost controlled at less than \$1 for each molecule. This capability paves the way for FEP predictions to be applied in new scenarios. In this section we will briefly discuss three possible examples of such new applications.

The first scenario is the combination of AI and scalable FEP applications. Here AI is used as a tool for fast idea generation with directed goals, e.g. for R-group substitutions, core hopping, or de novo molecule generation with defined criteria like, for instance, molecular weight, polar surface area, number of hydrogen bond donor and acceptor, synthesizability, and binding affinity, etc⁷⁰. Then thousands of molecules are selected

from the AI generation step, for FEP prediction with validated performance against a selected biological target. The predicted binding affinity data are then used in a next round of AI model refinement with reinforcement learning to generate new molecules for subsequent FEP evaluation leading to an automated iterative scheme that improves the quality of the candidate molecules generated. A recent paper from Schrödinger and our internal application of the combined AI and FEP approach suggest a consistent enrichment factor of 4-6 compared to [random selections](#) from bare AI generated molecules²⁵. However, further studies are required to demonstrate quantitatively how AI and FEP tools can be combined together in more intensive drug discovery scenarios. The second scenario we consider, is the application of the scalable ABFE prediction at the hit identification stage, which so far has seen only limited application compared to RBFE calculations due to the high cost and error of ABFE predictions³³. From a drug-design perspective, RBFE and ABFE represent two considerably different domains of applicability. RBFE is more appropriate for hit-to-lead and lead optimization, when near chemical accuracy (e.g., 1 kcal/mol) is critical to making correct ranking between the target and reference compounds; while in the hit finding process, for virtual screening ABFE accuracy need only be sufficient to separate binders from nonbinders. Hit identification is devoted to maximizing the number of hit series with confirmed activity to a validated biological target, from which drug-like compounds can be developed with the best chance. It is one of the most critical steps in drug discovery because a high quality of starting hits leaves more optimization space to trade-off with other essential properties. Typically, virtual screening based on molecular docking and

high throughput screening is used to select a diverse set of molecules from existing compound libraries. Thanks to our scalable implementation, ABFE can be used in combination with these widely used methods, and applied to thousands of selected compounds to further explore and prioritize the relevant chemical space. The performance of ABFE predictions suggests that addition of the evaluation of binding affinity to the screening process can lead to better and faster hit compounds identification. In a recent study, ABFE is applied to drug repurpose for identifying potent SARS-CoV-2 main protease inhibitors, from which an unprecedentedly high hit rate is achieved⁷¹. This demonstrates a promising potential to speed up hit identification by integrating the ABFE prediction in virtual screening. Further exploration is anticipated to demonstrate how scalable ABFE can help with better hit identification in more drug discovery programs.

The third scenario of scalable FEP application is virtual SAR evaluation at the hit-to-lead and lead optimization stages. The goal of the hit-to-lead stage is to identify two or three hit series with the best potential to be developed into drug-like leads. During this stage the binding affinity increases from the micro-molar (μM) range of typical hit compounds to nano-molar (nM) range of promising lead compounds. Lead optimization is then followed for further improvement of the binding affinity by one or two orders together with optimization of other essential pharmaceutical properties. During the hit-to-lead and lead optimization stages, SAR is constructed from DMTA cycles, from which more experimental data are generated. With validated prediction performance for given biological targets, FEP can thus be applied for virtual SAR

evaluation on a scale of thousands of compounds per project. This enables quick prioritization of proposed hypotheses and helps to identify most of the true negative compounds. In this scenario, FEP can be expected to speed up DMTA cycles with fewer compounds to make and test but with a larger success rate.

4. Remaining challenges for FEP applications in drug discovery projects

Despite the new opportunities for FEP applications in drug discovery projects discussed above, there are remaining challenges, some of which are as follows.

a. FEP for RNA target: To the best of our knowledge, both RBE and ABFE calculations have only been applied to protein targets. However, the number of druggable protein targets is very limited compared to the number of RNA targets. Multiple studies have shown RNA is a suitable small-molecule drug target⁷²⁻⁷⁴. However, FEP prediction targeting RNA faces several major challenges including the limited accuracy of RNA and related metal-ion FF⁷⁵⁻⁷⁸, the issue of proper sampling for flexible RNA structures, and the complications due to the highly charged environment for RNA and the counterions. This is a new scenario for future exploration of FEP applicability.

b. How to interpret FEP prediction results: From retrospective validation of FEP predictions, we know that it is unavoidable to have false negative and false positive results in most cases. False positive cases are not always just useless byproducts that only increase the computational cost. In fact, they can also be used to test the hypothesis of design ideas as invalidate chemical space. FEP can identify most true negative cases especially when there exist binding affinity cliffs. On the

other hand, false negatives are the most worrying cases because they have the power to derail a project by hampering serendipity. Thus it is tricky to interpret the FEP results under different circumstances and tolerance to the prediction error should be kept in mind.

c. The variability in FEP performance induced by different users: There are many variables that govern the FEP performance for given biological targets and chemical compounds, which includes the selection of reference complex structure, the treatment of biological targets and small molecules, the modeling details of systems, the choice and refinement of FF, the preparation of the systems before the FEP production runs, the atom mapping between the pair of compounds in RBFE or the selection of restraints in ABFE etc. These possibilities cannot all be covered in a user manual. Only experienced FEP experts can positively handle these variables to build a reliable FEP protocol within a given project timeline. However, most FEP users are not as experienced as the experts in real discovery projects. This creates a gap in FEP performance between different users. To alleviate this issue, more work dedicated to bridging the gap in user experience is required.

Conclusions

With progress in FEP methodology and protocol improvement, algorithm optimization, and GPU hardware development, FEP has been proven as a reliable in-silico method for both RBFE and ABFE predictions. Additionally, large-scale FEP calculations have become available in drug discovery projects in a more efficient and affordable way. These advancements, together with artificial intelligence (AI) techniques for goal-directed molecule generation and evaluation, enable more opportunities for FEP applications in the drug discovery stages of hit identification, hit-to-lead, and lead optimization with R-group substitutions, core hopping, and completely different molecules evaluation. Based on retrospective validation of FEP performance for given biological targets, scalable evaluation with FEP of 5,000 designed compounds can be completed in one week using 50-100 GPU cards for a cost approximately equivalent to that of the synthesis of one new compound. We anticipate that scalable FEP applications will become widely used in more drug discovery projects to accelerate the progress of phases from hit to pre-clinical candidate compound.

Reference

1. Nicolaou, C. A.; Brown, N., Multi-objective optimization methods in drug design. *Drug Discov Today Technol* **2013**, *10* (3), e427-35.
2. Schneider, G., Automating drug discovery. *Nature reviews. Drug discovery* **2018**, *17* (2), 97-113.
3. DiMasi, J. A.; Grabowski, H. G.; Hansen, R. W., Innovation in the pharmaceutical industry: New estimates of R&D costs. *J Health Econ* **2016**, *47*, 20-33.
4. Cook, D.; Brown, D.; Alexander, R.; March, R.; Morgan, P.; Satterthwaite, G.; Pangalos, M. N., Lessons learned from the fate of AstraZeneca's drug pipeline: a five-dimensional framework. *Nature reviews. Drug discovery* **2014**, *13* (6), 419-31.
5. Paul, S. M.; Mytelka, D. S.; Dunwiddie, C. T.; Persinger, C. C.; Munos, B. H.; Lindborg, S. R.; Schacht, A. L., How to improve R&D productivity: the pharmaceutical industry's grand challenge. *Nature reviews. Drug discovery* **2010**, *9* (3), 203-14.
6. Limongelli, V.; Bonomi, M.; Parrinello, M., Funnel metadynamics as accurate binding free-energy method. *Proceedings of the National Academy of Sciences* **2013**, *110* (16), 6358-6363.
7. Pan, A. C.; Xu, H.; Palpant, T.; Shaw, D. E., Quantitative Characterization of the Binding and Unbinding of Millimolar Drug Fragments with Molecular Dynamics Simulations. *J Chem Theory Comput* **2017**, *13* (7), 3372-3377.
8. Guvench, O.; MacKerell, A. D., Computational fragment-based binding site identification by ligand competitive saturation. *PLoS Comput Biol* **2009**, *5* (7), e1000435.
9. Rao, L.; Chi, B.; Ren, Y.; Li, Y.; Xu, X.; Wan, J., DOX: A new computational protocol for accurate prediction of the protein-ligand binding structures. *Journal of Computational Chemistry* **2016**, *37* (3), 336-344.
10. Mobley, D. L.; Gilson, M. K., Predicting Binding Free Energies: Frontiers and Benchmarks. *Annual Review of Biophysics* **2017**, *46* (1), 531-558.
11. Wang, L.; Chambers, J.; Abel, R., Protein-Ligand Binding Free Energy Calculations with FEP. *Methods in molecular biology (Clifton, N.J.)* **2019**, *2022*, 201-232.
12. Wang, E.; Sun, H.; Wang, J.; Wang, Z.; Liu, H.; Zhang, J. Z. H.; Hou, T., End-Point Binding Free Energy Calculation with MM/PBSA and MM/GBSA: Strategies and Applications in Drug Design. *Chemical Reviews* **2019**, *119* (16), 9478-9508.
13. Cai, C.; Wang, S.; Xu, Y.; Zhang, W.; Tang, K.; Ouyang, Q.; Lai, L.; Pei, J., Transfer Learning for Drug Discovery. *Journal of Medicinal Chemistry* **2020**, *63* (16), 8683-8694.
14. Bajorath, J.; Kearnes, S.; Walters, W. P.; Meanwell, N. A.; Georg, G. I.; Wang, S., Artificial Intelligence in Drug Discovery: Into the Great Wide Open. *Journal of Medicinal Chemistry* **2020**, *63* (16), 8651-8652.
15. Li, X.; Li, Z.; Wu, X.; Xiong, Z.; Yang, T.; Fu, Z.; Liu, X.; Tan, X.; Zhong, F.; Wan, X.; Wang, D.; Ding, X.; Yang, R.; Hou, H.; Li, C.; Liu, H.; Chen, K.; Jiang, H.; Zheng, M., Deep Learning Enhancing Kinome-Wide Polypharmacology Profiling: Model Construction and Experiment Validation. *Journal of Medicinal Chemistry* **2020**, *63* (16), 8723-8737.
16. Feinberg, E. N.; Joshi, E.; Pande, V. S.; Cheng, A. C., Improvement in ADMET Prediction with Multitask Deep Featurization. *Journal of Medicinal Chemistry* **2020**, *63* (16), 8835-8848.
17. Zwanzig, R. W., High-temperature equation of state by a perturbation method. I. Nonpolar gases. *J. Chem. Phys.* **1954**, *22* (8), 1420-1426.

18. Kirkwood, J. G., Statistical mechanics of fluid mixtures. *J Chem Phys* **1935**, *3* (5), 300-313.
19. Mobley, D. L.; Klimovich, P. V., Perspective: Alchemical free energy calculations for drug discovery. *The Journal of Chemical Physics* **2012**, *137* (23), 230901.
20. Wang, L.; Wu, Y.; Deng, Y.; Kim, B.; Pierce, L.; Krilov, G.; Lupyran, D.; Robinson, S.; Dahlgren, M. K.; Greenwood, J.; Romero, D. L.; Masse, C.; Knight, J. L.; Steinbrecher, T.; Beuming, T.; Damm, W.; Harder, E.; Sherman, W.; Brewer, M.; Wester, R.; Murcko, M.; Frye, L.; Farid, R.; Lin, T.; Mobley, D. L.; Jorgensen, W. L.; Berne, B. J.; Friesner, R. A.; Abel, R., Accurate and reliable prediction of relative ligand binding potency in prospective drug discovery by way of a modern free-energy calculation protocol and force field. *J Am Chem Soc* **2015**, *137* (7), 2695-703.
21. Liu, S.; Wang, L.; Mobley, D. L., Is Ring Breaking Feasible in Relative Binding Free Energy Calculations? *Journal of Chemical Information and Modeling* **2015**, *55* (4), 727-735.
22. Xu, H., Optimal Measurement Network of Pairwise Differences. *Journal of Chemical Information and Modeling* **2019**, *59* (11), 4720-4728.
23. Song, L. F.; Lee, T.-S.; Zhu, C.; York, D. M.; Merz, K. M., Using AMBER18 for Relative Free Energy Calculations. *Journal of Chemical Information and Modeling* **2019**, *59* (7), 3128-3135.
24. Kuhn, M.; Firth-Clark, S.; Tosco, P.; Mey, A. S. J. S.; Mackey, M.; Michel, J., Assessment of Binding Affinity via Alchemical Free-Energy Calculations. *Journal of Chemical Information and Modeling* **2020**, *60* (6), 3120-3130.
25. Ghanakota, P.; Bos, P. H.; Konze, K. D.; Staker, J.; Marques, G.; Marshall, K.; Leswing, K.; Abel, R.; Bhat, S., Combining Cloud-Based Free-Energy Calculations, Synthetically Aware Enumerations, and Goal-Directed Generative Machine Learning for Rapid Large-Scale Chemical Exploration and Optimization. *Journal of Chemical Information and Modeling* **2020**, *60* (9), 4311-4325.
26. Gapsys, V.; Pérez-Benito, L.; Aldeghi, M.; Seeliger, D.; van Vlijmen, H.; Tresadern, G.; de Groot, B. L., Large scale relative protein ligand binding affinities using non-equilibrium alchemy. *Chemical Science* **2020**, *11* (4), 1140-1152.
27. Lee, T. S.; Lin, Z.; Allen, B. K.; Lin, C.; Radak, B. K.; Tao, Y.; Tsai, H. C.; Sherman, W.; York, D. M., Improved Alchemical Free Energy Calculations with Optimized Smoothstep Softcore Potentials. *J Chem Theory Comput* **2020**, *16* (9), 5512-5525.
28. Boresch, S.; Tettinger, F.; Leitgeb, M.; Karplus, M., Absolute Binding Free Energies: A Quantitative Approach for Their Calculation. *The Journal of Physical Chemistry B* **2003**, *107* (35), 9535-9551.
29. Woo, H.-J.; Roux, B., Calculation of absolute protein-ligand binding free energy from computer simulations. *Proceedings of the National Academy of Sciences of the United States of America* **2005**, *102* (19), 6825-6830.
30. Gumbart, J. C.; Roux, B.; Chipot, C., Standard Binding Free Energies from Computer Simulations: What Is the Best Strategy? *Journal of Chemical Theory and Computation* **2013**, *9* (1), 794-802.
31. Aldeghi, M.; Bluck, J. P.; Biggin, P. C., Absolute Alchemical Free Energy Calculations for Ligand Binding: A Beginner's Guide. *Methods in molecular biology (Clifton, N.J.)* **2018**, *1762*, 199-232.
32. Li, Z.; Huang, Y.; Wu, Y.; Chen, J.; Wu, D.; Zhan, C.-G.; Luo, H.-B., Absolute Binding Free Energy Calculation and Design of a Subnanomolar Inhibitor of Phosphodiesterase-10. *Journal of Medicinal Chemistry* **2019**, *62* (4), 2099-2111.
33. Cournia, Z.; Allen, B. K.; Beuming, T.; Pearlman, D. A.; Radak, B. K.; Sherman, W., Rigorous Free Energy Simulations in Virtual Screening. *Journal of Chemical Information and Modeling* **2020**, *60*

(9), 4153-4169.

34. Lee, T. S.; Allen, B. K.; Giese, T. J.; Guo, Z.; Li, P.; Lin, C.; McGee, T. D., Jr.; Pearlman, D. A.; Radak, B. K.; Tao, Y.; Tsai, H. C.; Xu, H.; Sherman, W.; York, D. M., Alchemical Binding Free Energy Calculations in AMBER20: Advances and Best Practices for Drug Discovery. *J Chem Inf Model* **2020**.
35. Deflorian, F.; Perez-Benito, L.; Lenselink, E. B.; Congreve, M.; van Vlijmen, H. W. T.; Mason, J. S.; Graaf, C. d.; Tresadern, G., Accurate Prediction of GPCR Ligand Binding Affinity with Free Energy Perturbation. *Journal of Chemical Information and Modeling* **2020**.
36. Sakae, Y.; Zhang, B. W.; Levy, R. M.; Deng, N., Absolute Protein Binding Free Energy Simulations for Ligands with Multiple Poses, a Thermodynamic Path That Avoids Exhaustive Enumeration of the Poses. *Journal of Computational Chemistry* **2020**, *41* (1), 56-68.
37. Rufa, D. A.; Bruce Macdonald, H. E.; Fass, J.; Wieder, M.; Grinaway, P. B.; Roitberg, A. E.; Isayev, O.; Chodera, J. D., Towards chemical accuracy for alchemical free energy calculations with hybrid physics-based machine learning / molecular mechanics potentials. *bioRxiv* **2020**, 2020.07.29.227959.
38. Oshima, H.; Re, S.; Sugita, Y., Prediction of Protein-Ligand Binding Pose and Affinity Using the gREST+FEP Method. *Journal of Chemical Information and Modeling* **2020**.
39. Senapathi, T.; Suruzhon, M.; Barnett, C. B.; Essex, J.; Naidoo, K. J., BRIDGE: An Open Platform for Reproducible High-Throughput Free Energy Simulations. *Journal of Chemical Information and Modeling* **2020**.
40. Yang, Q.; Burchett, W.; Steeno, G. S.; Liu, S.; Yang, M.; Mobley, D. L.; Hou, X., Optimal designs for pairwise calculation: An application to free energy perturbation in minimizing prediction variability. *Journal of Computational Chemistry* **2020**, *41* (3), 247-257.
41. Zou, J.; Tian, C.; Simmerling, C., Blinded prediction of protein-ligand binding affinity using Amber thermodynamic integration for the 2018 D3R grand challenge 4. *Journal of computer-aided molecular design* **2019**.
42. Gregory, R.; Ellery, R.; Yuqing, D.; Chao, L.; Edward, H.; Robert, A.; Lingle, W., *Enhancing Water Sampling in Free Energy Calculations with Grand Canonical Monte Carlo*. 2020.
43. Paulsen, J. L.; Yu, H. S.; Sindhikara, D.; Wang, L.; Appleby, T.; Villaseñor, A. G.; Schmitz, U.; Shivakumar, D., Evaluation of Free Energy Calculations for the Prioritization of Macrocyclic Synthesis. *Journal of Chemical Information and Modeling* **2020**, *60* (7), 3489-3498.
44. Albanese, S. K.; Chodera, J. D.; Volkamer, A.; Keng, S.; Abel, R.; Wang, L., Is structure based drug design ready for selectivity optimization? *bioRxiv* **2020**, 2020.07.02.185132.
45. Edward, M.; Robert, M.; Daniel, S.; Ken, B.; Matthew, G.; Fabio, R.; Steven, D.; Steven, J.; Nicholas, B.; Tyler, D.; Phani, G.; Sayan, M.; Salma B., R.; Dawn M., T.; Robert, A.; Richard, F., *A Reliable and Accurate Solution to the Induced Fit Docking Problem for Protein-Ligand Binding*. 2020.
46. Samways, M. L.; Bruce Macdonald, H. E.; Essex, J. W., grand: A Python Module for Grand Canonical Water Sampling in OpenMM. *Journal of Chemical Information and Modeling* **2020**.
47. Zwanzig, R. W.; Kirkwood, J. G.; Oppenheim, I.; Alder, B. J., Statistical Mechanical Theory of Transport Processes. VII. The Coefficient of Thermal Conductivity of Monatomic Liquids. *The Journal of Chemical Physics* **1954**, *22* (5), 783-790.
48. Jorge, M.; Garrido, N. M.; Queimada, A. J.; Economou, I. G.; Macedo, E. A., Effect of the Integration Method on the Accuracy and Computational Efficiency of Free Energy Calculations Using Thermodynamic Integration. *Journal of Chemical Theory and Computation* **2010**, *6* (4),

1018-1027.

49. Shirts, M. R.; Chodera, J. D., Statistically optimal analysis of samples from multiple equilibrium states. *The Journal of Chemical Physics* **2008**, *129* (12), 124105.
50. Wang, L.; Berne, B. J.; Friesner, R. A., On achieving high accuracy and reliability in the calculation of relative protein-ligand binding affinities. *Proc Natl Acad Sci U S A* **2012**, *109* (6), 1937-42.
51. Maier, J. A.; Martinez, C.; Kasavajhala, K.; Wickstrom, L.; Hauser, K. E.; Simmerling, C., ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *Journal of Chemical Theory and Computation* **2015**, *11* (8), 3696-3713.
52. Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A., Development and testing of a general amber force field. *J Comput Chem* **2004**, *25* (9), 1157-74.
53. Smith, D. G. A.; Burns, L. A.; Sirianni, D. A.; Nascimento, D. R.; Kumar, A.; James, A. M.; Schriber, J. B.; Zhang, T.; Zhang, B.; Abbott, A. S.; Berquist, E. J.; Lechner, M. H.; Cunha, L. A.; Heide, A. G.; Waldrop, J. M.; Takeshita, T. Y.; Alenaizan, A.; Neuhauser, D.; King, R. A.; Simmonett, A. C.; Turney, J. M.; Schaefer, H. F.; Evangelista, F. A.; DePrince, A. E.; Crawford, T. D.; Patkowski, K.; Sherrill, C. D., Psi4NumPy: An Interactive Quantum Chemistry Programming Environment for Reference Implementations and Rapid Development. *Journal of Chemical Theory and Computation* **2018**, *14* (7), 3504-3511.
54. Bayly, C. I.; Cieplak, P.; Cornell, W.; Kollman, P. A., A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges: the RESP model. *The Journal of Physical Chemistry* **1993**, *97* (40), 10269-10280.
55. Zhang, P.; Wood, G. P. F.; Ma, J.; Yang, M.; Liu, Y.; Sun, G.; Jiang, Y. A.; Hancock, B. C.; Wen, S., Harnessing Cloud Architecture for Crystal Structure Prediction Calculations. *Crystal Growth & Design* **2018**, *18* (11), 6891-6900.
56. Yang, M.; Dybeck, E.; Sun, G.; Peng, C.; Samas, B.; Burger, V. M.; Zeng, Q.; Jin, Y.; Bellucci, M. A.; Liu, Y.; Zhang, P.; Ma, J.; Jiang, Y. A.; Hancock, B. C.; Wen, S.; Wood, G. P. F., Prediction of the Relative Free Energies of Drug Polymorphs above Zero Kelvin. *Crystal Growth & Design* **2020**, *20* (8), 5211-5224.
57. Roos, K.; Wu, C.; Damm, W.; Reboul, M.; Stevenson, J. M.; Lu, C.; Dahlgren, M. K.; Mondal, S.; Chen, W.; Wang, L.; Abel, R.; Friesner, R. A.; Harder, E. D., OPLS3e: Extending Force Field Coverage for Drug-Like Small Molecules. *J Chem Theory Comput* **2019**, *15* (3), 1863-1874.
58. 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. *Journal of the American Chemical Society* **2015**, *137* (7), 2695-2703.
59. Roos, K.; Wu, C.; Damm, W.; Reboul, M.; Stevenson, J. M.; Lu, C.; Dahlgren, M. K.; Mondal, S.; Chen, W.; Wang, L.; Abel, R.; Friesner, R. A.; Harder, E. D., OPLS3e: Extending Force Field Coverage for Drug-Like Small Molecules. *Journal of Chemical Theory and Computation* **2019**, *15* (3), 1863-1874.
60. Huang, X.; Cheng, C. C.; Fischmann, T. O.; Duca, J. S.; Yang, X.; Richards, M.; Shipps, G. W., Discovery of a Novel Series of CHK1 Kinase Inhibitors with a Distinctive Hinge Binding Mode. *ACS*

Medicinal Chemistry Letters **2012**, *3* (2), 123-128.

61. Wiener, J. J. M.; Wickboldt, A. T.; Wiener, D. K.; Lee-Dutra, A.; Edwards, J. P.; Karlsson, L.; Nguyen, S.; Sun, S.; Jones, T. K.; Grice, C. A., Discovery and SAR of novel pyrazole-based thioethers as cathepsin S inhibitors. Part 2: Modification of P3, P4, and P5 regions. *Bioorganic & Medicinal Chemistry Letters* **2010**, *20* (7), 2375-2378.
62. Zou, J.; Song, B.; Simmerling, C.; Raleigh, D., Experimental and Computational Analysis of Protein Stabilization by Gly-to-d-Ala Substitution: A Convolution of Native State and Unfolded State Effects. *Journal of the American Chemical Society* **2016**, *138* (48), 15682-15689.
63. Hauser, K.; Negron, C.; Albanese, S. K.; Ray, S.; Steinbrecher, T.; Abel, R.; Chodera, J. D.; Wang, L., Predicting resistance of clinical Abl mutations to targeted kinase inhibitors using alchemical free-energy calculations. *Communications Biology* **2018**, *1* (1), 70.
64. Zou, J.; Simmerling, C.; Raleigh, D. P., Dissecting the Energetics of Intrinsically Disordered Proteins via a Hybrid Experimental and Computational Approach. *The Journal of Physical Chemistry B* **2019**, *123* (49), 10394-10402.
65. Lu, W.; Apostol, I.; Qasim, M. A.; Warne, N.; Wynn, R.; Zhang, W. L.; Anderson, S.; Chiang, Y. W.; Ogin, E.; Rothberg, I.; Ryan, K.; Laskowski, M., Jr., Binding of amino acid side-chains to S1 cavities of serine proteinases. *J Mol Biol* **1997**, *266* (2), 441-61.
66. Duan, J.; Lupyan, D.; Wang, L., Improving the Accuracy of Protein Thermostability Predictions for Single Point Mutations. *Biophysical Journal* **2020**, *119* (1), 115-127.
67. Hauser, K.; Negron, C.; Albanese, S. K.; Ray, S.; Steinbrecher, T.; Abel, R.; Chodera, J. D.; Wang, L., Predicting resistance of clinical Abl mutations to targeted kinase inhibitors using alchemical free-energy calculations. *Communications Biology* **2018**, *1* (1).
68. Harder, E.; Damm, W.; Maple, J.; Wu, C.; Reboul, M.; Xiang, J. Y.; Wang, L.; Lupyan, D.; Dahlgren, M. K.; Knight, J. L.; Kaus, J. W.; Cerutti, D. S.; Krilov, G.; Jorgensen, W. L.; Abel, R.; Friesner, R. A., OPLS3: A Force Field Providing Broad Coverage of Drug-like Small Molecules and Proteins. *Journal of Chemical Theory and Computation* **2016**, *12* (1), 281-296.
69. Kramer, C.; Kalliokoski, T.; Gedeck, P.; Vulpetti, A., The Experimental Uncertainty of Heterogeneous Public Ki Data. *Journal of Medicinal Chemistry* **2012**, *55* (11), 5165-5173.
70. Thomas, B.; Josep, A.-P.; Hongming, C.; Christian, M.; Christian, T.; Ola, E.; Kostas, P.; Atanas, P., *REINVENT 2.0 – an AI Tool for De Novo Drug Design*. 2020.
71. Li, Z.; Li, X.; Huang, Y.-Y.; Wu, Y.; Liu, R.; Zhou, L.; Lin, Y.; Wu, D.; Zhang, L.; Liu, H.; Xu, X.; Yu, K.; Zhang, Y.; Cui, J.; Zhan, C.-G.; Wang, X.; Luo, H.-B., 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* **2020**, 202010470.
72. Connelly, Colleen M.; Moon, Michelle H.; Schneekloth, John S., The Emerging Role of RNA as a Therapeutic Target for Small Molecules. *Cell Chemical Biology* **2016**, *23* (9), 1077-1090.
73. Warner, K. D.; Hajdin, C. E.; Weeks, K. M., Principles for targeting RNA with drug-like small molecules. *Nature reviews. Drug discovery* **2018**, *17* (8), 547-558.
74. Disney, M. D., Targeting RNA with Small Molecules To Capture Opportunities at the Intersection of Chemistry, Biology, and Medicine. *Journal of the American Chemical Society* **2019**, *141* (17), 6776-6790.
75. Tan, D.; Piana, S.; Dirks, R. M.; Shaw, D. E., RNA force field with accuracy comparable to state-of-the-art protein force fields. *Proc Natl Acad Sci U S A* **2018**, *115* (7), E1346-e1355.
76. Dajnowicz, S.; Ghoreishi, D.; Modugula, K.; Damm, W.; Harder, E. D.; Abel, R.; Wang, L.; Yu, H.

- S., Advancing Free-Energy Calculations of Metalloenzymes in Drug Discovery via Implementation of LFMM Potentials. *Journal of Chemical Theory and Computation* **2020**.
77. Kührová, P.; Mlýnský, V.; Zgarbová, M.; Krepl, M.; Bussi, G.; Best, R. B.; Otyepka, M.; Šponer, J.; Banáš, P., Improving the Performance of the Amber RNA Force Field by Tuning the Hydrogen-Bonding Interactions. *Journal of Chemical Theory and Computation* **2019**, *15* (5), 3288-3305.
78. Lemkul, J. A.; MacKerell, A. D., Jr., Polarizable force field for RNA based on the classical drude oscillator. *J Comput Chem* **2018**, *39* (32), 2624-2646.
79. Suda, A.; Koyano, H.; Hayase, T.; Hada, K.; Kawasaki, K.-i.; Komiyama, S.; Hasegawa, K.; Fukami, T. A.; Sato, S.; Miura, T.; Ono, N.; Yamazaki, T.; Saitoh, R.; Shimma, N.; Shiratori, Y.; Tsukuda, T., Design and synthesis of novel macrocyclic 2-amino-6-arylpyrimidine Hsp90 inhibitors. *Bioorganic & Medicinal Chemistry Letters* **2012**, *22* (2), 1136-1141.
80. Sarver, P.; Acker, M.; Bagdanoff, J. T.; Chen, Z.; Chen, Y.-N.; Chan, H.; Firestone, B.; Fodor, M.; Fortanet, J.; Hao, H.; Hentemann, M.; Kato, M.; Koenig, R.; LaBonte, L. R.; Liu, G.; Liu, S.; Liu, C.; McNeill, E.; Mohseni, M.; Sendzik, M.; Stams, T.; Spence, S.; Tamez, V.; Tichkule, R.; Towler, C.; Wang, H.; Wang, P.; Williams, S. L.; Yu, B.; LaMarche, M. J., 6-Amino-3-methylpyrimidinones as Potent, Selective, and Orally Efficacious SHP2 Inhibitors. *Journal of Medicinal Chemistry* **2019**, *62* (4), 1793-1802.
81. Farnaby, W.; Koegl, M.; Roy, M. J.; Whitworth, C.; Diers, E.; Trainor, N.; Zollman, D.; Steurer, S.; Karolyi-Oezguer, J.; Riedmueller, C.; Gmaschitz, T.; Wachter, J.; Dank, C.; Galant, M.; Sharps, B.; Rumpel, K.; Traxler, E.; Gerstberger, T.; Schnitzer, R.; Petermann, O.; Greb, P.; Weinstabl, H.; Bader, G.; Zoephel, A.; Weiss-Puxbaum, A.; Ehrenhöfer-Wölfer, K.; Wöhrle, S.; Boehmelt, G.; Rinnenthal, J.; Arnhof, H.; Wiechens, N.; Wu, M.-Y.; Owen-Hughes, T.; Ettmayer, P.; Pearson, M.; McConnell, D. B.; Ciulli, A., BAF complex vulnerabilities in cancer demonstrated via structure-based PROTAC design. *Nature Chemical Biology* **2019**, *15* (7), 672-680.

Table 1. Comparison of MM and QM energy profiles for ligands of the eight protein targets²⁰

Biological targets	XFF initial parameters*		Parameters from system specific refinement	
	RMSE in kcal/mol	R ²	RMSE in kcal/mol	R ²
Bace	1.18	0.70	0.45	0.93
Jnk1	0.84	0.92	0.67	0.97
Tyk2	1.26	0.88	0.69	0.96
Thrombin	1.33	0.91	1.09	0.94
CDK2	1.84	0.59	0.59	0.90
MCL1	1.76	0.65	1.34	0.79
PTP1B	1.09	0.78	0.71	0.90
p38	1.03	0.79	0.63	0.93

*The larger RMSE for MCL1 and Thrombin partially due to deviation at the energy barrier regions along torsional profiles between QM and MM.

Table 2. Summary for the performance of XFEP and FEP+ (RMSE in kcal/mol)

Targets	Nr. of cmps	Nr. of pairs	XFF+XFEP			OPLS2.1 ^(a)			OPLS3e ^(b)		
			RMSE	R ²	τ	RMSE	R ²	τ	RMSE	R ²	τ
Bace	36	58	1.10	0.18	0.23	1.05	0.35	0.37	1.11	0.39	0.43
CDK2	16	25	0.95	0.56	0.56	1.14	0.14	0.27	1.02	0.32	0.43
Jnk1	21	34	0.86	0.39	0.46	1.01	0.34	0.45	0.88	0.36	0.39
MCL1	42	71	1.44	0.30	0.39	1.44	0.25	0.37	1.19	0.27	0.37
P38	34	56	1.05	0.74	0.69	1.06	0.60	0.59	0.89	0.57	0.55
PTP1B	23	49	1.18	0.56	0.55	1.27	0.41	0.51	0.66	0.82	0.74
Thrombin	11	16	0.65	0.05	0.16	0.98	0.15	0.12	1.09	0.28	0.44
Tyk2	16	24	0.50	0.83	0.71	0.95	0.47	0.51	0.66	0.71	0.66
Overall	199	333	1.11	0.47	0.49	1.17	0.35	0.44	0.98	0.46	0.49

(a). Ref. ²⁰(b). Ref. ⁵⁷

Table 3. Comparison of the calculated and experimental $\Delta\Delta G$ values for the change of binding free energy between SGPB and OMTLY3 caused by mutations (All units in kcal/mol).

Mutations	XFEP*	Experimental
ALA 18	3.15	3.04
PHE 18	1.18	1.37
GLY 18	4.57	5.02
PRO 18	8.02	8.47
VAL 18	1.48	3.04

*Overall performance, MUE: 0.55 kcal/mol, RMSE: 0.76 kcal/mol, R^2 : 0.97, Tau: 0.95.

Table 4. Comparison of the calculated and experimental $\Delta\Delta G$ values for the change of binding free energy between Axitinib and kinase Abl caused by single mutations (All units in kcal/mol).

Mutations	XFEP*	Experimental
L248V	0.84	0.32
T315V	-1.91	-1.73
F486S	0.14	0.05

*Overall performance, MUE: 0.26 kcal/mol, RMSE: 0.32 kcal/mol, R^2 : 0.98.

Table 5. The selectivity of compound CEP-701 in term of binding affinity calculated by XFEP and measured by experiments (All units in kcal/mol).

	XFEP	Experimental
MAP4K1/HPK1	0.0	0.0
MAP4K2/GCK	-1.2	-1.2
MAP4K3/GLK	0.4	0.6
MAP4K4/HGK	-1.1	-0.3
MAP4K5/KHS	-0.2	0.1

*Overall performance, MUE: 0.33 kcal/mol, RMSE: 0.51 kcal/mol, R^2 : 0.90.

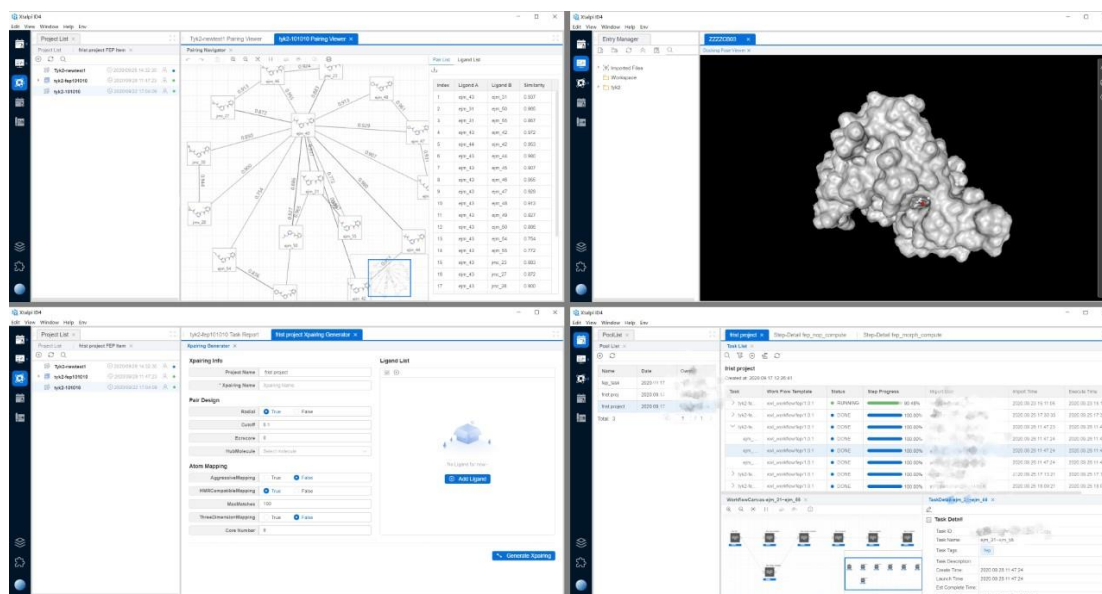


Figure 1. Examples of XFEP graphic user interface

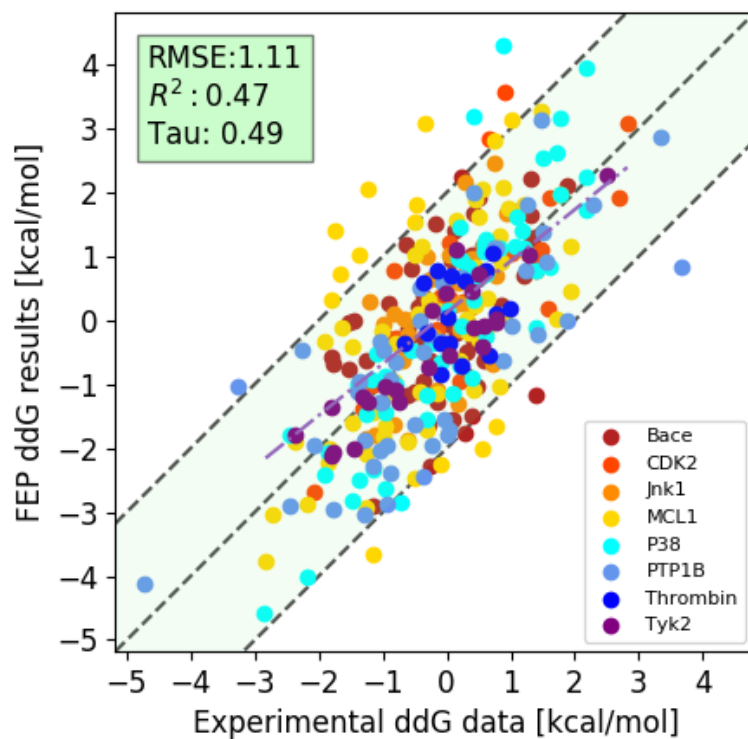


Figure 2. Correlation between FEP predicted relative binding free energies $\Delta\Delta G$ and experimental data for all eight systems. The dashed lines represent $y=x$ and $y=x\pm 2$, and the green area indicates that the FEP predicted $\Delta\Delta G$ are within 2 kcal/mol of their experimental values.

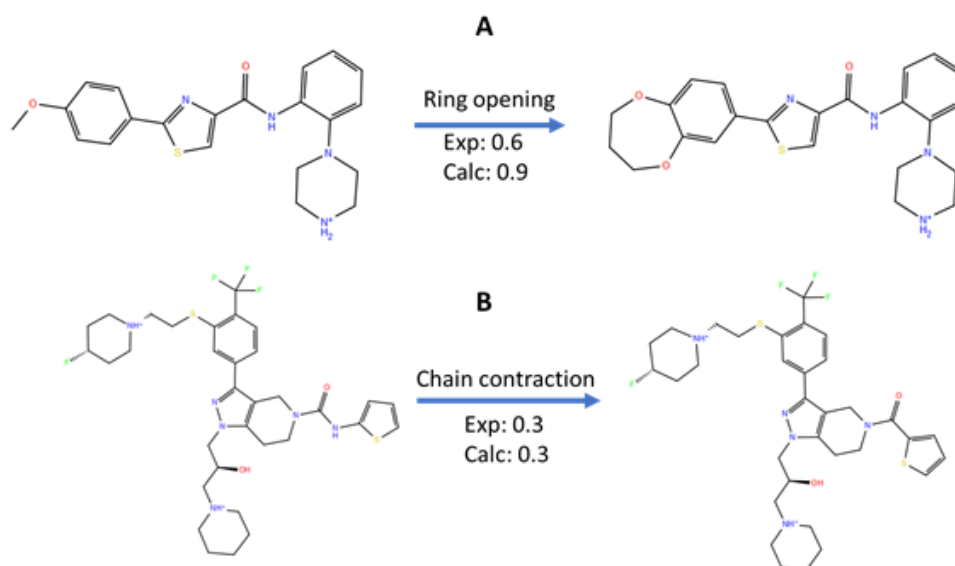


Figure 3. Examples of ring opening and chain contraction studied by XFEP. A) a ring closure transformation of two checkpoint kinase 1 ligands. B) a chain expansion transformation of two Cathepsin S ligands. Units are kcal/mol.

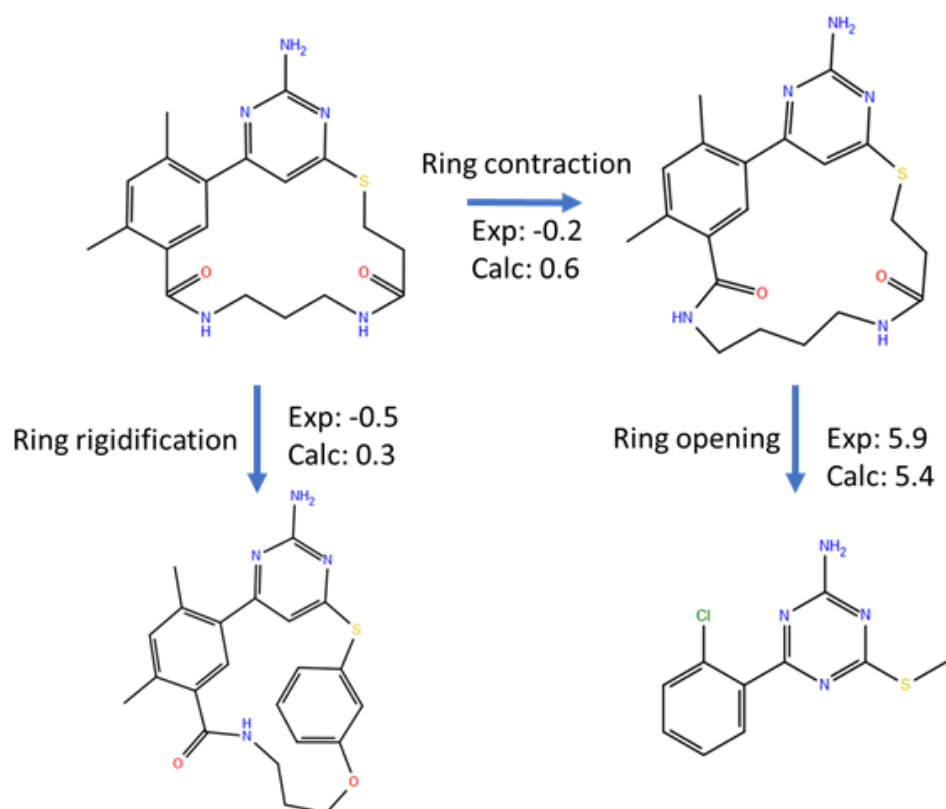


Figure 4. XFEP calculations on macrocyclic compounds of Hsp90⁷⁹. The transformations involve rigidification, contraction and ring-opening of the macrocycles.

Units are kcal/mol.

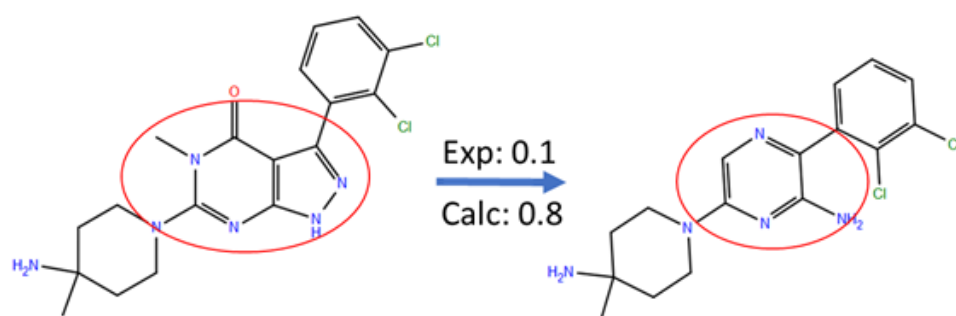


Figure 5. An example of FEP calculation on core-hopping using fragment-based design.

Core-hopping on the ligands of Tyrosine-protein phosphatase (SHP2)⁸⁰. Unit is kcal/mol.



42

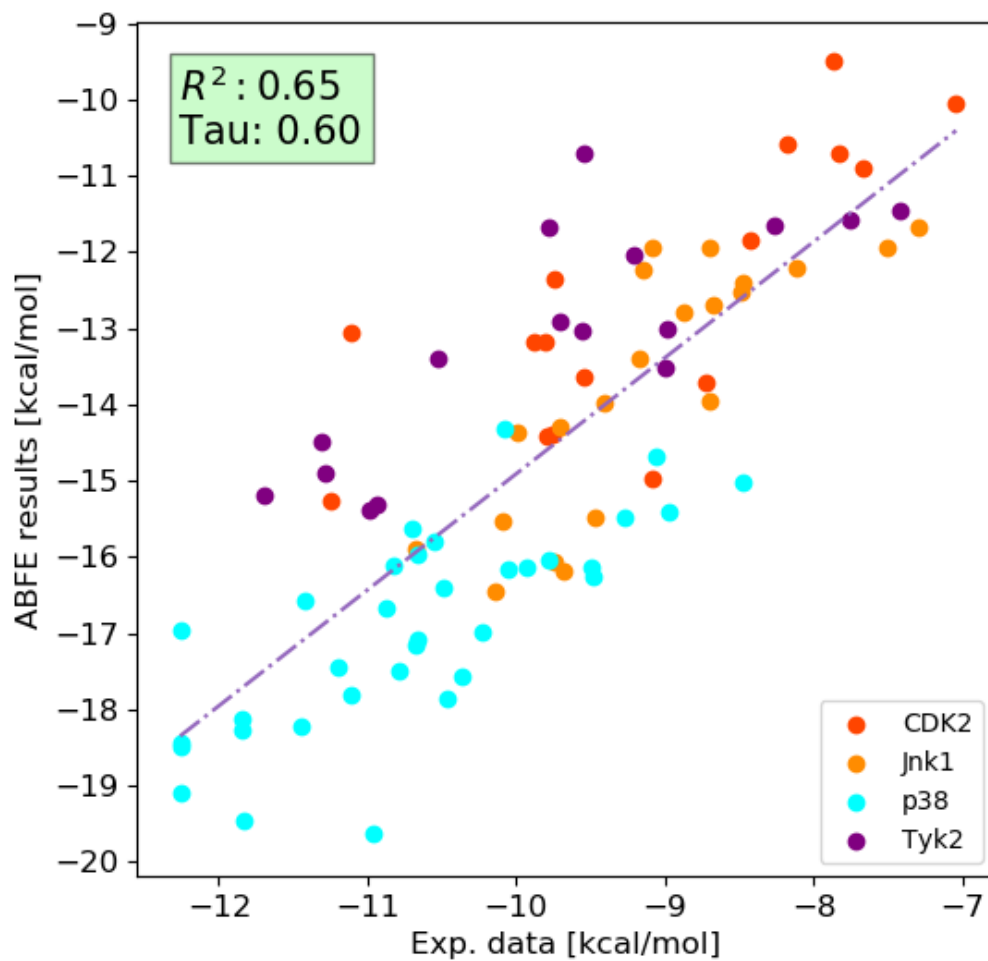


Figure 7. Correlation between ABFE predicted absolute binding free energies (ABFE) and experimental data for CDK2, Jnk1, p38 and Tyk2 compounds.

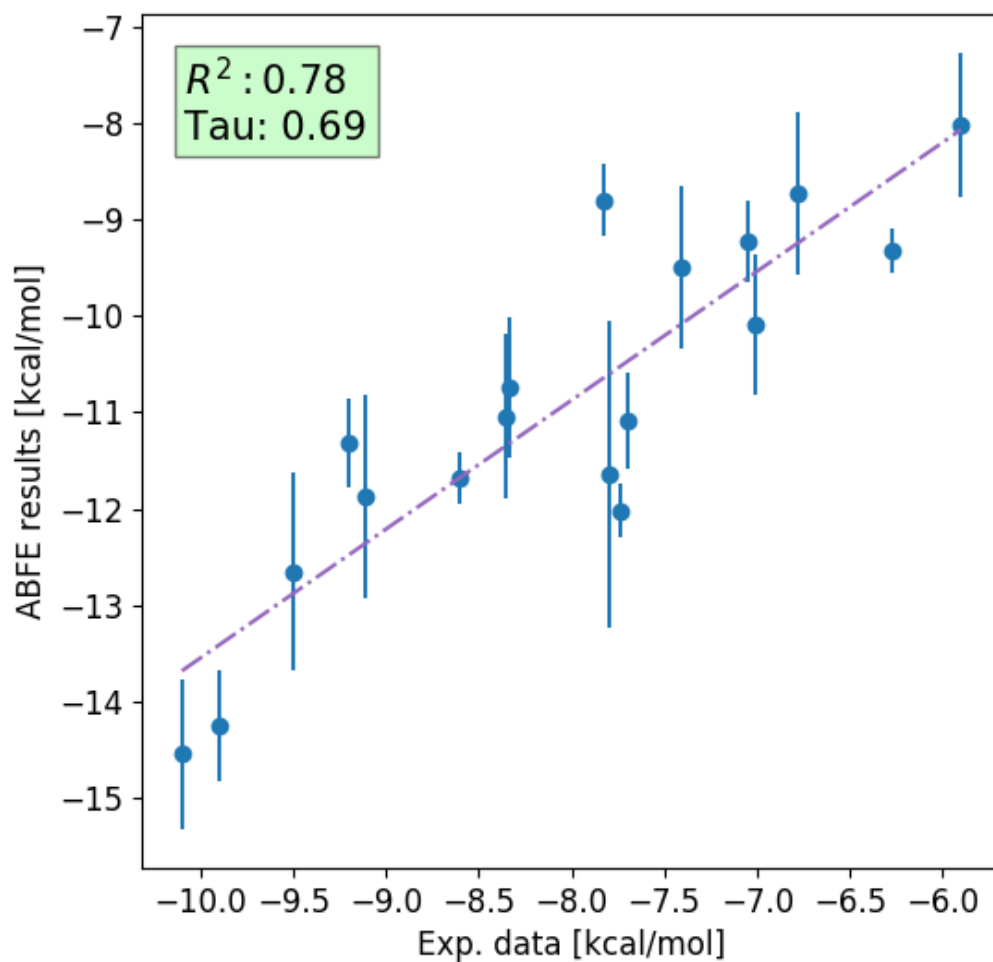


Figure 8. Correlation between ABFE predicted absolute binding free energies (ABFE) and experimental data for all BRD4 compounds.

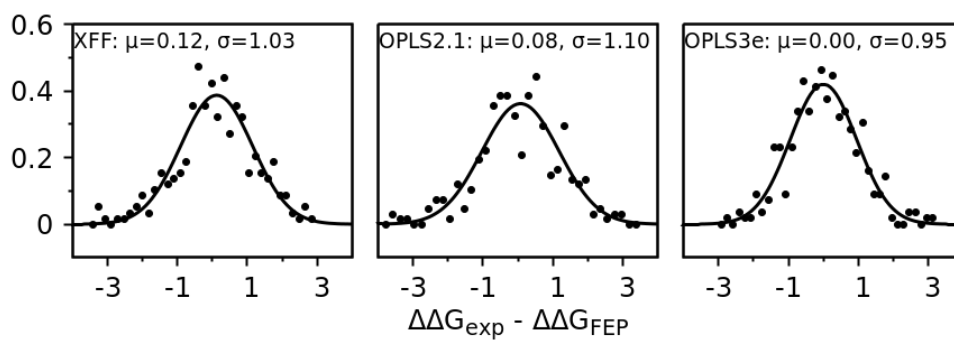


Figure 9. The probability density distribution of RBF predicted error $\Delta\Delta G_{\text{exp}} - \Delta\Delta G_{\text{FEP}}$ from different force fields. Solid line represents fitting to the 330 pairs with Gaussian distribution $N(\mu, \sigma)$.

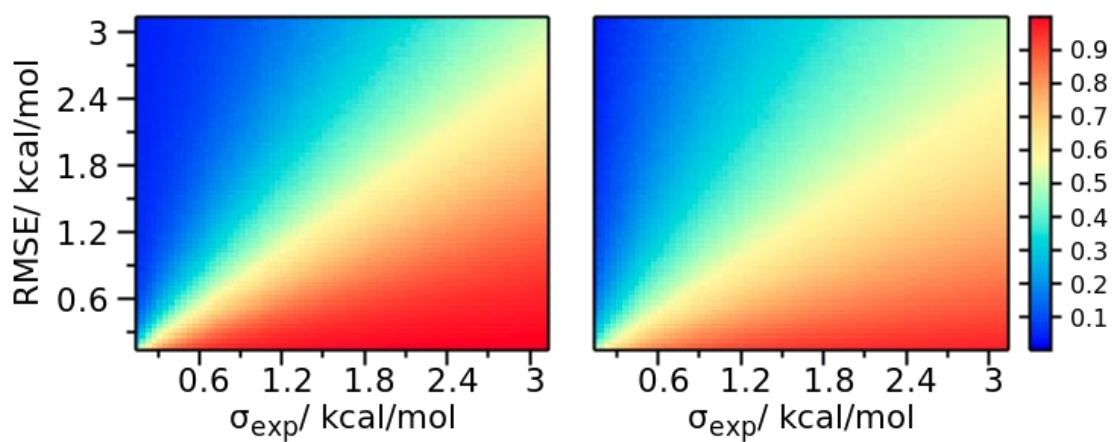


Figure 10. The distribution of R^2 and τ (different colors) along with the range of experimental values (represented by standard deviation of the mean value, σ_{exp} , in Gaussian distribution) and FEP predicted RMSE. Each point was averaged over 1000 independent generations for 30 pairs.