Enhancing hit discovery in virtual screening through accurate calculation of absolute protein-ligand binding free energies

Wei Chen,[1] Di Cui,[1] Steven V. Jerome,[2] Mayako Michino,[3,*] Eelke B. Lenselink,[4,*] David Huggins,[3] Alexandre Beautrait,[1] Jeremie Vendome,[1] Robert Abel,[1] Richard A. Friesner,[5] and Lingle Wang[1,*]

[1]Schrödinger, Inc., 1540 Broadway, 24th Floor, New York, NY 10036, United States
[2]Schrödinger, Inc., 10201 Wateridge Circle, Suite 220, San Diego, CA 92121, United States
[3]Tri-Institutional Therapeutics Discovery Institute, 413 E. 69th Street, New York, NY 10065, United States
[4]Galapagos NV, Generaal De Wittelaan L11 A3, 2800, Mechelen, Belgium
[5]Department of Chemistry, Columbia University, New York, NY 10027, United States

[*]Corresponding authors: lingle.wang@schrodinger.com, Bart.Lenselink@glpg.com, mmichino@tritdi.org

Abstract

In the hit identification stage of drug discovery, a diverse chemical space needs to be explored to identify initial hits. Contrary to empirical scoring functions, absolute protein-ligand binding free energy perturbation (ABFEP) provides a theoretically more rigorous and accurate description of protein-ligand binding thermodynamics and could in principle greatly improve the hit rates in virtual screening. In this work, we describe an implementation of an accurate and reliable ABFEP method in FEP+. We validated the ABFEP method on eight congeneric compound series binding to eight protein receptors including both neutral and charged ligands. For ligands with net charges, the alchemical ion approach is adopted to avoid artifacts in electrostatic potential energy calculations. The calculated binding free energies are highly correlated with experimental results with the weighted average of R^2 of 0.55 for the entire dataset and an overall RMSE of 1.1 kcal/mol when protein reorganization effect upon ligand binding was accounted for. Through ABFEP calculations using apo versus holo protein structures, we demonstrated that the protein conformational and protonation state changes between the apo and holo proteins are the main physical factors contributing to the protein reorganization free energy manifested by the overestimation of raw ABFEP calculated binding free energies using the holo structures of the proteins. Furthermore, we performed ABFEP calculations in three virtual screening applications for hit enrichment. ABFEP greatly improves the hit rates as compared to docking scores or other methods like metadynamics. The highly accurate ABFEP results demonstrated in this work position it as a useful tool to improve the hit rates in virtual screening, thus facilitate hit discovery.

Introduction

A primary objective of small molecule drug discovery is to design compounds that can tightly and selectively bind to a target protein. Accurate calculation of protein-ligand binding free energy is therefore of central importance in computational drug discovery. Benefiting from improved force fields and sampling algorithms and advanced hardware, rigorous free energy calculation by free energy perturbation (FEP) or related methods in explicit solvent simulations has dramatically improved the accuracy and begun to play an increasingly important role in modern computational drug discovery projects.[1,2,3] As an example, the FEP+ implementation of free energy calculations[4] has demonstrated a high level of accuracy in relative protein-ligand binding free energy calculations (RBFEP), with an overall root-mean-square error (RMSE) of about 1.1 kal/mol over a broad range of protein targets and ligand series.[5,6] It also enables the accurate modeling of very complex perturbations including scaffold hopping,[7] macrocyclization,[8] net-charge changes,[9] fragment linking,[10] and linker enumeration.[10] The high reliability and accuracy across a broad range of complex chemical modifications has also been validated in a large number of prospective studies in active drug discovery projects, positively impacting the projects through faster identification of novel potent chemical matters.[3] However, relative-binding free energy calculation through RBFEP can only be applied on congeneric series of ligands with similar binding modes and scaffolds, limiting its applications to only the hit-to-lead and lead optimization stages of drug discovery where the structure of the binding complex of an initial reference ligand with the target receptor was known.

During the hit discovery stage of drug discovery, a diverse chemical space needs to be explored to identify initial hits. As the current best practices, empirical scoring functions are used in virtual screening to dock a large library of compounds. Due to the limited accuracy of empirical scoring functions, the hit rate in virtual screening is usually very low, about 1-2% on average, with only a few confirmed hits for most screenings and sometimes not a single hit for challenging targets.[11,12] Absolute protein-ligand binding free energy calculation through free energy perturbation (ABFEP) provides a theoretically more rigorous description of protein-ligand binding thermodynamics, offering hope to dramatically improve the hit rates by rescoring the top compounds in virtual screening. However, due to the complexity for the implementation of ABFEP methods, the difficulty to converge the simulations to a level useful in practical applications, and the large computational cost associated in these calculations, accurate and reliable calculations of protein-ligand binding free energies through ABFEP in practical virtual screening settings for hit discovery have not been reported yet.

The first ABFEP method has been proposed decades ago through the construction of a non-physical alchemical pathway.[13,14,15] The method involves the calculation of the free energy to transfer the ligand from the solution to the gas phase, and the free energy to transfer the ligand from the protein binding pocket to the gas phase. The difference in the above two free energies corresponds to the absolute binding free energy of the ligand. Initial applications of the method have been focused on model systems, such as fragments binding to T4 lysozyme[16,17,18] and FK506-binding protein[19] and host-guest systems,[20,21] resulting in a reasonable accuracy with RMSE between the calculated and experimental binding free energies of 2-3 kcal/mol, with the goal to showcase the feasibility of the method. Due to the large complexity and computational cost, over a very long period of time, the majority of ABFEP literature on real protein-ligand systems only reported calculations of a small number of compounds binding to a handful of protein receptors, including FK506-binding proteins[22,23,24] and bromodomain-containing proteins.[25,26,27] The accuracies of the reported calculations varied, with the RMSE between calculation and experiment ranging from 1 to 3 kcal/mol.

Until very recently, with the great increase of computer power, ABFEP calculations on a medium-to-large number of drug-like compounds for multiple protein targets were reported. In one such study, Li et al. performed ABFEP simulations on 7 proteins and 101 congeneric ligands and reported surprisingly high accuracy with RMSEs of 0.6-1.5 kcal/mol and R^2 of 0.5-0.9 between calculated and experimental binding affinities, though the reported high accuracy was partially due to the removal of the systematic difference between calculated and experimental binding free energies (the RMSEs of the raw data are 0.9-5 kcal/mol).[28] In another study, Lin et al. performed ABFEP simulations for 5 proteins and 105 congeneric ligands, and obtained RMSEs of 3-6 kcal/mol and R^2 of 0.6-0.8.[29] Another study from Khalak et al. employed a non-equilibrium method incorporating the apo states of the proteins in the ABFEP calculations, and obtained RMSEs of 0.8-3 kcal/mol and R^2 of 0.02-0.76 on 7 proteins and 128 congeneric ligands.[30] In all these studies, ABFEP was applied on congeneric ligands where RBFEP worked better, with the goal

to validate the implementation. In addition, except for the work of Li et al.[28] with a few ligands carrying a net charge, all the other studies have focused exclusively on neutral ligands. Therefore, it is not clear how these methods would work in a practical virtual screening setting to score ligands with diverse structures and binding modes, particularly for ligands with net charges where the finite size effects[31] and strong electrostatic interactions between the protein and ligand are known to be prohibitively difficult to converge.

In this paper, we report an accurate and reliable ABFEP method implemented in the FEP+ program.[4] We validated the implementation on all the eight protein systems and 199 ligands from our previous RBFEP paper,[5] which was later used as the benchmark systems for free energy calculations by many groups. Four of the systems contain neutral ligands while the other four include ligands with net charges. To calculate the binding free energies of the charged ligands, the alchemical ion approach used for the charge-changing perturbations in RBFEP[9] was adopted in ABFEP as well. Using the holo conformation of the protein receptor, the raw ABFEP calculated binding free energies are systematically more negative (favorable) than experiment. This is expected since the apo versus holo protein conformational and/or protonation/tautomeric state changes induced upon ligand binding was not sampled in the relatively short simulations. This is further verified through ABFEP calculations using the apo conformation of the protein, where the calculated free energies are slightly more positive than experiment, providing evidence to elucidate a long-standing puzzle in the literature regarding how ABFEP calculations should be interpreted for real protein-ligand systems. After removing the systematic shift between the calculated and experimental binding free energies to account for the protein reorganization effect, the overall RMSE between calculation and experiment for the entire dataset is 1.1 kcal/mol with a weighted average R^2 of 0.55, comparable with the RBFEP results on the same dataset (RMSE of 0.9 kcal/mol and R^2 of 0.56). Comparing with previous ABFEP studies,[28,29,30] which have reported results on some of the systems, the accuracy of our results is comparable or better. We further validated the accuracy of our ABFEP implementation through three pilot applications in virtual screening to improve the hit rates. After docking a large library with millions or even billions of compounds, a short list with hundreds or thousands of compounds with best docking scores is generated and ABFEP is used to rescore the compounds in the short list to propose the final buy list.[32] On one virtual screening dataset taking from the literature for JAK2[33] and two virtual screening collaboration datasets for proprietary targets, ABFEP greatly improved the hit rates as compared to docking scores or alternative methods like metadynamics, to our knowledge the first validation of ABFEP to improve hit discovery in practical virtual screening settings.

Methods

Double decoupling method for ABFEP

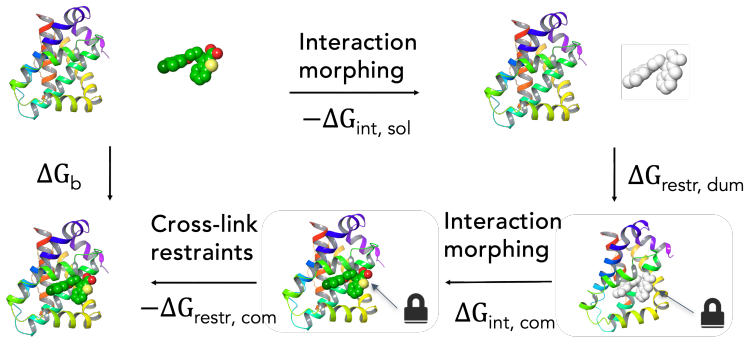

Figure 1: Thermodynamic cycle for ABFEP. The binding of the ligand to the protein receptor is decomposed into a few alchemical steps similar to the originally proposed double decoupling scheme.[14,15] Starting from the physical ligand in water, the vdw and electrostatic interactions within the ligand and between the ligand and water are slowly turned off, i.e., $-\Delta G_{int,sol}$ in the cycle; then the relative position and orientation of the dummy ligand with respect to the protein binding pocket are restrained through a set of cross-link restraints ($\Delta G_{restr,dum}$); in the third step, the intra-ligand and ligand-environment vdw and electrostatic interactions for the restrained ligand are slowly turned on in the protein binding pocket ($\Delta G_{int,com}$) followed by relaxing the cross-link restrains when ligand interactions are fully turned on ($-\Delta G_{restr,com}$). The binding free energy is the sum of the free energies of these processes. In our protocol, the free energies for the two horizontal legs, $-\Delta G_{int,sol}$ and $\Delta G_{int,com}$-$\Delta G_{restr,com}$, are calculated through two independent simulations in solvent and protein complex, respectively, while the free energy of the right vertical leg, $\Delta G_{restr,dum}$, is calculated analytically.

Our ABFEP protocol is similar to the originally proposed double decoupling scheme[14,15] with the thermodynamic cycle shown in Fig. 1. The binding of the ligand to the protein receptor is decomposed into a few alchemical steps. First, starting from the physical ligand in water, the vdw and electrostatic interactions within the ligand and between the ligand and water are slowly turned off until the ligand becomes dummy; second, the dummy ligand is attached to the protein binding pocket through a set of cross-link restraints similar to what was proposed by Boresch et al.;[15] in the third step, the intra-ligand and ligand-environment vdw and electrostatic interactions for the restrained ligand are slowly turned on in the protein binding pocket and the cross-link restrains are relaxed after that. The free energy to turn on/off the intra-ligand and ligand-water interactions ($-\Delta G_{int,sol}$) is calculated by one simulation in solvent, the free energy to turn on/off the intra-ligand and ligand-environment interactions and relaxing the restrain potentials in the protein binding pocket ($\Delta G_{int,com}$-$\Delta G_{restr,com}$) is calculated by another simulation in the protein binding pocket, and the free energy to restrain the relative position and orientation of the dummy ligand with respect to the protein binding pocket ($\Delta G_{restr,dum}$,) is calculated analytically with detailed derivations in the following section. The absolute binding free energy (ΔG_b) is the sum of these terms as follows:

$$\Delta G_b = \Delta G_{int,com} - \Delta G_{int,sol} + \Delta G_{restr,dum} - \Delta G_{restr,com} \tag{1}$$

Cross-link restraints for ABFEP

The relative position and orientation of the dummy ligand with respect to the protein binding pocket is restrained by a set of cross-link restrains originally proposed by Boresch et al.[15] Three protein atoms (a, b, and c) and three ligand atoms (A, B, and C) are selected for setting up cross-link restraints (Fig. 2). One distance r_{aA}, two angles θ_{Aab} and θ_{aAB}, and three dihedral angles ϕ_{baAB}, ϕ_{Aabc} and ϕ_{aABC}, are restrained by harmonic potentials. The free energy difference for adding cross-link restraints can be calculated as

$$\Delta G_{restr,dum} = -k_B T ln \frac{Z_{CL}}{Z_P Z_L}, \tag{2}$$

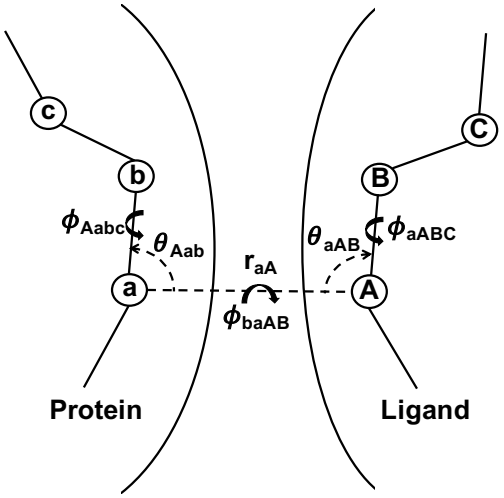

Figure 2: Cross-link restraints between the protein and ligand for ABFEP. Following the work of Boresch et al.,[15] three protein atoms (a, b, and c) and three ligand heavy atoms (A, B, and C) are selected. A distance r_{aA}, two angles θ_{Aab} and θ_{aAB}, and three dihedral angles ϕ_{baAB}, ϕ_{Aabc} and ϕ_{aABC} are restrained by harmonic potentials.

where Z_P, Z_L, and Z_{CL} are the partition functions for the protein, the free dummy ligand, and the dummy ligand-protein complex with the above cross-link restrains, respectively (right vertical leg in Fig. 1). Since the dummy ligand does not have interactions with the protein and water except for the cross-link restrains, an analytical formula for the above free energy can be easily derived:

$$\Delta G_{restr,dum} = -k_B T ln \frac{Z_r}{8\pi^2 V} = -k_B T ln \frac{Z^{dist}_{r,aA} Z^{ang}_{r,aAB} Z^{ang}_{r,Aab} Z^{dihed}_{r,baAB} Z^{dihed}_{r,Aabc} Z^{dihed}_{r,aABC}}{8\pi^2 V}, \tag{3}$$

where

$$\begin{aligned} Z^{dist}_r &= \int_0^\infty exp(-\beta K_r (r - r_0)^2) r^2 \, dr \\ &= \frac{r_0}{2\beta K_r} exp(-\beta K_r r_0^2) + \frac{\sqrt{\pi}}{4\beta K_r \sqrt{\beta K_r}} (1 + 2\beta K_r r_0^2)(1 + erf(\sqrt{\beta K_r} r_0), \end{aligned} \tag{4}$$

$$Z^{ang}_r = \int_0^\pi exp(-\beta K_\theta (\theta - \theta_0)^2) sin\theta d\theta \approx \sqrt{\frac{\pi}{\beta K_\theta}} exp(-\frac{1}{4\beta K_\theta}) sin\theta_0, \tag{5}$$

$$Z^{dihed}_r = \int_{\phi_0 - \pi}^{\phi_0 + \pi} exp(-\beta K_\phi (\phi - \phi_0)^2) d\phi = \sqrt{\frac{\pi}{\beta K_\phi}} erf(\pi \sqrt{\beta K_\phi}), \tag{6}$$

and V is 1660 $\mathring{A}^3$ for the standard state. In Eqns. 4-6, K_r, K_θ and K_ϕ are the force constants and r_0, θ_0 and ϕ_0 are equilibrium values for the distance, angle, and dihedral restraints, respectively.

Structure preparation

The input structures for ABFEP calculations were obtained from the Protein Data Bank (PDB)[34] and prepared by the Protein Preparation Wizard[35] in Maestro[36] with the default settings. For the eight congeneric series of ligands, the same crystal structures as that used for RBFEP benchmark[5] were also used for the ABFEP calculations: 4DJW[37] for BACE1, 1H1Q[38] for CDK2, 2GMX[39] for JNK1, 4HW3[40] for MCL1, 3FLN for P38, 2QBS[41] for PTP1B, 2ZFF for Thrombin and 4GIH[42] for TYK2. For seven of the systems, apo protein structures were available in PDB, and the following structures were used for ABFEP calculations using apo structures: 1SGZ[43] for BACE1, 1H27[44] for CDK2, 3O17 for JNK1, 6QB3[45] for MCL1, 1WFC[46] for P38, 2CM2[47] for PTP1B and 3D49 for Thrombin. The binding poses of compounds were taken from the previous RBFEP work.[5]

For the JAK2 virtual screening dataset, crystal structure with PDB ID 4FVR was used.[48] The top ranking poses from Glide SP docking[49] in the work of Cutrona et al.[33] were used as input for the ABFEP simulations. One compound (BI-D1870) has two enantiomers, and both of them were modeled. For both collaborations with the Tri-I TDI and Galapagos, an in-house structure (co-crystalized with a ligand) was used for the ABFEP rescoring. The structure was prepared using the Protein Preparation Wizard[35] in Maestro[36] with the default settings.

Simulation details

OPLS4 force field[6] was used for all simulations. In each ABFEP simulation, the protein-ligand complex was solvated in an orthorhombic SPC[50] water box. The buffer width was 5 Å for a neutral ligand and 8 Å for a charged ligand (a charged ligand was defined as a ligand with a net charge or charged groups). For charged ligands, the protein ligand complexes were neutralized by counter-ions and additional salt ions of 150 mM were added to mimic the buffer solution of experimental assay. The systems were then relaxed by a series of short molecular dynamics (MD) relaxations including: (1)100 ps Brownian Dynamics at 10 K with solute heavy atoms restrained (force constant 50 kcal/mol/Å^2) to relieve minor steric clashes; (2) 12 ps NVT simulation at 10 K with solute heavy atoms restrained; (3) 20 ps Grand Canonical Monte Carlo (GCMC) μVT simulations[51] at 300 K with solute heavy atoms restrained to solvate the binding pocket; (4) 20 ps GCMC μVT simulation at 300 K with protein backbone heavy atoms restrained. After relaxation, a 1-ns GCMC μVT simulation at 300 K was performed with protein backbone heavy atoms restrained.

To identify the optimal set of atoms for the protein-ligand cross-link restraints, the interactions between the protein and ligand during the 1-ns MD relaxation were analyzed and the set of atoms with most frequent hydrogen bond or salt-bridge interactions in MD were selected to be restrained. In particular, the hydrogen bond and salt bridge interactions between the protein and ligand occurred in any frame of the MD trajectories were identified, and their frequencies were collected. The frequencies of atom-based interactions were then summed and assigned to a ligand non-terminal heavy atom-protein residue pair as follows: (1) any protein atom in an interaction is assigned to the corresponding residue it belongs to; (2) any ligand atom in an interaction is assigned to a non-terminal heavy atom, which is bonded to at least two other heavy atoms. Say for example a terminal oxygen atom in a carboxylate group of the ligand forming a hydrogen bond with the protein is assigned to the carbon atom of the carboxylate group. If multiple ligand atom-protein residue pairs have hydrogen bond or salt bridge interaction frequencies of at least 50%, the ligand atom closest to the centroid of the ligand was selected as one of the atoms for the cross-link restraints (terminal groups in the ligand like SO_2 that can easily flip orientation were excluded). After the selection of the anchoring atom in the ligand (atom A in Fig. 2), two other ligand heavy atoms that atom A is bonded to (B and C in Fig. 2), and the three backbone atoms (N, C_α and C) of the protein residue forming hydrogen bond and/or salt-bridge with atom A (atoms a, b and c in Fig. 2) were selected for the cross-link restraints. To avoid co-linear geometry leading to the singularity in the dihedral angle restraints, we limited the set of restrained atoms to those with the four angles (θ_{Aab}, θ_{aAB}, θ_{abc} and θ_{ABC}) between 45 and 135 degrees. If none of the ligand atom-protein residue pairs with interaction frequency of at least 50% satisfied the above criteria, the ligand heavy atom closest to the centroid of the ligand, two of its bonded heavy atoms (atoms A, B and C in Fig. 2) and three protein C_α atoms (atoms a, b and c in Fig. 2), which satisfy the above angle requirement, were selected to be restrained. The force constant was 1 kcal/mol/Å^2 for the distance restraint, and 40 kcal/mol/rad^2 for the angle and dihedral restraints.

After the MD relaxation and the selection of atoms for the cross-link restraints, a representative structure from the MD trajectory was used as the input for the following FEP simulations. To select the representative structure, the mean value of each rotatable bond in the ligand sampled during the MD trajectory was calculated, and the representative structure had the ligand torsions closest to the corresponding mean values. For charged ligands, counter-ions and additional salt ions of 150 mM were again added the same as that in the MD relaxation, and the same alchemical ion approach introduced for charged perturbation in RBFEP[9] were also adopted for ABFEP. A total of 68 and 108 λ windows were used for neutral and charged ligands, respectively. Each replica was run for 5 ns. For the solvent leg FEP, the ligand was extracted from the representative structure selected above and then solvated in a SPC water box with 10 Å buffer width. Again, counter-ions and additional salt ions of 150 mM were added and the alchemical ion approach was utilized for charged ligands. 60 λ windows were used for all ligands and each replica was run for 5 ns.

To benchmark the accuracy of ABFEP versus RBFEP on the congeneric series of ligands, we also performed RBFEP calculations with OPLS4 on the eight congeneric series of ligands taken from Wang et

al.,[5] and the default RBFEP protocol as detailed in that paper was used for the RBFEP calculations. The RBFEP simulations lasted for 5 ns per replica.

To calculate the pK_a's of aspartic acids in the binding pocket of BACE1, protein residue mutation FEP[52] implemented in FEP+[4] was used. The free energy to mutate from the neutral to the charged ASP was calculated in the protein environment and for an isolated residue in solvent, and the difference corresponds to the shift of pKa due to protein environment, i.e., $\Delta\Delta$G = RTln10 (pK_a- 3.67), where 3.67 is the pK_a of an isolated Asp residue in solvent.[53] 24 λ windows were used and each replica was run for 20 ns.

Results and Discussion

Large-scale test on congeneric compound series for eight proteins

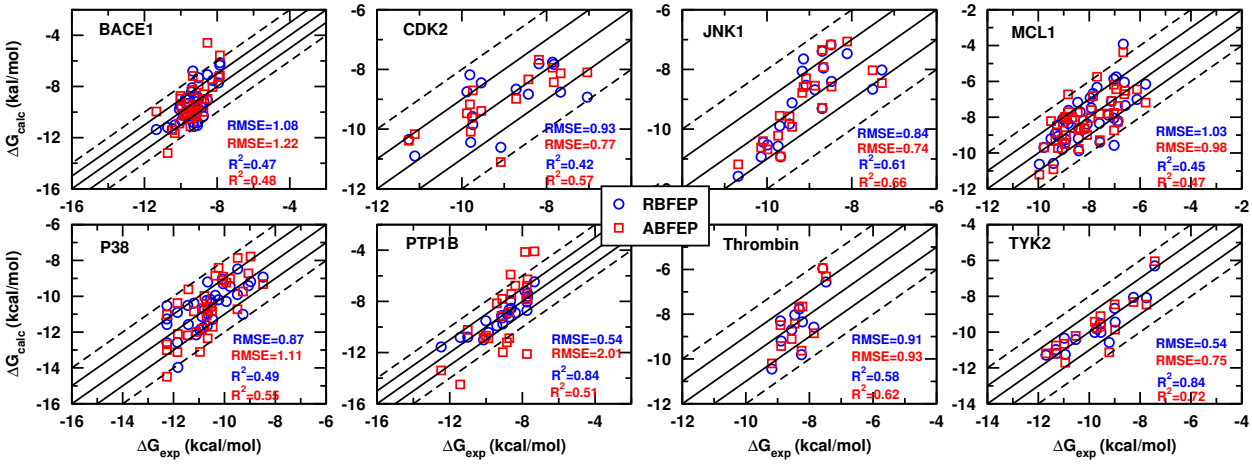

Figure 3: Comparison of calculated binding free energies versus experiment for both ABFEP and RBFEP. Blue circles and red squares are for RBFEP and ABFEP, respectively. The results of the first repeat of the ABFEP simulation are shown, and the results from the other two repeats were reported in supporting information. The raw ABFEP binding free energies were shifted to take into account the protein reorganization effect as detailed in the main text. Solid diagonal lines indicate the region within 1 kcal/mol of experimental values, while dashed diagonal lines indicate the region within 2 kcal/mol of experimental values. RMSE and R^2 are labeled.

We tested the ABFEP method on the entire dataset from the previous RBFEP benchmark[5] including a total number of 199 compounds binding to eight different protein systems. To gauge the convergence and consistency of the calculations, three replicas of independent ABFEP simulations with different random seeds were performed for each ligand. The ABFEP performance on these eight systems were presented in Fig. 3 (the results from the first replica were shown for simplicity, and results from the other two replicas are shown in Figs. S1 and S2) and summarized in Table I. Four series of ligands binding to BACE1, MCL1, PTP1B and Thrombin are charged while the other four are neutral. For comparison, the results from the previously established RBFEP method[5] were also shown. The raw ABFEP results for all these systems were more negative than experimental values, reflecting the miss of the protein reorganization effect for the structural difference between the apo protein and the holo complex due to limited sampling in the short ABFEP simulations. For proper comparison between ABFEP results and experiment, the protein reorganization contribution to the binding free energy for each system was estimated as the difference between the average of experimental binding free energies for all the ligands in that system and the average of the raw ABFEP binding free energies for the same set of ligands. The estimated protein reorganization contribution for each system was added to the raw ABFEP values for the final comparison with experimental data. The estimated protein reorganization contributions ($\Delta G_{\mathrm{prot-reorg}}$) are shown in Table I, and are ranged between 1.22 kcal/mol for CDK2 and 9.97 kcal/mol for PTP1B.

Taking protein reorganization contribution into account, the ABFEP calculated binding free energies for all systems agree very well with experiment (Fig. 3), with RMSE between calculation and experiment

Table I: Statistics of ABFEP and RBFEP results for congeneric compound series

Protein	# of ligands	Ligand charge	RBFEP		[a]ABFEP with holo-protein			[b]ABFEP with apo-protein		
			RMSE	R^2	[d]RMSE	R^2	$^c\Delta G_{\mathrm{prot-reorg}}$	[d]RMSE	R^2	$^c\Delta G_{\mathrm{prot-reorg}}$
BACE1	36	+1	1.08	0.47	1.21 ± 0.07	0.44 ± 0.04	2.91 ± 0.28	0.89 ± 0.06	0.26 ± 0.06	[e]-1.61 $\pm$ 0.27
CDK2	16	0	0.93	0.42	0.84 ± 0.06	0.50 ± 0.07	1.22 ± 0.20	1.05 ± 0.06	0.33 ± 0.10	-0.59 $\pm$ 0.10
JNK1	21	0	0.84	0.61	0.85 ± 0.13	0.63 ± 0.02	3.50 ± 0.12	0.64 ± 0.09	0.69 ± 0.09	2.72 $\pm$ 0.12
MCL1	42	-1	1.03	0.45	0.95 ± 0.06	0.53 ± 0.07	3.65 ± 0.15	1.03 ± 0.06	0.33 ± 0.04	2.48 $\pm$ 0.12
P38	34	0	0.87	0.49	1.09 ± 0.10	0.58 ± 0.06	5.27 ± 0.09	1.13 ± 0.07	0.39 ± 0.03	-1.64 $\pm$ 0.08
PTP1B	23	-2	0.54	0.84	1.78 ± 0.20	0.55 ± 0.03	9.97 ± 0.21	1.20 ± 0.06	0.50 ± 0.03	-1.12 $\pm$ 0.07
Thrombin	11	+1	0.91	0.58	1.01 ± 0.07	0.57 ± 0.07	1.82 ± 0.14	1.03 ± 0.19	0.48 ± 0.12	2.08 $\pm$ 0.08
TYK2	16	0	0.54	0.84	0.74 ± 0.02	0.74 ± 0.04	3.54 ± 0.17			
ALL	199		0.90	[f]0.56	1.13 ± 0.03	[f]0.55 ± 0.003		1.02 ± 0.02	[f]0.40 ± 0.02	

[a]ABFEP simulations using the crystal structures of holo-proteins. [b]ABFEP simulations using the crystal structures of apo-proteins. $^c\Delta G_{\mathrm{prot-reorg}}$: protein reorganization contribution estimated as the difference between the average of experimental binding free energies and the average of the raw ABFEP binding free energies for all ligands in a system. [d]RMSE was calculated after $\Delta G_{\mathrm{prot-reorg}}$ was added to the raw ABFEP binding free energies. $^e\Delta G_{\mathrm{prot-reorg}}$ for BACE1 using the apo protein structure was calculated after adding pK_{a} correction of 3.57 kcal/mol (see the main text for details) to raw ABFEP binding free energies. [f]Average R^2 weighted by the numbers of ligands in the eight systems for ABFEP with holo-proteins or the seven systems for ABFEP with apo-proteins. The ABFEP simulations were repeated for three times with different random seeds. The mean and standard deviation among the three repeats are reported in the table.

ranged between 0.74 kcal/mol for TYK2 system and 1.78 kcal/mol for PTP1B, and R^2 ranged between 0.44 for CDK2 and 0.74 for TYK2 (Table I). The overall RMSE and weighted average R^2 from ABFEP for the entire dataset is 1.13 kcal/mol and 0.55, respectively, slightly worse than the RBFEP performance (overall RMSE and weighted average R^2 of 0.90 kcal/mol and 0.56, respectively), due to much larger configurational space to sample in ABFEP. It should be noted that the protein reorganization contribution is a constant among all the ligands binding to the same protein conformation, thus does not affect the R^2 and rank ordering of the predictions. Our estimate of protein reorganization contribution minimizes the RMSE between prediction and experiment, but would not affect the prospective usage of ABFEP for compound selection and prioritization.

Among these eight systems, the protein reorganization contribution for PTP1B system of 9.97 kcal/mol is significantly larger than the other systems (in the range of 1.22 to 3.65 kcal/mol), and the overall RMSE for PTP1B system of 1.78 kcal/mol is also significantly worse than the others (in the range of 0.74 to 1.21 kcal/mol). To understand the possible reasons leading to the difference in the performance of ABFEP on PTP1B versus the other systems, we searched for an apo crystal structure of PTP1B (PDB ID 2CM2),[47] and compared it with the holo crystal structure (PDB ID 2QBS[41]) used in the ABFEP simulations (Fig. 4A). The large conformational changes in the WPD-loop (Thr177-Pro188) of the binding pocket from an open conformation for the apo protein to the closed conformation for the holo protein[47] could potentially explain the large protein reorganization contribution for this system. To validate this hypothesis, we repeated the ABFEP simulations for the PTP1B ligands using the apo conformation of the receptor, and the resulting calculated binding free energies were more positive than experiment values by about 1 kcal/mol (Table I). This again is expected as the closed WPD-loop conformation is preferred in the presence of ligands but could not be sampled in the ABFEP simulations using the apo structure of the protein with the WPD-loop in the open conformation. These results validated our hypothesis that the more negative binding free energies from the raw ABFEP results using the holo conformation of the protein as compared to experiment is due to the protein reorganization contribution between the apo and holo conformations of the receptor not sampled in the ABFEP simulations.

Surprisingly, the RMSE between the ABFEP calculated and experimental binding free energies (after accounting for the protein reorganization effect) for PTP1B ligands is greatly reduced to 1.2 kcal/mol using the apo conformation of the receptor (so does the variance of the ABFEP results among three independent repeats shown in Table I), suggesting that the simulations with the apo-structure converged much faster. We attribute the faster convergence of the calculated free energies using the apo structure to the much weaker hydrogen bond and salt bridge interactions between the two carboxylate groups conserved in all the ligands and surrounding charged residues in the binding pocket when the WPD-loop is open. In the holo structure, the two carboxyl groups in the ligands form multiple strong interactions with the protein including three salt bridges with Arg221 and Lys120, which were persistent in the ABFEP simulations (Fig. S3). In contrast, the two salt bridge interactions with Arg221 were partially lost in the open conformation of the WPD-loop in the apo structure when exposed to solvent.

Besides PTP1B, we also investigated the possible physical factors contributing to the protein re-

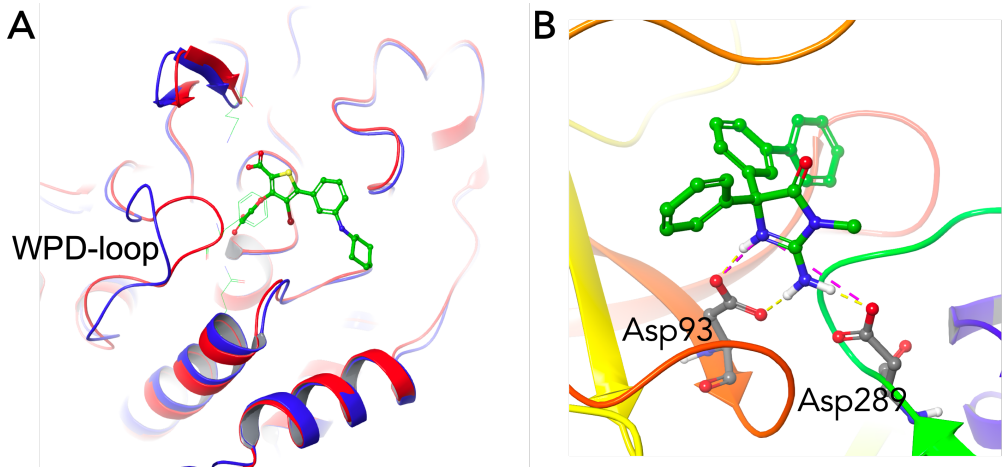

Figure 4: Conformational or protonation state change caused by ligand binding. **A.** Comparison between holo- and apo-crystal structures of PTP1B. The holo-crystal structure of PTP1B (red, PDB ID 2QBS[41]) is aligned to the apo-crystal structure (blue, PDB ID 2CM2[47]). The WPD-loop (Thr177-Pro188), which has the largest difference, is labeled. **B.** Aspartic acids in the binding pocket of BACE1. Two Asps are located at the binding pocket of BACE1 (PDB ID 4DJW[37]) and forming salt bridges with the ligand.

organization effects observed for other systems, although the magnitudes are much smaller than that for PTP1B (Table I). For seven out of the eight systems except for TYK2, we found apo crystal structures of the proteins from PDB. Repeating the ABFEP simulations using the apo structures, the raw ABFEP results of all systems except for Thrombin consistently got more positive as compared to that from the holo crystal structures (Table I), indicating that the conformational differences between the apo and holo structures are indeed the dominant factors contributing to the observed protein reorganization effects from the ABFEP calculations. For Thrombin, the apo and holo structures are almost identical, and the ABFEP results were not affected by the choice of input protein structures. For CDK2 and P38, the P-loop conformations are slightly different between the apo and holo structures, and similar to PTP1B the raw ABFEP results using the apo protein structures were slightly more positive than the experimental results. Therefore, the P-loop conformational changes between the apo and holo structures are fully responsible to the observed protein reorganization effects for these two systems.

For BACE1, ABFEP binding free energies using the apo protein structure are slightly more positive than that using the holo protein structure ($\Delta G_{\mathrm{prot-reorg}}$ is 2.91 and 1.96 kcal/mol for simulations with holo and apo structures, respectively), but do not fully explain the protein reorganization effects reflected in the ABFEP simulations. Inspecting the binding pocket, we notice that two aspartic acids are located at the BACE1 binding pocket, forming salt bridges with the ligand (Fig. 4B). Because of the strong salt bridge interactions, the two Asps should be deprotonated in the ligand bound form, which was the state used in the ABFEP simulations. However, in the apo form without ligand, the two ASPs may prefer an alternative protonation state. To verify that, we performed protein FEP to calculate pK_{a}'s of the two Asps. The resulting pK_{a} of Asp93 is 7.6 in the apo form, higher than experimental pH of 5, indicating the protonated form is preferred in the apo state. The corresponding penalty for the change of the protonation state from the protonated form in the apo state to the deprotonated form of the ligand bound state, estimated to be RTln10(pK_{a}- pH) = 3.6 kcal/mol, fully explained the observed protein reorganization effects in the ABFEP simulations.

For JNK1, MCL1 and Thrombin, ABFEP results using the apo protein structure are still slightly more negative than the experiments. Except for the conformational and protonation/tautomeric state changes between the apo and holo structures explored above, other factors could also contribute to the systematic overestimation of binding free energies from ABFEP as compared to experiment. For example, some experimental assays measured the IC_{50} instead of K_i, and the binding free energies converted directly from the experimental IC_{50}s using ΔG=RTln(IC_{50}) could systematically underestimate the real binding free energies. For the JNK1 system, the experimental IC_{50} and K_i were available for six compounds (Table S1),[39] and the differences in the free energies converted from IC_{50} versus that from K_i can be as large as 1.9 kcal/mol, comparable to the magnitude of the observed shift between the raw ABFEP results and

IC_{50} based experimental free energies (Table I). Post-translational modifications and the different protein constructs and/or buffer solutions in experiment versus that in the ABFEP simulations could also contribute to the systematic differences between ABFEP results and experimental measurements.

Several groups have previously reported ABFEP results on some of the systems we tested here.[28,29,30] The majority of these earlier studies were focused on the four protein systems with neutral ligands (CDK2, JNK1, P38 and TYK2), and only one paper reported results on Thrombin with charged ligands. Comparing with these earlier studies (Table II), after taking into account the protein reorganization effect for each method in the same way as we did above to remove the systematic difference between the raw ABFEP results and experimental measurements, for the four systems with neutral ligands, our ABFEP yielded RMSEs of 0.74-1.09 kcal/mol and R^2 of 0.5-0.74, while the work of Lin et al.[29] gave RMSEs of 0.86-1.11 kcal/mol and R^2 of 0.57-0.68, and the work from Khalak et al.[30] resulted in RMSEs of 0.76-1.17 kcal/mol and R^2 of 0.19-0.45, both slightly worse than our results. The protein reorganization effects from Khalak et al.[30] were very small for three systems, possibly because they incorporated the apo crystal structures into their ABFEP calculations. However, surprisingly, their raw ABFEP results for TYK2 (no apo crystal structure was available) were systematic more positive than experiments, which is opposite than expected based on possible protein conformational changes between apo and holo structures. Li et al.[28] reported ABFEP results on the three systems with neutral ligands (CDK2, JNK1 and TYK2) and one with charged ligands (Thrombin). Their results for three systems (CDK2, JNK1 and Thrombin) were comparable to ours (RMSEs of 0.59-0.75 kcal/mol and R^2 of 0.47-0.79 versus RMSEs of 0.84-1.01 kcal/mol and R^2 of 0.5-0.63 from ours), but were much worse than ours for TYK2 (RMSE of 1.32 kcal/mol and the R^2 of 0.52 as compared to RMSE of 0.74 kcal/mol and R^2 of 0.74 from ours). In addition, their ABFEP results on TYK2 ligands are more positive than experiment, contradictory to the expected protein reorganization effect between the apo and holo conformations. Overall, our ABFEP is competitive or slightly better in accuracy than previous benchmarks on the same systems and cover broader sets of ligands.

Table II: Comparison of ABFEP results with previous studies

Protein	ABFEP			Work of Lin et al.[29]			Work of Khalak et al.[30]			Work of Li et al.[28]		
	[b]RMSE	R^2	[a]$\Delta G_{\mathrm{prot-reorg}}$	[b]RMSE	R^2	[a]$\Delta G_{\mathrm{prot-reorg}}$	[b]RMSE	R^2	[a]$\Delta G_{\mathrm{prot-reorg}}$	[b] RMSE	R^2	[a]$\Delta G_{\mathrm{prot-reorg}}$
BACE1	1.21	0.44	2.91									
CDK2	0.84	0.50	1.22	1.11	0.62	3.50	1.13	0.19	-0.43	0.74	0.79	4.05
JNK1	0.85	0.63	3.50	1.03	0.68	4.61	0.76	0.45	0.32	0.75	0.49	2.97
MCL1	0.95	0.53	3.65									
P38	1.09	0.58	5.27	0.86	0.57	6.24	0.95	0.22	-0.03			
PTP1B	1.78	0.55	9.97									
Thrombin	1.01	0.57	1.82							0.59	0.47	5.02
TYK2	0.74	0.74	3.54	0.87	0.66	3.40	1.17	0.27	-2.59	1.32	0.52	-3.27

[a]$\Delta G_{\mathrm{prot-reorg}}$: protein reorganization contribution estimated as the difference between the average of experimental binding free energies and the average of the raw ABFEP binding free energies for all ligands in a system. [b]RMSE was calculated after $\Delta G_{\mathrm{prot-reorg}}$ was added to the raw ABFEP binding free energies.

Applications to hit discovery in virtual screening

Virtual screening (VS) campaigns within drug discovery projects are a natural application of absolute binding free energy calculations. Such campaigns routinely screen massive databases of diverse compounds efficiently using computational tools. Early enrichment of true binders is a critical measure of any in-silico screening funnel, given the costs associated with synthesis or purchase and experimental testing of compounds. The availability of accurate and scalable absolute binding calculations is already having a significant impact on real drug discovery projects within Schrödinger and elsewhere. Recently, Gilson and coworkers reported a work of improved enrichments using ABFEP to rescore top ranked docking compounds from the DUD-E dataset.[54] Here, we report three successful applications of our ABFEP method. In each case, ABFEP rescoring is performed retrospectively on buylists of compounds initially prioritized by docking or ligand-based shape screening. Post-processing with ABFEP leads to a dramatic improvement in the ranking of the confirmed hits, which is reflected in the enrichment statistics. Considering that these lists of tested compounds represent some of the highest scoring compounds according to the original screening method, ABFEP rescoring is essentially boosting extremely early enrichment. Results such as these suggest that ABFEP can reduce the number of compounds that must be purchased and tested in order to find hits. These cases are meant to be illustrative of the potential for application of absolute binding affinity calculations in hit discovery. A detailed study of prospective application of ABFEP in VS campaigns along

with a description of our workflow for integrating these calculations into a larger workflow is planned.

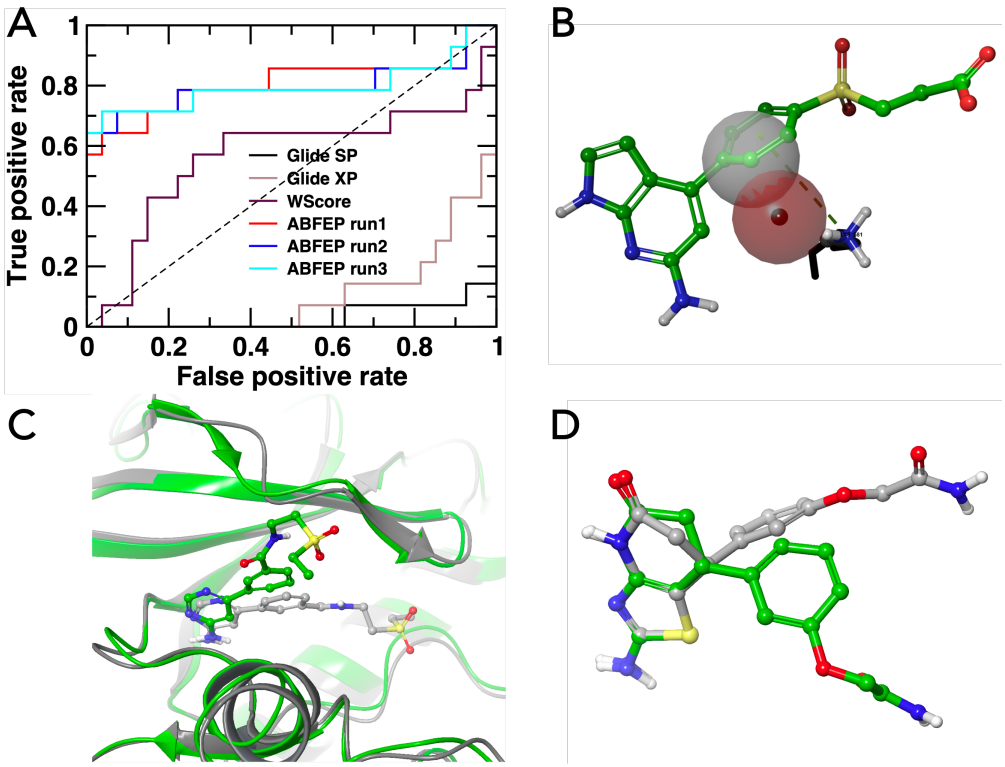

Figure 5: Virtual screening of JAK2. **A.** Hit enrichment of JAK2 ligands using different scoring methods. Black line: Glide SP; brown line: Glide XP; maroon line: WScore; red/blue/cyan lines: ABFEP. The results of Glide SP/XP and metadynamics are from Cutrona et al.[33] Three independent ABFEP simulations were performed with different random seeds. **B.** Desolvation penalty suppresses inactive score in WScore. WScore identifies a blocked charged sidechain (Lys587) upon docking of JAK198. A WaterMap water solvating this residue is displaced by a ligand aromatic ring without making forming a hydrogen bond to the affected residue. This results in a 4.0 kcal/mol penalty in the final score. **C.** Unstable binding of ligand JAK221 in the ABFEP simulations. Gray: starting structure; green: a structure with the partially dissociated ligand from the ABFEP simulation. **D.** Distinct conformation of ligand JAK209 in solvent from that in complex. Gray: a representative structure in solvent; green: a representative structure in complex.

JAK2. Jorgensen and coworkers reported the results of a virtual screen against the pseudokinase domain of JAK2 kinase using the Glide SP[49] docking program.[33] The 27 compounds selected based on Glide SP and Glide XP scores are all inactive, thus they are even worse than random selection at differentiating the 13 confirmed actives from the inactives (Fig. 5A and Table S2). The authors then conducted a retrospective analysis comparing several rescoring protocols on this list of 40 experimentally tested compounds, including MM/GBSA with the VSGB solvation model[55] and different metadynamics protocols. We applied our newest docking program, WScore[11] and ABFEP independently to this dataset. As seen in Fig. 5A, rescoring by WScore shows much improved enrichment over Glide SP and Glide XP with AUC of 0.57. ABFEP further improved the enrichment to AUC of 0.8. Three independent ABFEP simulations with different random seeds gave consistent results, showing robustness of our ABFEP method. We have also compared performance of WScore and ABFEP to the metadynamics protocol proposed by Jorgensen.[33] Metadynamics outperforms WScore docking, however ABFEP rescoring produces the highest enrichment of confirmed hits on this dataset (Fig. S4 and Table S2).

That WScore rescoring performs considerably better than Glide SP can be rationalized by considering the additional physics-based information in WScore as compared with Glide. WScore incorporates a detailed analysis of the binding site water network using a precomputed WaterMap.[56] Small movements of water molecules are sampled simultaneously with the query ligand to optimize the contribution of solvation

scoring terms to the final score. The concept of explicit water molecules is completely absent from Glide SP. For example, the compound JAK198 is experimentally inactive. This compound is the top ranked compound according to the initial Glide SP score. WScore redocking detects a desolvated charged LYS residue (581) following displacement of a water previously making a hydrogen bond to this residue (Fig. 5B). Burying a charged residue carries a significant 4 kcal/mol penalty in WScore. Consequently, WScore ranks this compound at position 14 out of 40 compounds.

While WScore is a notable improvement over Glide SP, this level of performance trails behind the metadynamics protocol recorded in the original study[33] (Fig. S4 and Table S2). Studying the ROC curve shown in Fig. 5A, we see that there are still a significant number of outranking inactive compounds, particularly in the earliest part of the list. WScore rescoring falls far short of the performance of the metadynamics protocols reported in the original study.[33] The top ranked compound after WScore rescoring is JAK221, an inactive compound. Fig. 5C shows the pose (gray) predicted by WScore docking. The WScore pose, which is very similar to the Glide pose, was used as input for ABFEP. The predicted binding mode was highly unstable in the simulation with the sulfonamide-terminated side of the compound completely moving away from the pocket in the ABFEP trajectory (one frame of the ABFEP trajectory is shown in green in Fig. 5C). Consequently, the average ABFEP ranking of this compound (using three independent runs) is 25/40 compared with WScore where it was ranked 1/40. In this case, ABFEP provides insight into the local stability of the predicted binding mode, addressing a key limitation of docking. Another example is JAK209, which is inactive but was ranked 3/40 by WScore. In the ABFEP simulation, JAK209 alone in solvent sampled multiple distinct conformations from that in complex. As shown in Fig. 5D, a representative structure from the solvent simulation of JAK209 (gray), the heterocycle adopts a different puckering state from that in complex (green). In addition, the dihedral angle between the heterocycle and the benzene ring is different between solvent and complex simulations. The large ligand strain energy missing in WScore estimation but accounted for in ABFEP led to the improved scoring accuracy in ABFEP.

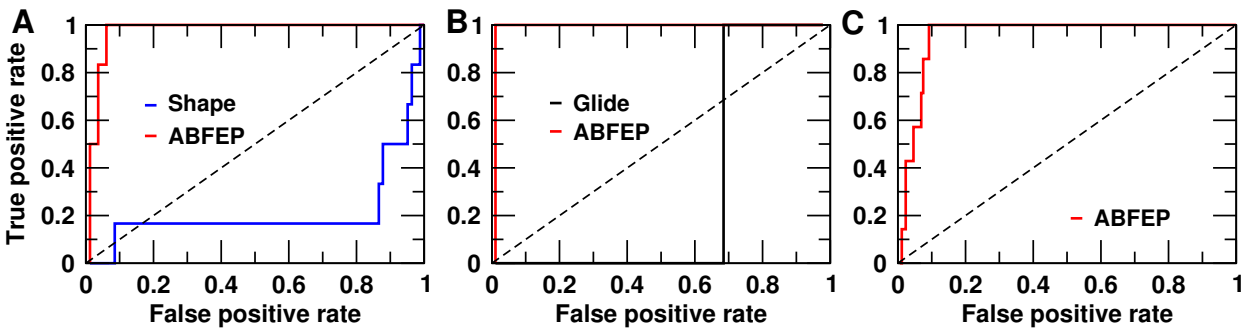

Figure 6: Virtual screening of the Tri-I TDI collaboration target. **A.** Hit enrichment of Track 1. Blue and red lines are the results of GPU-Shape and ABFEP, respectively. **B.** Hit enrichment of Track 2. Blue and red lines are the results of Glide and ABFEP, respectively. **C.** Hit enrichment for the combination of Tracks 1 and 2 using ABFEP.

Collaboration with the Tri-I TDI. Schrödinger conducted a joint hit discovery campaign with the Tri-I TDI against a proprietary target. In this campaign, multiple screening protocols were evaluated as independent experiments using the Enamine Real database. We focus on the two of these experiments, known here as Track 1 and Track 2. The protocols comprised the following: Track 1: the database was screened with GPU-Shape[57] and 88 compounds were selected for purchase. Of these 88 compounds, 6 actives were confirmed by functional assay. Track 2: The database was screened with Active-Learning Glide[58] and 93 compounds were selected for purchase. Of these 93 compounds, 1 active was confirmed by functional assay. Within the purchase list for Track 1, GPU-Shape is not able to effectively distinguish between inactive and active compounds, where the average rank of the actives was 68 (of 88). Rescoring with ABFEP produces a dramatic improvement in the ranks of the actives, with the average rank increasing to 6 (of 88). Fig. 6A shows the ROC curves associated with the original GPU-Shape ranking and the ranking after rescoring with ABFEP. With Track 2, only a single active was recovered by Glide docking score. Within the set of 93 tested compounds, the rank of this single hit was 64 (of 93). In contrast, the ABFEP rescoring

puts this lone hit at the very top of the ranking: 2 (of 93). ROC curves for Track 2 are shown in Fig. 6B. Finally, we combine the rescoring results from the two tracks to determine the performance using a single classification threshold in ABFEP. Enrichment for the combined screens is very good with an AUC of 0.95.

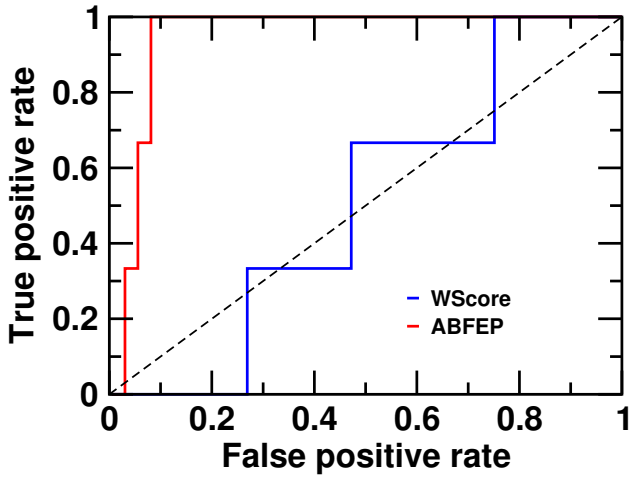

Figure 7: Hit enrichment of compounds identified in virtual screening against the Galapagos evaluation target. Blue and red lines are the ROC curves of WScore and ABFEP, respectively.

Galapagos software evaluation. Galapagos conducted a VS of the Enamine Real (2020-Q1; 1.2B compounds) database against Target B. The primary goal of this VS was two-fold: 1) to perform a head-to-head comparison of Glide versus WScore rescoring; 2) To identify novel hits with the desired properties, for which previous, smaller scale VS campaigns had failed. The compound database was screened using two different tracks here defined as Track 1 and Track 2. In Track 1, the entire database was screened first with Active-Learning Glide.[58] For Track 2, the database was initially screened with GPU-Shape.[57] For these two tracks, rescoring with Glide SP or WScore was compared. A selection of 200 top compounds by Glide SP across both tracks were selected for purchase. This was repeated with WScore. Due to overlap in the lists, a total of 370 compounds were ordered, and 324 delivered and tested using a biochemical assay. A total of four hits were confirmed by experiment, three from WScore and one from Glide. To further assess if ABFEP rescoring would have allowed us to purchase fewer compounds, ABFEP calculations were run on the top 200 compounds from WScore. Additional details concerning the virtual screening procedure and analysis are left to supporting information. On this set, WScore produces modest enrichment (AUC: 0.5) of the hit compounds. To recover all three compounds, it is necessary to go to position 151 in the list. The entire purchase list was later rescored by ABFEP, retrospectively. The ranks of the confirmed hits improve substantially with AUC increasing to 0.94. All three hits are recovered within the first 20 compounds (i.e. top 10%) sorted by ABFEP score. ROC curves comparing WScore and ABFEP enrichment are shown in Fig. 7. Individual ranks for the hit compounds according to docking and ABFEP are shown in Table S3. Once again, ABFEP rescoring delivers significantly improved enrichment over even advanced docking methods like WScore. These results indeed suggest that had ABFEP rescoring been applied prior to compound scoring, many fewer compounds could have been purchased to identify these hits.

Conclusions

We reported an implementation of the ABFEP method in the FEP+ program.[4] The ABFEP protocol utilizes the double decoupling scheme (Fig. 1), where the free energy to annihilate the ligand from the protein binding pocket into a restrained dummy ligand and the free energy to annihilate the ligand in solvent were calculated separately in two simulations. The translational and orientational entropy due to the restraints between the dummy ligand and the protein binding pocket (Fig. 2) is calculated by an analytical formula (Eqn. 3). For ligands with net charges, the alchemical ion approach introduced in the charge-changing perturbations in RBFEP[9] is adopted to resolve the finite-size effects and more replicas were used to converge the electrostatic interactions.

We validated our ABFEP protocol on the entire dataset from our previous RBFEP benchmark[5] including a total number of 199 ligands binding to eight protein targets. Different from previous ABFEP publications focusing mainly on the neutral ligands, four of our validation systems have charged ligands. The overall RMSE between ABFEP calculated and experimental binding free energies on the entire dataset is 1.13 kcal/mol and weighted average R^2 is 0.55, slightly worse than the well established RBFEP method[5] with RMSE of 0.9 kcal/mol and R^2 of 0.56 (Fig. 3 and Table I). This is expected due to the much larger configurational space to sample in ABFEP as compared to RBFEP. Comparing with earlier ABFEP publications,[29,30,28] which reported results on some of these systems, our ABFEP results are comparable or slightly better (Table II).

It should be noted that the raw ABFEP results are in general more negative than the experimental binding free energies for real protein-ligand systems. This is because the holo complexes with ligands bound most often have very different conformations and/or protonation states than the apo proteins, and the conformational and/or protonation state changes upon ligand binding, i.e., the protein reorganization contribution, can not be sampled in the short ABFEP simulations. We verified this hypothesis by comparing ABFEP calculations using the holo and apo protein structures and showing that ABFEP calculated binding free energies are more positive using the apo protein structures than using the holo structures (Table I). In addition, for four of the systems, with the apo protein structures and after taking into account the protonation state changes between the apo and holo structures, ABFEP calculated binding free energies are more positive than experiment, suggesting that the conformational and protonation state changes are the major factors contributing to the protein reorganization free energy. The hypothesis that the miss of protein reorganization free energy is the main reason for the systematic difference between ABFEP calculation and experimental binding free energy was suggested in the past,[59] and our calculations are the first attempt to validate this hypothesis.

Although the ABFEP results are usually more negative than experimental measurement, ABFEP can still give correct rank ordering of ligand binding considering that the protein reorganization contribution is a constant for ligands binding to the same protein conformation. In fact, a unique advantage of ABFEP as compared to RBFEP is its ability to rank order compounds with diverse scaffolds and binding modes, with the potential application to separate correct binding poses from the decoy poses or selecting actives from decoys for hit discovery in virtual screening. We tested our ABFEP method on three retrospective virtual screening datasets to explore its ability to rank order diverse compounds to improve hit rates. In all three cases, one for the JH2 domain of JAK2[33] and the other two for proprietary targets in collaboration, ABFEP was able to dramatically improve the hit rates, with AUC improved to 0.8 or more, significantly better than empirical Glide or WScore scoring functions or other alternative scoring method like metadynamics (Figs. 5-7.). Especially for the two collaboration projects, if ABFEP was available at the time these virtual screenings were run and was used for compound selection, the list of compounds for experimental test can be reduced by at least a factor of eight without losing any active compounds. These results further validate the high accuracy and reliability of our ABFEP method, in the challenging cases where diverse structures and binding modes need to be scored with high accuracy, proven to be extremely difficult for other methods. This is to our knowledge the first practical application of ABFEP to triage the compound list in virtual screening.

While the accuracy of ABFEP calculations demonstrated in these validations is exciting, a few important notes should be considered in prospective applications. First, due to the miss of protein reorganization effect discussed above, the raw ABFEP results usually over-estimate the binding free energies of the ligands. In the context of scoring ligands in virtual screening where no binding affinity information is available for any ligand, it should only be used to rank order the ligands. If experimental binding affinities

are available for a handful of ligands, the difference in the ABFEP results versus experimental data on these known ligands can be used to estimate the protein reorganization effect. As we demonstrated in the above validation, the absolute binding free energies of the ligands can be informed after accounting for protein reorganization effect. Second, due to the large configurational space to sample, ABFEP calculations are computationally much more expensive than RBFEP. For ranking congeneric series of ligands, RBFEP is still recommended both for cost benefit and for the superior accuracy. Third, although the convergence of ABFEP calculations is much improved with our optimized lambda schedules and sampling protocols, ligands with net charges involving strong salt-bridge interactions with the protein may still present challenges in some cases and further enhancements in the protocol is needed to make it more robust.

Supporting Information

Supplementary method description; Three tables and five figures; Two spreadsheets containing raw data of the large-scale test on congeneric compound series and the virtual screening of JAK2; The link for downloading the input structures for the congeneric compound series and the JAK2 virtual screening.

Notes

The authors declare the following competing financial interest(s): R.A.F. has a significant financial stake in, is a consultant for, and is on the Scientific Advisory Board of Schrödinger, Inc.

Acknowledgements

We thank Dr. Joe Kaus, Dr. Yuqing Deng and Dr. Yujie Wu for their helpful discussion. E.B.L. would like to thank all the Galapagos colleagues who were involved in this drug discovery project.

References

[1] Abel, R.; Wang, L.; Harder, E. D.; Berne, B. J.; Friesner, R. A. Advancing drug discovery through enhanced free energy calculations. *Acc. Chem. Res.* **2017**, *50*, 1625–1632.

[2] Abel, R.; Wang, L.; Mobley, D. L.; Friesner, R. A. A critical review of validation, blind testing, and real-world use of alchemical protein-ligand binding free energy calculations. *Curr. Top. in Med. Chem.* **2017**, *17*, 2577–2585.

[3] Schindler, C. E. M. et al. Large-scale assessment of binding free energy calculations in active drug discovery projects. *J. Chem. Inf. Model.* **2020**, *60*, 5457–5474.

[4] *Schrödinger Release 2021-3: FEP+*; Schrödinger, LLC: New York, NY, 2021.

[5] Wang, L. et al. 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.

[6] Lu, C.; Wu, C.; Ghoreishi, D.; Chen, W.; Wang, L.; Damm, W.; Ross, G. A.; Dahlgren, M. K.; Russell, E.; Bargen, C. D. V.; Abel, R.; Friesner, R. A.; Harder, E. D. OPLS4: improving force field accuracy on challenging regimes of chemical space. *J. Chem. Theory Comput.* **2021**, *17*, 4291–4300.

[7] Wang, L.; Deng, Y.; Wu, Y.; Kim, B.; LeBard, D. N.; Wandschneider, D.; Beachy, M.; Friesner, R. A.; Abel, R. Accurate modeling of scaffold hopping transformations in drug discovery. *J. Chem. Theory Comput.* **2017**, *13*, 42–54.

[8] Yu, H. S.; Deng, Y.; Wu, Y.; Sindhikara, D.; Rask, A. R.; Kimura, T.; Abel, R.; Wang, L. Accurate and reliable prediction of the binding affinities of macrocycles to their protein targets. *J. Chem. Theory Comput.* **2017**, *13*, 6290–6300.

[9] Chen, W.; Deng, Y.; Russell, E.; Wu, Y.; Abel, R.; Wang, L. Accurate calculation of relative binding free energies between ligands with different net charges. *J. Chem. Theory Comput.* **2018**, *14*, 6346–6358.

[10] Yu, H. S.; Modugula, K.; Ichihara, O.; Kramschuster, K.; Keng, S.; Abel, R.; Wang, L. General theory of fragment linking in molecular design: Why fragment linking rarely succeeds and how to improve outcomes. *J. Chem. Theory Comput.* **2021**, *17*, 450–462.

[11] Murphy, R. B.; Repasky, M. P.; Greenwood, J. R.; Tubert-Brohman, I.; Jerome, S.; Annabhimoju, R.; Boyles, N. A.; Schmitz, C. D.; Abel, R.; Farid, R.; Friesner, R. A. WScore: a flexible and accurate treatment of explicit water molecules in ligand–receptor docking. *J. Med. Chem.* **2016**, *59*, 4364–4384.

[12] Sliwoski, G.; Kothiwale, S.; Meiler, J.; Edward W. Lowe, J. Computational methods in drug discovery. *Pharmacol. Rev.* **2014**, *66*, 334–395.

[13] Jorgensen, W. L.; Buckner, J. K.; Boudon, S.; TiradoRives, J. Efficient computation of absolute free energies of binding by computer simulations: application to the methane dimer in water. *J. Chem. Phys.* **1988**, *89*, 3742.

[14] 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.

[15] Boresch, S.; Tettinger, F.; Leitgeb, M.; Karplus, M. Absolute binding free energies: a quantitative approach for their calculation. *J. Phys. Chem. B* **2003**, *107*, 9535–9551.

[16] Deng, Y.; Roux, B. Calculation of standard binding free energies: aromatic molecules in the T4 lysozyme L99A mutant. *J. Chem. Theory Comput.* **2006**, *2*, 1255–1273.

[17] Mobley, D. L.; Graves, A. P.; Chodera, J. D.; McReynolds, A. C.; Shoichet, B. K.; Dill, K. A. Predicting absolute ligand binding free energies to a simple model site. *J. Mol. Biol.* **2007**, *371*, 1118–1134.

[18] Boyce, S. E.; Mobley, D. L.; Rocklin, G. J.; Graves, A. P.; Dill, K. A.; Shoichet, B. K. Predicting ligand binding affinity with alchemical free energy methods in a polar model binding site. *J. Mol. Biol.* **2009**, *394*, 747–763.

[19] 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*, 3372–3377.

[20] Lee, J.; Tofoleanu, F.; Pickard, F. C.; König, G.; Huang, J.; Damjanović, A.; Baek, M.; Seok, C.; Brooks, B. R. Absolute binding free energy calculations of CBClip host–guest systems in the SAMPL5 blind challenge. *J. Comput. Aided Mol. Des.* **2017**, *31*, 71–85.

[21] Laury, M. L.; Wang, Z.; Gordon, A. S.; Ponder, J. W. Absolute binding free energies for the SAMPL6 cucurbit[8]uril host–guest challenge via the AMOEBA polarizable force field. *J. Comput. Aided Mol. Des.* **2018**, *32*, 1087–1095.

[22] Jayachandran, G.; Shirts, M. R.; Park, S.; Pande, V. S. Parallelized-over-parts computation of absolute binding free energy with docking and molecular dynamics. *J. Chem. Phys.* **2006**, *125*, 084901.

[23] Wang, J.; Deng, Y.; Roux, B. Absolute binding free energy calculations using molecular dynamics simulations with restraining potentials. *Biophys. J.* **2006**, *91*, 2798–2814.

[24] Fujitani, H.; Tanida, Y.; Matsuura, A. Massively parallel computation of absolute binding free energy with well-equilibrated states. *Phys. Rev. E* **2009**, *79*, 021914.

[25] Aldeghi, M.; Heifetz, A.; Bodkin, M. J.; Knapp, S.; Biggin, P. C. Accurate calculation of the absolute free energy of binding for drug molecules. *Chem. Sci.* **2016**, *7*, 207–218.

[26] Aldeghi, M.; Heifetz, A.; Bodkin, M. J.; Knapp, S.; Biggin, P. C. Predictions of ligand selectivity from absolute binding free energy calculations. *J. Am. Chem. Soc.* **2017**, *139*, 946–957.

[27] Gapsys, V.; Yildirim, A.; Aldeghi, M.; Khalak, Y.; van der Spoel, D.; de Groot, B. L. Accurate absolute free energies for ligand–protein binding based on non-equilibrium approaches. *Commun. Chem.* **2021**, *4*, 1–13.

[28] Li, Z.; Huang, Y.; Wu, Y.; Chen, J.; Wu, D.; Zhan, C. G.; Luo, H. B. Absolute binding free energy calculation and design of a subnanomolar inhibitor of phosphodiesterase-10. *J. Med. Chem.* **2019**, *62*, 2099–2111.

[29] Lin, Z.; Zou, J.; Liu, S.; Peng, C.; Li, Z.; Wan, X.; Fang, D.; Yin, J.; Gobbo, G.; Chen, Y.; Ma, J.; Wen, S.; Zhang, P.; Yang, M. A cloud computing platform for scalable relative and absolute binding free energy predictions: new opportunities and challenges for drug discovery. *J. Chem. Inf. Model.* **2021**, *61*, 2720–2732.

[30] Khalak, Y.; Tresdern, G.; Aldeghi, M.; Baumann, H. M.; Mobley, D. L.; de Groot, B.; Gapsys, V. Alchemical absolute protein–ligand binding free energies for drug design. *Chem. Sci.* **2021**,

[31] Rocklin, G. J.; Mobley, D. L.; Dill, K. A.; Hünenberger, P. H. Calculating the binding free energies of charged species based on explicit-solvent simulations employing lattice-sum methods: an accurate correction scheme for electrostatic finite-size effects. *J. Chem. Phys.* **2013**, *139*, 184103.

[32] Cournia, Z.; Allen, B. K.; Beuming, T.; Pearlman, D. A.; Radak, B. K.; Sherman, W. Rigorous Free Energy Simulations in Virtual Screening. *J. Chem. Inf. Model.* **2020**, *60*, 4153–4169.

[33] Cutrona, K. J.; Newton, A. S.; Krimmer, S. G.; Tirado-Rives, J.; Jorgensen, W. L. Metadynamics as a postprocessing method for virtual screening with application to the pseudokinase domain of JAK2. *J. Chem. Inf. Model.* **2020**, *60*, 4403–4415.

[34] Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. The Protein Data Bank. *Nucleic Acids Res.* **2000**, *28*, 235–242.

[35] Madhavi Sastry, G.; Adzhigirey, M.; Day, T.; Annabhimoju, R.; Sherman, W. Protein and ligand preparation: parameters, protocols, and influence on virtual screening enrichments. *J. Comput. Aided Mol. Des.* **2013**, *27*, 221–234.

[36] *Schrödinger Release 2021-3: Maestro*; Schrödinger, LLC: New York, NY, 2021.

[37] Cumming, J. N. et al. Structure based design of iminohydantoin BACE1 inhibitors: identification of an orally available, centrally active BACE1 inhibitor. *Bioorg. Med. Chem. Lett.* **2012**, *22*, 2444–2449.

[38] Davies, T. G. et al. Structure-based design of a potent purine-based cyclin-dependent kinase inhibitor. *Nat. Struct. Biol.* **2002**, *9*, 745–749.

[39] Szczepankiewicz, B. G. et al. Aminopyridine-based c-Jun N-terminal kinase inhibitors with cellular activity and minimal cross-kinase activity. *J. Med. Chem.* **2006**, *49*, 3563–3580.

[40] Friberg, A.; Vigil, D.; Zhao, B.; Daniels, R. N.; Burke, J. P.; Garcia-Barrantes, P. M.; Camper, D.; Chauder, B. A.; Lee, T.; Olejniczak, E. T.; Fesik, S. W. Discovery of potent myeloid cell leukemia 1 (Mcl-1) inhibitors using fragment-based methods and structure-based design. *J. Med. Chem.* **2012**, *56*, 15–30.

[41] Wilson, D. P. et al. Structure-based optimization of protein tyrosine phosphatase 1B inhibitors: from the active site to the second phosphotyrosine binding site. *J. Med. Chem.* **2007**, *50*, 4681–4698.

[42] Liang, J. et al. Lead identification of novel and selective TYK2 inhibitors. *Eur. J. Med. Chem.* **2013**, *67*, 175–187.

[43] Hong, L.; Tang, J. Flap position of free memapsin 2 (β-secretase), a model for flap opening in aspartic protease catalysis. *Biochemistry* **2004**, *43*, 4689–4695.

[44] Lowe, E. D.; Tews, I.; Cheng, K. Y.; Brown, N. R.; Gul, S.; Noble, M. E. M.; Gamblin, S. J.; Johnson, L. N. Specificity determinants of recruitment peptides bound to phospho-CDK2/cyclin A. *Biochemistry* **2002**, *41*, 15625–15634.

[45] Luptak, J.; Bista, M.; Fisher, D.; Flavell, L.; Gao, N.; Wickson, K.; Kazmirski, S.; Howard, T.; Rawlins, P.; Hargreaves, D.; IUCr, Antibody fragments structurally enable a drug-discovery campaign on the cancer target Mcl-1. *Acta Crystallogr. D Struct. Biol.* **2019**, *75*, 1003–1014.

[46] Wilson, K. P.; Fitzgibbon, M. J.; Caron, P. R.; Griffith, J. P.; Chen, W.; McCaffrey, P. G.; Chambers, S. P.; Su, M. S.-S. Crystal structure of p38 mitogen-activated protein kinase. *J. Biol. Chem.* **1996**, *271*, 27696–27700.

[47] Ala, P. J. et al. Structural basis for inhibition of protein-tyrosine phosphatase 1B by isothiazolidinone heterocyclic phosphonate mimetics. *J. Biol. Chem.* **2006**, *281*, 32784–32795.

[48] Bandaranayake, R. M.; Ungureanu, D.; Shan, Y.; Shaw, D. E.; Silvennoinen, O.; Hubbard, S. R. Crystal structures of the JAK2 pseudokinase domain and the pathogenic mutant V617F. *Nat. Struct. Mol. Biol.* **2012**, *19*, 754–759.

[49] Friesner, R. A.; Banks, J. L.; Murphy, R. B.; Halgren, T. A.; Klicic, J. J.; Mainz, D. T.; Repasky, M. P.; Knoll, E. H.; Shelley, M.; Perry, J. K.; Shaw, D. E.; Francis, P.; Shenkin, P. S. Glide: A new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. *Journal of Medicinal Chemistry* **2004**, *47*, 1739–1749.

[50] Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; Hermans, J. In *Intermolecular Forces*; Pullman, B., Ed.; Springer, Dordrecht, 1981; pp 331–342.

[51] Ross, G. A.; Russell, E.; Deng, Y.; Lu, C.; Harder, E. D.; Abel, R.; Wang, L. Enhancing water sampling in free energy calculations with grand canonical Monte Carlo. *J. Chem. Theory Comput.* **2020**, *16*, 6061–6076.

[52] Steinbrecher, T.; Zhu, C.; Wang, L.; Abel, R.; Negron, C.; Pearlman, D.; Feyfant, E.; Duan, J.; Sherman, W. Predicting the effect of amino acid single-point mutations on protein stability–large-scale validation of MD-based relative free energy calculations. *J. Mol. Biol.* **2017**, *429*, 948–963.

[53] Thurlkill, R. L.; Grimsley, G. R.; Scholtz, J. M.; Pace, C. N. pK values of the ionizable groups of proteins. *Protein Sci.* **2006**, *15*, 1214–1218.

[54] Feng, M.; Heinzelmann, G.; Gilson, M. K. Absolute binding free energy calculations improve enrichment of actives in virtual compound screening. *Sci. Rep.* **2022**, *12*, 1–11.

[55] Li, J.; Abel, R.; Zhu, K.; Cao, Y.; Zhao, S.; Friesner, R. A. The VSGB 2.0 model: A next generation energy model for high resolution protein structure modeling. *Proteins* **2011**, *79*, 2794–2812.

[56] Abel, R.; Young, T.; Farid, R.; Berne, B. J.; Friesner, R. A. Role of the active-site solvent in the thermodynamics of factor Xa ligand binding. *J. Am. Chem. Soc.* **2008**, *130*, 2817–2831.

[57] Sastry, G. M.; Dixon, S. L.; Sherman, W. Rapid shape-based ligand alignment and virtual screening method based on atom/feature-pair similarities and volume overlap scoring. *J. Chem. Inf. Model.* **2011**, *51*, 2455–2466.

[58] Yang, Y.; Yao, K.; Repasky, M. P.; Leswing, K.; Abel, R.; Shoichet, B. K.; Jerome, S. V. Efficient exploration of chemical space with docking and deep learning. *J. Chem. Theory Comput.* **2021**, *17*, 7106–7119.

[59] Mobley, D. L.; Chodera, J. D.; Dill, K. A. Confine-and-release method: obtaining correct binding free energies in the presence of protein conformational change. *J. Chem. Theory Comput.* **2007**, *3*, 1231–1235.

For Table of Contents Only