JCIM JOURNAL OF CHEMICAL INFORMATION AND MODELING

pubs.acs.org/jcim Article

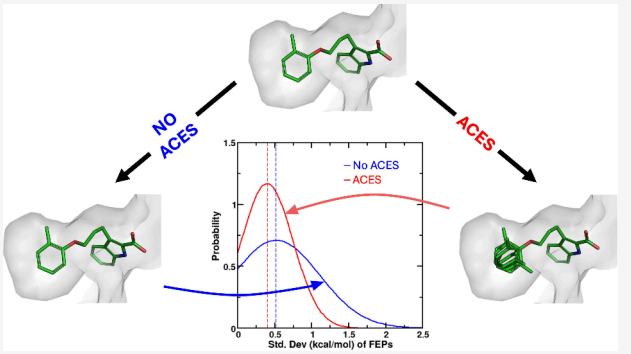
Improvements in Precision of Relative Binding Free Energy Calculations Afforded by the Alchemical Enhanced Sampling (ACES) Approach

Hsu-Chun Tsai, James Xu, Zhenyu Guo, Yinhui Yi, Chuan Tian, Xinyu Que, Timothy Giese, Tai-Sung Lee, Darrin M. York,* Abir Ganguly,* and Albert Pan*

Cite This: *J. Chem. Inf. Model.* 2024, 64, 7046−7055 Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: Accurate *in silico* predictions of how strongly small molecules bind to proteins, such as those afforded by relative binding free energy (RBFE) calculations, can greatly increase the efficiency of the hit-to-lead and lead optimization stages of the drug discovery process. The success of such calculations, however, relies heavily on their precision. Here, we show that a recently developed alchemical enhanced sampling (ACES) approach can consistently improve the precision of RBFE calculations on a large and diverse set of proteins and small molecule ligands. The addition of ACES to conventional RBFE calculations lowered the average hysteresis by over 35% (0.3−0.4 kcal/mol) and the average replicate spread by over 25% (0.2−0.3 kcal/mol) across a set of 10 protein targets and 213 small molecules while maintaining similar or improved accuracy. We show in atomic detail how ACES improved convergence of several representative RBFE calculations through enhancing the sampling of important slowly transitioning ligand degrees of freedom.



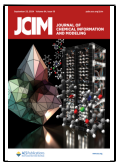
1. INTRODUCTION

Computational assays for protein−small molecule potency, such as calculations based on the relative binding free energy (RBFE) approach, have emerged as powerful tools for accelerating the hit-to-lead and lead optimization stages of drug discovery. They rely on physics-based atomistic models that take advantage of nonphysical ("alchemical") transformations[1,2] to estimate the relative binding affinities of congeneric small molecules to protein receptors. Historically plagued by inaccuracies and poor precision due to issues related to force fields, inadequate simulation times, and complicated setup,[1−7] these calculations have seen tremendous development in recent years, including advances in force field development,[8−13] computational hardware architectures,[14] high-performance software implementations,[15−19] and streamlined workflows.[20−25]

Despite recent advancements, however, RBFE calculations are still expensive computations that are often affected by poor convergence and large variances in free energy estimates (especially between independent replicates) that limit reproducibility and predictive capability. Key contributors to this variability are (1) instabilities along the alchemical transformation pathway that can produce first-order phase transition behavior in the λ-dimension[26] and (2) the inability to sample relevant, but slow, conformational degrees of freedom during practically accessible simulation time scales.[27]

Many enhanced sampling approaches that aim at improving convergence and precision in free energy calculations have been proposed, including metadynamics,[28] enveloping distribution sampling (EDS/λ-EDS),[29,30] parallel or simulated tempering, orthogonal space random walk (OSRW),[31] thermodynamic integration with enhanced sampling (TIES),[32] and several methods based on replica exchange (RE) framework, such as multiple-replica adaptive biasing force,[33−35] replica exchange with solute-tempering (REST and REST2),[36−38] and the recently developed alchemical enhanced sampling (ACES) approach.[39] Both REST/REST2 and ACES belong to the category of "Generalized Ensemble Monte Carlo" methods that achieve enhanced sampling in the alchemical dimension by scaling specific intra- and intermolecular potential energy terms for a select group of atoms. These scaled states are then connected to the real states by Hamiltonian replica exchange networks.

Received: March 15, 2024
Revised: August 14, 2024
Accepted: August 19, 2024
Published: September 3, 2024



The ACES method uses robust alchemical transformation pathways to connect real and enhanced sampling states using new smoothstep softcore potentials, nonlinear Hamiltonian mixing, and flexible λ-scheduling capabilities[40,41] that have been integrated into the GPU-accelerated free energy simulation engine in AMBER[18] and available in AMBER22.[42] ACES relies on a dual-topology framework that can allow it to overcome local "hot-spot" problems encountered with REST/REST2.[36−38]

Here, we focus on the usefulness of the ACES approach and demonstrate that ACES noticeably increases the precision of RBFE calculations across a large and diverse dataset containing 10 different targets and over 200 ligands. For the entire dataset, we performed multiple independent replicates of RBFE calculations both with and without ACES, and for each target, we observed an increase in precision ranging from 0.1 to 1.8 kcal/mol.

The rest of the article is organized as follows: in section 2, we briefly recapitulate the salient features of the ACES approach. In section 3, we provide details related to the computational approaches adopted in this work. In section 4, we discuss the broad dataset considered in this work and present the key results obtained with and without the use of enhanced sampling. In section 4.1, we discuss, in atomic detail, three case studies that illustrate the importance of enhanced sampling, and in particular ACES, in RBFE calculations. Finally, in section 5, we conclude by summarizing the key take home messages.

2. THE ALCHEMICAL ENHANCED SAMPLING (ACES) APPROACH

As discussed in the Introduction, there is a rich literature that describes enhanced sampling (or importance sampling) methods in the context of alchemical free energy simulations.[2,30−39,41] The alchemical enhanced sampling (ACES) method[39] is an integrated approach incorporating the recently developed robust softcore potentials,[40] flexible control of dummy states,[43] interaction energies,[44] and the Hamiltonian replica exchange (HREMD) framework[45,46] in alchemical transformation simulations. While the concept of enhanced sampling methods leveraging alchemical transformations and dual-topology frameworks has been explored for many years,[47−49] developing practical and reliable methods for predicting protein−ligand binding affinities has presented challenges and remains an active area of research and software development.[48,50,51]

The strength of the ACES method lies in the precise implementation of three fundamental elements.

- Creation of localized (focused) enhanced sampling states through tuning of intra- and intermolecular energy terms for selected groups of atoms in the softcore region.
- Design of robust alchemical transformation pathways to connect real and enhanced sampling end states using new smoothstep softcore potentials, λ-dependent weight functions, and flexible λ-scheduling capabilities.
- Construction of efficient HREMD networks to facilitate Boltzmann sampling of the real state end points and maintain equilibrium between windows along the alchemical transformation pathways.

The first element creates a fictitious "enhanced sampling" state with a barrier-reducing potential energy, whereas the second and third elements work together to provide a mechanism to rigorously and efficiently connect the conformational ensembles of the real state and enhanced sampling state end points using a Hamiltonian replica exchange (HREMD) framework.[45,46] Specifically, with ACES,[39] the enhanced sampling dummy state is one where all nonbonded interactions with the environment are turned off using recently introduced smoothstep softcore potentials,[40] and in addition, all internal electrostatic interactions as well as torsion angle and 1−4 Lennard-Jones terms for rotatable single bonds are turned off to avoid kinetic traps. Other internal energy terms are retained, as they help to reduce the volume of relevant phase space that needs to be sampled. Efficient HREMD networks are created using a recently introduced optimized phase-space overlap method.[41] In an RBFE calculation between two ligands "A" and "B", as the ACES region of the real state for ligand A is being annihilated with respect to its interactions with the environment and transformed into an enhanced sampled dummy state,[43] the ACES region of the real state for ligand B is being created. This concerted counterdiffusion of alchemical states produces minimal rearrangement of the environment along the λ path. Further, the computational overhead of ACES is negligible relative to conventional alchemical free energy simulations with HREMD. Overall, the ACES method provides a convincing and comprehensive framework for enhanced sampling in molecular simulations. It can be employed as a standalone method for efficient exploration of conformational space or integrated as an important sampling technique within alchemical free energy simulations, offering improved accuracy and reliability in predicting protein−ligand binding affinities.

3. METHOD

The crystal structures on which the RBFE calculations were based for the various proteins are: 4DJW for BACE,[52] 1H1Q for CDK2,[53] 2GMX for JNK1,[54] 4HW3 for MCL1,[55] 3FLY for p38,[56] 2QBS for PTP1B,[57] 2ZFF for Thrombin,[58] 4GIH for Tyk2,[59,60] 7EW9 for KRAS-G12D[61,62] and 7LMD for SARS-CoV-2.[63] The PDBs were processed for MD simulations using various programs available as part of AmberTools,[64] particularly *pdb4amber*, for cleaning PDB structures, and *tleap*, for building topology and parameter files. In certain cases, specific active site titratable residues were assigned protonation states based on visual inspection. For a given protein, 3D structures of the target ligands were generated based on a maximum common substructure-based 3D alignment approach with respect to the reference cocrystal ligand.

The protein systems were modeled using the AMBER ff14SB force field, while the ligands were parametrized using an in-house force field parametrization workflow that generates bespoke *ab initio*-based force field parameters for small molecules. In this workflow, each ligand is optimized at the QM level and then am1-bcc partial charges are derived for the optimized structure. To obtain the torsional parameters, first the torsional profile for each rotatable bond in the molecule is calculated at the QM level. For each torsion, 12 constrained geometry optimizations are performed at intervals of 30°. Thereafter, a linear least-squares fitting is performed to derive MM torsional parameters that closely reproduce the QM profiles. The QM calculations are performed at the 6-31G**/B3LYP level of theory. The protein−ligand complexes and isolated ligands were immersed in separate truncated octahedral boxes of TIP3P waters,[65] with buffer distances of

https://doi.org/10.1021/acs.jcim.4c00464
J. Chem. Inf. Model. 2024, 64, 7046−7055

10 and 18 Å, respectively. The Monte Carlo barostat and Langevin thermostat[66] with a friction constant of 2.0 ps^{-1} were used to maintain constant temperature and pressure, respectively. SHAKE[67,68] and hydrogen mass repartitioning (HMR)[69] were applied to both protein−ligand complex and ligand topologies. Nonbonded interactions were computed directly within an 8 Å cutoff, and long-range electrostatic interactions were evaluated with the particle mesh Ewald (PME) method[70,71] using approximately a 1 Å grid spacing. All simulations employed a 4 fs time step.

An in-house (TandemAI) FEP workflow was employed to set up the RBFE calculations. Each system was subjected to a rigorous equilibration protocol having the following steps: (a) full system minimization, (b) heating from 0 to 100 K at constant volume and temperature (NVT) ensemble over 2 ps, followed by MD at constant pressure and temperature (NPT) ensemble at 100 K for 2 ps, (c) heating from 100 to 200 K at constant NVT over 2 ps, followed by MD at constant NPT at 200 K for 2 ps, and (d) heating from 200 to 300 K at constant NVT over 2 ps, followed by MD at constant NPT at 298 K for 2 ps. For (b−d), positional restraints were imposed on all protein and ligand heavy atoms. (e) MD at constant NPT at 298 K without restraints for 5 ns. The end structures from the above protocol were used to generate the hybrid-dual topologies necessary for RBFE calculations.

All RBFE calculations were performed using 12 λ windows that were determined using a newly developed approach for predicting optimal λ schedules for RBFE calculations.[41] In this approach, a series of short simulations are first performed with a larger number of uniformly spaced λ windows (25 in this work) that are then analyzed to generate a 2D map of the phase-space overlap in the λ space. The 2D map $O(\lambda, \lambda')$ represents a predictive measure of the phase-space overlap involving the intervals defined by λ and λ'. The final set of 12 λ windows is then obtained by optimizing an alchemical pathway on this 2D map such that the phase-space overlaps are equal between the λ intervals. This optimized λ schedule, maximizing the phase-space overlap between neighboring windows, leads to a higher number of end-to-end round trips in HREMD simulations, which in turn increases the efficiency of ACES.

To obtain the initial structures for the different λ windows, a fast sequential equilibration was performed starting from $\lambda = 0$, in which at each λ, the system was first minimized, heated to 300 K in 60 ps at constant NVT, and equilibrated at constant NPT for 60 ps. The end structure was used as an initial seed for the next λ window. Finally, production runs of 5 ns were carried out in the NVT ensemble. In both the "ACES" and "no ACES" sets of simulations, HREMD was employed, and replica exchanges were attempted every 625 fs. The transformations were performed with the one-step concerted protocol, using the modified SSC(2) softcore potentials ($m = n = 2$, $\alpha^{LJ} = 0.5$, $\alpha^{Coul} = 1$).[40] The ACES method[39] has been implemented using the dual-topology framework in AMBER. Here, the transforming region is separated into a "common core" (CC) set of atoms that share common coordinates and a set of "softcore" (SC) atoms that have separable coordinates. It is in the selection of the SC atoms that enables ACES to achieve enhanced sampling through the dual topology by counterpoised creation and annihilation of the SC atoms of the end states. All simulations were performed using the PMEMD.cuda program that is available as part of the official AMBER22 release[42] augmented by the AMBER Drug Discovery Boost package.[44] Network-wide free energy analysis including cycle closure constraints[72] was conducted using `FE-ToolKit` distributed with the latest version of AmberTools.[64] Error bars for mean unsigned error (MUE), root mean squared error (RMSE), spread, and hysteresis represent standard errors of these metrics within each dataset. The "overall" values along with their error bars are averages of each of these metrics over the entire dataset, weighted by the size of each individual dataset. The following equations are used to calculate the weighted average and error bar for a particular metric, X:

$$\bar{X}_{wtd} = \sum_{i=1}^{n} W_i X_i \tag{1}$$

$$(S_{\bar{x}})_{wtd} = \sqrt{\frac{\sum_{i=1}^{n} W_i (X_i - \bar{X}_{wtd})^2}{n-1}} \tag{2}$$

where $\bar{X}_{wtd}$ is the weighted average, W_i is the weight of a particular dataset calculated as the ratio of the size of that dataset to the overall dataset, $(S_{\bar{x}})_{wtd}$ is the weighted error bar, and n is the number of datasets.

The error bars for R^2 were calculated using a bootstrapping approach in which for a particular dataset, at a given time, an R^2 value was calculated by choosing any one of the replicate runs for each edge within that dataset. The process was repeated 200 times to obtain the standard error of the distribution.

4. RESULTS AND DISCUSSION

We considered a diverse dataset containing 213 ligands across 10 different targets forming 345 perturbations. Specifically, these targets include the BACE, CDK2, JNK1, MCL1, p38, thrombin, Tyk2, and PTP1B proteins that were part of a well-known validation study of the FEP+ program from Schrödinger.[25] Additionally, we included two other pharmaceutically significant targets: KRAS-G12D, a member of the KRAS family of proteins that is an important cancer target,[61,62] and the main protease (M^{pro}) of the SARS-CoV-2 virus.[63]

To assess the impact of ACES, we performed RBFE computations on the entire dataset with and without the use of the enhanced sampling approach. As discussed in the Method section, ACES relies on the creation of focused enhanced sampling states by scaling certain inter- and intramolecular energy terms of atoms assigned to the SC region and then connecting the enhanced and real state conformational ensembles using alchemical transformation pathways within an HREMD network. In this study, in addition to all electrostatic and Lennard-Jones (LJ) energy terms between the SC region and the environment, the internal electrostatic and torsion energy terms associated with rotatable bonds for the atoms in the SC region were also scaled.

While the choice of the atoms to be targeted for enhanced sampling is problem specific, a good rule-of-thumb is to choose a minimal set of atoms that can distinguish between different and potentially important conformational states. In practice, our automated workflow assigns entire chemical moieties up to the nearest rotatable bond to the SC region (in our case, atoms selected for the SC region correspond to those targeted for enhanced sampling), even if only a part of the chemical moiety changes during the perturbation. This choice facilitates enhanced sampling of plausible conformations of the entire moiety, which may include high-barrier ring flipping or rotamer transitions. We compared the results obtained from RBFE calculations with ACES to those obtained from an

https://doi.org/10.1021/acs.jcim.4c00464
J. Chem. Inf. Model. 2024, 64, 7046−7055

analogous set of RBFE calculations without ACES. The only difference between these calculations is that the internal electrostatic and torsion energy terms of the SC region are not scaled, which prevents the creation of the alchemically enhanced states. In both sets, we calculated multiple replicates for each RBFE edge in both forward and reverse directions to obtain precision estimates.

The results from the RBFE computations on the various targets with and without ACES are summarized in Table 1. Across the different datasets, all three metrics related to the accuracy of the RBFE computations, specifically the Pearson correlation (R^2), mean unsigned error (MUE) and root mean squared error (RMSE), obtained from simulations using ACES are either comparable to or slightly better (higher R^2, lower MUE, RMSE) than those obtained using simulations without ACES, suggesting that with ACES, the accuracy of RBFE calculations is maintained. Raw $\Delta\Delta G$s obtained from RBFE computations with or without ACES achieve an overall MUE between the predicted and experimental free energies of approximately 1 kcal/mol across the entire dataset (MUE of 1.02 ± 0.10 without ACES, MUE of 0.91 ± 0.10 with ACES). Indeed, the MUE is less than 1.5 kcal/mol for 8 out of 10 targets, with the exceptions being KRAS-G12D and SARS-CoV-2, for both of which the maximum MUE is less than 2 kcal/mol. Overall, these results highlight the predictive nature of FEP calculations and are comparable to those obtained by Schrödinger's FEP+ and other FEP platforms.[25,73]

The average spread and hysteresis, both measures of precision in these calculations, are calculated as the average of the absolute difference between the minimum and maximum RBFEs obtained for a given perturbation from independent calculations and the average of the absolute difference between independent RBFEs computed in the forward direction and in the reverse direction for a given perturbation, respectively. With ACES, we observe a systematic decrease in the spreads for every target in the dataset, with the calculated spread on each target being, on average, 0.4 kcal/mol lower than those calculated from simulations without ACES. The most noticeable change is observed for KRAS-G12D, where the spread decreases by more than 1.5 kcal/mol with the use of ACES. The overall average spread and hysteresis across the entire dataset are 1.02 ± 0.15 and 0.87 ± 0.13, respectively, in RBFEs without ACES. With the use of ACES, the overall average hysteresis is reduced by 37% (from 0.87 ± 0.13 to 0.55 ± 0.05 kcal/mol), and the overall average spread is reduced by 25% (from 1.02 ± 0.10 to 0.76 ± 0.06 kcal/mol). This observation is further highlighted in Figure 1a, which compares the distributions of the standard deviations in RBFEs obtained with and without ACES on the entire dataset. The distributions in the ACES simulations are clearly left-shifted, with the mean shifted by 0.1 kcal/mol, which is consistent with Table 1. This improvement in precision afforded by ACES, while it may seem numerically small, is a significant result; as noted earlier, the "no ACES" simulations are identical to the ACES simulations, except for the scaling of the specific SC internal energy terms. For example, the "no ACES" simulations were still performed with the latest smoothstep softcore potentials[40] and involved conventional HREMD. Improvement across the board on such a diverse set of targets over the existing state-of-the-art demonstrates the broad applicability of the ACES approach in improving precision in RBFE calculations.

Table 1. Summary Table for Raw $\Delta\Delta G$ (kcal/mol) with and without ACES[a]

	$\Delta\Delta G$ (no ACES)					$\Delta\Delta G$ (ACES)				
system	R^2	MUE	RMSE	spread	hysteresis	R^2	MUE	RMSE	spread	hysteresis
BACE	0.11 ± 0.07	1.01 ± 0.12	1.35 ± 0.12	0.90 ± 0.12	1.04 ± 0.11	0.31 ± 0.09	0.72 ± 0.09	0.99 ± 0.09	0.79 ± 0.08	0.49 ± 0.04
CDK2	0.42 ± 0.14	0.89 ± 0.10	1.02 ± 0.10	1.50 ± 0.26	1.07 ± 0.14	0.56 ± 0.14	0.66 ± 0.09	0.80 ± 0.09	1.01 ± 0.12	0.68 ± 0.08
JNK1	0.27 ± 0.14	0.86 ± 0.13	1.11 ± 0.13	0.85 ± 0.21	0.68 ± 0.19	0.36 ± 0.16	0.76 ± 0.10	0.95 ± 0.10	0.68 ± 0.10	0.41 ± 0.07
MCL1	0.25 ± 0.09	1.32 ± 0.13	1.71 ± 0.13	0.83 ± 0.11	0.56 ± 0.07	0.28 ± 0.09	1.26 ± 0.14	1.68 ± 0.14	0.62 ± 0.08	0.49 ± 0.06
p38	0.65 ± 0.07	0.83 ± 0.09	1.05 ± 0.09	0.94 ± 0.10	0.70 ± 0.09	0.65 ± 0.07	0.76 ± 0.07	0.91 ± 0.07	0.79 ± 0.09	0.55 ± 0.05
thrombin	0.03 ± 0.10	0.73 ± 0.14	0.92 ± 0.14	1.03 ± 0.18	0.55 ± 0.07	0.11 ± 0.14	0.64 ± 0.12	0.78 ± 0.12	0.80 ± 0.16	0.46 ± 0.08
Tyk2	0.82 ± 0.10	0.39 ± 0.06	0.52 ± 0.06	0.68 ± 0.11	0.40 ± 0.06	0.80 ± 0.11	0.43 ± 0.06	0.54 ± 0.06	0.43 ± 0.05	0.30 ± 0.03
PTP1B	0.66 ± 0.12	1.02 ± 0.13	1.32 ± 0.13	1.06 ± 0.17	1.15 ± 0.16	0.64 ± 0.13	1.02 ± 0.14	1.38 ± 0.14	0.75 ± 0.09	0.76 ± 0.11
KRAS-G12D	0.77 ± 0.25	1.76 ± 0.23	1.89 ± 0.23	3.35 ± 0.85	2.34 ± 0.65	0.81 ± 0.18	1.89 ± 0.32	2.12 ± 0.32	1.57 ± 0.27	1.09 ± 0.25
SARS-CoV-2	0.35 ± 0.18	1.71 ± 0.21	1.86 ± 0.21	1.15 ± 0.17	1.59 ± 0.65	0.51 ± 0.19	1.44 ± 0.20	1.60 ± 0.20	0.87 ± 0.12	0.55 ± 0.09
overall	-	1.02 ± 0.10	1.28 ± 0.12	1.02 ± 0.15	0.87 ± 0.13	-	0.91 ± 0.10	1.17 ± 0.13	0.76 ± 0.06	0.55 ± 0.05

[a]Results presented here are from three independent trials. For each dataset, the column "spread" corresponds to the absolute difference between the maximum and minimum $\Delta\Delta G$ of an edge calculated from the three independent trials, averaged over the dataset. The column "hysteresis" corresponds to the average absolute difference between the $\Delta\Delta G$ of an edge from each of the three trials to the $\Delta\Delta G$ of that edge calculated in the reverse direction, averaged over the dataset.

https://doi.org/10.1021/acs.jcim.4c00464
J. Chem. Inf. Model. 2024, 64, 7046−7055

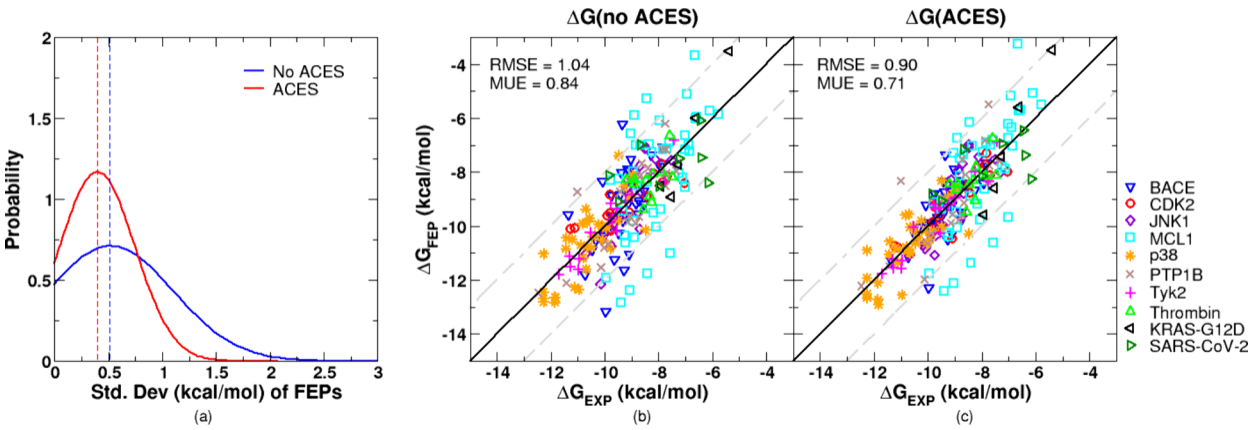


Figure 1. (a) The distribution of standard deviation values among three replicates for the entire dataset[25,61−63] with and without ACES. (b) Correlation plot of calculated and experimental binding free energy values for the entire dataset[25,61−63] without ACES. (c) Correlation plot of calculated and experimental binding free energy values for the entire dataset[25,61−63] with ACES.

We employed a rigorous variational network-wide MBAR analysis framework with cycle closure constraints (CCC) to convert the calculated $\Delta\Delta G$ values into absolute ΔG values. The network-wide MBAR with the CCC approach ensures that for each thermodynamic cycle in the network, all the participating free energies ($\Delta\Delta G$s) sum up to zero, and the free energy corrections arising from this process are distributed optimally among the various raw $\Delta\Delta G$s. For a more comprehensive understanding of this approach, readers are referred to the detailed work by Giese et al.[72] Figure 1b,c shows that the absolute ΔG values of all ligands from the various targets across the dataset are mostly comparable from calculations with and without ACES (Table S1 in the Supporting Information).

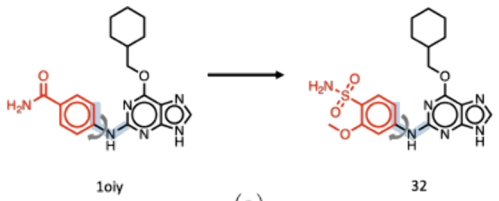
4.1. Case Studies. In this subsection, we highlight several problematic transformations in detail. These transformations posed challenges in traditional HREMD simulations, as they exhibited convergence issues with different simulations, yielding disparate free energy values. We found that the incorporation of ACES, however, proved to be instrumental in improving the convergence of these transformations.

4.1.1. Case Study 1: 1oiy → 32 in CDK2. The first case study is the perturbation between 1oiy and 32 in CDK2. We focused on analyzing the structural and conformational aspects that contributed to the observed outlier in the absence of ACES. The structure of this perturbation is depicted in Figure 2a, where the red colored atoms represent the softcore region. During our simulations without ACES, we identified an outlier with a calculated $\Delta\Delta G$ = 1.07 kcal/mol among three trials (see more details in the Supporting Information). This outlier arose due to the particular conformational behavior of the 1oiy ligand in CDK2. Specifically, 1oiy was observed to become trapped in a local minimum, forming a hydrogen bond between the ligand and an aspartic acid (ASP) residue in the protein. This interaction greatly favored the transformation, leading to the exceptionally favorable binding free energy change (Figure 2b). To further investigate the conformational sampling and its influence on the outcome, we examined the torsion distribution in the blue region shown in Figure 2a. The torsion distribution was obtained from three independent trials, and the corresponding histogram is displayed in Figure 2c. In the absence of ACES, the torsion distribution exhibited a propensity to remain in an *anti*-conformation (around 180°) in the dummy state. Consequently, the conformation in the real state also became stuck in an *anti*-conformation due to the limited conformational sampling. With the incorporation of ACES, a remarkable difference was observed in the torsion distribution. With ACES, the torsion angles were equally distributed across all angles, indicating a more diverse sampling in the dummy state. This enhanced sampling allowed for a broader exploration of the conformational space, enabling the ligand to escape from the *anti*-conformational trap and sample a wider range of conformations in the real state.

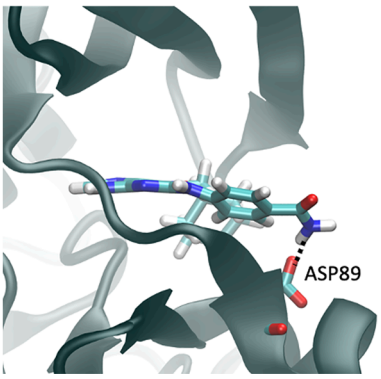
4.1.2. Case Study 2: 28 → 47 in MCL1. In the second case study involving the perturbation between 28 and 47 in MCL1, we examined the impact of ACES on the conformational behavior and the resulting calculated binding free energy values. This perturbation involves two distinct conformations, namely *syn* and *anti*, in ligand 47, as illustrated in Figure 3b. In the absence of ACES, an outlier among three independent trials with a calculated $\Delta\Delta G$ of 2.01 kcal/mol was identified. In contrast, the other two trials yielded $\Delta\Delta G$ values of −0.40 and −0.76 kcal/mol, as described in more detail in the Supporting Information. To understand the origin of the outlier and the influence of conformational sampling, we monitored the torsion distribution in the blue region of Figure 3a across three trials. In run 1, ligand 47 remained in the *anti*-conformation, while in run 2 and run 3, it predominantly adopted the *syn*-conformation. This conformational bias was reflected in the real state for each run, resulting in distinct calculated binding free energy values (see the top panel of Figure 3c. However, with the incorporation of ACES, a significant improvement in conformational sampling was observed. The torsion angles became equally distributed across all angles in the dummy state, as depicted in the bottom panel of Figure 3c. This enhanced sampling allowed for a more diverse exploration of the conformational space, enabling the ligand to escape any conformational traps and sample a wider range of conformations in the real state. Consequently, the incorporation of ACES led to more precise and consistent results for perturbations between 28 and 47 in MCL1. By promoting more diverse sampling in the dummy state, ACES facilitated the exploration of multiple conformational states, resulting in improved accuracy and reliability in the calculated binding free energy values.

4.1.3. Case Study 3: 23467 → 23469 in PTP1B. In the third case study involving the perturbation between 23467 and 23469 in the PTP1B protein target, we observed notable

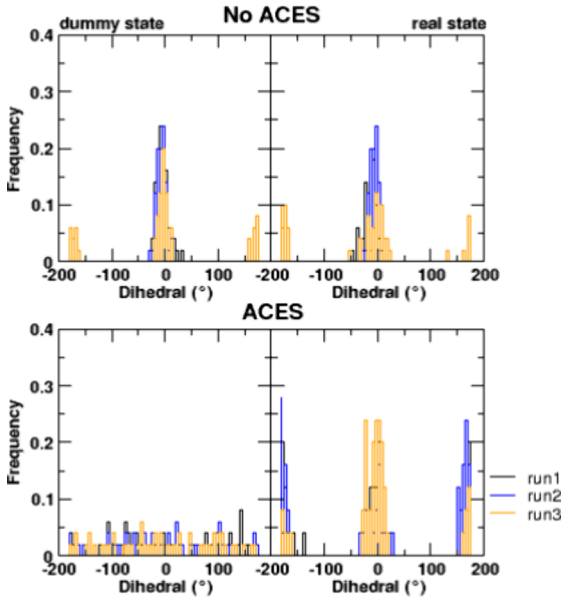
https://doi.org/10.1021/acs.jcim.4c00464
J. Chem. Inf. Model. 2024, 64, 7046−7055



(a)



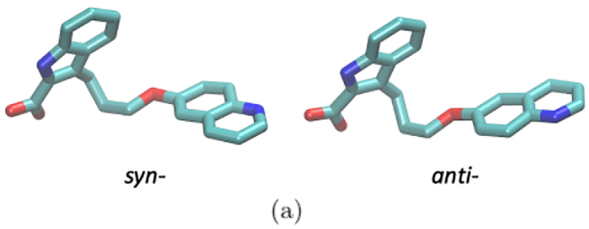
(b)



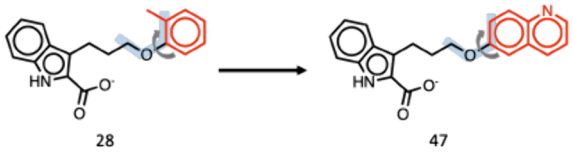
(c)

Figure 2. (a) Illustration of the CC/SC regions and torsion angle of the phenyl ring of the 1oiy ligand. The red colored atoms represent the softcore region, and the blue shadow regions highlight the torsion angle of the phenyl ring of the ligands. (b) Illustration of the interaction between 1oiy ligand and CDK2 protein target. (c) Torsion distribution with and without ACES in CDK2 protein target. The figure presents the torsion distribution plots obtained from three independent trials, comparing the results with and without the use of ACES. The top panel shows the torsion distributions without ACES, while the bottom panel illustrates the distributions with ACES. The left column displays the torsion distribution in the dummy state, and the right column depicts the distribution in the real state.

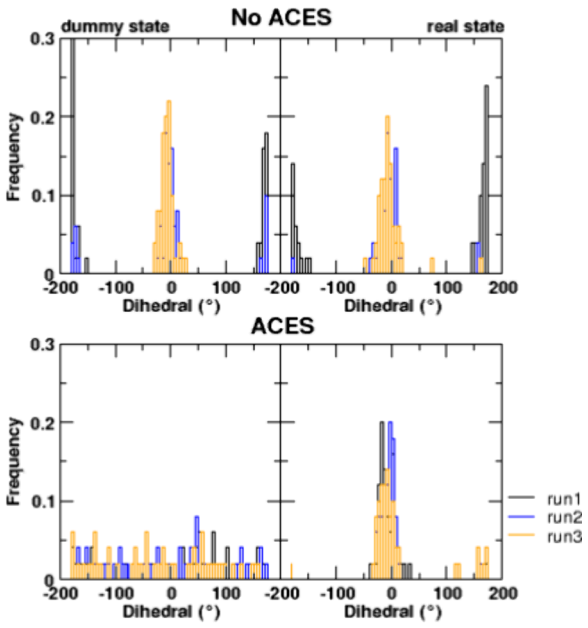
differences in the calculated $\Delta\Delta G$ values with and without the ACES method. Without ACES, the average calculated $\Delta\Delta G$ was 2.04 ± 1.41 kcal/mol, significantly deviating from the experimental value of −0.38 kcal/mol by 2.42 kcal/mol.



(a)



(b)



(c)

Figure 3. (a) Illustration of the common core/softcore (CC/SC) regions and torsion angle of the quinoline ring of ligand 47. The red colored atoms represent the softcore region, and the blue shadow regions highlight the torsion angle of the ligands. (b) Illustration of *syn*- and *anti*-conformations of ligand 47 in MCL1. (c) Torsion distribution with and without ACES in MCL1 protein target. The figure presents the torsion distribution plots obtained from three independent trials, comparing the results with and without the use of ACES. The top panel shows the torsion distributions without ACES, while the bottom panel illustrates the distributions with ACES. The left column displays the torsion distribution in the dummy state, while the right column depicts the distribution in the real state.

However, in the presence of ACES, the average calculated $\Delta\Delta G$ improved to −0.20 ± 0.21 kcal/mol, reducing the difference to only 0.18 kcal/mol compared to the experimental value. To gain insights into the underlying reasons for these discrepancies, we examined the torsion distribution in the blue region of Figure 4a in the PTP1B protein target for each independent run with and without ACES, as shown in Figure 4b. Without ACES, the torsion distribution exhibited a high concentration in the *anti*-conformation in both the dummy and real states. This suggests that the system remained trapped in the initial structure, limiting its conformational exploration. In

https://doi.org/10.1021/acs.jcim.4c00464
J. Chem. Inf. Model. 2024, 64, 7046−7055

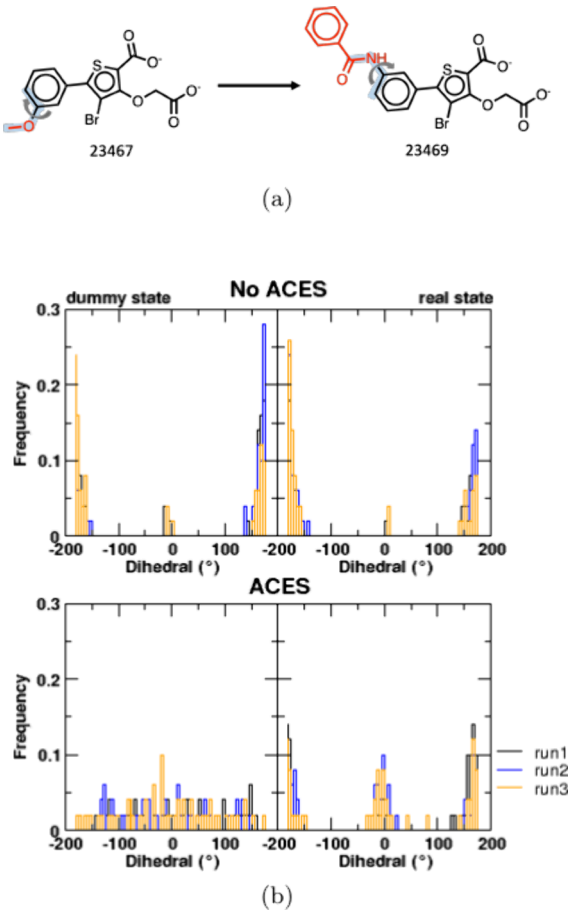


Figure 4. (a) Illustration of the common core/softcore (CC/SC) regions in perturbations of 23467 and 23469 ligands. The red colored atoms represent the softcore region, and the blue shadow regions highlight the torsion angle of the ligands. (b) Torsion distribution with and without ACES in PTP1B protein target. The figure presents the torsion distribution plots obtained from three independent trials, comparing the results with and without the use of ACES. The top panel shows the torsion distributions without ACES, while the bottom panel illustrates the distributions with ACES. The left column displays the torsion distribution in the dummy state, while the right column depicts the distribution in the real state.

contrast, with the incorporation of ACES, the torsion angles were evenly distributed across all angles in the dummy state. This balanced distribution allowed for more diverse sampling in the enhanced sampling state, facilitating the exploration of a wider range of conformations in the real state. Consequently, the torsion distribution in the real state became more similar among the three independent runs, indicating improved precision in the results. By leveraging the enhanced sampling capabilities of ACES, the PTP1B protein target was able to overcome the limitations of being trapped in a specific conformation and achieve a more accurate representation of the true thermodynamics. The more precise results obtained with ACES demonstrate its effectiveness in enhancing the convergence and accuracy of the calculated $\Delta\Delta G$ values, bridging the gap between computational predictions and experimental observations.

5. CONCLUSION

In this study, the impact of the ACES approach on the accuracy and precision of RBFE calculations was explored by performing RBFE calculations on a broad dataset containing 10 protein targets and over 200 ligands, with and without the use of the ACES. Each RBFE computation was repeated 3 times in the forward direction and once in the reverse direction to ensure reproducibility of the results and assess precision in these calculations. Incorporation of ACES was found to have a minor impact on the overall accuracy of the computed RBFEs; the MUE, RMSE, and correlation (R^2) to experimental data were comparable between RBFEs performed with and without ACES. However, the agreement between independent replicates of RBFEs was found to be systematically higher across all targets for calculations performed with ACES. While this result does not lead to a quantitative estimate of the difference in precision in RBFEs computed with and without ACES, consistently lower hysteresis and spread in RBFEs with ACES across all targets strongly suggest that ACES contributes to improving the precision of the RBFE calculations.

Three case studies were highlighted to investigate problematic transformations in detail. In the first case study involving the perturbation between 1oiy and 32 in CDK2, the impact of conformational behavior on the results was analyzed. The incorporation of ACES facilitated more diverse conformational sampling, leading to improved accuracy and reliability in the calculated binding free energy values. In the second case study involving the perturbation between 28 and 47 in MCL1, the influence of ACES on conformational behavior and calculated binding free energy values was examined. The results showed that ACES allowed for a broader exploration of the conformational space, resulting in more precise and consistent results. In the third case study involving the perturbation between 23467 and 23469 in PTP1B, significant differences were observed in the calculated $\Delta\Delta G$ values with and without ACES. The incorporation of ACES improved the agreement between calculated and experimental values, indicating its effectiveness in enhancing the convergence and accuracy of the predictions. Additionally, an efficient RBFE protocol was proposed to improve the computational efficiency. Instead of conducting three independent trials for each perturbation, the protocol involved one forward-direction run and one reverse-direction run. The results obtained from this more efficient protocol showed accuracy and precision comparable to those of the original protocol while significantly reducing computational resources. Overall, the study demonstrated the benefits of incorporating ACES and employing cycle closure constraint analysis in improving the accuracy, precision, and reliability of the relative binding free energy calculations. The proposed efficient RBFE protocol provided a streamlined approach without compromising the quality of the computational predictions, offering a time- and resource-efficient alternative for drug discovery applications.

■ ASSOCIATED CONTENT

Data Availability Statement

Details related to the RBFE protocols employed in this study are provided in the Method section. All protein and ligand starting structures, including AMBER topology files (that include protein and ligand forcefield parameters) and restart files and relevant AMBER input files are made available in PDB, SDF, PARM7, RST7 and text formats, respectively, on GitHub (https://github.com/tandemai-inc/aces_submission_data) and are also deposited in Zenodo (DOI: 10.5281/zenodo.13305152).

https://doi.org/10.1021/acs.jcim.4c00464
J. Chem. Inf. Model. 2024, 64, 7046−7055

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jcim.4c00464.

Summary table for ΔG (kcal/mol) with and without ACES, summary table for raw $\Delta\Delta G$ (kcal/mol) with and without ACES, 2D structures of KRAS-G12D and SARS-CoV-2 datasets, and computational cost comparison between HREMD and ACES (PDF)

Energy data for all datasets (XLSX)

AUTHOR INFORMATION

Corresponding Authors

Darrin M. York − *Laboratory for Biomolecular Simulation Research, Institute for Quantitative Biomedicine, and Department of Chemistry and Chemical Biology, Rutgers, The State University of New Jersey, Piscataway, New Jersey 08854, United States;* orcid.org/0000-0002-9193-7055; Email: Darrin.York@rutgers.edu

Abir Ganguly − *TandemAI, New York, New York 10036, United States;* orcid.org/0000-0002-0630-1109; Email: abir.ganguly@tandemai.com

Albert Pan − *TandemAI, New York, New York 10036, United States;* orcid.org/0000-0001-5050-5603; Email: albert.pan@tandemai.com

Authors

Hsu-Chun Tsai − *TandemAI, New York, New York 10036, United States; Laboratory for Biomolecular Simulation Research, Institute for Quantitative Biomedicine, and Department of Chemistry and Chemical Biology, Rutgers, The State University of New Jersey, Piscataway, New Jersey 08854, United States;* orcid.org/0000-0001-7027-5649

James Xu − *TandemAI, New York, New York 10036, United States;* orcid.org/0009-0005-7417-7137

Zhenyu Guo − *TandemAI, New York, New York 10036, United States;* orcid.org/0009-0001-7250-4895

Yinhui Yi − *TandemAI, New York, New York 10036, United States;* orcid.org/0009-0005-3854-4833

Chuan Tian − *TandemAI, New York, New York 10036, United States;* orcid.org/0000-0002-3518-7237

[#]Xinyu Que − *TandemAI, New York, New York 10036, United States;* orcid.org/0000-0002-4057-3255

Timothy Giese − *Laboratory for Biomolecular Simulation Research, Institute for Quantitative Biomedicine, and Department of Chemistry and Chemical Biology, Rutgers, The State University of New Jersey, Piscataway, New Jersey 08854, United States;* orcid.org/0000-0002-0653-9168

Tai-Sung Lee − *Laboratory for Biomolecular Simulation Research, Institute for Quantitative Biomedicine, and Department of Chemistry and Chemical Biology, Rutgers, The State University of New Jersey, Piscataway, New Jersey 08854, United States;* orcid.org/0000-0003-2110-2279

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jcim.4c00464

Notes

The authors declare no competing financial interest.

[#]The work was done while X.Q. was working at TandemAI.

ACKNOWLEDGMENTS

The authors are grateful for the financial support provided by the National Institutes of Health (No. GM107485 to D.M.Y.). Computational resources were provided in part by the Extreme Science and Engineering Discovery Environment (XSEDE), which is supported by National Science Foundation grant ACI-1548562 (supercomputer Expanse at SDSC through allocation CHE190067), and by the Texas Advanced Computing Center (TACC) at the University of Texas at Austin (supercomputer Longhorn through allocation CHE20002). The authors would also like to thank Thomas Watson at TandemAI for assistance with organizing data shared on Github and Zenodo.

REFERENCES

(1) Jorgensen, W. L. Efficient drug lead discovery and optimization. *Acc. Chem. Res.* **2009**, *42*, 724−733.

(2) York, D. M. Modern Alchemical Free Energy Methods for Drug Discovery Explained. *ACS Phys. Chem. Au* **2023**, *3*, 478−491.

(3) 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**, *60*, 5595−5623.

(4) 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.

(5) Mey, A. S.J.S.; Allen, B. K.; Bruce Macdonald, H. E.; Chodera, J. D.; Hahn, D. F.; Kuhn, M.; Michel, J.; Mobley, D. L.; Naden, L. N.; Prasad, S.; Rizzi, A.; Scheen, J.; Shirts, M. R.; Tresadern, G.; Xu, H. Best Practices for Alchemical Free Energy Calculations [Article v1.0]. *Living Journal of Computational Molecular Science* **2020**, *2*, 18378−18429.

(6) Cournia, Z.; Chipot, C.; Roux, B.; York, D. M.; Sherman, W. Free Energy Methods in Drug Discovery-Introduction. *ACS Symp. Ser.* **2021**, *1397*, 1−38.

(7) Chodera, J. D.; Mobley, D. L.; shirts, M. R.; Dixon, R. W.; Branson, K.; Pande, V. S. Alchemical free energy methods for drug discovery: progress and challenges. *Curr. Opin. Struct. Biol.* **2011**, *21*, 150−160.

(8) 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*, 1157−1174.

(9) Träg, J.; Zahn, D. Improved GAFF2 Parameters for Fluorinated Alkanes and Mixed Hydro- and Fluorocarbons. *J. Mol. Model.* **2019**, *25*, 39.

(10) Vanommeslaeghe, K.; MacKerell, A. D. Automation of the CHARMM General Force Field (CGenFF) I: Bond Perception and Atom Typing. *J. Chem. Inf. Model.* **2012**, *52*, 3144−3454.

(11) Vanommeslaeghe, K.; Raman, E. P.; MacKerell, A. D. Automation of the CHARMM General Force Field (CGenFF) II: Assignment of Bonded Parameters and Partial Atomic Charges. *J. Chem. Inf. Model.* **2012**, *52*, 3155−3168.

(12) 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*, 1863−1874.

(13) Lu, C.; Wu, C.; Ghoreishi, D.; Chen, W.; Wang, L.; Damm, W.; Ross, G. A.; Dahlgren, M. K.; Russell, E.; Von Bargen, C. D.; 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.

(14) Stone, J. E.; Hardy, D. J.; Ufimtsev, I. S.; Schulten, K. GPU-accelerated Molecular Modeling Coming of Age. *J. Mol. Graphics Modell* **2010**, *29*, 116−125.

(15) Salomon-Ferrer, R.; Gotz, A. W.; Poole, D.; Le Grand, S.; Walker, R. C. Routine microsecond molecular dynamics simulations with AMBER on GPUs. 2. Explicit solvent Particle Mesh Ewald. *J. Chem. Theory Comput.* **2013**, *9*, 3878−3888.

(16) Lindert, S.; Bucher, D.; Eastman, P.; Pande, V.; McCammon, J. A. Accelerated Molecular Dynamics Simulations with the AMOEBA

Polarizable Force Field on Graphics Processing Units. *J. Chem. Theory Comput.* **2013**, *9*, 4684−4691.
(17) Giese, T. J.; York, D. M. A GPU-Accelerated Parameter Interpolation Thermodynamic Integration Free Energy Method. *J. Chem. Theory Comput.* **2018**, *14*, 1564−1582.
(18) Lee, T.-S.; Cerutti, D. S.; Mermelstein, D.; Lin, C.; LeGrand, S.; Giese, T. J.; Roitberg, A.; Case, D. A.; Walker, R. C.; York, D. M. GPU-Accelerated Molecular Dynamics and Free Energy Methods in Amber18: Performance Enhancements and New Features. *J. Chem. Inf. Model.* **2018**, *58*, 2043−2050.
(19) Kutzner, C.; Kniep, C.; Cherian, A.; Nordstrom, L.; Grubmuller, H.; de Groot, B. L.; Gapsys, V. GROMACS in the Cloud: A Global Supercomputer to Speed Up Alchemical Drug Design. *J. Chem. Inf. Model.* **2022**, *62*, 1691−1711.
(20) Ganguly, A.; Tsai, H.-C.; Fernández-Pendás, M.; Lee, T.-S.; Giese, T. J.; York, D. M. AMBER Drug Discovery Boost Tools: Automated Workflow for Production Free-Energy Simulation Setup and Analysis (ProFESSA). *J. Chem. Theory Comput.* **2022**, *62*, 6069−6083.
(21) Zhang, H.; Kim, S.; Giese, T. J.; Lee, T.-S.; Lee, J.; York, D. M.; Im, W. CHARMM-GUI Free Energy Calculator for Practical Ligand Binding Free Energy Simulations with AMBER. *J. Chem. Inf. Model* **2021**, *61*, 4145−4151.
(22) Klimovich, P. V.; Mobley, D. L. A Python Tool to Set Up Relative Free Energy Calculations in GROMACS. *J. Comput. Aided Mol. Des.* **2015**, *29*, 1007−1014.
(23) Loeffler, H. H.; Michel, J.; Woods, C. FESetup: Automating Setup for Alchemical Free Energy Simulations. *J. Chem. Inf. Model* **2015**, *55*, 2485−2490.
(24) Fu, H.; Gumbart, J. C.; Chen, H.; Shao, X.; Cai, W.; Chipot, C. BFEE: A User-Friendly Graphical Interface Facilitating Absolute Binding Free-Energy Calculations. *J. Chem. Inf. Model* **2018**, *58*, 556−560.
(25) Wang, L.; Wu, Y.; Deng, Y.; Kim, B.; Pierce, L.; Krilov, G.; Lupyan, D.; Robinson, S.; Dahlgren, M. K.; Greenwood, J.; Romero, D. L.; Masse, C.; Knight, J. L.; Steinbrecher, T.; Beuming, T.; Damm, W.; Harder, E.; Sherman, W.; Brewer, M.; Wester, R.; Murcko, M.; Frye, L.; Farid, R.; Lin, T.; Mobley, D. L.; Jorgensen, W. L.; Berne, B. J.; Friesner, R. A.; Abel, R. Accurate and Reliable Prediction of Relative Ligand Binding Potency in Prospective Drug Discovery by Way of a Modern Free-Energy Calculation Protocol and Force Field. *J. Am. Chem. Soc.* **2015**, *137*, 2695−2703.
(26) Pal, R. K.; Gallicchio, E. Perturbation potentials to overcome order/disorder transitions in alchemical binding free energy calculations. *J. Chem. Phys.* **2019**, *151*, 124116.
(27) Fu, H.; Shao, X.; Cai, W.; Chipot, C. Taming Rugged Free Energy Landscapes Using an Average Force. *Acc. Chem. Res.* **2019**, *52*, 3254−3264.
(28) Laio, A.; Parrinello, M. Escaping Free-energy Minima. *Proc. Natl. Acad. Sci. U.S.A* **2002**, *99*, 12562−12566.
(29) Perthold, J. W.; Petrov, D.; Oostenbrink, C. Toward Automated Free Energy Calculation with Accelerated Enveloping Distribution Sampling (A-EDS). *J. Chem. Inf. Model.* **2020**, *60*, 5395−5406.
(30) König, G.; Glaser, N.; Schroeder, B.; Kubincová, A.; Hünenberger, P. H.; Riniker, S. An Alternative to Conventional λ-Intermediate States in Alchemical Free Energy Calculations: λ-Enveloping Distribution Sampling. *J. Chem. Inf. Model.* **2020**, *60*, 5407−5423.
(31) Zheng, L.; Chen, M.; Yang, W. Random walk in orthogonal space to achieve efficient free-energy simulation of complex systems. *Proc. Natl. Acad. Sci. U.S.A* **2008**, *105*, 20227−20232.
(32) Bieler, N. S.; Hünenberger, P. H. Orthogonal sampling in free-energy calculations of residue mutations in a tripeptide: TI versus λ-LEUS. *J. Comput. Chem.* **2015**, *36*, 1686−1697.
(33) Sugita, Y.; Okamoto, Y. Replica-exchange molecular dynamics method for protein folding. *Chem. Phys. Lett.* **1999**, *314*, 141−151.
(34) Zhou, R.; Berne, B. J. Can a continuum solvent model reproduce the free energy landscape of a β-hairpin folding in water? *Proc. Natl. Acad. Sci. U.S.A* **2002**, *99*, 12777−12782.
(35) Comer, J.; Phillips, J. C.; Schulten, K.; Chipot, C. Multiple-Replica Strategies for Free-Energy Calculations in NAMD: Multiple-Walker Adaptive Biasing Force and Walker Selection Rules. *J. Chem. Theory Comput.* **2014**, *10*, 5276−5285.
(36) Liu, P.; Kim, B.; Friesner, R. A.; Berne, B. J. Replica exchange with solute tempering: A method for sampling biological systems in explicit water. *Proc. Natl. Acad. Sci. U.S.A* **2005**, *102*, 13749−13754.
(37) Wang, L.; Friesner, R. A.; Berne, B. J. Replica Exchange with Solute Scaling: A More Efficient Version of Replica Exchange with Solute Tempering (REST2). *J. Phys. Chem. B* **2011**, *115*, 9431−9438.
(38) Kamiya, M.; Sugita, Y. Flexible Selection of the Solute Region in Replica Exchange with Solute Tempering: Application to Protein-folding Simulations. *J. Chem. Phys.* **2018**, *149*, No. 072304.
(39) Lee, T.-S. L.; Tsai, H.-C. T.; Ganguly, A.; York, D. M. ACES: Optimized Alchemically Enhanced Sampling. *J. Chem. Theory Comput.* **2023**, *19*, 472−487.
(40) Tsai, H.-C.; Lee, T.-S.; Ganguly, A.; Giese, T. J.; Ebert, M. C.; Labute, P.; Merz, K. M. J.; York, D. M. AMBER Free Energy Tools: A New Framework for the Design of Optimized Alchemical Transformation Pathways. *J. Chem. Theory Comput.* **2023**, *19*, 640−658.
(41) Zhang, S.; Giese, T. J.; Lee, T.-S.; York, D. M. Alchemical Enhanced Sampling with Optimized Phase Space Overlap. *J. Chem. Theory Comput.* **2024**, *20*, 3935−3953.
(42) Case, D. A.; Aktulga, H. M.; Belfon, K.; Ben-Shalom, I. Y.; Berryman, J.; Brozell, S. R.; Cerutti, D. S.; Cheatham, T. E.; Cruzeiro, V. W. D.; Darden, T. A.; Duke, R. E.; Giambasu, G.; Gilson, M. K.; Gohlke, H.; Goetz, A. W.; Harris, R.; Izadi, S.; Izmailov, S. A.; Kasavajhala, K.; Kaymak, M. C.; King, E.; Kovalenko, A.; Kurtzman, T.; Lee, T. S.; LeGrand, S.; Li, P.; Lin, C.; Liu, J.; Luchko, T.; Luo, R.; Machado, M.; Man, V.; Manathunga, M.; Merz, K. M.; Miao, Y.; Mikhailovskii, O.; Monard, G.; Nguyen, H.; O'Hearn, K. A.; Onufriev, A.; Pan, F.; Pantano, S.; Qi, R.; Rahnamoun, A.; Roe, D.; Roitberg, A.; Sagui, C.; Schott-Verdugo, S.; Shajan, A.; Shen, J.; Simmerling, C. L.; Skrynnikov, N. R.; Smith, J.; Swails, J.; Walker, R. C.; Wang, J.; Wang, J.; Wei, H.; Wolf, R. M.; Wu, X.; Xiong, Y.; Xue, Y.; York, D. M.; Zhao, S.; Kollman, P. A. *AMBER22*; University of California, San Francisco, CA, USA, 2022.
(43) Fleck, M.; Wieder, M.; Boresch, S. Dummy Atoms in Alchemical Free Energy Calculations. *J. Chem. Theory Comput.* **2021**, *17*, 4403−4419.
(44) Lee, T.-S.; Tsai, H.-C.; Ganguly, A.; Giese, T. J.; York, D. M. Robust, Efficient and Automated Methods for Accurate Prediction of Protein-Ligand Binding Affinities in AMBER Drug Discovery Boost. In *Free Energy Methods in Drug Discovery: Current State and Future Directions*; Armacost, K. A., Thompson, D. C., Eds.; ACS Symposium Series, Vol. *1397*; American Chemical Society, 2021; p 161−204.
(45) Hritz, J.; Oostenbrink, C. Hamiltonian Replica Exchange Molecular Dynamics Using Soft-core Interactions. *J. Chem. Phys.* **2008**, *128*, 144121.
(46) Luitz, M. P.; Zacharias, M. Protein-Ligand Docking Using Hamiltonian Replica Exchange Simulations with Soft Core Potentials. *J. Chem. Inf. Model.* **2014**, *54*, 1669−1675.
(47) Torrie, G. M.; Valleau, J. P. Nonphysical Sampling Distributions in Monte Carlo Free- energy Estimation: Umbrella Sampling. *J. Comput. Phys.* **1977**, *23*, 187−199.
(48) Min, D.; Li, H.; Li, G.; Bitetti-Putzer, R.; Yang, W. Synergistic Approach to Improve "Alchemical" Free Energy Calculation in Rugged Energy Surface. *J. Chem. Phys.* **2007**, *126*, 144109.
(49) Baumann, H. M.; Dybeck, E.; McClendon, C. L.; Pickard, F. C., 4th; Gapsys, V.; Perez-Benito, L.; Hahn, D. F.; Tresadern, G.; Mathiowetz, A. M.; Mobley, D. L. Broadening the Scope of Binding Free Energy Calculations Using a Separated Topologies Approach. *J. Chem. Theory Comput.* **2023**, *19*, 5058−5076.
(50) Decherchi, S.; Cavalli, A. Thermodynamics and Kinetics of Drug-Target Binding by Molecular Simulation. *Chem. Rev.* **2020**, *120*, 12788−12833.

https://doi.org/10.1021/acs.jcim.4c00464
J. Chem. Inf. Model. 2024, 64, 7046−7055

(51) Wu, X. W.; Xu, L.-Y.; Li, E.-M.; Dong, G. Application of Molecular Dynamics Simulation in Biomedicine. *Chem. Biol. Drug Des.* **2022**, *99*, 789−800.
(52) Cumming, J. N.; Smith, E. M.; Wang, L.; Misiaszek, J.; Durkin, J.; Pan, J.; Iserloh, U.; Wu, Y.; Zhu, Z.; Strickland, C.; Voigt, J.; Chen, X.; Kennedy, M. E.; Kuvelkar, R.; Hyde, L. A.; Cox, K.; Favreau, L.; Czarniecki, M. F.; Greenlee, W. J.; McKittrick, B. A.; Parker, E. M.; Stamford, A. W. Structure based design of iminohydantoin BACE1 inhibitors: Identification of an orally available, centrally active BACE1 inhibitor. *Bioorg. Med. Chem. Lett.* **2012**, *22*, 2444−2449.
(53) Hardcastle, I. R.; Arris, C. E.; Bentley, J.; Boyle, F. T.; Chen, Y.; Curtin, N. J.; Endicott, J. A.; Gibson, A. E.; Golding, B. T.; Griffin, R. J.; Jewsbury, P.; Menyerol, J.; Mesguiche, V.; Newell, D. R.; Noble, M. E. M.; Pratt, D. J.; Wang, L.-Z.; Whitfield, H. J. N^2-Substituted O^6-Cyclohexylmethylguanine Derivatives: Potent Inhibitors of Cyclin-Dependent Kinases 1 and 2. *J. Med. Chem.* **2004**, *47*, 3710−3722.
(54) Szczepankiewicz, B. G.; Kosogof, C.; Nelson, L. T. J.; Liu, G.; Liu, B.; Zhao, H.; Serby, M. D.; Xin, Z.; Liu, M.; Gum, R. J.; Haasch, D. L.; Wang, S.; Clampit, J. E.; Johnson, E. F.; Lubben, T. H.; Stashko, M. A.; Olejniczak, E. T.; Sun, C.; Dorwin, S. A.; Haskins, K.; Abad-Zapatero, C.; Fry, E. H.; Hutchins, C. W.; Sham, H. L.; Rondinone, C. M.; Trevillyan, J. M. Aminopyridine-Based c-Jun N-Terminal Kinase Inhibitors with Cellular Activity and Minimal Cross-Kinase Activity. *J. Med. Chem.* **2006**, *49*, 3563−3580.
(55) 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.* **2013**, *56*, 15−30.
(56) Goldstein, D. M.; Soth, M.; Gabriel, T.; Dewdney, N.; Kuglstatter, A.; Arzeno, H.; Chen, J.; Bingenheimer, W.; Dalrymple, S. A.; Dunn, J.; Farrell, R.; Frauchiger, S.; La Fargue, J.; Ghate, M.; Graves, B.; Hill, R. J.; Li, F.; Litman, R.; Loe, B.; McIntosh, J.; McWeeney, D.; Papp, E.; Park, J.; Reese, H. F.; Roberts, R. T.; Rotstein, D.; San Pablo, B.; Sarma, K.; Stahl, M.; Sung, M.-L.; Suttman, R. T.; Sjogren, E. B.; Tan, Y.; Trejo, A.; Welch, M.; Weller, P.; Wong, B. R.; Zecic, H. Discovery of 6-(2,4-Difluorophenoxy)-2-[3-hydroxy-1-(2-hydroxyethyl)propylamino]-8-methyl-8H-pyrido-[2,3-d]pyrimidin-7-one (Pamapimod) and 6-(2,4-Difluorophenoxy)-8-methyl-2-(tetrahydro-2H-pyran-4-ylamino)pyrido[2,3-d]pyrimidin-7(8H)-one (R1487) as Orally Bioavailable and Highly Selective Inhibitors of p38α Mitogen-Activated Protein Kinase. *J. Med. Chem.* **2011**, *54*, 2255−2265.
(57) Wilson, D. P.; Wan, Z.-K.; Xu, W.-X.; Kirincich, S. J.; Follows, B. C.; Joseph-McCarthy, D.; Foreman, K.; Moretto, A.; Wu, J.; Zhu, M.; Binnun, E.; Zhang, Y.-L.; Tam, M.; Erbe, D. V.; Tobin, J.; Xu, X.; Leung, L.; Shilling, A.; Tam, S. Y.; Mansour, T. S.; Lee, J. L. 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.
(58) Baum, B.; Mohamed, M.; Zayed, M.; Gerlach, C.; Heine, A.; Hangauer, D.; Klebe, G. More than a Simple Lipophilic Contact: A Detailed Thermodynamic Analysis of Nonbasic Residues in the S1 Pocket of Thrombin. *J. Mol. Biol.* **2009**, *390*, 56−69.
(59) Liang, J.; Tsui, V.; Van Abbema, A.; Bao, L.; Barrett, K.; Beresini, M.; Berezhkovskiy, L.; Blair, W. S.; Chang, C.; Driscoll, J.; Eigenbrot, C.; Ghilardi, N.; Gibbons, P.; Halladay, J.; Johnson, A.; Kohli, P. B.; Lai, Y.; Liimatta, M.; Mantik, P.; Menghrajani, K.; Murray, J.; Sambrone, A.; Xiao, Y.; Shia, S.; Shin, Y.; Smith, J.; Sohn, S.; Stanley, M.; Ultsch, M.; Zhang, B.; Wu, L. C.; Magnuson, S. Lead identification of novel and selective TYK2 inhibitors. *Eur. J. Med. Chem.* **2013**, *67*, 175−187.
(60) Liang, J.; van Abbema, A.; Balazs, M.; Barrett, K.; Berezhkovsky, L.; Blair, W.; Chang, C.; Delarosa, D.; DeVoss, J.; Driscoll, J.; Eigenbrot, C.; Ghilardi, N.; Gibbons, P.; Halladay, J.; Johnson, A.; Kohli, P. B.; Lai, Y.; Liu, Y.; Lyssikatos, J.; Mantik, P.; Menghrajani, K.; Murray, J.; Peng, I.; Sambrone, A.; Shia, S.; Shin, Y.; Smith, J.; Sohn, S.; Tsui, V.; Ultsch, M.; Wu, L. C.; Xiao, Y.; Yang, W.; Young, J.; Zhang, B.; Zhu, B.-y.; Magnuson, S. Lead Optimization of a 4-Aminopyridine Benzamide Scaffold To Identify Potent, Selective, and Orally Bioavailable TYK2 Inhibitors. *J. Med. Chem.* **2013**, *56*, 4521−4536.
(61) Wang, X.; Allen, S.; Blake, J. F.; Bowcut, V.; Briere, D. M.; Calinisan, A.; Dahlke, J. R.; Fell, J. B.; Fischer, J. P.; Gunn, R. J.; Hallin, J.; Laguer, J.; Lawson, J. D.; Medwid, J.; Newhouse, B.; Nguyen, P.; O'Leary, J. M.; Olson, P.; Pajk, S.; Rahbaek, L.; Rodriguez, M.; Smith, C. R.; Tang, T. P.; Thomas, N. C.; Vanderpool, D.; Vigers, G. P.; Christensen, J. G.; Marx, M. A. Identification of MRTX1133, a Noncovalent, Potent, and Selective $KRAS^{G12D}$ Inhibitor. *J. Med. Chem.* **2022**, *65*, 3123−3133.
(62) Mao, Z.; Xiao, H.; Shen, P.; Yang, Y.; Xue, J.; Yang, Y.; Shang, Y.; Zhang, L.; Li, X.; Zhang, Y.; Du, Y.; Chen, C.-C.; Guo, R.-T.; Zhang, Y. KRAS(G12D) Can Be Targeted By Potent Inhibitors via Formation of Salt Bridge. *Cell Discovery* **2022**, *8*, 5.
(63) Han, S. H.; Goins, C. M.; Arya, T.; Shin, W.-J.; Maw, J.; Hooper, A.; Sonawane, D. P.; Porter, M. R.; Bannister, B. E.; Crouch, R. D.; Lindsey, A. A.; Lakatos, G.; Martinez, S. R.; Alvarado, J.; Akers, W. S.; Wang, N. S.; Jung, J. U.; Macdonald, J. D.; Stauffer, S. R. Structure-Based Optimization of ML300-Derived, Noncovalent Inhibitors Targeting the Severe Acute Respiratory Syndrome Coronavirus 3CL Protease (SARS-CoV-2 $3CL^{pro}$). *J. Med. Chem.* **2022**, *65*, 2880−2904.
(64) Case, D. A.; Aktulga, H. M.; Belfon, K.; Cerutti, D. S.; Cisneros, G. A.; Cruzeiro, V. W. D.; Forouzesh, N.; Giese, T. J.; Gotz, A. W.; Gohlke, H.; Izadi, S.; Kasavajhala, K.; Kaymak, M. C.; King, E.; Kurtzman, T.; Lee, T.-S.; Li, P.; Liu, J.; Luchko, T.; Luo, R.; Manathunga, M.; Machado, M. R.; Nguyen, H. M.; O'Hearn, K. A.; Onufriev, A. V.; Pan, F.; Pantano, S.; Qi, R.; Rahnamoun, A.; Risheh, A.; Schott-Verdugo, S.; Shajan, A.; Swails, J.; Wang, J.; Wei, H.; Wu, X.; Wu, Y.; Zhang, S.; Zhao, S.; Zhu, Q.; Cheatham, T. E., 3rd; Roe, D. R.; Roitberg, A.; Simmerling, C.; York, D. M.; Nagan, M. C.; Merz, K. M., Jr AmberTools. *J. Chem. Inf. Model.* **2023**, *63*, 6183−6191.
(65) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of Simple Potential Functions for Simulating Liquid Water. *J. Chem. Phys.* **1983**, *79*, 926−935.
(66) Hoover, W. G.; Ree, F. H. Use of Computer Experiments to Locate the Melting Transition and Calculate the Entropy in the Solid Phase. *J. Chem. Phys.* **1967**, *47*, 4873−4878.
(67) Miyamoto, S.; Kollman, P. A. SETTLE: An Analytical Version of the SHAKE and RATTLE Algorithm for Rigid Water Models. *J. Comput. Chem.* **1992**, *13*, 952−962.
(68) Ryckaert, J. P.; Ciccotti, G.; Berendsen, H. J. C. Numerical Integration of the Cartesian Equations of Motion of a System with Constraints: Molecular Dynamics of n-alkanes. *J. Comput. Phys.* **1977**, *23*, 327−341.
(69) Hopkins, C. W.; Le Grand, S.; Walker, R. C.; Roitberg, A. E. Long-Time-Step Molecular Dynamics through Hydrogen Mass Repartitioning. *J. Chem. Theory Comput.* **2015**, *11*, 1864−1874.
(70) Darden, T.; York, D.; Pedersen, L. Particle Mesh Ewald: An N·log(N) Method for Ewald Sums in Large Systems. *J. Chem. Phys.* **1993**, *98*, 10089−10092.
(71) Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. A Smooth Particle Mesh Ewald Method. *J. Chem. Phys.* **1995**, *103*, 8577−8593.
(72) Giese, T. J.; York, D. M. Variational Method for Networkwide Analysis of Relative Ligand Binding Free Energies with Loop Closure and Experimental Constraints. *J. Chem. Theory Comput.* **2021**, *17*, 1326−1336.
(73) 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.

https://doi.org/10.1021/acs.jcim.4c00464
J. Chem. Inf. Model. 2024, 64, 7046−7055