JCTC Journal of Chemical Theory and Computation

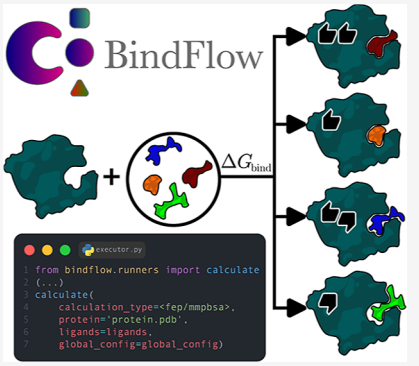
pubs.acs.org/JCTC Article

BindFlow: A Free, User-Friendly Pipeline for Absolute Binding Free Energy Calculations Using Free Energy Perturbation or MM(PB/GB)SA

Alejandro Martínez León,* Lucas Andersen, and Jochen S. Hub*

Cite This: J. Chem. Theory Comput. 2026, 22, 1198−1213 Read Online

ABSTRACT: We present BindFlow, a Python-based software for automated absolute binding free energy (ABFE) calculations at the free energy perturbation (FEP) or at the molecular mechanics Poisson−Boltzmann/generalized Born surface area [MM(PB/GB)SA] level of theory. BindFlow is free, open-source, user-friendly, and easily customizable, runs on workstations or distributed computing platforms, and provides extensive documentation and tutorials. BindFlow uses GROMACS as a molecular dynamics engine and provides built-in support for the small-molecule force fields GAFF, OpenFF, and Espaloma, as well as support for user-provided custom force fields. We test BindFlow by computing affinities for 139 receptor−ligand pairs, involving eight different targets, including six soluble proteins, one membrane protein, and one nonprotein host−guest system. We find that the agreement of BindFlow predictions with experiments is overall similar to gold standards in the field. Interestingly, we find that MM(PB/GB)SA achieves correlations that, for some systems and force fields, approach those obtained with FEP while requiring only a fraction of the computational cost. This study establishes BindFlow as a validated and accessible tool for ABFE calculations.



INTRODUCTION

During the early stages of drug discovery, a large number of chemical compounds are evaluated with the aim of finding potent binders, a process that is both time-consuming and resource-intensive.[1,2] Thus, computational models are widely used to estimate the binding free energy of small molecules to biological targets in order to prioritize compounds for follow-up synthesis and experimental evaluation.[3−6] In the 1950s, methods collectively known as alchemical free energy perturbation were introduced.[7] The term "free energy perturbation" (FEP) referred originally to a specific class of alchemical methods[8] but has since been more generally applied to alchemical binding free energy methods, a notion followed in this study. FEP enabled the estimation of free energy differences at a fraction of the computational cost compared with conventional molecular dynamics (MD) simulations. Nevertheless, FEP remained computationally demanding for high-throughput studies, rationalizing the development of computationally more efficient yet more approximate end-point free energy techniques such as the molecular mechanics Poisson−Boltzmann/generalized Born surface area [MM(PB/GB)SA] methods,[9] which achieved varying degrees of success.[10] Recent advances in biomolecular simulation methods[6,11] combined with the growth of computational power have enabled increasingly accurate affinity predictions,[6,12] making both FEP and MM(PB/GB)SA calculations more routine.

Binding free energy estimation has been categorized into relative (RBFE) and absolute binding free energy (ABFE) calculations.[11] RBFE is typically preferred for ranking congeneric molecular series[13] or for evaluating the effects of mutations on ligand binding.[14] Alchemical RBFE methodologies have seen substantial advancements, resulting in numerous implementations.[15−29] These developments have enabled alchemical RBFE methods to achieve remarkable accuracy, often with a root mean-squared error (RMSE) of 1−2 kcal mol^{-1} relative to experimental values,[30] making them highly effective for drug discovery applications.[13−15,31−38]

ABFE calculations yield the binding free energy relative to a standard state of the compound in solution.[39] ABFE calculations are particularly useful for studying highly diverse molecular sets,[4,40−42] for binding pose validation,[43,44] or for multitarget selectivity prediction.[45,46] Alchemical ABFE methods have demonstrated their power in addressing complex challenges across various projects,[4,34,40,41,45−49] and large-scale comparisons with experimental data have revealed that

Received: December 5, 2025
Revised: December 27, 2025
Accepted: December 29, 2025
Published: January 7, 2026

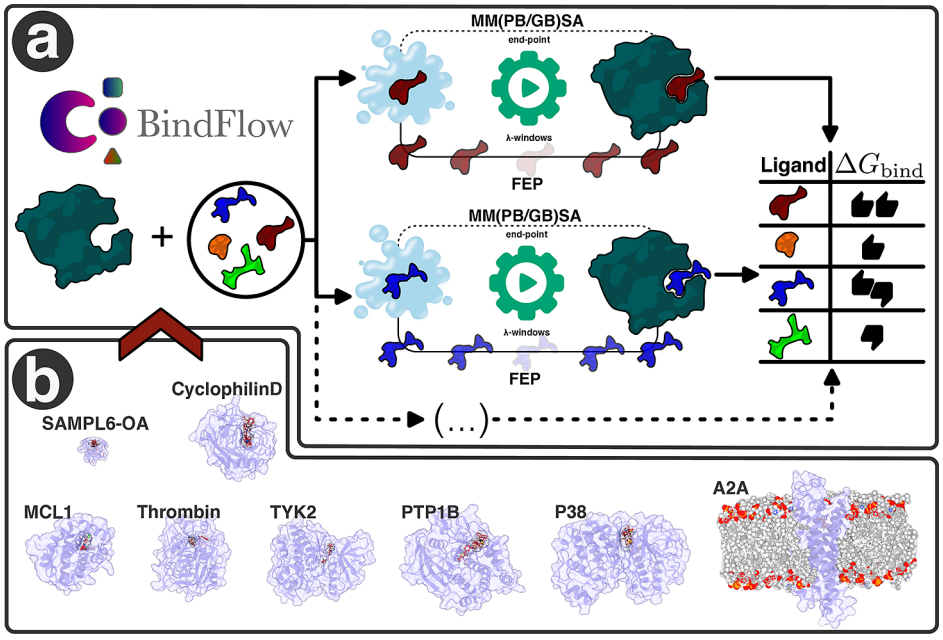


Figure 1. (a) Schematic representation of the BindFlow workflow and (b) the systems used in this study for validating ABFE calculations.

alchemical ABFE predictions are approaching experimental accuracy.[50]

However, the preparation, execution, and analysis of ABFE simulations are tedious, system-dependent processes that frequently rely on expert intervention. Fully automating the binding free energy pipeline can mitigate these challenges, enabling high-throughput calculations for drug discovery campaigns while improving reproducibility and accessibility. We suggest that a robust and accessible ABFE workflow should meet six properties: (i) open-source commitment to ensure transparency and enable community-driven method and code development; (ii) efficiency by a flexible and error-tolerant use of computational resources, in particular in distributed platforms such as high-performance computing (HPC) clusters; (iii) reliability, as demonstrated through extensive benchmarking across sets of diverse ligands and receptors, including challenging targets; (iv) flexibility for expert users to test effects of using different force fields or diverse simulation and deployment parameters; (v) multiligand end-to-end automation, reducing human error and setup time and increasing reproducibility; and (vi) accessibility, supported by comprehensive documentation and tutorials. While several powerful ABFE workflows have been established, they typically emphasize only a subset of these properties.

Commercial solutions such as FEP+[47] have set high standards for reliability and automation in the pharmaceutical industry,[47] and other commercial platforms like XFEP[17] have contributed to broader adoption. However, their closed-source design and licensing costs restrict accessibility for broad academic use and do not support the methodological development by the scientific community. In contrast, academic and open-source tools[23,29,51−61] have played a key role in democratizing ABFE calculations and expanding the ecosystem of automated free energy workflows. However, available academic tools are limited to diverse aspects. Certain tools have been validated using rather small receptor−ligand data sets or provide limited or even nonexisting documentation or tutorials. Other tools automate only part of the workflow, such as the setup of a simulation system and topologies but not the deployment to distributed computing environments or data analysis. Certain tools are focused on specific force fields, use a nonfree MD engine, or support limited options for expert users for customizing MD parameters or the λ-schedule. A brief overview of available tools is provided in the Supporting Discussion. In summary, to the best of our knowledge, no available software for ABFE calculations commits to all six properties outlined above.

Here, we aim to close this gap by introducing *BindFlow*, a free, user-friendly, open-source software package that provides a multiligand, end-to-end automated ABFE workflow and offers fine-grained control over simulation and deployment parameters (Figure 1a). BindFlow implements two ABFE methods: (i) the end-point free energy method MM(PB/GB)SA,[9] based on a simulation of the receptor−ligand complex only (also referred to as the single-trajectory approach), and (ii) a double-decoupling alchemical free energy method that uses thermodynamic integration (TI)[62] or the multistate Bennett acceptance ratio (MBAR)[63] frameworks for free energy estimation, following Alibay et al.[5] and Ries et al.[52] BindFlow uses GROMACS as the MD engine. BindFlow is highly user-friendly and provides extensive tutorials and documentation. It natively supports the small-molecule force fields GAFF, OpenFF, and Espaloma, as well as any GROMACS-compatible force field. At present, it schedules tasks on either a local desktop or SLURM-based distributed computing environments while also providing scope for incorporating additional HPC platforms in the future. By scheduling its tasks with Snakemake,[64] BindFlow efficiently uses the available hardware and is resilient against rare hardware failures or simulation instabilities. For advanced users, this allows full control over the pipeline and MD settings and provides various options for further customization. BindFlow has been forked from *ABFE_workflow*[52] but resolves its technical restrictions and introduces diverse new functionalities and numerous options for customization.

https://doi.org/10.1021/acs.jctc.5c02026
J. Chem. Theory Comput. 2026, 22, 1198−1213

Below, we first present implementation principles and Application Programming Interface (API) concepts of BindFlow. Next, we validate BindFlow by computing ABFEs for 139 diverse ligands binding to eight different receptors, including six soluble proteins, one membrane protein, and one host−guest system (Figure 1b). The calculations include both FEP and MM(PB/GB)SA methods using three widely used open-source small-molecule force fields: GAFF-2.11,[65] OpenFF-2.0.0,[66] and Espaloma-0.3.1,[67] referred to as GAFF, OpenFF, and Espaloma in the following. Statistical uncertainties are rigorously assessed by running at least three independent replicates for each free energy calculation workflow.

Overall, our results show that BindFlow delivers a predictive performance on par with established gold standards in the field. Strikingly, we observe that for certain targets and force fields, the computationally efficient MM(PB/GB)SA method achieves correlations with the experiment that approach those of the more rigorous FEP calculations—yet at only a fraction of the computational cost. By fully automating the binding free energy workflow and minimizing user intervention, BindFlow enhances not only efficiency but also reproducibility. In addition, BindFlow scales efficiently across HPC environments, enabling large simulation campaigns to be executed with high throughput and resource utilization. Together, these features make BindFlow a practical, accessible, and reliable platform for large-scale binding affinity prediction with strong potential for accelerating modern drug discovery campaigns.

THEORY

The theories of FEP and MM(PB/GB)SA have been reviewed frequently;[9,10,12,63,68−70] hence, we provide only a brief summary of the underlying concepts.

MM(PB/GB)SA

Following the thermodynamic cycle in Figure 2a, MM(PB/GB)SA estimates the binding free energy ΔG_{bind} as follows[9,10,69,70]

$$\begin{aligned}\Delta G_{\mathrm{bind}} &= -(\Delta G_{\mathrm{S}}^{\mathrm{R}} + \Delta G_{\mathrm{S}}^{\mathrm{L}}) + \Delta G_{\mathrm{bind_v}} + \Delta G_{\mathrm{S}}^{\mathrm{RL}} \\ \Delta G_{\mathrm{bind_v}} &= G_{\mathrm{v}}^{\mathrm{RL}} - G_{\mathrm{v}}^{\mathrm{R}} - G_{\mathrm{v}}^{\mathrm{L}}\end{aligned} \tag{1}$$

Here, $\Delta G_{\mathrm{S}}^{\mathrm{X}}$ denotes the solvation free energy for species X (X = R, L, or RL), where superscripts R, L, and RL denote the receptor, ligand, and receptor−ligand complex, respectively. The solvation free energy for species X is decomposed into polar and nonpolar terms, $\Delta G_{\mathrm{S}}^{\mathrm{X}} = \Delta G_{\mathrm{PB/GB}}^{\mathrm{X}} + \Delta G_{\mathrm{SA}}^{\mathrm{X}}$. $\Delta G_{\mathrm{PB/GB}}^{\mathrm{X}}$ is obtained by solving the Poisson−Boltzmann (PB) equation or using the generalized Born (GB) model,[10] while $\Delta G_{\mathrm{SA}}^{\mathrm{X}}$ is derived from the solvent-accessible surface area (SASA).[71]

$\Delta G_{\mathrm{bind_v}}$ denotes the binding free energy in vacuum, computed from the free energies $G_{\mathrm{v}}^{\mathrm{X}}$ of the receptor, ligand, or receptor−ligand complex. The free energies are estimated via

$$G_{\mathrm{v}}^{\mathrm{X}} = E_{\mathrm{MM}}^{\mathrm{X}} - TS^{\mathrm{X}} \tag{2}$$

where $E_{\mathrm{MM}}^{\mathrm{X}}$ is the molecular mechanics (MM) potential energy of species X comprising bonded and nonbonded terms, as defined by the MM force field. T is the temperature and S^{X} is the entropy of species X. By defining the free energy of species X as

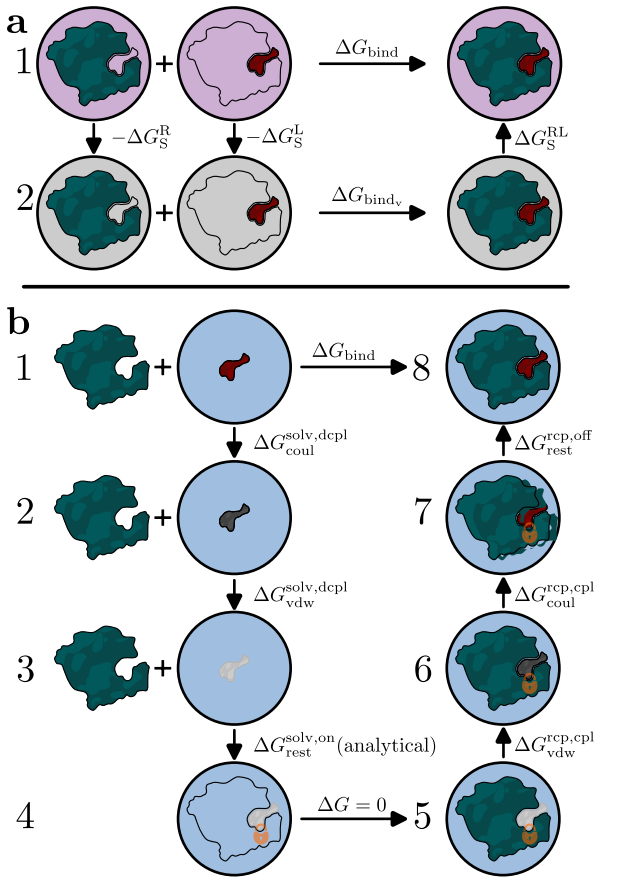


Figure 2. Thermodynamic cycles for (a) MM(PB/GB)SA and (b) FEP. $\Delta G_{\mathrm{S}}^{\mathrm{X}}$ denotes the solvation free energy for species X, where superscripts R, L, and RL denote the receptor, ligand, and receptor−ligand complex, respectively. $\Delta G_{\mathrm{bind_v}}$ is the binding free energy in vacuum. Superscripts "solv" and "rcp" indicate states with the ligand in solvent or in the receptor, respectively. Superscripts "dcpl" and "cpl" indicate decoupling or coupling processes, while "coul" and "vdw" specify transitions of Coulomb or Lennard-Jones interactions, respectively. Superscripts "on/off" specify the activation or deactivation of Boresch restraints (subscript "rest"). The light pink, gray, and light blue backgrounds represent the implicit solvation model, vacuum, and explicit water model, respectively.

$$G^{\mathrm{X}} = E_{\mathrm{MM}}^{\mathrm{X}} + \Delta G_{\mathrm{PB/GB}}^{\mathrm{X}} + \Delta G_{\mathrm{SA}}^{\mathrm{X}} - TS^{\mathrm{X}} \tag{3}$$

ΔG_{bind} may be written as the free energy difference between products and reactants

$$\Delta G_{\mathrm{bind}} = G^{\mathrm{RL}} - G^{\mathrm{R}} - G^{\mathrm{L}}$$

thereby rationalizing by the subscripts of eq 3 the name MM(PB/GB)SA.[9,10,69,70]

BindFlow computes the energy contributions to ΔG_{bind} by averaging over multiple MD simulation frames using the package gmx_MMPBSA.[69] The entropic term may be estimated using normal-mode analysis, interaction entropy (IE),[72] or a cumulant approximation to the second order of the exponential average (C2).[73−75] These entropy estimations have been implemented by gmx_MMPBSA[69] and may therefore be used by BindFlow.

BindFlow employs a single-trajectory protocol; that is, all contributions from the ligand, receptor, or receptor−ligand complex are computed from MD simulations of the complex. This approach reduces the computational cost and facilitates

https://doi.org/10.1021/acs.jctc.5c02026
J. Chem. Theory Comput. 2026, 22, 1198−1213

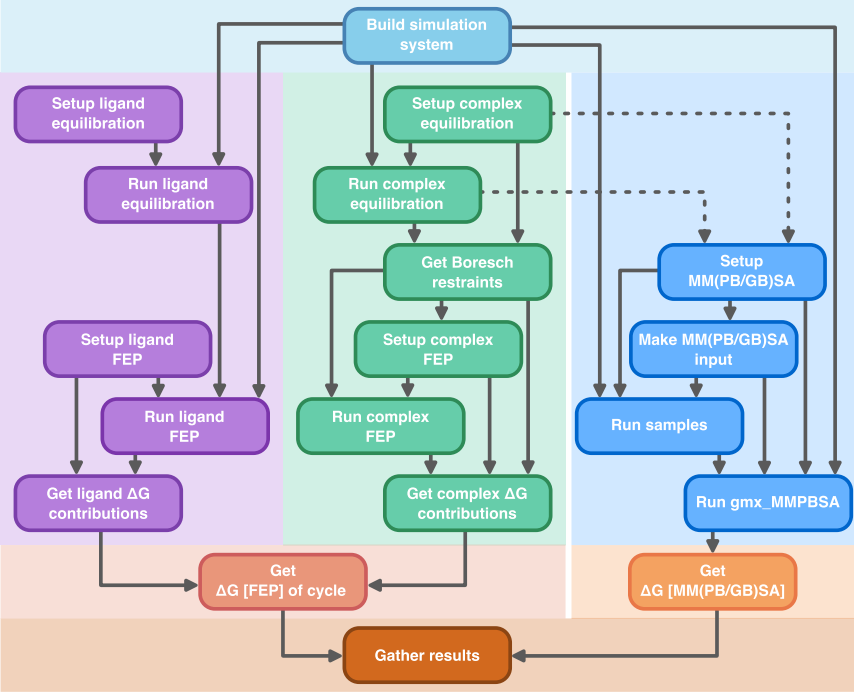


Figure 3. Task dependency graph for BindFlow. Each rectangle represents a task (or group of tasks), while arrows indicate dependencies between them. Some tasks are shared between methodologies. Purple boxes correspond to ligand-in-solvent FEP simulations, and green boxes correspond to receptor−ligand complex FEP simulations. The darkest blue boxes denote MM(PB/GB)SA simulations, which reuse some tasks from the receptor−ligand complex FEP workflow. All three approaches share the initial system building step (light blue) and converge at the final result gathering stage to report the binding free energy, ΔG_{bind}.

the cancellation of bonded terms in $E^{\mathrm{X}}_{\mathrm{MM}}$. However, the approach neglects contributions from structural changes of the ligand or receptor upon binding.[10,69]

FEP

Alchemical FEP estimates binding free energies by constructing a thermodynamic cycle. Such cycles involve decoupling the ligand from the solvent (Figure 2b, Steps 1−3) and coupling it back in the binding site (Figure 2b, Steps 5−7) in the presence of restraints. The restraints maintain the ligand in the binding pocket, thereby improving sampling and unambiguously defining the reference standard state. BindFlow employs Boresch restraints.[39] The free energy for activating Boresch restraints in the solvent is calculated analytically (Figure 2b, Steps 3−4) and the free energy for removing the restraints in the receptor environment is derived numerically (Figure 2b, Steps 7−8).

Free energy differences ΔG between two states (e.g., between states 2 and 1 in Figure 2b) are estimated from simulations across a series of intermediate alchemical states, specified by a parameter λ. BindFlow calculates ΔG values using either TI[68] or the MBAR[63] method. TI computes the ΔG values by integrating the mean force along the λ parameter[68]

$$\Delta G_{1\to 2} = \int_{\lambda_1}^{\lambda_2} \left\langle \frac{\partial U(\lambda)}{\partial \lambda} \right\rangle_{\lambda} \mathrm{d}\lambda \tag{4}$$

where $\langle \cdots \rangle_{\lambda}$ denotes the ensemble average at a given value λ and $U(\lambda)$ is the potential energy. In practice, the integral is solved numerically by using a finite number of λ points. MBAR estimates ΔG values using information from all of the alchemical states. MBAR requires solving the following set of equations self-consistently to obtain the reduced free energies f_i[63]

$$e^{-f_i} = \sum_{j=1}^{M} \sum_{n=1}^{N_j} \frac{e^{-u_i(\mathbf{x}_{j,n})}}{\sum_{k=1}^{M} N_k e^{-u_k(\mathbf{x}_{j,n})+f_k}} \tag{5}$$

Here, M is the number of λ-states, N_j is the number of samples from λ-state j, and $u_i(\mathbf{x}_{j,n})$ is the reduced potential energy of the nth sample $\mathbf{x}_{j,n}$ from the simulation of λ-state j evaluated with the energy function of λ-state i.

IMPLEMENTATION

General Concept of BindFlow

As illustrated in Figure 3, BindFlow streams the entire binding free energy pipeline, including building of the simulations systems, definition of force fields and MD parameters, careful multistep equilibration routines, launching of productions simulations, and analysis. For FEP calculations, specifically, production simulations include the definition of Boresch restraints, as well as the setup, equilibration, and launching of all λ-windows. The pipeline is organized into tasks with well-defined dependencies, which are deployed to the computing environment. The environment may be a desktop computer, a parallelized high-performance computer, or another distributed computing environment. Task scheduling is carried out by Snakemake,[64] a robust task manager that has been widely used for defining complex bioinformatics analyses but has to our knowledge been less used by the MD community. Snakemake constructs a direct acyclic graph that specifies the dependencies among BindFlow tasks to be executed asynchronously, thereby optimally using the available hardware. BindFlow uses GROMACS[76] as a molecular dynamics engine.

https://doi.org/10.1021/acs.jctc.5c02026
J. Chem. Theory Comput. 2026, 22, 1198−1213

BindFlow may compute ΔG_{bind} from a protein PDB file, together with a set of ligands provided as MOL files. However, for more complex systems including cofactors or membranes, additional input files may be provided.

The pipeline is implemented into a single function *calculate* within the module *runners*. The function *calculate* takes general parameters such as the protein and ligand structures and, through the keyword argument *global_config*, more specific definitions such as the computational environment or parameters for simulations analysis. Scheme 1 presents a minimalistic Python script for running an FEP pipeline in an HPC environment, while Listing S1 presents a more extended example for running an MM(PB/GB)SA pipeline on a desktop computer. Parameters passed with *global_config* may be specified within Python (Scheme 1) or, more conveniently, using JSON or YAML files (Listing S1). While BindFlow implements default workflows, the user has full control over equilibration and production settings, GROMACS parameters, and task deployment (Listing S7). Full documentation is provided online. Two classes for deployment are provided, namely, for a desktop computer and for the SLURM queuing system;[77] however, alternative deployment managers may be added by the user via the abstract base class *Scheduler*, for instance, for using cloud computing services. Thus, BindFlow is highly user-friendly yet allows extensive customization by the user.

Scheme 1. Minimal Python code for running an FEP campaign with BindFlow, as specified by *calculation_type*. The *protein* specifies the path to the protein PDB file, and the *ligands* specifies the list of ligand MOL files. The dictionary *global_config* specifies options for the computing environment.

```python
from bindflow.runners import calculate

ligands = [
    'path/to/ligand1.mol',
    'path/to/ligand2.mol'
    'path/to/ligand3.mol'
]

global_config = {
    'cluster': {
        'options': {
            'calculation': {
                'partition': 'deflt',
                'time': '2-00:00:00',
                'gpus': 1,
                'memb': '5G'
            }
        }
    }
}

calculate(
    calculation_type='fep',
    protein='path/to/protein.pdb',
    ligands=ligands,
    global_config=global_config)
```

BindFlow has been forked from ABFE_Workflow[52] yet has been largely rewritten with the following aims in mind: options for extensive customization, efficient resource utilization, built-in support for several small-molecule force fields, flexibility with respect to user-provided force fields, comprehensive documentation, implementation of MM(PB/GB)SA, handling of cofactors, and support for membrane proteins and nonprotein receptors. BindFlow is released as an open source under the GPL-3.0 license.

Input Files and Force Fields

For a typical workflow, the protein structure is provided as a PDB file and a set of ligands as MOL files, in which the ligand coordinates are aligned with the protein binding pocket. Optionally, cofactors are provided as MOL files. Running membrane protein systems requires additional care, as described in the online documentation. To enable the use of custom force fields or to simulate complex receptors, the user may provide GROMACS topology (TOP) and structure files (GRO), allowing BindFlow to operate with any GROMACS-compatible force field or simulation system. Examples for system component definitions are given in Listings S2, S3, S4, S5, and S6. More technical details are provided in the online documentation.

To achieve reliable ABFE results, accurate initial structural models are crucial with respect to the correct receptor−ligand arrangement and correct tautomeric, isomeric, and protonation states.[78] While BindFlow does not aim to solve these upstream challenges, it provides basic tools for structural cleanup, for instance, for resolving missing atoms or correcting atom names, via the pdbfixer[79] and pdb2gmx[76] software tools.

By default, proteins are described with Amber99sb-ildn,[80] membranes with SLipids2020,[81] cofactors and ligands with OpenFF-2.0.0,[66] and water molecules with the TIP3P model.[82] Other protein force fields may be selected from the GROMACS distribution or from a user-specified path. Alternative water models can be chosen. The small-molecule force fields OpenFF,[66] GAFF,[65] and Espaloma[67] are natively supported by BindFlow via the TOFF[83] package. Because the choice of the force field may influence the accuracy of ABFE calculations, users are encouraged to consult the literature when selecting a force field for their system. For example, Hahn et al.[84] recently reviewed small-molecule force fields in the context of FEP calculations.

RESULTS

Validation Set Involving 139 Receptor−Ligand Pairs and Three Small-Molecule Force Fields

Comprehensive benchmarks are computationally costly, which explains their limited presence in academic ABFE workflows. In developing BindFlow, however, we placed particular emphasis on rigorously validating its FEP and MM(PB/GB)SA methodologies. To this end, we computed the binding affinity of 139 receptor−ligand pairs for which high-quality experimental data are available (Figure 1b). The receptors included six soluble proteins (cyclophilin D, MCL1, thrombin, TYK2, PTP1B, P38), the transmembrane GPCR protein A2A, and the nonprotein host−guest system SAMPL6-OA that has been used for a binding affinity prediction challenge.[86] Affinity data for P38, PTP1B, TYK2, thrombin, and MCL1 have been widely utilized for validating RBFE[15−18] and ABFE calculations,[17,47,87] allowing us to compare BindFlow results with the literature (see below). These systems involve various

https://doi.org/10.1021/acs.jctc.5c02026
J. Chem. Theory Comput. 2026, 22, 1198−1213

Table 1. Simulation Systems Used for Validating BindFlow

System	Number of Atoms	Force Field		Number of Ligands	Cofactor (s)	Dynamic Range [kcal mol^{-1}]	References
		Protein	Ligands				
P38	86,376	Amber14sb	Espaloma-0.3.1 GAFF-2.11 OpenFF-2.0.0	29	3H_2O	3.80	Hahn et al.[30]
A2A	84,996	Amber14sb	Espaloma-0.3.1 GAFF-2.11 OpenFF-2.0.0	10	Na^+	2.69	Deflorian et al.[13]
PTP1B	74,246	Amber14sb	Espaloma-0.3.1 GAFF-2.11 OpenFF-2.0.0	22	4H_2O	5.17	Hahn et al.[30]
TYK2	66,425	Amber14sb	Espaloma-0.3.1 GAFF-2.11 OpenFF-2.0.0	13	2H_2O	3.47	Hahn et al.[30]
Thrombin	49,471	Amber14sb	Espaloma-0.3.1 GAFF-2.11 OpenFF-2.0.0	23	3H_2O	5.87	Hahn et al.[30]
MCL1	34,829	Amber14sb	Espaloma-0.3.1 GAFF-2.11 OpenFF-2.0.0	25	none	4.19	Hahn et al.[30]
Cyclophilin D	31,861	Amber99sb-ildn	Espaloma-0.3.1 GAFF-2.11 OpenFF-2.0.0	10	none	8.49	Alibay et al.,[5] Ries et al.[52]
SAMPL6-OA	9,204	Espaloma-0.3.1	Espaloma-0.3.1	7	none	3.79	Rizzi et al.,[85] Isik et al.[86]

challenges for binding affinity calculations. (i) Some ligands binding to PTP1B, thrombin, MCL1, or SAMPL6-OA are charged (Figures S18, S20, S21, S23). (ii) Ligands binding to PTP1B or P38 may proceed via slow induced-fit conformational transitions, suggesting that simulations may not sufficiency sample the apo state and, thereby, partly miss the free energy cost for the induced-fit transition.[47] (iii) Binding affinities may be influenced by cofactors: for P38, PTP1B, TYK2, and thrombin, crystallographic water molecules may interact with the ligand (Table 1, sixth column). In A2A, a sodium ion interacts with the ligand, suggesting that imprecise modeling of a salt bridge by the force field may add considerable uncertainty. (iv) Among the ten ligands binding to cyclophilin D (Figure S22), four comprise more than 30 heavy atoms, which could potentially increase sampling challenges, while (v) eight ligands contain exotic ring structures, posing challenges for the accurate force field representation.[5,52,53] Thus, our validation sets involve highly diverse receptor−ligand pairs that cover various common challenges during binding affinity calculations. All ligands are shown in Figures S16−S23.

We computed binding affinities for these 139 receptor−ligand pairs in triplicate using FEP or MM(PB/GB)SA, except for the cyclophilin D set, where five FEP replicates were used instead. For all protein targets, we computed binding affinities using three popular small-molecule force fields: GAFF-2.11,[65] OpenFF-2.0.0,[66] and Espaloma-0.3.1.[67] For the SAMPL6-OA set, only Espaloma-0.3.1 was used.[67] Cyclophilin D was described with the Amber99sb-ildn force field to allow comparison with previous studies.[5,52,53] Among our MM(PB/GB)SA calculations, we compared results from the Poisson−Boltzmann model with results from the generalized Born model and, for each solvation model, evaluated the effects from using either no entropy contribution or using the IE or C2 entropy contribution. Together, our validation sets comprise 1,269 FEP and 7,254 MM(PB/GB)SA calculations.

We quantified the agreement between calculated ΔG_{calc} and experimental ΔG_{exp} affinities using the Pearson ρ, Kendall τ, and Spearman r_S correlation coefficients. Here, Pearson ρ quantifies the linear correlation between calculated and experimental values, accounting for magnitudes rather than just ranks, although it remains sensitive to outliers. In contrast, Kendall τ and Spearman r_S quantify agreement in ranking and are far less sensitive to outliers. Whereas Spearman r_S quantifies whether the ligand ranking according to ΔG_{exp} agrees with the ligand ranking according to ΔG_{calc}, Kendall τ quantifies the concordance between pairs of ligands. Upon comparison of correlation coefficients from different receptors, it is critical to note that Pearson ρ is highly sensitive to the dynamic range among the ligands, whereas Kendall τ and Spearman r_S are less sensitive to the dynamic range (Table 1, seventh column). In this study, we strongly rely on Kendall τ as a key quality measure for ΔG_{bind} calculations since it has been shown to be a robust estimator,[88] while reporting Pearson ρ and Spearman r_S as complementary measures. In addition to the correlation coefficients, we computed the RMSE, mean signed error (MSE), and mean unsigned error (MUE) of ΔG_{calc} relative to ΔG_{exp} (see the Supporting Information methods). To focus on the global correlation between ΔG_{calc} and ΔG_{exp} across all receptor−ligand pairs, we furthermore subtracted from each set its corresponding MSE from ΔG_{calc} to yield offset-corrected ΔG^{oc}_{calc} values and, respectively, offset-corrected RMSE (ocRMSE, see the Methods section). Thereby, ocRMSE ignores systematic offsets between ΔG_{calc} and ΔG_{exp}, for instance, owing to insufficient sampling of the apo state or owing to a systematic bias in receptor−ligand interactions as modeled by the force field. ocRMSE values may be compared with results from RBFE calculations since the

https://doi.org/10.1021/acs.jctc.5c02026
J. Chem. Theory Comput. 2026, 22, 1198−1213

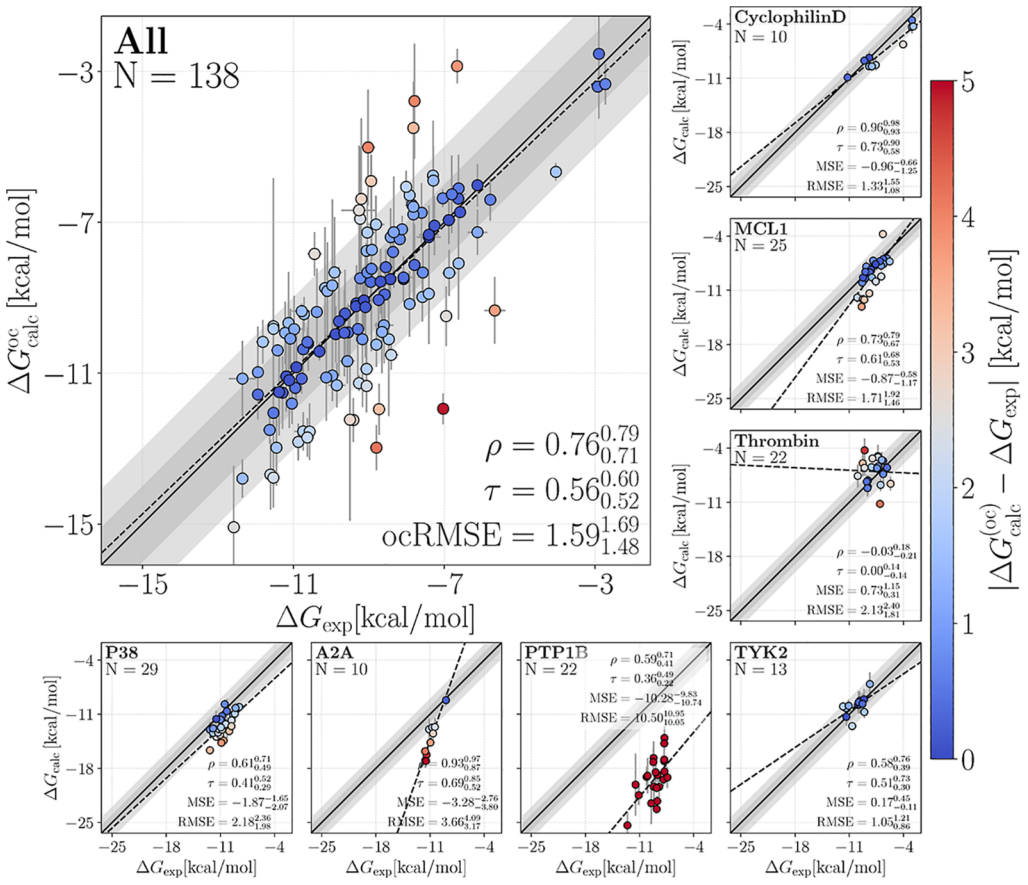


Figure 4. Calculated affinities ΔG_{calc} (or offset-corrected calculated affinities, ΔG^{oc}_{calc}) versus experimental affinities ΔG_{exp} from FEP with GAFF-2.11. Results show ΔG_{calc} for individual set (small panels, see labels for protein name and number of ligands N) and ΔG^{oc}_{calc} collected from all sets (large panel). Insets show Pearson ρ, Kendall τ, MSE, and RMSE (or ocRMSE)—the last two in kcal mol^{-1}—for each data set with their corresponding 68% confident interval. Colors of dots indicate the absolute deviation between ΔG_{calc} (or ΔG^{oc}_{calc}) and ΔG_{exp} (color bar). Dark and light gray diagonal regions indicate 1 or 2 kcal mol^{-1} deviations, respectively. Dashed lines are linear fits shown to guide the eye. A single outlier (lig_4, thrombin) has been removed. Error bars show uncertainties obtained via three independent replicates. Figure style has been inspired by ref 87.

latter are blind to such systematic offsets. Uncertainties of the statistical measures were derived by bootstrapping among the ligands for each set. Thus, critically, the errors bars for statistical measures are not only caused by limited sampling but furthermore stem from the limited number of 7 to 29 ligands per receptor.

Comparison of FEP Results with Experiments

Figure 4 presents correlation plots between calculated affinities ΔG_{calc} obtained with GAFF and experimental affinities ΔG_{exp}. Values are listed either for each of the seven protein targets individually (small panels) or for all seven targets combined (large panel). The aggregated plot displays the data after applying set-specific offset corrections. Insets report Pearson ρ, Kendall τ, MSE, and RMSE for each data set, as well as ocRMSE for the aggregated data. Correlation plots obtained with OpenFF or Espaloma are shown in Figures S2 and S3. Notably, we removed a single extreme outlier given by the binding of ligand lig_4 to thrombin (Figure S20) from the statistical quality measures to avoid a bias from a single receptor−ligand pair (see Figure 5). Lig_4 comprises a cationic amidinium moiety (−C−$(NH_2)_2^+$) forming a salt bridge to Asp-189 of thrombin (Figure S24). We speculate that the greatly overestimated affinity of lig_4 may be a consequence of an overly stable amidinium−carboxylate bridge modeled by the three force fields.

For most receptors and force fields, the affinities were overestimated, as shown by mostly negative MSE values between typically −4 and 0 kcal mol^{-1} (small panels of Figures 4, S2, and S3; see also Figure 5d and Table S1). Such negative MSE values have been rationalized by insufficient sampling of the apo state, thereby partly missing the free energy cost of the induced-fit conformations transition upon ligand binding.[47] An extreme case was PTP1B with MSE values between −14 and −10 kcal mol^{-1}, pointing to a major free energy cost for the induced-fit conformational transition that was not captured during FEP simulation, in line with the value of approximately −10 kcal mol^{-1} reported by Chen et al.[47]

Statistical quality measures for the agreement with experiments greatly depended on the protein target and the force field, with RMSE values far below 2 kcal mol^{-1} for cyclophilin D with GAFF up to approximately 14 kcal mol^{-1} for PTP1B with Espaloma and OpenFF These variations were likewise reflected by Kendall τ, which spanned 0.82 for cyclophilin D with Espaloma down to poor τ values for thrombin. Pearson ρ and MSE values reported similar trends. The wide range of correlations obtained with different receptors and different force fields are summarized by Figure 5 and Table S1.

https://doi.org/10.1021/acs.jctc.5c02026
J. Chem. Theory Comput. 2026, 22, 1198−1213

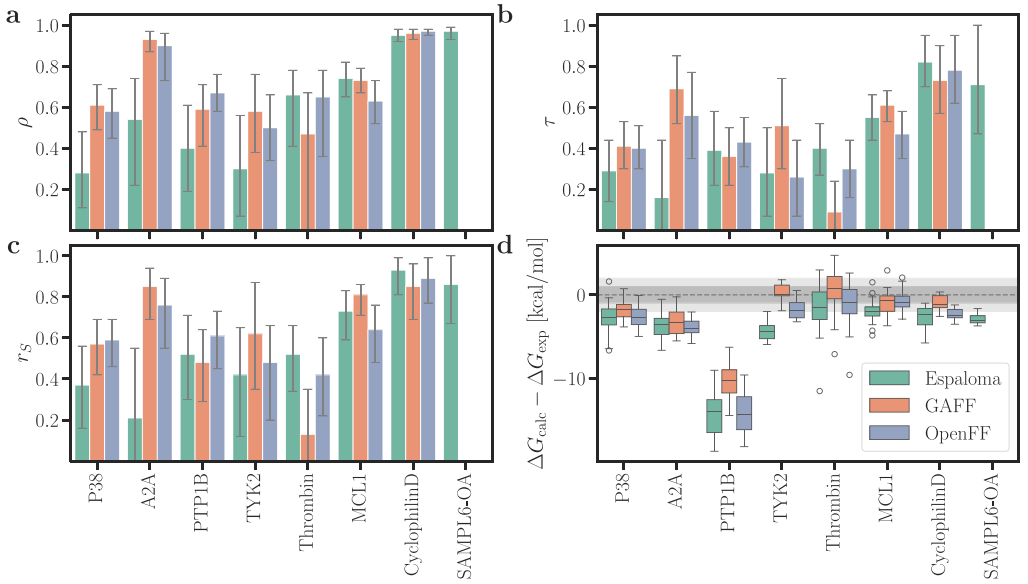


Figure 5. (a) Pearson ρ, (b) Kendall τ, (c) Spearman r_S, and (d) deviations between calculated and experimental binding free energies $\Delta G_{calc} - \Delta G_{exp}$ from FEP using Espaloma-0.3.1 (green), GAFF-2.11 (orange), and OpenFF-2.0.0 (blue). Results are shown for seven different data sets (labels at abscissa). Error bars represent the 68% confident interval obtained from bootstrapping among the ligands from each set. For $\Delta G_{calc} - \Delta G_{exp}$, dark and light gray horizontal regions indicate 1 or 2 kcal mol^{-1} deviations, respectively. Box plots present the median (50th percentile) as a line within the box, and the lower and upper box edges correspond to the first (25th percentile) and third (75th percentile) quartiles, respectively. Whiskers extend to the smallest and largest data points within the 1.5 × interquartile range (IQR = Q3 − Q1), while outliers beyond this range are shown as circles.

Each reported binding free energy corresponds to the average across three independent replicas (five for cyclophilin D), with uncertainties represented by the standard error of the mean (SEM). For systems such as MCL1 and P38, SEM values are small—typically in the range 0.5 to 1 kcal mol^{-1}, as visible from the small error bars in the small panels of Figure 4. In contrast, PTP1B shows markedly larger SEMs, reflecting convergence challenges likely caused by occasional sampling of apo-like conformations. This highlights both the difficulty of capturing the induced-fit transition in PTP1B within typical simulation times and the utility of replica-based SEMs for revealing such uncertainties.

Comparison of FEP Results from BindFlow with the Literature

Ross et al.[89] evaluated the reproducibility of experimental RBFE values and reported an overall RMSE (weighted by number of ligands on each set) of 0.91 kcal mol^{-1} (95% CI: [0.83, 1.11]) and a weighted average Kendall τ = 0.71 (95% CI: [0.65, 0.74]). Thus, upon comparison of ΔG_{calc} with ΔG_{exp} in this study, the considerable intrinsic uncertainty of ΔG_{exp} must be kept in mind. Accordingly, an RMSE of ∼1 kcal mol^{-1} and a Kendall τ of ∼0.7 between ΔG_{calc} and ΔG_{exp} would indicate excellent agreement.

In addition, Ross et al.[89] validated FEP+ software for RBFE calculations in combination with the OPLS4 force field[90] against protein−ligand sets different from the ones used here. Beyond the differences in validation sets, their study employed RBFEs, whereas we performed ABFEs, which are generally more challenging. For these reasons, a direct, quantitative comparison of the RMSE and Kendall τ is not meaningful. Nevertheless, placing the results side by side helps contextualize the current state of the art in FEP calculations. FEP+ achieved an RMSE of 1.25 (95% CI: [1.17, 1.33]) and a Kendall τ of 0.51 (95% CI: [0.48, 0.55]). With Espaloma, GAFF, and OpenFF, we obtained slightly higher ocRMSE values, which may indicate that the closed-source OPLS4 force field[90] performs somewhat better than the currently freely available force fields. However, the Kendall τ values obtained here were even slightly higher than those from FEP+, suggesting that our ABFE calculations are competitive in terms of ligand ranking.

Next, we systematically compared ABFE results from BindFlow with FEP+ by restricting the analysis to ligands that were present in both the data set of Chen et al.[47] and our BindFlow simulations, resulting in 89 ligands across the P38, PTP1B, TYK2, and MCL1 sets. Both BindFlow and FEP+ systematically overestimated affinities for PTP1B (Figure S7, top row). Thus, to enable a quantitative comparison of ABFE values, we removed PTP1B from the data sets and focused the comparison on the remaining 67 ligands. FEP+ yielded systematically stronger binding affinities on these three sets compared to BindFlow, as quantified by the MSE of $-4.29^{-4.16}_{-4.42}$ kcal mol^{-1} for FEP+, versus $-2.74^{-2.53}_{-2.95}$, $-1.10^{-0.93}_{-1.28}$, and $-1.79^{-1.62}_{-1.98}$ kcal mol^{-1} for Espaloma, GAFF, and OpenFF, respectively (Figure S7, middle row). Keeping the PTP1B set and removing the constant offset for each set by subtracting the set-specific MSE from ΔG_{calc}, FEP+ yielded considerably lower ocRMSE values compared to BindFlow when used with any of the three force fields. FEP+ also yielded higher correlation coefficients (Figure S7, bottom row). Thus, in light of these data, FEP+ may achieve superior ranking at the cost of overestimating the absolute binding affinity. We speculate that the superior ranking by FEP+ is primarily a consequence of using the OPLS4 force field, although alternative sampling algorithms may also play a role. Within this subset of ligands, GAFF achieved better agreement than Espaloma and OpenFF with experimental data, with ocRMSE = $1.52^{1.64}_{1.40}$ kcal mol^{-1} and $\tau = 0.53^{0.58}_{0.48}$ (Figure S7, bottom row).

https://doi.org/10.1021/acs.jctc.5c02026
J. Chem. Theory Comput. 2026, 22, 1198−1213

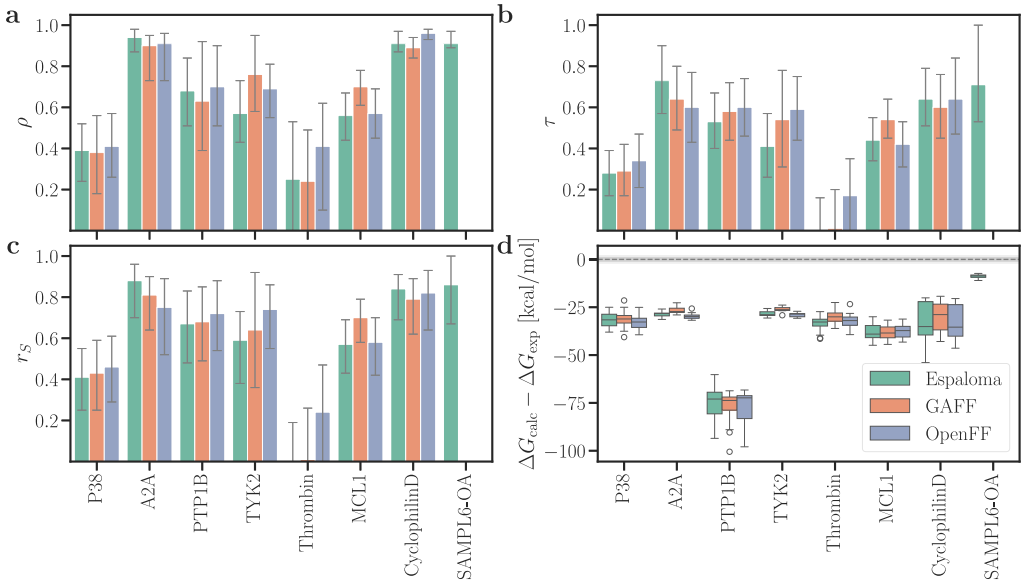


Figure 6. (a) Pearson ρ, (b) Kendall τ, (c) Spearman r_S, and (d) deviations between calculated and experimental binding free energies $\Delta G_{calc} - \Delta G_{exp}$ from MMGBSA results without entropy contribution. Presentation style according to Figure 5.

Deflorian et al.[13] reported RBFE calculations for the A2A system using FEP+ software. We converted their RBFE values into ABFEs by referencing the experimental binding affinity of ligand 4k and correcting for the MSE. From their data, we obtained ocRMSE = $0.65^{0.79}_{0.47}$ kcal mol^{-1} and τ = $0.78^{0.90}_{0.65}$. Evidently, since all ligands were similarly composed of three aromatic rings (Figure S17), RBFE required relatively small perturbations, rationalizing the high τ and exceptionally low ocRMSE. Although our ABFE calculations involved by far larger perturbations, we obtained decent agreement (Figure S10) with GAFF (ocRMSE: $1.64^{1.87}_{1.35}$ kcal mol^{-1}, τ = $0.69^{0.85}_{0.52}$, Figure 4) and with OpenFF (ocRMSE: $1.02^{1.19}_{0.80}$ kcal mol^{-1}, τ = $0.56^{0.76}_{0.33}$, Figure S2). Only with Espaloma, we obtained a rather poor τ of only $0.16^{0.44}_{-0.12}$ (Figure S3), which might indicate inaccurate modeling of the ligand interactions with the Na^+ cofactor inside the A2A binding pocket.

The cyclophilin D set was first introduced by Alibay et al.[5] for ABFE benchmarking and was recently employed to validate ABFE_workflow[52] and A3FE.[53] Excellent agreement with experiments was found across the three pipelines—ABFE_workflow, A3FE, and BindFlow—underscoring that the cyclophilin D set was the least challenging receptor−ligand set considered in this study (Figures S8 and 5).

The P38 and TYK2 systems have been used by several authors for validating binding free energy calculations using different methodologies.[17,47,49,51,87] Figure S9 compares results from BindFlow with previous studies.[17,47,49,51,87] Here, Lin et al.[17] and Chen et al.[47] achieved the best ranking of ligands, though at the cost of overestimating the absolute binding free energies, as shown by large negative MSE. Among our results, GAFF achieved comparable ranking at a less negative MSE.

Taken together, the reasonable agreement with previous data validates BindFlow's FEP pipeline and demonstrates competitive agreement with experimental data for a wide range of ligands and receptors.

Comparison of MM(PB/GB)SA with Experiments

MM(PB/GB)SA provides a computationally efficient alternative to FEP. Specifically, the MM(PB/GB)SA pipeline used here employed only ∼3 ns of MD simulations, corresponding to a 75-fold lower computing cost compared to the FEP pipeline used here. However, whereas the MMGBSA calculation from the MD frames is highly efficient, the MMPBSA calculation may take considerable additional computing time owing to the cost of Poisson−Boltzmann calculations. As expected from previous studies,[10] absolute binding free energies obtained with MMGBSA (Figures 6d, S11d and S12d) or MMPBSA (Figures S13d, S14d, and S15d) revealed poor agreement with experimental values. For instance, MMGBSA without entropy contribution overestimated the absolute binding affinities to protein targets by approximately 25 to 75 kcal mol^{-1} (Figure 6d). Even upon correcting the binding affinity with a target-specific offset, ocRMSE values remained large (Figures S4−S6).

However, MMGBSA without entropy contribution achieved overall excellent Pearson, Kendall, and Spearman correlation coefficients for A2A, cyclophilin D, and SAMPL6-OA while still reaching good correlation coefficients for PTP1B, TYK2, and MCL1 (Figures S4−S6, and 6a−c). In contrast, correlation coefficients for P38 or thrombin were rather poor. The three force fields GAFF, OpenFF, and Espaloma achieved similar ranking. Thus, while MMGBSA is not suitable for obtaining absolute binding free energies (Figure 6d), it may provide for certain receptors reasonable ranking among ligands and, thereby, serve as a useful tool for screening large data sets of potential binders.

Using the Poisson−Boltzmann instead of the generalized Born solvation model had only moderate effects on the correlation coefficients (compare Figures S13 with 6), suggesting that the computationally more efficient generalized Born model yields a good starting point for computational studies. Using the IE entropy contribution had only a small effect on the correlation coefficients (compare Figure S15 with S12), whereas the C2 entropy contribution decreased the correlation coefficients (compare Figure S14 with S11). Hence, in line with previous findings,[91] entropy contributions should be used with care or tested for the receptor of interest because entropy contributions may deteriorate the ranking by MM(PB/GB)SA.

https://doi.org/10.1021/acs.jctc.5c02026
J. Chem. Theory Comput. 2026, 22, 1198−1213

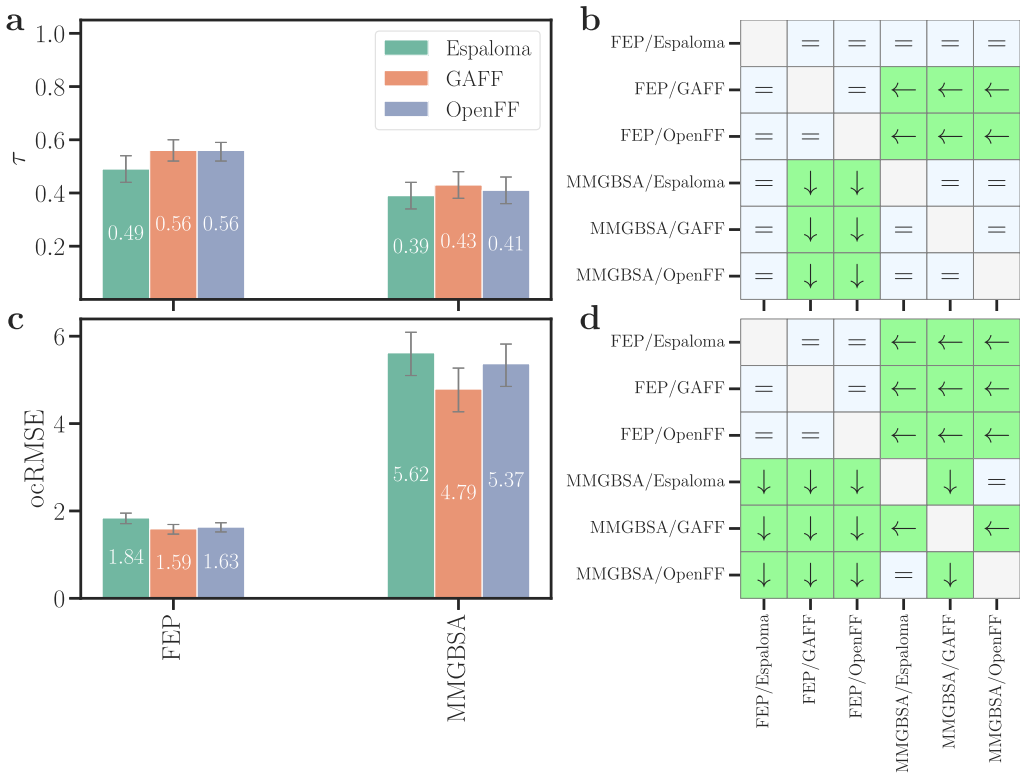


Figure 7. Summary of agreement between FEP or MMGBSA calculations and experimental data quantified by (a/b) Kendall τ and (b/c) offset-corrected root-mean-squared error (ocRMSE) in kcal mol^{-1}, summarized from all receptors. Error bars show 68% confident intervals obtained from bootstrapping among the ligands from all sets. Pairwise significance difference matrices are for (b) τ and (d) ocRMSE. An arrow indicates a statistically significant difference pointing to the combination of method/force field with the better statistical metric. Equal symbols indicate an insignificant difference. The set SAMPL6-OA is excluded to ensure consistence in the analysis across force fields. The extreme outlier lig_4 from the thrombin set was excluded.

Ligand Ranking by FEP versus MMGBSA and by Different Force Fields

Figure 7 summarizes the overall performance of FEP versus MMGBSA among the three force fields, as collected from all receptor−ligand pairs, thus including easy receptors such as cyclophilin D and challenging receptors such as thrombin (see also Table S1). Here, the matrices in Figure 7b/d visualize whether one combination of method/force field, such as FEP/Espaloma, outperforms another combination, such as MMGBSA/GAFF (see the Methods section for the applied significance test). The following conclusions are drawn from the analysis: (i) In terms of ocRMSE, FEP by far outperformed MMGBSA (Figure 7c/d). These findings suggest that errors caused by major approximations underlying MMGBSA remain after correction of ΔG_{calc} by the MSE. (ii) In terms of ranking by Kendall τ, FEP overall outperformed MMGBSA (Figure 7a/b). An exception was given by FEP/Espaloma that showed only insignificant differences relative to those of MMGBSA (Figure 7b). (iii) For a given method (FEP or MMGBSA), the three force fields showed no statistically significant differences (Figure 7b/d), except among the poor ocRMSE by MMGBSA. Thus, simulations with additional ligands will be required to test whether the slightly lower τ and slightly larger ocRMSE obtained with Espaloma are significant or whether they were caused by the limited set of receptor−ligand pairs considered in this study.

Computational Costs

BindFlow's completion time mostly depends on (i) the aggregated MD simulation time, as defined by the number of ligands and replicas, simulation times, and number of λ-windows for FEP or samples for MM(PB/GB)SA, together with (ii) the performance of the GROMACS MD engine, which depends on parameters such as hardware architecture, system size, or integration time step.

By modeling BindFlow execution within a queue-based scheduling environment similar to SLURM,[77,92] we estimated completion times for common multiligand campaigns. Details are presented in the Performance Section of the BindFlow online documentation. For example, for the thrombin system and the protocol described in this study, a cluster with 200 Nvidia RTX A4000 GPUs, each equipped with 10 CPU cores of Ryzen Threadripper PRO 3975WX, would complete 580 FEP or 39,965 MMGBSA calculations within 1 week. On average, MMGBSA completed in a 75-fold shorter time than FEP.

However, these values may vary considerably upon using alternative scheduling strategies, for instance, involving more λ-windows in combination with shorter simulation times per window, or by adhering to nonlinear instead of equidistant λ-window spacing.[93] In addition, we here used relatively large water compartments by keeping a distance of 1.5 nm between the receptor and simulation box surface, suggesting that tuning the number of water molecules may provide a route for optimizing the computational cost in future campaigns.

DISCUSSION

We introduced BindFlow, a free and user-friendly software program for ABFE calculations. BindFlow offers fully

https://doi.org/10.1021/acs.jctc.5c02026
J. Chem. Theory Comput. 2026, 22, 1198−1213

automated default pipelines for FEP and MM(PB/GB)SA, requiring minimal user intervention while providing extensive configuration options for advanced users. Comprehensive online tutorials and documentation support accessibility. The software is developed according to best coding practices, providing transparency, reproducibility, and a solid foundation for future community-driven development. By scheduling its tasks with Snakemake, BindFlow optimizes the use of distributed platforms, is resilient to rare failures of individual simulations, and enables seamless interruption and continuation of large-scale multiligand campaigns. It supports globular and membrane proteins as well as nonprotein host−guest systems, provided that a GROMACS-compatible topology can be generated. Thereby, BindFlow integrates each of the six suggested properties for ABFE solutions, as outlined in the Introduction section, into a single workflow (see also the Supporting Discussion).

We validated BindFlow by computing ΔG_{bind} values for 139 ligands binding to one of eight different receptors using FEP together with MBAR or using MM(PB/GB)SA. Overall, we found that BindFlow yields competitive ΔG_{bind} values compared to the standards in the field. As expected from previous studies, the accuracy of ΔG_{bind} values and ligand ranking greatly depend on the target and the chemical moieties of the specific ligand.

For the 139 receptor−ligand pairs, in terms of accuracy of ΔG_{bind} predictions quantified by ocRMSE, FEP by far outperformed MM(PB/GB)SA. In terms of ligand ranking, FEP still outperformed MMGBSA significantly; however, for several receptors studied here, ranking by MMGBSA was remarkably good. Only for P38 and thrombin did MMGBSA achieve poor ranking. Thus, our study underscores that MMGBSA may provide a computationally inexpensive method for prescreening large sets of ligands prior to validation by the more expensive FEP. However, future studies will need to establish whether favorable correlations between MM(PB/GB)SA and experiments, as observed here, also hold for noncongeneric sets of ligands.

Beyond the benchmark simulations presented here, BindFlow has recently been applied successfully for combining ∼60,000 MM(PB/GB)SA calculations with Bayesian active learning;[42] for interpreting X-ray crystallographic data from the antiviral drug tecovirimat binding to the phospholipase F13 from monkeypox virus;[44] and as a module to automate the system preparation phase of a typical MD simulation.[94] These use cases highlight that BindFlow is ready for production use.

BindFlow will be useful to systematically test and improve small-molecule force fields. For instance, we spotted that the affinity of lig_4 binding to thrombin was greatly overestimated by 7 to 13 kcal mol^{-1}, possibly caused by an overly strong amidinium−aspartate salt bridge (Figure S24). This interpretation is supported by the fact that larger partial charges used by Espaloma correlated to an even stronger computed affinity. Since amidinium moieties are quite common in drugs, these findings point toward options for further force field refinements.

Future implementations may further enhance the capabilities of BindFlow. (i) Enhanced sampling techniques, such as Hamiltonian replica exchange,[95−97] may accelerate convergence and, thereby, reduce the computational cost. Enhanced sampling may, specifically, improve the sampling of the apo state and lead to smaller MSE values. (ii) The adaptive allocation of computational resources may focus sampling to λ windows with increased sampling challenges.[53] (iii) Charged ligand systems (e.g., MCL1, thrombin, and PTP1B) showed satisfactory accuracy in our validations. However, finite-size effects in charge-imbalanced systems, primarily due to the use of particle-mesh Ewald method,[98,99] remain a concern.[100,101] Charge corrections[102,103] such as the coalchemical ion approach[12] represent viable solutions to address this issue. (iv) Water exchange between the binding site and bulk solvent represents a slow sampling process that is difficult to capture during the short MD simulations performed in FEP.[12] Monte Carlo-based water swap methods in a grand canonical ensemble,[104,105] methods based on inhomogeneous fluid solvation theory,[106] or specialized hydration-shell generators such as SOLVATE[78,107] have shown promise in treating the explicit water. (v) Additional flavors of end-point free energy methods, as alternative to the single-trajectory MM(PB/GB)SA approach used here, may be readily implemented, for instance, involving three-trajectory techniques, quantum corrections,[108] or central limit free energy perturbation.[109] Future versions of BindFlow may include such methodologies.

CONCLUSIONS

We presented BindFlow, an open-source and user-friendly pipeline for ABFE calculations that integrates both rigorous FEP and efficient end-point methods [MM(PB/GB)SA] into a single unified framework. Validation on 139 receptor−ligand pairs showed that BindFlow achieves predictive performance comparable to standards in the field while offering full automation and extensive customization. FEP achieved more accurate ΔG_{bind} predictions and ligand ranking than MM(PB/GB)SA. However, MM(PB/GB)SA provided remarkable ligand rankings for several systems at a fraction of the computational cost, highlighting its value for large-scale computational prescreening campaigns. Thus, we anticipated that the combination of MM(PB/GB)SA and FEP will be powerful for balancing efficiency and accuracy in future BindFlow applications. BindFlow scales efficiently across HPC environments, enabling high-throughput calculations with an excellent resource utilization. BindFlow is designed to serve both nonexpert and expert users: it provides a straightforward, plug-and-play interface with extensive online documentations and tutorials that allows users to perform ABFE calculations without detailed technical knowledge, while also offering advanced configuration options for expert users who wish to customize specific aspects of the workflow. By combining automation, flexibility, accessibility, documentation, and HPC-oriented scalability, BindFlow lowers the entry barrier for routine ABFE calculations and provides a platform for modern drug discovery or for systematically improving small-molecule force fields.

METHODS

Building Simulation Systems

BindFlow version 0.13.post24 was used. Simulations followed largely the current default settings of BindFlow. The setup protocols are fully automated yet may be easily adapted by the user via the *global_config* keyword, which is a Python dictionary. This dictionary may be conveniently defined with a YAML file (Listing S7). The simulations presented here were set up as follows.

The task *Build Simulation System* (Figure 3) assembles the simulation systems and defines force fields. Starting structures were taken from previous studies.[5,13,30,52,85,86] Soluble proteins, SAMPL6-OA, and ligands were placed in an octahedral simulation box, keeping

https://doi.org/10.1021/acs.jctc.5c02026
J. Chem. Theory Comput. 2026, 22, 1198−1213

a distance of 1.5 nm between the solute and box boundary. The systems were solvated by explicit water with the GROMACS solvate module and approximately 150 mM NaCl was added, thereby neutralizing the system. A minimum distance of 1 nm between ions and nonsolvent molecules was used. The structure of the membrane protein A2A was taken from ref 13. BindFlow does not build simulation systems of membrane proteins. Thus, the A2A system was built using CHARMM-GUI[110] by placing the protein in a simulation box of a hexagonal prism and by embedding it in a membrane of 172 1-palmitoyl-2-oleoyl-*sn*-glycero-3-phosphocholine (POPC) lipids.

Ligand interactions were described with Espaloma-0.3.1,[67] GAFF-2.11,[65] or OpenFF-2.0.0.[66] Ligand topologies were generated within BindFlow using the TOFF Python library.[83] Cyclophilin D was described with Amber99sb-ildn,[80] as used by previous studies.[5,52,53] All other proteins were described with Amber14sb,[111] using OpenMM[112] and ParmEd[113] for generating force field topologies. The host molecule octa acid of the SAMPL6-OA system was described with Espaloma-0.3.1,[67] and the topologies were generated with TOFF.[83] Water was modeled with the TIP3P model,[82] and ion parameters were taken from the definitions provided with Amber99sb-ildn.[80]

Simulation Parameters

The tasks *Setup Complex Equilibration*, *Setup Ligand Equilibration*, *Setup Complex FEP*, *Setup Ligand FEP*, and *Setup Complex MM(PB/GB)SA* (Figure 3) define MD parameters and set up directory structures for the initial equilibration, FEP, or MM(PB/GB)SA simulations.

All simulations were carried out with GROMACS, version 2022.4.[76] The geometry of water acting as a cofactor was constrained with LINCS,[114] while the geometry of all other water molecules was constrained with SETTLE.[115] All other bonds were constrained with LINCS[114] if not stated otherwise. Hydrogen mass repartitioning was used with a mass repartition factor of 2.5. During production simulations, a time step of 4 fs was used. Dispersive interactions and short-range repulsion were modeled by using a Lennard-Jones potential with a 1 nm cutoff. Electrostatic interactions were calculated with the particle-mesh Ewald[98,99] method, applying a real-space cutoff of 1 nm.

The temperature was maintained at 298.15 K. For the initial equilibration of membrane protein−ligand complexes, the temperature was maintained using velocity rescaling (τ_t = 1 ps). Here, the protein and Na^+ cofactor were coupled to the same thermostat. During all other simulations, the temperature was maintained by using Langevin dynamics (τ_t = 2 ps). The pressure was maintained at 1 bar for membrane protein−ligand complexes using semi-isotropic stochastic cell rescaling[116] with a time constant of τ_p = 5 ps during the initial equilibration. The same barostat was used during the perturbation phase for membrane protein−ligand complexes, but the pressure was maintained at 1 atm (1.01325 bar) and τ_p = 2 ps was used. For all other simulations, the pressure was controlled at 1 atm with the isotropic Berendsen barostat during equilibration (τ_t = 1 ps)[117] and the Parrinello−Rahman barostat during production (τ_p = 2 ps).[118]

Initial Multistep Equilibration Protocol

After minimizing the energy with the steepest-descent algorithm, each system was equilibrated with a multistep protocol (see Supporting Information methods).

FEP Calculations

Following Alibay et al.[5] and Ries et al.,[52] ΔG_{bind} from FEP was computed according the thermodynamic cycle shown in Figure 2b. Coulomb interactions of the ligand in solvent were decoupled over 11 λ points (Figure 2b, transition 1 → 2), followed by decoupling of Lennard-Jones interactions over 11 λ points (transition 2 → 3). Both, inter- and intramolecular interactions were decoupled. A 10 ns equilibrium simulation together with the software MDRestraintsGenerator[119] was used to obtain the optimal Boresch restraints,[39] the free energy cost for introducing restraints to the ligand in solvent (transition 3 → 4), and the initial frame for the complex decoupling simulations. Inter- and intramolecular Lennard-Jones interactions of the ligand in the receptor were activated over 21 λ points (transition 5 → 6), followed by 11 λ points for activating the Coulomb interactions (transition 6 → 7). Finally, Boresch restraints were removed for the ligand in the receptor over 11 λ points (transition 7 → 8).

For each λ window, the system was energy minimized and equilibrated with a three-step protocol (see Supporting Information methods). Each window was simulated for 10 ns using a stochastic dynamics integrator. During simulations that decouple Lennard-Jones interactions, a soft-core potential (α = 0.5, σ = 0.3, power = 1) was used. Free energy differences were computed with MBAR.[63] As a control, free energies were additionally computed using TI.[68] BindFlow reports a warning if results from MBAR and TI differ by more than 0.5 kcal mol^{-1} as such differences may indicate poor convergence or simulation instabilities. Alchemlyb-2.0.0 was used for MBAR and TI evaluations.[120]

The binding free energy was computed via (Figure 2B)

$$\Delta G_{bind} = \Delta G_{coul}^{solv,dcpl} + \Delta G_{vdw}^{solv,dcpl} + \Delta G_{rest}^{solv,on} + \Delta G_{vdw}^{rcp,cpl} + \Delta G_{coul}^{rcp,cpl} + \Delta G_{rest}^{rcp,off} \quad (6)$$

Here, $\Delta G_{coul}^{solv,dcpl}$ and $\Delta G_{vdw}^{solv,dcpl}$ denote free energies for decoupling (dcpl) Coulomb and Lennard-Jones interactions of the ligand in solvent (solv), respectively. $\Delta G_{rest}^{solv,on}$ denotes the free energy for turning on Boresch restraints (rest) for the ligand in solvent, which is computed analytically. Here, $\Delta G_{rest}^{solv,on}$ is defined such that the decoupled state of the ligand in solvent corresponds to the standard state with a ligand concentration of 1 mol/L.[39,119] $\Delta G_{vdw}^{rcp,cpl}$ and $\Delta G_{coul}^{rcp,cpl}$ denote the free energy for activating (coupling, cpl) Lennard-Jones and Coulomb interactions for the ligand in the receptor (rcp), respectively. $\Delta G_{rest}^{rcp,off}$ denotes the free energy cost for turning off the Boresch restraints for the ligand in the receptor.

BindFlow reports the statistical error obtained from MBAR[63] together with Gaussian error propagation. However, owing to autocorrelations, this value greatly underestimates the true uncertainty of ΔG_{bind}. Thus, we carried out the whole pipeline including setup, equilibration, and FEP in three (five for Cyclophilin D[5]) independent replicates and report the respective SEM as error bars in the correlation plots.

MM(PB/GB)SA Calculations

The initial multistep equilibration for MM(PB/GB)SA calculation was carried out as described in the Supporting Information methods. Following Su et al.,[91] MM(PB/GB)SA values were computed from multiple short simulations, rather than from a single long simulation. Accordingly, 20 starting frames were taken from a 950 ps equilibrium simulation, using one frame every 50 ps. From each frame, a 100 ps simulation was carried out by writing a frame every 5 ps. Here, we follow Genheden and Ryde,[10] who recommended using an output frequency of 1 to 10 ps. MM(PB/GB)SA values were computed with the gmx_MMPBSA software,[69] yielding 20 samples of the binding affinity. Cofactors were defined as a part of the receptor for the MM(PB/GB)SA calculation. For each complex, the pipeline was run in three independent replicates. Binding affinities reported here were computed by averaging over 60 samples from the three replicates with 20 samples per replica, and the error was derived as the SEM and reported as error bars in correlation plots. PB and GB models were used, each without an entropy contribution or using the IE or C2 entropy contribution. The main text reports results from the GB model without entropy contribution, while all other results are provided in the Supporting Information.

■ ASSOCIATED CONTENT

Data Availability Statement

BindFlow is free software published under the GPL-3.0 license. The code is currently hosted at https://github.com/ale94mleon/BindFlow. Documentation is available at https://bindflow.readthedocs.io. Scripts and input files re-

https://doi.org/10.1021/acs.jctc.5c02026
J. Chem. Theory Comput. 2026, 22, 1198−1213

quired for reproducing our results and analysis are accessible at: https://github.com/ale94mleon/bindflow-api-paper-si.

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jctc.5c02026.

Supporting Information includes the following: supporting methods and discussion; example code for an MM(PB/GB)SA calculation of a membrane system using BindFlow (Listing S1); examples of optional dictionaries used to define system components in BindFlow (Listing S2−S6); list of user-configurable options passed to *global_config* (Listing S7); disk space usage of BindFlow simulations for FEP and MM(PB/GB)SA workflows, including effects of data compression (Figure S1); correlation plots for FEP using OpenFF or Espaloma force fields and for MMGBSA without entropy contribution using GAFF, OpenFF, or Espaloma force fields (Figures S2−S6); comparison of BindFlow FEP results with literature data (Figure S7−S10); statistical metrics for MMGBSA with C2 or IE contribution (Figures S11 and S12); statistical metrics for MMPBSA with and without entropy contribution (Figures S13−S15); chemical structures of ligands with reported experimental affinities for all validation sets (Figures S16−S23); salt bridge interaction between lig_4 and Asp-189 in the thrombin complex (Figure S24); and summary table of agreement between calculated and experimental binding free energies across all validation sets (Table S1) (PDF)

AUTHOR INFORMATION

Corresponding Authors

Alejandro Martínez León − *Theoretical Physics and Center for Biophysics, Universität des Saarlandes, 66123 Saarbrücken, Germany;* orcid.org/0009-0005-6162-7303; Email: alejandro.martinezleon@uni-saarland.de

Jochen S. Hub − *Theoretical Physics and Center for Biophysics, Universität des Saarlandes, 66123 Saarbrücken, Germany;* orcid.org/0000-0001-7716-1767; Email: jochen.hub@uni-saarland.de

Author

Lucas Andersen − *Theoretical Physics and Center for Biophysics, Universität des Saarlandes, 66123 Saarbrücken, Germany*

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jctc.5c02026

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

We thank the authors of ABFE_Workflow for publishing the software under a free license and in particular Benjamin Ries and Aniket Magarkar for helpful discussions. We thank Mario Sergio Valdés-Trasanco for support with the gmx_MMPBSA software. This research was funded by the European Union's Horizon 2020 research and innovation program under the Marie Skłodowska Curie Grant 860592 and by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) via grant INST 256/539-1.

REFERENCES

(1) Paul, S. M.; Mytelka, D. S.; Dunwiddie, C. T.; Persinger, C. C.; Munos, B. H.; Lindborg, S. R.; Schacht, A. L. How to improve R&D productivity: the pharmaceutical industry's grand challenge. *Nat. Rev. Drug Discovery* **2010**, *9*, 203−214.

(2) DiMasi, J. A.; Grabowski, H. G.; Hansen, R. W. Innovation in the pharmaceutical industry: new estimates of R&D costs. *J. Health Econ.* **2016**, *47*, 20−33.

(3) Cournia, Z.; Chipot, C.; Roux, B.; York, D. M.; Sherman, W. *Free Energy Methods in Drug Discovery: Current state and Future Directions*; ACS Publications, 2021; pp 1−38.

(4) Li, Z.; Li, X.; Huang, Y.-Y.; Wu, Y.; Liu, R.; Zhou, L.; Lin, Y.; Wu, D.; Zhang, L.; Liu, H.; et al. Identify potent SARS-CoV-2 main protease inhibitors via accelerated free energy perturbation-based virtual screening of existing drugs. *Proc. Natl. Acad. Sci. U.S.A.* **2020**, *117*, 27381−27387.

(5) Alibay, I.; Magarkar, A.; Seeliger, D.; Biggin, P. C. Evaluating the use of absolute binding free energy in the fragment optimization process. *Commun. Chem.* **2022**, *5*, 105.

(6) Qian, R.; Xue, J.; Xu, Y.; Huang, J. Alchemical Transformations and Beyond: Recent Advances and Real-World Applications of Free Energy Calculations in Drug Discovery. *J. Chem. Inf. Model.* **2024**, *64*, 7214−7237.

(7) Zwanzig, R. W. High-temperature equation of state by a perturbation method. I. Nonpolar gases. *J. Chem. Phys.* **1954**, *22*, 1420−1426.

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

(9) Kollman, P. A.; Massova, I.; Reyes, C.; Kuhn, B.; Huo, S.; Chong, L.; Lee, M.; Lee, T.; Duan, Y.; Wang, W.; Donini, O.; Cieplak, P.; Srinivasan, J.; Case, D. A.; Cheatham, T. E. Calculating structures and free energies of complex molecules: Combining molecular mechanics and continuum models. *Acc. Chem. Res.* **2000**, *33*, 889−897.

(10) Genheden, S.; Ryde, U. The MM/PBSA and MM/GBSA methods to estimate ligand-binding affinities. *Expet Opin. Drug Discov.* **2015**, *10*, 449−461.

(11) Muegge, I.; Hu, Y. Recent Advances in Alchemical Binding Free Energy Calculations for Drug Discovery. *ACS Med. Chem. Lett.* **2023**, *14*, 244−250.

(12) Mey, A. 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]. *Comp. Mol. Sci.* **2020**, *2*, 1−52.

(13) Deflorian, F.; Perez-Benito, L.; Lenselink, E. B.; Congreve, M.; van Vlijmen, H. W.; Mason, J. S.; Graaf, C. d.; Tresadern, G. Accurate prediction of GPCR ligand binding affinity with free energy perturbation. *J. Chem. Inf. Model.* **2020**, *60*, 5563−5579.

(14) Aldeghi, M.; Gapsys, V.; de Groot, B. L. Accurate estimation of ligand binding affinity changes upon protein mutation. *ACS Cent. Sci.* **2018**, *4*, 1708−1718.

(15) 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.

(16) Kuhn, M.; Firth-Clark, S.; Tosco, P.; Mey, A. S.; Mackey, M.; Michel, J. Assessment of binding affinity via alchemical free-energy calculations. *J. Chem. Inf. Model.* **2020**, *60*, 3120−3130.

(17) Lin, Z.; Zou, J.; Liu, S.; Peng, C.; Li, Z.; Wan, X.; Fang, D.; Yin, J.; Gobbo, G.; Chen, Y.; et al. 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.

(18) Moore, J. H.; Margreitter, C.; Janet, J. P.; Engkvist, O.; de Groot, B. L.; Gapsys, V. Automated relative binding free energy calculations from SMILES to ΔΔG. *Commun. Chem.* **2023**, *6*, 82.

(19) Carvalho Martins, L.; Cino, E. A.; Ferreira, R. S. PyAutoFEP: An Automated Free Energy Perturbation Workflow for GROMACS Integrating Enhanced Sampling Methods. *J. Chem. Theory Comput.* **2021**, *17*, 4262−4273.
(20) Zavitsanou, S.; Tsengenes, A.; Papadourakis, M.; Amendola, G.; Chatzigoulas, A.; Dellis, D.; Cosconati, S.; Cournia, Z. FEPrepare: A Web-Based Tool for Automating the Setup of Relative Binding Free Energy Calculations. *J. Chem. Inf. Model.* **2021**, *61*, 4131−4138.
(21) Gowers, R. J.; Alibay, I.; Swenson, D. W.; Henry, M. M.; Ries, B.; Baumann, H. M.; Eastwood, J. R. B.; Mitchell, J. A.; Dotson, D.; Horton, J. T.; Thompson, M. The Open Free Energy library. 2024; https://github.com/openfreeEnergy/openfe.
(22) 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.; et al. Alchemical Binding Free Energy Calculations in AMBER20: Advances and Best Practices for Drug Discovery. *J. Chem. Inf. Model.* **2020**, *60*, 5595−5623.
(23) Heinzelmann, G.; Gilson, M. K. Automation of absolute protein-ligand binding free energy calculations for docking refinement and compound evaluation. *Sci. Rep.* **2021**, *11*, 1116.
(24) Homeyer, N.; Gohlke, H. FEW: A workflow tool for free energy calculations of ligand binding. *J. Comput. Chem.* **2013**, *34*, 965−973.
(25) Jespers, W.; Esguerra, M.; Åqvist, J.; Gutiérrez-De-Terán, H. Qligfep: An automated workflow for small molecule free energy calculations in Q. *J. Cheminf.* **2019**, *11*, 26.
(26) Matricon, P.; Ranganathan, A.; Warnick, E.; Gao, Z. G.; Rudling, A.; Lambertucci, C.; Marucci, G.; Ezzati, A.; Jaiteh, M.; Dal Ben, D.; Jacobson, K. A.; Carlsson, J. Fragment optimization for GPCRs by molecular dynamics free energy calculations: Probing druggable subpockets of the A2A adenosine receptor binding site. *Sci. Rep.* **2017**, *7*, 6398.
(27) Lenselink, E. B.; et al. Predicting Binding Affinities for GPCR Ligands Using Free-Energy Perturbation. *ACS Omega* **2016**, *1*, 293−304.
(28) Hedges, L.; Mey, A.; Laughton, C.; Gervasio, F.; Mulholland, A.; Woods, C.; Michel, J. BioSimSpace: An interoperable Python framework for biomolecular simulation. *J. Open Source Softw.* **2019**, *4*, 1831.
(29) Kim, S.; Oshima, H.; Zhang, H.; Kern, N. R.; Re, S.; Lee, J.; Roux, B.; Sugita, Y.; Jiang, W.; Im, W. CHARMM-GUI Free Energy Calculator for Absolute and Relative Ligand Solvation and Binding Free Energy Simulations. *J. Chem. Theory Comput.* **2020**, *16*, 7207−7218.
(30) Hahn, D. F.; Bayly, C. I.; Boby, M. L.; Bruce Macdonald, H. E.; Chodera, J. D.; Gapsys, V.; Mey, A. S. J. S.; Mobley, D. L.; Benito, L. P.; Schindler, C. E. M.; Tresadern, G.; Warren, G. L. Best Practices for Constructing, Preparing, and Evaluating Protein-Ligand Binding Affinity Benchmarks [Article v1.0]. *Comp. Mol. Sci.* **2022**, *4*, 1−35.
(31) Zhang, C. H.; et al. Potent Noncovalent Inhibitors of the Main Protease of SARS-CoV-2 from Molecular Sculpting of the Drug Perampanel Guided by Free Energy Perturbation Calculations. *ACS Cent. Sci.* **2021**, *7*, 467−475.
(32) Tresadern, G.; Velter, I.; Trabanco, A. A.; Van den Keybus, F.; MacDonald, G. J.; Somers, M. V.; Vanhoof, G.; Leonard, P. M.; Lamers, M. B.; Van Roosbroeck, Y. E. M.; Buijnsters, P. J. [1,2,4]Triazolo[1,5-a]pyrimidine Phosphodiesterase 2A Inhibitors: Structure and Free-Energy Perturbation-Guided Exploration. *J. Med. Chem.* **2020**, *63*, 12887−12910.
(33) O'Donovan, D. H.; Gregson, C.; Packer, M. J.; Greenwood, R.; Pike, K. G.; Kawatkar, S.; Bloecher, A.; Robinson, J.; Read, J.; Code, E.; et al. Free energy perturbation in the design of EED ligands as inhibitors of polycomb repressive complex 2 (PRC2) methyltransferase. *Bioorg. Med. Chem. Lett.* **2021**, *39*, 127904.
(34) Papadourakis, M.; Sinenka, H.; Matricon, P.; Hénin, J.; Brannigan, G.; Pérez-Benito, L.; Pande, V.; van Vlijmen, H.; de Graaf, C.; Deflorian, F.; et al. Alchemical Free Energy Calculations on Membrane-Associated Proteins. *J. Chem. Theory Comput.* **2023**, *19*, 7437−7458.
(35) Majellaro, M.; Jespers, W.; Crespo, A.; Núñez, M. J.; Novio, S.; Azuaje, J.; Prieto-Diaz, R.; Gioé, C.; Alispahic, B.; Brea, J.; et al. 3,4-Dihydropyrimidin-2(1*H*)-ones as Antagonists of the Human A_{2B} Adenosine Receptor: Optimization, Structure−Activity Relationship Studies, and Enantiospecific Recognition. *J. Med. Chem.* **2020**, *64*, 458−480.
(36) Keränen, H.; de Terán, H. G.; Åqvist, J. Structural and energetic effects of A_{2A} adenosine receptor mutations on agonist and antagonist binding. *PLoS One* **2014**, *9*, No. e108492.
(37) Xue, B.; et al. Development and Comprehensive Benchmark of a High-Quality AMBER-Consistent Small Molecule Force Field with Broad Chemical Space Coverage for Molecular Modeling and Free Energy Calculation. *J. Chem. Theory Comput.* **2024**, *20*, 799−818.
(38) Nie, Z.; et al. Accelerated In Silico Discovery of SGR-1505: A Potent MALT1 Allosteric Inhibitor for the Treatment of Mature B-Cell Malignancies. *J. Med. Chem.* **2025**, *68*, 23977−23992.
(39) 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.
(40) Feng, M.; Heinzelmann, G.; Gilson, M. K. Absolute binding free energy calculations improve enrichment of actives in virtual compound screening. *Sci. Rep.* **2022**, *12*, 13640.
(41) Michino, M.; Beautrait, A.; Boyles, N. A.; Nadupalli, A.; Dementiev, A.; Sun, S.; Ginn, J.; Baxt, L.; Suto, R.; Bryk, R.; Jerome, S. V.; Huggins, D. J.; Vendome, J. Shape-Based Virtual Screening of a Billion-Compound Library Identifies Mycobacterial Lipoamide Dehydrogenase Inhibitors. *ACS Bio and Med. Chem. Au* **2023**, *3*, 507−515.
(42) Andersen, L.; Rausch-Dupont, M.; Martínez León, A.; Volkamer, A.; Hub, J. S.; Klakow, D. Accelerating ligand discovery by combining Bayesian optimization with MMGBSA-based binding affinity calculations. *bioRxiv* **2025**, bioRxiv:2025.06.22.660936.
(43) 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.
(44) Vernuccio, R.; et al. Structural insights into tecovirimat antiviral activity and poxvirus resistance. *Nat. Microbiol.* **2025**, *10*, 734−748.
(45) 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.
(46) Lindahl, E.; Friedman, R. Estimation of Absolute Binding Free Energies for Drugs That Bind Multiple Proteins. *J. Chem. Inf. Model.* **2025**, *65*, 3431−3438.
(47) Chen, W.; Cui, D.; Jerome, S. V.; Michino, M.; Lenselink, E. B.; Huggins, D. J.; Beautrait, A.; Vendome, J.; Abel, R.; Friesner, R. A.; Wang, L. Enhancing Hit Discovery in Virtual Screening through Absolute Protein-Ligand Binding Free-Energy Calculations. *J. Chem. Inf. Model.* **2023**, *63*, 3171−3185.
(48) Araujo, S. C.; Maltarollo, V. G.; Almeida, M. O.; Ferreira, L. L.; Andricopulo, A. D.; Honorio, K. M. Structure-based virtual screening, molecular dynamics and binding free energy calculations of HIT candidates as ALK-5 inhibitors. *Molecules* **2020**, *25*, 264.
(49) 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.
(50) Fu, H.; Zhou, Y.; Jing, X.; Shao, X.; Cai, W. Meta-Analysis Reveals That Absolute Binding Free-Energy Calculations Approach Chemical Accuracy. *J. Med. Chem.* **2022**, *65*, 12970−12978.
(51) Li, P.; Pu, T.; Mei, Y. FEP-SPell-ABFE: An Open-Source Automated Alchemical Absolute Binding Free-Energy Calculation Workflow for Drug Discovery. *J. Chem. Inf. Model.* **2025**, *65*, 2711−2721.
(52) Ries, B.; Alibay, I.; Anand, N. M.; Biggin, P. C.; Magarkar, A. Automated Absolute Binding Free Energy Calculation Workflow for Drug Discovery. *J. Chem. Inf. Model.* **2024**, *64*, 5357−5364.
(53) Clark, F.; Robb, G. R.; Cole, D. J.; Michel, J. Automated Adaptive Absolute Binding Free Energy Calculations. *J. Chem. Theory Comput.* **2024**, *20*, 7806−7828.

(54) Heinzelmann, G.; Huggins, D. J.; Gilson, M. K. BAT2: an Open-Source Tool for Flexible, Automated, and Low Cost Absolute Binding Free Energy Calculations. *J. Chem. Theory Comput.* **2024**, *20*, 6518−6530.
(55) Santiago-McRae, E.; Ebrahimi, M.; Sandberg, J. W.; Brannigan, G.; Hénin, J. Computing absolute binding affinities by Streamlined Alchemical Free Energy Perturbation [Article v1.0]. *Comp. Mol. Sci.* **2023**, *5*, 1−25.
(56) Fu, H.; Chen, H.; Blazhynska, M.; Goulard Coderc de Lacam, E.; Szczepaniak, F.; Pavlova, A.; Shao, X.; Gumbart, J. C.; Dehez, F.; Roux, B.; Cai, W.; Chipot, C. Accurate determination of protein:ligand standard binding free energies from molecular dynamics simulations. *Nat. Protoc.* **2022**, *17*, 1114−1141.
(57) Fu, H.; Chen, H.; Cai, W.; Shao, X.; Chipot, C. BFEE2: Automated, Streamlined, and Accurate Absolute Binding Free-Energy Calculations. *J. Chem. Inf. Model.* **2021**, *61*, 2116−2123.
(58) 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.
(59) Dakka, J.; Farkas-Pall, K.; Turilli, M.; Wright, D. W.; Coveney, P. V.; Jha, S. Concurrent and adaptive extreme scale binding free energy calculations. *Proc. 14th Int. Conf. eScience* **2018**, 189−200.
(60) Loeffler, H. H.; Michel, J.; Woods, C. FESetup: Automating Setup for Alchemical Free Energy Simulations. *J. Chem. Inf. Model.* **2015**, *55*, 2485−2490.
(61) OpenFE Development Team OPenFE: Setting up and Running Absolute Binding Free Energy Calculations. https://docs.openfree.energy/en/latest/tutorials/abfe_tutorial.html, 2025; (accessed Nov 19, 2025).
(62) Kirkwood, J. G. Statistical mechanics of fluid mixtures. *J. Chem. Phys.* **1935**, *3*, 300−313.
(63) Shirts, M. R.; Chodera, J. D. Statistically optimal analysis of samples from multiple equilibrium states. *J. Chem. Phys.* **2008**, *129*, 124105.
(64) Mölder, F.; Jablonski, K. P.; Letcher, B.; Hall, M. B.; Tomkins-Tinch, C. H.; Sochat, V.; Forster, J.; Lee, S.; Twardziok, S. O.; Kanitz, A.; Wilm, A.; Holtgrewe, M.; Rahmann, S.; Nahnsen, S.; Köster, J. Sustainable data analysis with Snakemake. *F1000Research* **2021**, *10*, 33.
(65) 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.
(66) Behara, P. K.; et al. Development and Benchmarking of Open Force Field 2.0.0: The Sage Small Molecule Force Field. *J. Chem. Theory Comput.* **2023**, *19*, 3251−3275.
(67) Wang, Y.; Fass, J.; Kaminow, B.; Herr, J. E.; Rufa, D.; Zhang, I.; Pulido, I.; Henry, M.; Bruce Macdonald, H. E.; Takaba, K.; Chodera, J. D. End-to-end differentiable construction of molecular mechanics force fields. *Chem. Sci.* **2022**, *13*, 12016−12033.
(68) Chipot, C.; Pohorille, A. *Free Energy Calculations*; Springer, 2007; Vol. *86*.
(69) Valdés-Tresanco, M. S.; Valdés-Tresanco, M. E.; Valiente, P. A.; Moreno, E. gmx_MMPBSA: a new tool to perform end-state free energy calculations with GROMACS. *J. Chem. Theory Comput.* **2021**, *17*, 6281−6291.
(70) Wang, E.; Sun, H.; Wang, J.; Wang, Z.; Liu, H.; Zhang, J. Z.; Hou, T. End-point binding free energy calculation with MM/PBSA and MM/GBSA: strategies and applications in drug design. *Chem. Rev.* **2019**, *119*, 9478−9508.
(71) Lee, B.; Richards, F. M. The interpretation of protein structures: estimation of static accessibility. *J. Mol. Biol.* **1971**, *55*, 379.
(72) Duan, L.; Liu, X.; Zhang, J. Z. Interaction entropy: A new paradigm for highly efficient and reliable computation of protein−ligand binding free energy. *J. Am. Chem. Soc.* **2016**, *138*, 5722−5728.
(73) Sun, Z.; Yan, Y. N.; Yang, M.; Zhang, J. Z. H. Interaction entropy for protein-protein binding. *J. Chem. Phys.* **2017**, *146*, 124124.
(74) Menzer, W. M.; Li, C.; Sun, W.; Xie, B.; Minh, D. D. L. Simple Entropy Terms for End-Point Binding Free Energy Calculations. *J. Chem. Theory Comput.* **2018**, *14*, 6035−6049.
(75) Ekberg, V.; Ryde, U. On the Use of Interaction Entropy and Related Methods to Estimate Binding Entropies. *J. Chem. Theory Comput.* **2021**, *17*, 5379−5391.
(76) Abraham, M. J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J. C.; Hess, B.; Lindahl, E. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **2015**, *1*, 19−25.
(77) Yoo, A. B.; Jette, M. A.; Grondona, M. Slurm.: Simple linux utility for resource management. In *Workshop on job scheduling strategies for parallel processing*; Springer, 2003; pp 44−60.
(78) Behera, S.; Hahn, D. F.; Wilson, C. J.; Marsili, S.; Tresadern, G.; Gapsys, V.; de Groot, B. L. Quantification of the Impact of Structure Quality on Predicted Binding Free Energy Accuracy. *J. Chem. Inf. Model.* **2025**, *65*, 6927−6938.
(79) Eastman, P.; PDBFixer. https://github.com/openmm/pdbfixer, 2023; (accessed Nov 19, 2025).
(80) Lindorff-Larsen, K.; Piana, S.; Palmo, K.; Maragakis, P.; Klepeis, J. L.; Dror, R. O.; Shaw, D. E. Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins: Struct., Funct., Bioinf.* **2010**, *78*, 1950−1958.
(81) Grote, F.; Lyubartsev, A. P. Optimization of Slipids Force Field Parameters Describing Headgroups of Phospholipids. *J. Phys. Chem. B* **2020**, *124*, 8784−8793.
(82) 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.
(83) Martínez León, A. *TOFF: Topology from Open Force Fields*, 2023.
(84) Hahn, D. F.; Gapsys, V.; de Groot, B. L.; Mobley, D. L.; Tresadern, G. Current State of Open Source Force Fields in Protein-Ligand Binding Affinity Predictions. *J. Chem. Inf. Model.* **2024**, *64*, 5063−5076.
(85) Rizzi, A.; Murkli, S.; McNeill, J. N.; Yao, W.; Sullivan, M.; Gilson, M. K.; Chiu, M. W.; Isaacs, L.; Gibb, B. C.; Mobley, D. L.; Chodera, J. D. Overview of the SAMPL6 host-guest binding affinity prediction challenge. *J. Comput.-Aided Mol. Des.* **2018**, *32*, 937−963.
(86) Isik, M.; Rizzi, A.; Mobley, D. L.; Bergazin, T. D.; Shirts, M. *SAMPL6*, 2024.
(87) Khalak, Y.; Tresadern, G.; Aldeghi, M.; Baumann, H. M.; Mobley, D. L.; de Groot, B. L.; Gapsys, V. Alchemical absolute protein-ligand binding free energies for drug design. *Chem. Sci.* **2021**, *12*, 13958−13971.
(88) Croux, C.; Dehon, C. Influence functions of the Spearman and Kendall correlation measures. *Stat. Methods Appl.* **2010**, *19*, 497−515.
(89) Ross, G. A.; Lu, C.; Scarabelli, G.; Albanese, S. K.; Houang, E.; Abel, R.; Harder, E. D.; Wang, L. The maximal and current accuracy of rigorous protein-ligand binding free energy calculations. *Commun. Chem.* **2023**, *6*, 222.
(90) Lu, C.; Wu, C.; Ghoreishi, D.; Chen, W.; Wang, L.; Damm, W.; Ross, G. A.; Dahlgren, M. K.; Russell, E.; Von Bargen, C. D.; et al. OPLS4: Improving Force Field Accuracy on Challenging Regimes of Chemical Space. *J. Chem. Theory Comput.* **2021**, *17*, 4291−4300.
(91) Su, P. C.; Tsai, C. C.; Mehboob, S.; Hevener, K. E.; Johnson, M. E. Comparison of radii sets, entropy, QM methods, and sampling on MM-PBSA, MM-GBSA, and QM/MM-GBSA ligand binding energies of F. tularensis enoyl-ACP reductase (FabI). *J. Comput. Chem.* **2015**, *36*, 1859−1873.
(92) Abhishek, M. K.; Rao, D. R.; Subrahmanyam, K. Priority Queue-Based Framework for Allocation of High Performance Computing Resources. *IOT with Smart Systems* **2023**, *312*, 317−327.
(93) Wu, Z.; Konig, G.; Boresch, S.; Cossins, B. P. Optimizing Absolute Binding Free Energy Calculations for Production Usage. *J. Chem. Theory Comput.* **2025**, *21*, 8330−8340.
(94) Meola, A.; et al. Structural basis of poxvirus fusion regulation and anti-A16/G9 antibody-mediated neutralization and protection. *Cell* **2025**, *188*, 6266−6282.

(95) Wang, K.; Chodera, J. D.; Yang, Y.; Shirts, M. R. Identifying ligand binding sites and poses using GPU-accelerated Hamiltonian replica exchange molecular dynamics. *J. Comput.-Aided Mol. Des.* **2013**, *27*, 989−1007.
(96) Chodera, J. D.; Shirts, M. R. Replica exchange and expanded ensemble simulations as Gibbs sampling: Simple improvements for enhanced mixing. *J. Chem. Phys.* **2011**, *135*, 194110.
(97) Sugita, Y.; Kitao, A.; Okamoto, Y. Multidimensional replica-exchange method for free-energy calculations. *J. Chem. Phys.* **2000**, *113*, 6042−6051.
(98) 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.
(99) 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.
(100) Wu, Z.; Biggin, P. C. Correction Schemes for Absolute Binding Free Energies Involving Lipid Bilayers. *J. Chem. Theory Comput.* **2022**, *18*, 2657−2672.
(101) 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.
(102) 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.
(103) Öhlknecht, C.; Lier, B.; Petrov, D.; Fuchs, J.; Oostenbrink, C. Correcting electrostatic artifacts due to net-charge changes in the calculation of ligand binding free energies. *J. Comput. Chem.* **2020**, *41*, 986−999.
(104) Ross, G. A.; Bodnarchuk, M. S.; Essex, J. W. Water sites, networks, and free energies with grand canonical Monte Carlo. *J. Am. Chem. Soc.* **2015**, *137*, 14930−14943.
(105) Bruce Macdonald, H. E.; Cave-Ayland, C.; Ross, G. A.; Essex, J. W. Ligand binding free energies with adaptive water networks: two-dimensional grand canonical alchemical perturbations. *J. Chem. Theory Comput.* **2018**, *14*, 6586−6597.
(106) Haider, K.; Cruz, A.; Ramsey, S.; Gilson, M. K.; Kurtzman, T. Solvation structure and thermodynamic mapping (SSTMap): an open-source, flexible package for the analysis of water in molecular dynamics trajectories. *J. Chem. Theory Comput.* **2018**, *14*, 418−425.
(107) Grubmüller, H.; Groll, V. *Solvate*; Max Planck Institute for Multidisciplinary Sciences, 2025; https://www.mpinat.mpg.de/grubmueller/solvate.
(108) Wang, X.; Yang, H.; Wang, M.; Huai, Z.; Sun, Z. Virtual screening of cucurbituril host-guest complexes: Large-scale benchmark of end-point protocols under MM and QM Hamiltonians. *J. Mol. Liq.* **2024**, *407*, 125245.
(109) Ruiz-Blanco, Y. B.; Sanchez-Garcia, E. CL-FEP: An End-State Free Energy Perturbation Approach. *J. Chem. Theory Comput.* **2020**, *16*, 1396−1410.
(110) Jo, S.; Kim, T.; Iyer, V. G.; Im, W. CHARMM-GUI: A web-based graphical user interface for CHARMM. *J. Comput. Chem.* **2008**, *29*, 1859−1865.
(111) Maier, J. A.; Martinez, C.; Kasavajhala, K.; Wickstrom, L.; Hauser, K. E.; Simmerling, C. ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *J. Chem. Theory Comput.* **2015**, *11*, 3696−3713.
(112) Eastman, P.; Swails, J.; Chodera, J. D.; McGibbon, R. T.; Zhao, Y.; Beauchamp, K. A.; Wang, L. P.; Simmonett, A. C.; Harrigan, M. P.; Stern, C. D.; Wiewiora, R. P.; Brooks, B. R.; Pande, V. S. OpenMM 7: Rapid development of high performance algorithms for molecular dynamics. *PLoS Comput. Biol.* **2017**, *13*, No. e1005659.
(113) Shirts, M. R.; Klein, C.; Swails, J. M.; Yin, J.; Gilson, M. K.; Mobley, D. L.; Case, D. A.; Zhong, E. D. Lessons learned from comparing molecular dynamics engines on the SAMPL5 dataset. *J. Comput.-Aided Mol. Des.* **2017**, *31*, 147−161.
(114) Hess, B.; Bekker, H.; Berendsen, H. J.; Fraaije, J. G. LINCS: A Linear Constraint Solver for molecular simulations. *J. Comput. Chem.* **1997**, *18*, 1463−1472.
(115) 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.
(116) Bernetti, M.; Bussi, G. Pressure control using stochastic cell rescaling. *J. Chem. Phys.* **2020**, *153*, 114107.
(117) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A.; Haak, J. R. Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* **1984**, *81*, 3684−3690.
(118) Parrinello, M.; Rahman, A. Crystal structure and pair potentials: A molecular-dynamics study. *Phys. Rev. Lett.* **1980**, *45*, 1196−1199.
(119) Alibay, I. *IAlibay/MDRestraintsGenerator: Mdrestraintsgenerator v0.2.0*, 2022.
(120) Wu, Z.; et al. alchemlyb: the simple alchemistry library. *J. Open Source Softw.* **2024**, *9*, 6934.

https://doi.org/10.1021/acs.jctc.5c02026
J. Chem. Theory Comput. 2026, 22, 1198−1213