Published in final edited form as:

Living J Comput Mol Sci. 2020 ; 2(1): . doi:10.33011/livecoms.2.1.18378.

Best Practices for Alchemical Free Energy Calculations [Article v1.0]

Antonia S. J. S. Mey[1,*], **Bryce K. Allen**[2], **Hannah E. Bruce Macdonald**[3], **John D. Chodera**[3,*], **David F. Hahn**[9], **Maximilian Kuhn**[1,10], **Julien Michel**[1], **David L. Mobley**[4,*], **Levi N. Naden**[5], **Samarjeet Prasad**[6], **Andrea Rizzi**[2,7], **Jenke Scheen**[1], **Michael R. Shirts**[8,*], **Gary Tresadern**[9], **Huafeng Xu**[2]

[1]EaStCHEM School of Chemistry, David Brewster Road, Joseph Black Building, The King's Buildings, Edinburgh, EH9 3FJ, UK;

[2]Silicon Therapeutics, Boston, MA, USA;

[3]Computational and Systems Biology Program, Sloan Kettering Institute, Memorial Sloan Kettering Cancer Center, New York NY, USA;

[4]Departments of Pharmaceutical Sciences and Chemistry, University of California, Irvine, Irvine, USA;

[5]Molecular Sciences Software Institute, Blacksburg VA, USA;

[*] **For correspondence:** antonia.mey@ed.ac.uk (ASJSM); john.chodera@choderalab.org (JDC); dmobley@mobleylab.org (DLM); michael.shirts@colorado.edu (MRS).

Author Contributions

ASJSM: Coordinated the document, contributed to most sections, and co-designed Figs. 2, 3, 4, 5, 6, 14, and created Figs. 7, 1, 13, 10 and replotted 11 and 5.

BA: Helped write the uncertainty estimation, stopping conditions, and output analysis sections and created figure 9.

HBM Contributed to Sec. 8.7 and Fig. 14 and helped edit the paper.

JDC: Wrote Sec. 8.2 and 8.1 discussed structure and design of the whole document, suggested Figs. 1 and 8.

DFH Contributed to Sec. 10 and helped edit the paper.

MK: Contributed to Sec. 8, provided the data for figure 13, compiled the dataset for Sec. 11 and helped edit the paper.

JM: Contributed to Sec. 4.2, 6, 7.2.3, 8.3, 8.4, and 9

DLM: Contributed to the outline, drafted some of the sections, gave ideas on figures, and helped edit the paper.

LNN: Helped write the simulation length, stopping conditions, and information saving section. Edited and reviewed alchemical path section.

SP Wrote Sec. 10.

AR: Created figure 12, contributed to sections 3 and 7, and helped edit the paper.

JS: Created Figs. 1, 2, 3, 4, 6, 14, and an initial draft of 5. Wrote Sec. 8.7, the checklist Sec. 12, and contributed to general formatting discussions and editing.

MRS: Helped create figure 7, wrote Sec. 7.2.3 describing choices for alchemical pathways and parts of 8 on the analysis for free energy calculations. Reviewed and edited text throughout.

GT: Contributed to Sec. 1 and 5, and helped edit the paper.

HX: Contributed Sec. 4.4, to Sec. 7.1.1, and to Sec. 8.5. For a more detailed description of author contributions, see the GitHub issue tracking and changelog at https://github.com/alchemistry/alchemical-best-practices.

Other Contributions

Julia E. Rice participated in the original discussion of the document at the Best Practices in Molecular Simulation Workshop Hosted by at NIST, Gaithersburg, MD, August 24th–25th, 2017. Marieke Schor proofread the manuscript.

For a more detailed description of contributions from the community and others, see the GitHub issue tracking and changelog at https://github.com/alchemistry/alchemical-best-practices.

This LiveCoMS document is maintained online on GitHub at https://github.com/alchemistry/alchemical-best-practices; to provide feedback, suggestions, or help improve it, please visit the GitHub repository and participate via the issue tracker.

Potentially Conflicting Interests

JM is a current member of the Scientific Advisory Board of Cresset. MK is employed by Cresset who commercially distribute a software for performing alchemical free energy calculations. MRS is a Open Science Fellow and consultant for Silicon Therapeutics. JDC is a current member of the Scientific Advisory Board of OpenEye Scientific Software and a consultant to Foresite Laboratories.

[6]National Institutes of Health, Bethesda, MD, USA;

[7]Tri-Institutional Training Program in Computational Biology and Medicine, New York, NY, USA;

[8]University of Colorado Boulder, Boulder, CO, USA;

[9]Computational Chemistry, Janssen Research & Development, Turnhoutseweg 30, Beerse B-2340, Belgium;

[10]Cresset, Cambridgeshire, UK

Abstract

Alchemical free energy calculations are a useful tool for predicting free energy differences associated with the transfer of molecules from one environment to another. The hallmark of these methods is the use of “bridging” potential energy functions representing *alchemical* intermediate states that cannot exist as real chemical species. The data collected from these bridging alchemical thermodynamic states allows the efficient computation of transfer free energies (or differences in transfer free energies) with orders of magnitude less simulation time than simulating the transfer process directly. While these methods are highly flexible, care must be taken in avoiding common pitfalls to ensure that computed free energy differences can be robust and reproducible for the chosen force field, and that appropriate corrections are included to permit direct comparison with experimental data.

In this paper, we review current best practices for several popular application domains of alchemical free energy calculations performed with equilibrium simulations, in particular relative and absolute small molecule binding free energy calculations to biomolecular targets.

1 What are alchemical free energy methods?

Alchemical free energy calculations compute free energy differences associated with transfer processes, such as the binding of a small molecule to a receptor, the transfer of a small molecule from an aqueous to apolar phase [1], or the effects of protein side chain mutations on binding affinities or thermostabilities. These calculations use non-physical[1] intermediate states in which the chemical identity of some portion of the system (such as a small molecule ligand or protein sidechain) is changed by modifying the potential governing the interactions with the environment for the atoms being modified, inserted, or deleted.

Fig. 1 illustrates common free energy changes that may be difficult to compute with unbiased molecular dynamics methods, but are more tractable with alchemical methods. In alchemical simulations, the introduction of intermediate *alchemical states* that bridge the high-probability regions of configuration space between two physical endstates of interest, permits the robust computation of free energy for large transformations. Alchemical calculations can be used in a variety of scenarios, such as:

[1] Here, the non-physical nature of the transformation is referred to as “alchemical”, a term coined by Tembre and McCammon in Ref. [2].

- computing the free energy of a conformational change for a molecule with a high barrier to interconversion (Fig. 1A);
- computing partition (log *P*) or distribution (log *D*) coefficients between environments (Fig. 1B) [3, 4]
- determining partitioning between compartments into membranes (Fig. 1C) [5].

Furthermore, alchemical calculations are frequently used to estimate changes in free energies upon modifying a ligand or protein:

- a protein residue can be alchemically mutated to probe the impact on binding affinity (Fig. 1D)[6, 7] or changes in protein thermostability [8–11];
- the entire ligand can be alchemically transferred from protein to solvent in an absolute binding free energy calculation (Fig. 1E) [12–14];
- small alchemical modifications can be made between chemically related ligands to estimate relative differences in binding free energies (Fig. 1F) [15–19].

After an alchemical calculation is performed, which generally involves multiple simulations at a variety of alchemical states, the data must be analyzed to compute an estimate of the free energy for the transformation of interest. Early work used simple but statistically suboptimal estimators for this: free energy perturbation (FEP) used a simple (but highly biased) estimator based on the Zwanzig relation [1] or numerical quadrature via thermodynamic integration (TI), for which the theory dates back the better part of a century but with the first computational applications emerging in the 1980's and 90's [20–24]. More recent developments have seen new, highly efficient statistical estimators that make better use of all the data, often building on the more efficient and less biased Bennett acceptance ratio (BAR) [25], producing multistate generalizations [26] or removing the need for global equilibrium [27–29].

Subsequent work in the 2000s led to improved implementations of alchemical methods in popular biomolecular simulation packages [15, 30–35]. This foundational work, combined with the methodological, technological, and hardware improvements of the last 5–10 years, has led to an explosion of interest and direct commercial application of these technologies [15, 19, 36–39].

As the field of molecular simulation can now routinely access microsecond timescales with the aid of GPUs [40], and millisecond timescales appear to soon be within reach, accurate alchemical calculations on even more challenging problems will become reasonable to perform. In the meantime, today's users may find it difficult to get started with these complex calculations whilst also keeping up with the fast pace of change. This Best Practices guide provides current recommendations and tips for users of all experience. Updates and suggestions are welcomed via our GitHub repository at https://github.com/alchemistry/alchemical-best-practices.

2 Prerequisites and Scope

This Best Practices guide focuses on providing a good starting point for new practitioners and a reference for experienced practitioners. For this purpose we provide a convenient checklist (Sec. 12) to help ensure all calculations comply with currently-understood best practices for alchemical simulation and analysis. Where the best practices are currently not certain, we highlight areas where further research is needed to identify an unambiguous recommendation. This guide can also serve as a set of best practices to ensure simulation robustness and reproducibility which reviewers may wish to consider as they evaluate papers.

We assume that novice practitioners have at least moderate experience with molecular simulation concepts and use of simulation packages. Furthermore, basic familiarity with the principles of molecular mechanics, molecular dynamics simulations, statistical mechanics, and the biophysics of protein-ligand association are essential. If you feel unfamiliar with some of these concepts, good starting points can be found in these references [41–44].

While reading this Best Practices guide, it is important to bear in mind *this is not a review* of all free energy calculation methods at the cutting edge of current research. Instead this guide aims to answer the following questions:

- Is my problem suitable for an alchemical calculation?
- How do I select an appropriate alchemical protocol?
- What software tools are available to perform alchemical calculations?
- How should I analyze my data and report uncertainties?

Some other background information may be needed depending on the nature of the alchemical project. For example, often, if binding poses are not known, docking calculations can be used to generate an initial small molecule binding pose to start alchemical simulations. This will require some basic familiarity on how to perform docking to generate reasonable simulation starting points [45].

As some of the theoretical background can seem daunting, we do, however, provide a guide to the essential theory behind alchemical free energy calculations in Sec. 3. In the remainder of this paper, we will cover topics that are key to the preparation (Sec. 6), choice and use of correct protocols (Sec. 7), and finally the best practices that should be used in the analysis of alchemical calculations (Sec. 8). Particular focus will be given to aspects of the molecular simulations which are unique to alchemical calculations—these include the calculation of transfer free energies (hydration free energies, partition coefficients, etc.), and binding free energies (absolute and relative). We primarily focus on free energy calculations using simulations performed at *equilibrium* in this Best Practices guide, as best practices for these are more developed, and non-equilibrium techniques may warrant their own guide as such practices evolve.

While we try to address as many methods and practices as possible, the field of free energy calculations is broad, and there are many advanced topics that are left to future Best

Practices documents focusing on specific issues. Below, we provide a non-exhaustive list of topics we have *not* addressed, along with some references to provide starting points on these more advanced topics:

- covalent inhibition [46]
- free energies of mutation of protein side chains [7, 10]
- nonspecific binding or multiple binding sites [47]
- approximate and often less accurate endpoint free energy methods such as MM-PBSA [48] and LIE [49]
- Free energy methods that extract the ligand using geometric order parameters and potential of mean force methods [50]
- forcefield dependence for protein, ligand, ions, cosolvents, and co-factors. A number of different studies have looked at the influence of force fields and it is assumed the user has made an appropriate choice for the system under study [51–53].
- non-equilibrium free energy calculations [18]
- Free energy calculations using QM/MM methods [54–56].
- Free energy calculations using machine learning methods [57–59]

For convenience we have also compiled a list of common acronyms and common symbols used throughout this paper.

3 Statistical mechanics demonstrates why alchemical free energy calculations work

Why would you want to run an alchemical free energy calculation and why do they work? In this section, we use the example of relative free energy calculations to sketch the theory of alchemical simulations and illustrate their utility. The emphasis here is placed on bridging theoretical foundations and intuition. A rigorous derivation of the standard (absolute) free energy of binding using the principles of statistical mechanics can be found in Gilson's classic work [60].

3.1 Simulating binding events of receptor-drug systems can be computationally expensive

Suppose you want to compute the binding affinity, or free energy of binding, of a ligand L to a receptor R, given by:

$$R + L \rightleftharpoons RL\,. \tag{1}$$

The binding constant (K_b°) is given by the law of mass action as the ratio of concentrations of product $[RL]$ and reactants $[R]$, $[L]$:

$$K_b^{\circ} = c^{\circ}\frac{[RL]}{[L][R]}. \quad (2)$$

The standard state concentration c° depends on the reference state, but it is usually set to 1 mol/L assuming a constant pressure of 1 atm (see also Sec. 7.1.2). Eq. 2 also holds for dilute solutions if thermodynamic activities can be approximated by concentrations [60]. Thus, the standard Gibbs free energy of binding ΔG_{bind} is given by:

$$\Delta G_{\text{bind},L} = -k_B T \ln K_b^{\circ}, \quad (3)$$

where k_B is the Boltzmann constant and T the temperature of the system. Note that, unless otherwise specified, we use ΔG throughout the paper to refer to the *standard* free energy of binding (or solvation) (see also Sec. 7.1.2), which is often indicated in other works with ΔG° to differentiate it from non-standard free energies. This is in line with current literature on alchemical calculations, where the standard free energy is normally the only quantity of interest. Furthermore, we will use the term configuration for a single set of position vectors and occasionally use the term conformation to refer to a set of configurations that represent a metastable state.

The free energy of binding can be expressed as a ratio of partition functions—The law of mass action in Eq. 2 is not directly applicable to typical molecular simulations as they normally include a single receptor/ligand in a small box (i.e., using large concentrations) [61]. Instead, a natural, though generally very computationally expensive, way to estimate the equilibrium constant is by directly simulating several binding and unbinding events and computing the probability of finding the receptor-ligand system in the bound state, $P(RL)$, or the unbound state, $P(R + L)$. Assuming the volume change upon binding to be negligible, which is often the case at 1 atm due to the incompressibility of water, then the Gibbs free energy $\Delta G_{\text{bind},L}$ is approximately equal to the Helmholtz free energy $\Delta A_{\text{bind},L}$, and we can simulate the system in a box of volume V to obtain [61]

$$\Delta G_{\text{bind},L} \approx \Delta A_{\text{bind},L} = -k_B T\left(\ln\frac{P(RL)}{P(R+L)} + \ln\left(c^{\circ} N_{\text{Av}} V\right)\right), \quad (4)$$

where N_{Av} is the Avogadro number, and the last term corrects for the simulated concentration being different than the standard concentration. Let Γ_{bound} and Γ_{unbound} be the set of receptor-ligand configurations $\vec{q}$ that we consider bound and unbound respectively. The probability of a configuration $\vec{q}$ is given by the Boltzmann probability density function

$$P(\vec{q}) = \frac{\exp(-\beta U(\vec{q}))}{\int_{\Gamma} \exp(-\beta U(\vec{q}))d\vec{q}}, \quad (5)$$

where $\beta = (k_B T)^{-1}$ is the inverse temperature, $U(\vec{q})$ is the potential energy of configuration $\vec{q}$, and the integration is over the set of all possible configurations accessible in the simulation box volume Γ, with Γ_{bound}, $\Gamma_{\text{unbound}} \subset \Gamma$. If the simulation is long enough, we expect the fraction of configurations $\vec{q}$ found in the bound state to converge to

$$P(RL) = \int_{\Gamma_{\text{bound}}} P(\vec{q})d\vec{q} = \frac{\int_{\Gamma_{\text{bound}}} \exp(-\beta U(\vec{q}))d\vec{q}}{\int_{\Gamma} \exp(-\beta U(\vec{q}))d\vec{q}}. \tag{6}$$

After similar considerations for $P(R + L)$, we find that the ratio of visited bound and unbound conformations, in the limit of long simulations, should converge to

$$\frac{P(RL)}{P(R+L)} = \frac{\int_{\Gamma_{\text{bound}}} \exp(-\beta U(\vec{q}))d\vec{q}}{\int_{\Gamma_{\text{unbound}}} \exp(-\beta U(\vec{q}))d\vec{q}} = \frac{Z(RL)}{Z(R+L)}, \tag{7}$$

where we have defined the *configurational integral* or *configurational partition function* as $Z(\text{state}) \equiv \int_{\Gamma_{\text{state}}} \exp(-\beta U(\vec{q}))d\vec{q}$.

Simulating binding events is computationally expensive—While simulating binding events has been used to estimate binding affinities [61, 62] or to get insights into the binding pathways and kinetics of receptor-ligand systems [63–67], the computational cost of these calculations is usually dominated by the rate of dissociation, which can be on the microsecond timescale even for millimolar binders [62] and reaches the microsecond to second timescale for a typical drug [68, 69]. Depending on system size and simulation settings, common molecular dynamics software packages can reach a few hundreds of ns/day using currently available high-end GPUs [70, 71], making these type of calculations unappealing and irrelevant on a pharmaceutical drug discovery timescale. Other methods compute the free energy of binding by building potential of mean force profiles along a reaction coordinate [50, 72–74], but these methods require prior knowledge of a high-probability binding pathway, which is not easily available, especially in the prospective scenarios typical of the drug development process.

3.2 Alchemical free energy calculations yield predictions that do not require direct simulation of binding/unbinding events

In many cases, the quantity of interest is the change in binding affinity between a compound A and a related compound B (e.g., by modifying one of the drug scaffold's substituents, see (Fig. 1F)), which, by using Eq. 4 and 7 is given by

$$\begin{aligned} \Delta\Delta G_{\text{bind},\,AB} &= \Delta G_{\text{bind},\,B} - \Delta G_{\text{bind},\,A} \\ &\approx -k_B T\left(\ln\frac{Z(RB)}{Z(R+B)} - \ln\frac{Z(RA)}{Z(R+A)}\right). \end{aligned} \tag{8}$$

Note that the terms involving the standard concentration cancel out when we assume that the volume is identical for A and B. Predictions of $\Delta\Delta G_{\text{bind},AB}$ with non-alchemical methods generally require long simulations of both ligands, possibly through different binding pathways. Alchemical relative free energy calculations avoid the need to simulate binding and unbinding events by making use of the fact that the free energy is a state function and exploiting the thermodynamic cycle illustrated in Fig. 2. This is apparent after rewriting Eq. 8 as

$$\begin{aligned}\Delta\Delta G_{\text{bind, }AB} &\approx -k_BT\left(\ln\frac{Z(RB)}{Z(RA)} - \ln\frac{Z(R+B)}{Z(R+A)}\right)\\ &= -k_BT\left(\ln\frac{Z(RB)}{Z(RA)} - \ln\frac{Z(B)}{Z(A)}\right)\\ &= \Delta G_{\text{bound}} - \Delta G_{\text{unbound}},\end{aligned} \tag{9}$$

where $\Delta G_{\text{bound/unbound}}$ is the free energy of mutating A to B in the bound/unbound state. Eq. 9 and Fig. 2 tell us that the difference in free energy of binding between toluene (A) and benzyl alcohol (B) can be computed by running two independent calculations estimating the free energy cost of mutating A into B in the binding pocket (ΔG_{bound}) and in solvent ($\Delta G_{\text{unbound}}$), saving us the need to simulate the physical binding process of the two compounds. In particular, the second line of Eq. 9 is a consequence of $\Delta G_{\text{unbound}}$ being independent of the presence of the receptor in the simulation box as the definition of the unbound state assumes receptor and ligand to be at a sufficient distance for them to have no energetic interactions. Note that, when A and B have a different number of atoms, the factors $\ln\frac{Z(RB)}{Z(RA)}$ and $\ln\frac{Z(B)}{Z(A)}$ in Eq. 9 appear both to have factors with units of volume in the logarithms, but these factors exactly cancel between the terms.

How are alchemical transformations performed in practice?—In practice, the mutation of A to B is carried out by introducing one or more parameters $\vec{\lambda}$ controlling the potential energy function $U(\vec{q};\vec{\lambda})$ such that the potential of compounds A and B is recovered at two particular values $\vec{\lambda}_A$ and $\vec{\lambda}_B$. Briefly, this is achieved by simulating a "chimeric" molecule composed of enough atoms to represent both A and B. A subset of the energetic terms in $U(\vec{q};\vec{\lambda})$ is then modulated by $\vec{\lambda}$ so that at $\vec{\lambda}_A$, the atoms that form molecule A are activated and those belonging exclusively to B are non-interacting "dummy atoms", while the opposite occurs at $\vec{\lambda}_B$ (see Sec. 7.1.1 for details).

We can rigorously account for fluctuations in other thermodynamic parameters such as changes in volume V when simulating at constant pressure p or changes in number of molecules N_i of species i at constant chemical potential μ_i (e.g., number of waters or ions) by introducing the *reduced potential* [26]

$$u(\vec{q};\vec{\lambda}) \equiv \beta\left[U(\vec{q};\vec{\lambda}) + pV(\vec{q}) + \sum_i \mu_i N_i(\vec{q}) + \cdots\right]. \tag{10}$$

Here, the collection of thermodynamic and alchemical parameters $\{\beta, \vec{\lambda}, p, \mu, \ldots\}$ defines a *thermodynamic state*. In the context of alchemical calculations, in which the thermodynamic states vary only in their value of $\vec{\lambda}$, these are also referred to as *alchemical states*. The free energy of mutating A to B in any environment (ΔG_{env} e.g., binding site, solvent) can then be computed as

$$\Delta G_{\mathrm{env}} = -k_B T \ln \frac{Z\left(\vec{\lambda}_B\right)}{Z\left(\vec{\lambda}_A\right)} = -k_B T \ln \frac{\int_{\Gamma_{\mathrm{env}}} \exp\left(u\left(\vec{q}; \vec{\lambda}_B\right)\right) d\vec{q}}{\int_{\Gamma_{\mathrm{env}}} \exp\left(u\left(\vec{q}; \vec{\lambda}_A\right)\right) d\vec{q}}, \tag{11}$$

over the configurational space of the environment (Γ_{env}). While it is generally not feasible to compute the two partition functions $Z(\vec{\lambda})$, several estimators have been devised to robustly estimate the ratio of partition functions in Eq. 11 (see Sec. 8.3) from a set of configurations usually collected with MD simulations from the thermodynamic states defined at $\vec{\lambda}_A$ and $\vec{\lambda}_B$ and intermediates thereof.

Why do alchemical calculations need unphysical intermediate states?—While it is theoretically possible to estimate the ratio of partition functions from samples collected only at states $\vec{\lambda}_A$ and $\vec{\lambda}_B$, the efficiency of the free energy estimators rapidly decreases as the phase-space overlap between the two states also decreases [75, 76]. Roughly, the phase-space overlap between two thermodynamic states measures the degree to which high-probability configurations (i.e., those with very negative potential energy) in one state are also high-probability configurations in the other state (see Sec. 8.5 and Fig. 7).

Equilibrium free energy calculations, our focus here, solve the problem of having poor overlap between the states of interest by introducing multiple intermediate alchemical states at values $\vec{\lambda}_A = \vec{\lambda}_0, \vec{\lambda}_1, \cdots, \vec{\lambda}_K = \vec{\lambda}_B$ so that each pair of consecutive states $\vec{\lambda}_K, \vec{\lambda}_{K+1}$ share good overlap. Each intermediate state models a ligand that is neither A nor B but a interpolation of the two. Many estimators (e.g., exponential reweighting (EXP) [1] and Bennett's acceptance ratio (BAR) [25, 77]) can then be used to compute the free energy as

$$\Delta G_{\mathrm{env}} = k_B T \sum_{k=0}^{K-1} \Delta f\left(\vec{\lambda}_k, \vec{\lambda}_{k+1}\right), \tag{12}$$

from samples collected at all the alchemical states $\{\vec{\lambda}_k\}$, where Δf is the *unitless free energy difference*

$$\Delta f\left(\vec{\lambda}_k, \vec{\lambda}_{k+1}\right) = f\left(\vec{\lambda}_{k+1}\right) - f\left(\vec{\lambda}_k\right) = -\ln \frac{Z\left(\vec{\lambda}_{k+1}\right)}{Z\left(\vec{\lambda}_k\right)}. \tag{13}$$

While this strategy usually results in sampling thermodynamic states whose Boltzmann distributions are very similar, thus collecting information that is to some degree redundant, some estimators, such as the Multistate Bennett acceptance ratio (MBAR) [26], can exploit similarities between states to improve the precision of the estimates. This is achieved by using the configurations sampled at all alchemical states $\{\vec{\lambda}_k\}$ to compute the free energy difference $\Delta f\left(\vec{\lambda}_i, \vec{\lambda}_j\right)$ between any pair of states i, j (see Sec. 8.3).

Non-equilibrium free energy techniques provide an alternate approach to this problem, driving λ between states, but these are not our focus here. [18, 78–80]

How do absolute free energy calculations differ from relative?—While absolute and relative free energy calculations have subtle differences in their practical applications (e.g., use of restraints, handling of the standard state), the fundamental ideas and concepts of relative free energy approaches remain unaltered in other types of alchemical calculations. Absolute binding, hydration, and partition free energies still use thermodynamic cycles that enable computing transfer free energies without actually simulating the physical transfer from one environment to another.

The main difference in these approaches lies instead in the thermodynamic cycle to which this strategy is applied. For example, a typical thermodynamic cycle for an alchemical absolute binding free energy calculation is represented in Fig. 6. In this case, two independent calculations compute the free energy of removing the interactions between the ligand and its environment in solvent or in the binding site respectively through a series of intermediate states in which the energy terms are only partially deactivated.

4 What can be expected from alchemical simulations?

When starting an alchemical free energy project, a key first step is to decide whether free energy calculations are really the right tool. Particularly, count the cost of your project: Can you even hope to tackle the problem with available resources and, if successful, will it be worth it in terms of human and computational cost?

4.1 How accurate are alchemical free energy calculations?

We first note that the accuracy of any free energy calculation method will depend on the quality of the underlying force field. Therefore, any description of the force field for the molecules under study must be carefully checked to be sufficiently accurate to experiment. In particular, one must make sure that if using automatically generated molecular descriptions from either one's own workflow or some other computational chemistry program, there are no obvious problems in these files either through errors in the workflow, or lack of chemical coverage in the data used to construct the molecular description. Torsional parameters, in particular can be misassigned or improperly parameterized.

Alchemical free energy calculations involving small molecules seem to achieve, in favorable cases, root mean square (RMS) errors relative to experiment around 1–2 kcal/mol depending on force field, system, and a variety of other factors such as simulation time, sampling method, and whether the calculations employed are absolute or relative. A small selection of example datasets and case studies can be found in Sec. 11 at the end of this document. However, the domain of applicability is a significant concern [38, 39], especially for relative calculations, which typically require a high quality and usually experimental bound structure of a closely related ligand as a starting point. Additional factors such as slow protein or ligand rearrangements, uncertainties in ligand binding mode, or charged ligands can make these calculations far less reliable and more of a research effort.

It is worth noting that the accuracy of free energy calculations is highly variable across different protein targets, and likely across different ligand chemotypes as well. For instance, FEP+ with OPLS3 achieves an RMSE of 0.62 kcal/mol for a set of 21 compounds binding to JNK1 kinase, but an RMSE of 1.05 kcal/mol for a set of 34 compounds binding to P38α kinase [81]. Furthermore, perturbations for the same chemotype in different pockets of the BACE enzyme gave varied errors [82]. Here the errors refer to the difference in ΔG derived from calculated $\Delta\Delta G$'s while fitting a constant offset to best reproduce the experimental binding free energies for known compounds [15]. Each $\Delta\Delta G$ is associated with a particular free energy calculation or transformation, which can be thought of as an edge in the graph spanning the compound series, see examples of such graphs in Fig. 5.

The fact that we can analyze both ΔG values and $\Delta\Delta G$ values raises an important question about analysis – which calculated values should we assess? It is important to be clear on what error to report: ΔG after shifting by a constant to minimize the RMSE, unshifted ΔG, $\Delta\Delta G$ of computed edges, or $\Delta\Delta G$ of all edges. (See recommendations for reporting best practices, Sec. 8.7.) Additionally, as it is possible to perform calculations on a set of ligands using different pairwise comparisons of molecules, the performance of the method may be biased based on which pairs of comparisons are performed. Additionally, it is possible that the error associated with the relative free energy between a two ligands that was not directly computed, but can be deduced using one or more thermodynamic paths involving other ligands will likely be more uncertain. Given the need to understand the performance of the system with alchemical free energy calculations, we recommend that retrospective studies for a particular target and a particular chemical series be performed for each application case.

4.2 How reproducible are alchemical free energy calculations?

We restrict our analysis here to repetitions of the same calculation performed with precisely the same force field, as different force fields used to describe the same molecule can lead to wide differences in free energies in some cases. Even within this restriction, finite computing resources necessarily limit the generated number of uncorrelated samples of potential energy surfaces, and therefore alchemical free energy calculations only give free energy estimates to within finite precision. An important consideration is how reproducible alchemical free energy calculations are in practice. In simple cases such as absolute hydration free energies of small organic molecules, or relative hydration free energy calculations between structurally similar small organic molecules, it should be possible to obtain highly precise estimates with a given software package, i.e.with a sample standard deviation under 0.01 kcal/mol [83]. For more complex use cases such as protein-ligand binding free energies the repeatability is often substantially worse [83]. A good practice is to perform two or three runs of the same perturbation to assess precision with a given protocol, using different initial velocities. The sample standard deviation will give a crude estimate of the reliability of the estimates, and whether the precision is sufficient for the problem at hand. When practical, a more stringent test is to use different input coordinates for each repeat run as well as different velocities.

Note that these types of statistical differences concern calculations carried out with a single software package, but simulation package variations can introduce additional discrepancies. Such issues of reproducibility of free energy calculations across different simulation packages have attracted attention recently [51, 83]. Greater variability is expected between packages due to methodological differences such as integrators, thermostats, barostats, treatment of long-range electrostatics, and potentially other factors. For absolute and relative hydration free energies of small organic molecules a variability of ca. 0.2 kcal/mol between popular simulation packages has been reported [51]. In the recent SAMPL6 SAMPLing challenge a larger variability of 0.3 to 1.0 kcal/mol was noted in the computed absolute binding free energies of host/guest systems even though the study sought to use identical input and simulation parameters [83] and, in many cases, single-point energies were identical or nearly so. Further work is needed to ensure reproducibility of alchemical free energy calculations across different software implementations to guarantee that force-field development efforts lead to transferable potential energy functions.

4.3 Is my problem suitable for alchemical free energy calculations?

Before even planning free energy calculations to study binding to a particular target, it is important to assess what is known about the system and its timescales and its suitability for free energy calculations, as well as the *purpose* of the calculations and the amount of available computer resources. In some cases, predicting accurate binding free energies for a particular target might be *more* challenging than simply measuring them! This is often the case when dealing with database screening problems, where compounds might be easily and quickly available commercially for testing and free energy calculations could consume far more resources. Free energy calculations thus typically only appeal when (slow or costly) synthesis would be required or experiments are otherwise cost-prohibitive.

Sometimes, however, free energy calculations can provide answers that are not readily available from experiments. For example, type II kinase inhibitors selectively bind to different kinases in the so-called DFG-out conformations [84]. The selectivity of such inhibitors may be attributed either to their differential binding to different kinases in the DFG-out conformations, or to different stability of the DFG-out conformations of different kinases.

Let K_C be the equilibrium constant between DFG-in and DFG-out conformations of one kinase, and K_D^* be the dissociation constant of a type II inhibitor against this kinase, the apparent binding constant of this inhibitor against this kinase is then

$$K_D = K_D^* \frac{1 + K_C}{K_C} \tag{14}$$

Since binding experiments cannot resolve K_D^* and K_C individually, such experiments cannot address the basis of selectivity of the type II inhibitors. Absolute binding free energy calculations, in contrast, can take advantage of the slow kinetics of DFG-in/out conversion, and estimate the conformation-specific binding constant K_D^*, thus yielding clues as to the source of selectivity.

4.4 Is the expected accuracy of the computation sufficient?

The requisite level of accuracy is another important consideration. If the goal is to guide lead optimization when many compounds will be synthesized, free energy calculations can be appealing even with accuracies in the 1–2 kcal/mol range [85], but if the number of compounds to be synthesized is very small, this accuracy may not be enough to provide much value.

Here we provide a simple estimate of the value provided by alchemical free energy calculations in lead optimization. Let $P(\Delta\Delta G)$ be the probability distribution of the changes in the binding free energies of a new set of molecules during one round of lead optimization, and let $P(\Delta\Delta G^{\dagger}|\Delta\Delta G)$ be the conditional probability of the binding free energy change computed by the free energy calculations, $\Delta\Delta G^{\dagger}$, given the actual change $\Delta\Delta G$. The latter conditional probability can be modeled by a normal distribution

$$P\left(\Delta\Delta G^{\dagger} \mid \Delta\Delta G\right) = \frac{1}{\sqrt{2\pi\sigma^2}}\exp\left(-\frac{\left(\Delta\Delta G^{\dagger} - \Delta\Delta G\right)^2}{2\sigma^2}\right), \tag{15}$$

where σ signifies the accuracy of free energy calculations. Here we assume that there is no systematic bias in the free energy calculations, i.e., on average, the free energy change computed by free energy calculations agrees with the actual free energy change. Additional analysis of this type is presented in Brown et al. [86]

In lead optimization guided by free energy calculations, we will likely only synthesize and experimentally test molecules that are predicted to have favorable free energy changes. We are thus interested in how often a molecule predicted to bind stronger actually turns out to bind stronger. In other words, we are interested in the conditional probability:

$$P\left(\Delta\Delta G < 0 \mid \Delta\Delta G^{\dagger} < 0\right). \tag{16}$$

For illustrative purposes, consider a proposed set of new molecules, and assume that the changes proposed in these molecules yield a set of relative binding free energies that follow a normal distribution. That is, assume that the standard deviation in the relative binding free energies for the changes represented is $RT \ln 5$ (corresponding to a 5-fold change in the binding affinities), and that 1 in 10 new molecules have increased binding affinity ($\Delta\Delta G \leq 0$). Under such assumptions, the conditional probability in Eq. 16 can be easily computed.

If the accuracy of a collection of free energy calculations is $\sigma = 1$ kcal/mol, $P(\Delta\Delta G < 0|\Delta\Delta G^{\dagger} < 0) = 0.35$, which means that out of every 10 molecules selected for predicted favorable free energy change, on average 3.5 molecules will have actual favorable free energy change. In other words, selection by free energy calculations yields 3.5 times more molecules of improved affinities than selection without free energy calculations under these assumptions.

Available computational resources and timescales of motion also factor into this initial analysis. An individual free energy calculation involves simulations at many different intermediate states (perhaps 20–40 or more) and each of these must typically be long enough to capture the relevant motions in the system. If such motions are microsecond events or longer, the computational cost of running 20–40 microsecond or longer simulations for each of *N* ligands will likely be prohibitive for most users with today's hardware. On the other hand, if key motions are fast and minimal (as is often assumed in practice), much shorter simulations may be sufficient.

4.5 Can I afford the calculation?

Furthermore, are available computational resources sufficient that throughput will be reasonable compared to needs of experimental collaborators working on this system? How many ligands (*N*) can you afford to handle given your computational resources? As cloud computing becomes more available, in-house GPU clusters may not be necessary if calculations are not run on a regular basis. This analysis should be done up front as part of "counting the cost" of involvement in a particular project. In some cases, the analysis may conclude that free energy calculations will not be feasible for the proposed problem. Here, by "cost", we refer not just to financial cost of the calculations relative to experiments, but also time – can the calculations be run faster than experiments are done? How will the relevant resource and opportunity costs factor in? Both computation and experiment require human time, supplies (of different sorts), and equipment. In the extreme limit, for example, it would not make sense to spend a month running a binding free energy calculation if the equivalent experiment could be done in a day with resources already on hand. Such issues should be considered before deciding to conduct binding free energy calculations.

4.6 Is an exploratory study what I want?

An additional consideration is how much is known about your particular target, ligand binding modes in the target, and any relevant motions – essentially, has it been studied enough to know whether it might be suitable for free energy calculations? It is important to know if the system has hardly been studied, because should the initial calculations perform poorly, the effort may turn into an attempt to understand the relevant sampling, force field, or system preparation problems.

If you are unsure whether your project is feasible, as mentioned above, one recommended option is to conduct a short exploratory study to assess tractability for a small number of ligands. This can be sufficient to get an initial idea of feasibility and accuracy of the calculations for the proposed target [37].

5 How should alchemical simulations be applied to drug discovery?

Many practitioners expect alchemical methods to provide valuable guidance for drug discovery, and to exhibit accuracy superior to most alternative approaches for suitable targets [87]. Successful application in industry may require considerable knowledge of the "domain of applicability" of free energy calculations – where they work well and where they will not [39]. Successful application also requires robust protocols for preparing, submitting and analysing alchemical calculations. In this regard, the issues mentioned in the previous section such as understanding the suitability and timescales to capture the structure activity relationships (SAR), and performing up-front tests of performance are all relevant to drug discovery applications. Without venturing too far into details of system setup, which is beyond the scope of this article, we highlight some critical factors affecting accuracy and successful application.

5.1 Capturing experimental conditions

The calculations aim to capture the alchemical change from one ligand to another as accurately as possible. Therefore, it is necessary to consider details of the experimental setup, such as pH. Biological assays are usually run at neutral pH but this is not always the case. For example, some enzymes exhibit pH-dependent activity and assays may thus be done in conditions other than neutral pH. Therefore, computational protein and ligand preparation protocols should reflect experimental pH.

The formal charge and/or tautomeric state of the small molecules can change within a series of analogs, necessitating care in treatment. Additionally, medicinal chemistry efforts might deliberately modify the pKa of a series to modify drug properties, requiring explicit efforts to incorporate these changes into alchemical calculations.

To ensure modeling matches experiment, we also need to accurately prepare and simulate the same system – which requires understanding what protein construct is used in the bioassay. For instance, does the X-ray structure that is to be used for the calculations match the construct used for screening (i.e. only the catalytic domain vs. full length, monomer vs. dimer, etc.) [88]? Also, were certain co-factors or partner proteins required in the bioassay?

5.2 Is my binding mode accurate?

As also mentioned, good performance of alchemical calculations requires an accurate representation of the ligand binding mode, usually from a high quality X-ray crystal structure. If more than one structure is available, the modeler should pay attention to choose the most suitable. The quality of the structure can be a concern, and the reader is referred to work of Warren et al. for a detailed discussion of choosing optimal structures for structure-based modeling [89].

It is also useful to study the structure activity relationship and understand the expected impact of any mutations on the binding site, such as whether side chain movement in the protein will be required, and whether there is evidence of this in any alternative X-ray structures of the same protein. Often, only one protein and water configuration is used for a series of alchemical calculations, so this needs to be capable of accommodating the smallest through to largest ligands in a way that allows stable and well behaved simulations. This can provide a practical limit on the alchemical changes that are feasible, though a simple work-around can be to separate compounds into sub-series for different calculations.

If multiple structures are available there is some evidence the higher affinity complex can give better match to experiment [90], at least in some cases. However, ligands and proteins can also undergo unexpected changes in binding mode for related ligands, which can make these issues more complex to deal with [16].

5.3 Input setup and scale of calculations

In a drug discovery setting it is normal to consider dozens (or more) of ligands and it is necessary to align them in the binding site. There is no detailed study of how different alignment approaches may affect results, but the user should be aware of some practical considerations. Tools are available to compare the ligands and build the combined topologies that define the changes between one ligand and another [34, 91, 92]. In simple terms, providing poor alignment to these tools will make this job harder. Docking with restraints is often beneficial in this regard. Particularly, fixing the 3D spatial position of the scaffold using maximal common substructure (MCSS) restrained docking can help provide well aligned input for the topology generation. Nevertheless, in this case careful attention is still needed to ensure consistency of alignment for identical substituents. Another alternative is to manually edit the same core and add/modify the changing substituents. This provides assurances that coordinates for the non-perturbed portion of the structure remain identical and aromatic substituents, for instance, have consistent dihedral angles. However, it is not feasible for many compounds and therefore automation is desirable.

Finally, the role of water in ligand binding is not always well understood and it can be crucial to capture the changes in binding site solvation during ligand binding. Can crystallographic waters be retained? Do they clash with some of the larger ligands used in the alchemical perturbation? See Sec. 6.1 for different strategies that can be applied to dealing with waters. Generally, before launching large numbers of alchemical free energy calculations it is always recommended to test the system using classical MD simulations and limited numbers of alchemical perturbations. Metrics such as ligand and protein RMSD and RMSF can be inspected, along with visual inspection of simulations, to ensure the system is stable and likely to be suitable for alchemical calculations.

Running binding free energy calculations in a drug discovery application will typically require the use of software or tools to facilitate the large number of calculations. Commercial implementations such as FEP+, OpenEye Tools, or Flare allow for a fast setup and deployment to GPU hardware in minutes, but may have limited ability to customize calculations [15, 19]. Commercial tools can be expensive in some cases, but non-commercial tools are becoming more straight forward to use to run alchemical free energy calculations [17–19, 34, 91–93].

For relative free energy calculations, various graph topologies or maps of calculations are possible, and choices may depend on the target application. For instance, if the goal is to accurately assess the relative binding energy of a small number of compounds, possibly with challenging syntheses, the map of perturbations should contain as many connections between compounds as affordable. However, when running calculations on hundreds of compounds a so called *star-map* (see Fig. 5A) can be used that just contains one connection per compound: perturbing every compound to a central ligand, typically the crystal structure ligand [94]. In this way the top-ranking examples can be readily identified and submitted to additional calculations in a second round. Alternatively, if the goal is to achieve the smallest possible error with minimal computational expense, certain graph topologies provide benefits [95, 96]

5.4 Making predictions, understanding errors

For prospective drug discovery applications there are several other considerations including understanding likely errors and taking selection bias into account.

It is crucial when proposing compounds for synthesis to have some idea of the underlying error or uncertainty in the predictions. A retrospective assessment can give an indication of prospective performance for similar molecules [97]. Beyond this, several parameters provide useful indicators of performance. For example error estimates provided by free energy estimators that are too large can highlight poorly converged simulations [90]. Hysteresis, either within cycles in the perturbation network or between forward and backward perturbations can be checked [98] to indicate problematic perturbations involved in cycles connecting many compounds (See also Secs. 7.1.1 and 8.5). Once synthesis and testing of compounds is complete a standard strategy is to look back at how the calculations performed. In this regard it is important to consider the issue of selection bias upfront. It is tempting to only synthesize the compounds predicted to be most active, thus a narrow range of calculated activity is tested that imposes limits on the statistical assessment of performance, ideally example molecules from across the range of predicted activity can be assessed or corrections can be applied based on previous recommendations [99]. For a more detailed discussion on checking the robustness of your alchemical free energy calculation see also Sec. 8.5.

In summary, the successful use of alchemical calculations, particularly for drug discovery, requires working in the domain of applicability, using a high quality X-ray structure of the target bound to compounds in the series, and testing the approach retrospectively to ensure the system setup is well-behaved. Always assess your confidence in the resulting predictions and communicate this when discussing with experimentalists. Consider performing repeat calculations for at least some of the perturbations in the study. There are many accounts of success of alchemical calculations, the methods show good performance towards the goal of binding free energy prediction. However, it is important to have realistic expectations.

Structure based drug design projects are often capable of improving potency relatively quickly, even with only limited application of computational approaches and the range of activity narrows to just two-to-three log units. It may seem hard to have impact with substantially different, more potent, stand-out compounds in this scenario, but binding free energy predictions can still be extremely useful for ensuring activity is maintained as other properties are optimized. An interesting cost benefit analysis has shown the value of activity prediction, see discussion above and articles such as [85]. From a drug discovery point of view, alchemical calculations are expanding their domain of applicability, and there are reports of success using homology models [100] and GPCRs [101, 102] for instance, as well as enabling charge change and scaffold hopping [103, 104], but these systems are undoubtedly more difficult. In the meantime, use cases are expanding to resistance prediction, selectivity prediction, solubility prediction – an exciting future for alchemical calculations [6, 105, 106].

6 Simulation prerequisites

Alchemical free energy protocols as discussed below (Sec. 7) are defined for a specific type of free energy calculation, i.e. a free energy of binding or a free energy of hydration. Different types of simulations require different choices for ligands, solvent, and host molecules (in the case of the estimation of free energies of binding).

6.1 Free energies of binding

In principle, in the limit of sufficient configurational sampling, the free energy changes estimated from an alchemical free energy calculation should be independent of the system's initial coordinates. However, in practice, because simulations are of finite duration (typically 1–100 ns per state at present), this is only true for certain classes of alchemical free energy calculations such as relative or absolute free energies of hydration of small and relatively rigid organic molecules. Protein-ligand complexes typically exhibit slowly relaxing degrees of freedom that significantly exceed the duration of an alchemical free energy calculation, and host-guest calculations can be susceptible to these issues as well, depending on timescale and system. It is therefore generally important to carefully select input coordinates to obtain satisfactory results. The following questions may be relevant before diving into the simulation setup.

- Do I have one or multiple good receptor structures? (e.g. a good resolution X-ray crystal of the protein target)
- Do I have information on one or all of the ligand binding sites? (e.g. an X-ray structure)
- Should I include buried waters, or other small molecules that can be found in an X-ray structure?
- Are my ligands part of a congeneric series? (i.e. simple R group substitutions around the same scaffold)

Are there good X-ray structures available?—As with any simulation, care should be taken in selecting available X-ray structures in the Protein DataBank [107]. In some cases it may be wise to choose multiple starting structures to account for variability in receptor conformations as well as the accuracy of available X-ray structures. Typically, clustering of receptor structures can be used to identify different receptor conformations near the binding site, as well as assessing relevant side chain placements from the X-ray structure, see for example [16]. In terms of set up and other choices, following general best practice guidelines is advisable [41].

Many free energy calculations focus on a congeneric series of ligands, which can make these calculations suitable for relative free energy protocols (see Sec. 7). For relative calculations, some care has to be taken selecting binding poses for these ligands. Generally, a common assumption for a congeneric series is that the binding mode is conserved. Therefore, if an X-ray structure of one of the ligands is available, this should be used to position the ligands in the putative binding site in an energetically reasonable conformation without steric or electrostatic mismatch with the receptor. Checking the X-ray structure versus the experimental electron densities is important, as the position of part of the ligand or important sidechains may be based on the interpretation of the crystallographer rather than the available electron density, especially in cases of missing density. For example, looking at a cyclohexane ring density, a chair configuration is vastly more likely than that of a boat and, if a boat configuration is present in the structure, it may be worth inspecting the density to ensure it adequately supports this choice.

Are you prepared to deal with any binding mode challenges?—Generally, binding modes within congeneric series are conserved [108], however, exceptions exist [109, 110], as discussed in more detail in Sec. 7.2.6. Certain functional groups may be particularly prone to this due to symmetries or near symmetries. One such issue involves a 180 degree flip in the dihedral angle of an aromatic ring, or five-membered ring leading to a different spatial position of ortho- or meta- substituents that otherwise should overlap within a series. The 180 degree flip of the ring may not occur enough during simulations (due to steric obstructions) to overcome bias due to the starting configuration. Another scenario may be equatorial and axially substituted saturated rings (e.g. cyclohexane derivatives). This situation may be addressed by explicitly modelling different binding modes of the same ligand and combining later computed free energy differences for different binding modes into a relative free energies of binding [111].

Have you considered stereoisomers and enantiomers?—Congeneric series can contain stereoisomers or enantiomers which can bind very differently, resulting in large errors if treated incorrectly. For racemates, the relative abundance of each stereoisomer is normally not known. Therefore, the experimental activity associated with just one stereoisomer/enantiomer is more uncertain. However, the modeling typically uses just the bioactive conformation that best fits the active site. Clearly this introduces potential for larger errors compared to experiment. Nevertheless, if all compounds in the congeneric series are racemic, originating from similar synthetic procedures with an expected similar abundance of stereoisomers, then the differences may cancel and the trend in calculated and observed binding energies may be robust. Despite this, we can see that care and further testing is needed in this scenario, and the quality of the predictions may suffer. Additionally, unexpected changes in what stereoisomer binds experimentally, if they occur, could pose significant challenges for modelling efforts.

Conserved binding site waters can play an important role in binding free energies—Binding site water molecules may form water mediated protein-ligand interactions which can pose challenges whenever exchange with bulk water is slow compared to simulation timescales. This happens typically in buried binding sites. Overlaying multiple protein X-ray structures can identify conserved or additional water molecules that can be useful to include in calculations. In cases where water molecules are known to play an important role in the binding, software implementations that use water sampling facilitated by Grand Canonical Monte Carlo methods may be useful [112]. Other tools such as WaterMap or open source equivalents (SSTMap, GIST, and others) can be used to define water structure for systems with no experimental evidence of water sites [113]. Well-known protein systems with water mediated ligand interactions are for example:

HSP90 which formed part of the D3R grand challenge 2015 [16], A2A [114], MUP [115], [101], and others [116].

Protonation states depend on the pH of the experimental assay—Care should be taken when preparing ligands and proteins to match the pH of the experimental assay, if known. As mentioned above in Sec. 5.1, the pH of the assay can differ from neutral pH and will determine the protonation states of the proteins and ligands. Since the pKa of reference amino acid sidechain residues is known, but can vary in the protein environment, many different tools have emerged for predicting sidechain pKa in proteins, such as the H++ server, ProPKa, APBS, and Maestro [117–120]. Strongly acidic (Glu, Asp) or basic (Arg, Lys) sidechains can reliably be predicted to be ionized, but care is still needed as the local environment can modify expected ionization states, such as the catalytic Asp dyad in proteases. Histidine is notoriously more difficult to predict as its pKa suggests it ionizes closer to the experimental pH range. For ligands, often the pKa needs to be determined, if it is not known experimentally. There are many different available tools for this purpose, but common choices may be propKa [118, 121], Chemicalize (https://chemicalize.com/welcome), or Maestro [120]. Still, accurate pKa prediction for small molecules remains a challenging problem, even with dedicated tools [122]. While often it can be assumed that the protonation state of a ligand and protein will remain the same as the ligand binds, some care needs to be taken with systems where the protonation state may change upon binding [123]. BACE [124], for example, famously undergoes a protonation state change on ligand binding.

Congeneric series often need alignment—Input coordinates for a congeneric series may be generated by docking calculations, or by ligand alignment using MCSS algorithms. The latter tends to produce alignments that are more conserved and more consistent free energy changes across a dataset, but will struggle to yield reasonable results for relative binding free energy calculations that involve a significant binding mode rearrangement. This may also lead to steric clashes with the receptor coordinates of the reference ligand if structural rearrangements are needed to accommodate different members of the congeneric series. Small steric clashes may be resolved during subsequent simulation equilibration prior to data collection, but there is a risk that the complex relaxes to an alternative metastable state.

An additional consideration arises for single topology relative free energy calculations. In this class of alchemical free energy calculations it is necessary to generate a molecular topology that may describe the initial and final states of the perturbation (see Fig. 3). In cases where the end states have high topological similarity and high structural overlap this is relatively straightforward and typically handled by use of MCSS calculations. In situations where the end state topologies differ significantly, or where there is relatively little spatial overlap between the two end states, some user intervention may be necessary to produce a satisfactory input topology.

If the binding site location is uncertain but the structure of the receptor is well defined and plausible binding sites are identified, it may be more useful to choose an absolute free energy protocol to compute the standard free energy of binding of the ligand to a set of binding sites. This requires the user to prepare input files describing the bound conformation in different putative binding sites [125]. The apparent binding free energy of the ligand may be obtained by combining the individual binding site free energies, which also indicate where the ligand is more likely to bind. In this case a docking program can generate initial structures. Different commercial and non-commercial tools are available, such as rDock [126], Autodock Vina [127], Glide [128], or Flare, to name a few [19].

If the putative binding sites are not apparent, for instance due to significant induced-fit effects, it may be challenging to obtain meaningful free energies of binding. One may have to account for the free energy cost of forming a binding site in the target receptor which may not be feasible on alchemical simulation timescales.

6.2 Free energies of hydration or partition coefficients

Preliminary considerations necessary for using free energy methods to compute partition coefficients are generally more straight forward. For example, a 3D minimised structure of a solute can be generated with a simple tool such as openBabel and solvated to prepare the input to compute a free energy of hydration [129]. However, in these cases a careful choice of forcefield for the organic solvent model, as well as water model is essential. See for example [3, 4] for a good discussion of these choices. And, while sampling problems might seem to be a non-issue for small molecules, this is not always the case; e.g. even the hydroxyl orientation on neutral carboxylic acids can occasionally pose a challenge [130, 131].

7 What simulation protocol should I choose?

Alchemical free energy calculations can be grouped into two main categories, “absolute” (see Fig. 6) and “relative”[2] (see Fig. 2), which differ in whether they compute properties for a single molecule (absolute) or compare properties of different, usually closely related, molecules (relative). To use binding as a concrete example, in absolute binding free energy calculations, we compute the binding free energy of a ligand to an individual receptor relative to a standard reference concentration. In contrast, in relative binding free energy calculations, we compare the binding free energy of two related ligands to determine the potency difference.

7.1 Absolute and relative free energy calculations have important differences

Many of the issues around simulation setup and protocol choice for alchemical calculations are common, but there are some differences between absolute and relative calculations. We will consider protocol differences before treating the common elements.

7.1.1 Choices unique to relative free energy calculations

Topologies: A critical first step is to determine which approach to use for atom mapping between end state molecules. Often this is predetermined by the choice of simulation software. Typically it is possible to chose between a *dual topology* versus a *single topology*

[2] The distinction is a bit of a misnomer, since both compute ratios of partition functions relative to another state and in that sense are relative, while neither computes an absolute free energy.

The distinction between single and dual topology can be illustrated by considering a hypothetical transformation from molecule A to molecule B, where both atoms share a common substructure but differ in their substituents; in particular, consider a transformation of benzene to benzyl alcohol shown in Fig. 3. In this case the common substructure between the two molecules is the benzene ring, though in practice substructure may be selected to be larger depending the mapping chosen, as we discuss below.

In single topology calculations, the overall transformation is set up to involve as few additional atoms as possible, so benzene would be typically changed into benzyl alcohol by first changing one of the hydrogens into a carbon. The site of this transformation will also be the future home of two additional hydrogen atoms bound to the new carbon, so these must initially be present as non-interacting atoms called “dummy atoms”, which retain their bonded interactions but do not interact with the rest of the system. Bond parameters as well as partial charges between the changing atoms are adjusted accordingly between the initial and final state. Thus, in a single topology calculation, atoms may change their type, ensuring minimal dummy atoms are created. This is illustrated in the left arm of Fig. 3.

In contrast, in a dual topology alchemical free energy calculation, *no atoms are allowed to change type* [44, 132]. This means that the benzene to benzyl alcohol transformation involves starting with benzene plus the non-interacting dummy atoms making up the hydroxymethyl group, then passing through an intermediate state where some atoms are partially interacting—particularly, those atoms which are becoming dummy atoms or ceasing to be dummy atoms [133]. The transformation finally culminates in a state where benzyl alcohol is present along with the additional dummy atom which was previously a corresponding hydrogen of the benzene. Fig. 3‘s right branch depicts how such a dual topology works.

As far as we are aware, the distinction between single and dual topology approaches is made primarily for two main reasons. First, this choice affects the alchemical pathway followed, and thus may affect convergence properties—though we are not yet aware of a study of the relative efficiency and merits of these two approaches. Additionally, historically, some simulation packages implemented only one approach and not the other, meaning that the distinction was functional.

Some additional terms have also been employed to talk about these different intermediate pathways. Particularly, some studies refer to a “hybrid” topology approach to free energy calculations [10, 18, 92], though this term may not yet have achieved widespread use. In this case, “hybrid” seems to indicate that the set up of these free energy calculations involves a hybrid of the two molecules, and much of what is done in these studies uses a single topology approach [18].

One final approach, so far in its infancy, has been called “separated topologies” and essentially consist of two absolute free energy calculations in opposite directions at the same time, turning one molecule’s interactions with the environment off, while turning the other molecule’s interaction on [134, 135].

Software packages vary in their use of single or dual topology approaches; for example, AMBER TI uses a dual topology approach, while BioSimSpace uses a single topology approach. Please make sure to check what approach is used with your software package of choice, or whether it supports your choice of approach (GROMACS and GROMOS, for example, support both). To our knowledge, efficiency differences have not been thoroughly explored, though conventional wisdom suggests that fewer dummy atoms are better, as introducing or removing atomic sites is usually more difficult, requiring more intermediate steps [85, 136].

Atom mapping: Once a particular approach to the topology is selected, a crucial next step is to identify the common atoms which will not be perturbed. Rigorously, this process typically comprises a MCSS search of the molecules involved to identify the common substructure—though the parameters of the MCSS search will differ depending on whether single or dual topology calculations are planned. Specifically, with a single topology approach in mind, atom types are allowed to change, so a permissive MCSS search can be done, whereas with dual topology a more strict search is required.

There are different tools that allow the generation of MCSS matches as well as single topology input. A large number of software tools can compute MCSS matches using different cheminformatics packages. Some rely on RDKit [137], such as pmx [92], LOMAP [136], FESetup [91] and partially BioSimSpace [34], while others such as fkckombu [138] are standalone tools. Schrödinger's FEP+ planning tool was originally based on a version of LOMAP, and it also uses MCSS matching as well as 3D considerations to plan the network of single topology calculations between molecules [15].

MCSS searches can be relatively time consuming, so if the goal is to assess a library of ligands to identify promising pairs for relative calculations, it can be helpful to use faster approaches such as shape similarity to perform an initial similarity assessment and then use MCSS only to identify final mappings for relative calculations [139–141]. The MCSS approach, though relatively standard, takes into account only topological similarity. It is possible that changes in binding mode could actually require a different choice of mapping, so in some cases mappings may need to be planned differently depending on 3D positioning of atoms in space. Visual inspection prior to simulation is recommended to ensure that the mapping criteria correspond to the expected binding mode. If the mapping protocol returns simulations that correspond to different binding modes of a ligand within a perturbation map, this can cause large hysteresis.

Single topology relative calculations and calculations based on substructure searches only work if in fact the ligands share a common substructure, e.g. are part of a congeneric series, see Fig. 4. If no common substructure is shared, then alternative dual or separated topology free energy calculations are needed. One would co-localize a pair of compounds in a binding site, exclude their interactions with one another, and compute the relative binding free energy by turning one molecule on from being dummy atoms while turning the other off. To our knowledge no general pipeline for such calculations yet exists and this would likely remain a research problem. Using an absolute free energy approach instead seems more promising in such a case.

Ring breaking and forming.: Relative free energy calculations for ring breaking and forming are particularly challenging/problematic (see Fig. 4B), in part because relative calculations rely on the free energy contributions of dummy atoms canceling between different legs of the thermodynamic cycle, which may not be true whenever dummy atoms are involved in rings [142]. Some approaches have attempted to address this [143] but a general solution is not yet in mainstream use. Still, FEP+ implements one solution.

Perturbation maps: Based on the input ligand series, a perturbation map or network can be planned. Recent heuristics have shown the more connected the perturbation network the better. However, there is a way to optimize network structure while minimizing the number of perturbations that need to be computed reducing the resulting computational cost [95, 96]. Sometimes the introduction of intermediates that are not part of the original congeneric series are essential to avoid ring breaking, or to deal with perturbations that would otherwise result in large numbers of atoms being inserted or deleted. Some commercial tools have good underlying heuristics but may fail with complicated input, needing user validation in particular when dealing with chiral compounds.

In some cases, during the lead optimization stage, or for very large datasets that would benefit from rougher initial free energy ranking, or in cases where perturbations would be rather large, a star shaped network as seen in Fig. 5 A is used. However, adding redundancy into the network means that a better error analysis can be carried out by looking at cycle closure errors as discussed in sec. 8.5, with an example given in Fig. 5B.

Methods from experimental design have been applied to the construction of the perturbation maps. Yang et al. [95] showed how to optimize the perturbation map by selecting a fixed number of calculations from the pairwise perturbations so that the resulting set of calculations minimize the total variance. Xu [96] showed how to optimize the perturbation map by allocating different amounts of simulation time to different pairwise perturbations so as to minimize the total variance, given the total simulation time of all the perturbation calculations. Both approaches lead to substantial reduction in the statistical error of the estimated free energies.

Constraints and relative free energy calculations: One issue which requires particular care is the use of constraints. Commonly, bonds involving hydrogen are constrained to a fixed length using algorithms such as SHAKE or LINCS, allowing the use of longer timesteps [144]. However, in single topology relative free energy calculations, the atoms involved might be mutated to other atom types—for example, in a mutation of methane to methanol, one hydrogen might become an oxygen atom. The bonds with such atoms might not have any constraints, or if all bond are constrained, would have constraints of different lengths. Some molecular dynamics engines are not set up to recognize this change, or at least not to correctly include contributions to the free energy from changing constraints/-constraint length, so results for a transformation can be erroneous. At present the most general solution to this problem is simply to avoid the use of constraints (and thus use a smaller timestep if necessary, usually of around 1 fs) in any relative free energy calculation involving a transformation of a constrained bond. Individual software programs and settings can handle such issues. For example, bonds can be transformed in both GROMACS and GROMOS, because contributions of LINCS (GROMACS) and SHAKE (GROMOS) constrained bonds are added to the $\frac{dH}{d\lambda}$ term [145–148]. However, the constraints are not taken into account when calculating energy differences to other intermediate states as the effect is entropic, not energetic. User manuals should carefully checked for how these effects are included if a constrained bond changes in length. We do note that if one performs calculations of the changing constraint in both solution and in the binding site, then the errors often cancel substantially, but the cancellation is not guaranteed.

7.1.2 Absolute free energy calculations must handle the standard state and use restraints—Absolute free energy calculations involve completely removing the interactions between the ligand or solute and its environment, taking it to a non-interacting state that may or may not retain intramolecular non-bonded interactions. This non-interacting state can then be shifted between environments—from the protein to water, or from one solution to another—without changing its free energy other than that due to the changing volume of the simulations, and then interactions can be restored in the new environment.

Absolute free energies are by definition reported with respect to a specific reference or standard state, which effectively determines the arbitrary point at which the free energy is 0. The role of the standard state is particularly evident from the expression of the binding free energy between a receptor R and ligand L

$$\Delta G = -k_BT\ln\left(c^\circ K_b\right) = -k_BT\ln\left(c^\circ\frac{[RL]}{[L][R]}\right). \quad (17)$$

Here, the reference state concentration c° converts the binding constant K_b into a dimensionless quantity expressed in reference concentration units. It should be noted that ignoring the term c° is equivalent to assuming a reference concentration of 1 D^{-1}, where D are the units used to express K_b and would thus cause the value of ΔG to vary with the choice of the units. It is convenient to define a standard state at a constant pressure of 1 atm and where each chemical species (i.e., R, L, and RL) in the reaction solvent has a concentration of c° = 1 M = 1 molecule/1660 Å^3 but do not interact with other molecules of R, L, or RL.

<u>Handling the standard state in absolute free energy calculations.:</u> For solvation free energy calculations, handling the standard state is typically straightforward, and treating it correctly simply means ensuring that the non-interacting solute still occupies essentially the same volume as the solute in the interacting system. So typically in such cases no special care is required to ensure the correct standard state, as long as the *experimental* data being analyzed uses the same standard state. If this is not the case, a simple entropic correction to the free energy of $k_BT\ln(V_f/V_i) = k_BT\ln(C_i/C_f)$ to the experimental data is needed.

For binding, however, the situation is more complex and requires special care. Because the simulations are typically performed using restraints and at concentrations that are different from 1 M, the expression of the free energy requires the following correction [60] (see an example of such a thermodynamic cycle in Fig. 6)

$$\Delta G^{\circ}_{\mathrm{restr}} = -k_B T \ln\left(c^{\circ} V_L\right) - k_B T \ln\left(\frac{\xi_L}{8\pi^2}\right), \tag{18}$$

where V_L and ξ_L are respectively the volume of the translational and rotational degrees of freedom of the non-interacting ligand in the simulation box. When no restraints are used, the non-interacting ligand is free to translate and rotate in the simulation box (i.e., $V_L = V_{\mathrm{box}}$ and $\xi_L = 8\pi^2$), and the rotational term is zero. A sufficiently thorough exploration of the simulation box by the non-interacting ligand is, however, required for the formula to be valid. This is typically hard to achieve as the exploration process is governed by diffusion, and weak transient nonspecific binding will occur at other sites on the protein. The addition of a restraint limits the volume available to the non-interacting ligand, thus speeding the convergence of the sampling. In addition, when enhanced sampling methods such as Hamiltonian λ exchange are used (see Sec. 7.2.4), the use of a restraint is typically necessary as it keeps the ligand in the binding site in the interacting state (see also Sec. 7.2.1) and generally reduce the round-trip time of replicas. When restraints are employed, the values of V_L and ξ_L are restraint-dependent, but for commonly employed restraints, these can be usually easily computed analytically or numerically by solving the relevant integral.

Several choices of restraints are possible.: In practice, a variety of types of restraints are common, from simple harmonic distance restraints between the ligand and the protein [149], to flat-bottom restraints which work similarly but only exert a force if the ligand leaves a specific region [150]. Because these restraints do not limit the rotational degrees of freedom of the ligand, the rotational term entering the correction in Eq. 18 is zero.

Alternatively, a set of restraints proposed by Boresch have also commonly been employed, where all six rigid-body degrees of freedom governing the orientation of the ligand relative to the receptor are restrained [151, 152]. Further restraints, such as on the overall ligand RMSD have also been used [72].

In principle, all of these forms will yield correct binding free energies in the limit of adequate sampling if their effects and connection to the standard state are correctly handled, but they have different strengths and weaknesses. For example, with more involved restraints, sampling at intermediate $\vec{\lambda}$ values will usually not need to be as extensive but more computational effort must go to computing the free energy to turn on the restraints. Additionally, such restraints would typically keep the ligand from exploring alternative binding modes. This restriction may be undesirable when using Hamiltonian λ exchange or expanded ensemble techniques where allowing the ligand to exchange binding modes when it is non-interacting could provide sampling benefits [153]. More specifically, flat-bottom restraints might allow a ligand to explore multiple binding sites, and harmonic restraints allow exploration of multiple binding modes within a site, while Boresch restraints only allow a single binding mode within a single site. See additional discussion of the possibility of multiple binding modes in Sec. 7.2.6 below.

Many choices of restraints involve selecting reference atoms. Again, in principle this choice is unimportant given adequate simulation time but practical considerations may be important. The choice is likely especially important with Boresch-style restraints, where some relative placements of reference atoms are likely to be numerically unstable; additionally, ligand reference atoms should likely be in a part of the molecule which defines the binding orientation well, rather than in a floppy solvent-exposed tail, for example.

7.2 Absolute and relative calculations deal with some of the same issues

7.2.1 Handling weak binders and high dissociation rates—In binding free energy calculations, only the conformations in which the receptor and ligand form a bound complex should be sampled from the bound states (Sec. 3). Determining what the bound states actually are can be challenging for weakly bound ligands. For tightly bound ligands, virtually all reasonable definitions of the bound state will lead to be equivalent free energies, since the partition function will be dominated by a relatively small number of low-energy poses.

For weak binders, this simplification breaks down. In fact, the correct bound state may depend on the type of experiment performed. For example, isothermal titration calorimetry (ITC) or surface plasmon resonance (SPR) measurements effectively define a binding state that includes all ligand comformations that are complexed with the protein, regardless of where on the protein they bind. In contrast, fluorescence polarization competition assays measure binding to only a single location, where the ligand of interest displaces a competing binder. Therefore, care must be taken to ensure that a reasonable definition of the binding site is used [153].

In absolute calculations, this need to explicitly define a binding site applies to the fully interacting state in the complex leg of the thermodynamic cycle (top-right state in Fig. 6), while in relative calculations the binding site must be defined at both end states of the complex leg (top- and bottom-right states in Fig. 2). In principle, this requires defining which conformations are considered to be bound before running the calculation, but it is common practice to start the simulation with the ligand already placed in the binding site and rely on kinetic trapping to maintain the bound complex.

However, this strategy of using kinetic trapping to maintain the bound complex can fail when the dissociation rate of the ligand has the same or smaller order of magnitude than the length of the simulation. This is typical of weak binders such as fragments binding shallow pockets with μM-mM affinities [62, 154]. In the case of weak binders, using a flat-bottom or harmonic restraint between receptor and ligand in the bound state(s) can prevent dissociations [83, 154]. We stress that this type of restraint is normally avoided as it generally introduces bias in the free energy estimate, which is why the restraint is usually activated only in the intermediate states in absolute calculations. The bias can be corrected through reweighting schemes [83], but this post-processing step can be avoided if a flat-bottom restraint is used and the ligand is never sampled while hitting the potential wall during the simulation in the bound state as the numerical correction will be exactly zero. It is important to note that the spring constant and/or radius parameters of the restraint effectively determine which conformations are considered to be bound. As a consequence, these parameters must be tuned to the system so that only the binding site is accessible to the ligand. Again, this step is particularly important for weak binders as their free energy of binding is known to be more sensitive to the definition of the binding site [60].

In absolute calculations, this restraint can substitute or be added to the restraint used to handle the standard state correction (Sec. 7.1.2). In the latter case, however, care must be taken when computing the standard state correction. When multiple restraining potentials are active in the non-interacting state, the correction can generally be computed only through numerical integration. Alternatively, one can adopt a protocol that removes the bound-state restraint in the non-interacting state. Finally, even for tight binders, dissociation events can be enhanced by methods such as Hamiltonian replica exchange [153, 155, 156] and expanded ensemble [157, 158], especially in absolute free energy calculations using harmonic or flat-bottom restraints. In the latter case, dissociations can be averted simply by increasing the spring constant and/or reducing the radius of the restraint potential to prevent the exploration of ligand conformations outside the binding site in the decoupled state (bottom-right state in Fig. 6) that could be propagated to the bound state.

7.2.2 Changes in net charge can be challenging/problematic.—If the net charge of the system changes as the alchemical variable changes during the calculation, this can pose major challenges. Specifically, finite-size effects can introduce significant charge-dependent artifacts into computed binding free energies, in part because typical schemes for long-range electrostatics (including PME and reaction field) do not handle free energy contributions from such changes effectively or as they would be handled in a hypothetical macroscopic bulk solution [159–161].

There are two main potential solutions to avoid artifacts due to changes in net charge: avoiding changing the net charge, and correcting for the introduced artifacts.

Many relative free energy planning tools have been set up to avoid changing the net charge of the systems considered, including LOMAP [136] and Schrödinger’s FEP+ [15]. Absolute free energy calculations can also potentially avoid changing the charge of the system by making a charge perturbation of equal and opposite sign elsewhere in the system; for example, as a charged ligand is removed, a charged counterion of opposite sign could also be removed, or one of the same sign could be inserted. This is sometimes referred to as an “alchemical ion” approach for dealing with the needed charge change, and is also employed by the Yank free energy package [153]. Charge corrections have also been explored, and are potentially a viable solution to this problem [162] where artifacts introduced by finite-size effects are corrected numerically [103, 160]. However, application of such corrections typically remains less common than the use of a coalchemical ion. A third approach has been proposed by Gapsys et al. [163] which uses a double-system/single box setup.

When free energy calculations *do* need to change the charge of a ligand or solute, the literature does not yet seem to indicate what approach should be preferable, so considerable care should be taken. We are not yet aware of a careful comparison of charge corrections versus other approaches such as decoupling an ion at the same time, so in our view the issue of proper handling of charge mutations in the context of alchemical calculations remains a research problem, and several papers give good guides for exploring the problem further [159–161].

7.2.3 The importance of the alchemical pathway—

Both absolute and relative calculations must choose an alchemical pathway connecting initial and final states. In principle, because of the path independence of the free energy, any arbitrary pathway will give the correct free energy change, but the choice of pathway will greatly affect the efficiency of the calculations. Some choices are particularly crucial—for example, transformations involving insertions or deletions of atoms should employ a soft-core potential path for Lennard-Jones or other interactions with repulsive interactions that go to infinite energy at small radius [164–166].

The key consideration for choosing alchemical pathways is that the intermediate states that a given pathway produces should sample configurational ensembles that change as slowly as possible as $\overrightarrow{\lambda}$ changes, while still managing to go from the initial state to the final state as $\overrightarrow{\lambda}$ goes from 0 to 1.

Another way of stating this is that intermediate states should sample molecular configurations that are as similar as possible to their neighboring states. The more similar the configurations are between intermediate states, the lower the statistical uncertainty is in the estimate of free energy between intervals. This can be proven directly from the BAR and MBAR formulas [25, 43], though the exact same principles apply for TI. For a ‘good’ path to work and give a sequence of states with maximally similar configurations, sufficient similarity in potential energies is required. Fig. 7A and B illustrate this. Fig. 7A shows in a pictorial way a soft-core potential can be applied across different $\overrightarrow{\lambda}$s. Fig. 7B illustrates the potential energy distributions at the different $\overrightarrow{\lambda}$ intermediates, with sufficient overlap between neighboring $\overrightarrow{\lambda}$ states to ensure that reweighting estimators such as MBAR can be used for analysis (see Sec. 8.3). The actual transformation is best handled with soft-core potentials of the form shown in Fig. 7 C and B, with more details given below.

So what are the options to adjust the potentials between the two end states based on $\overrightarrow{\lambda}$? The simplest possible alchemical pathway is a *linear* pathway:

$$U(\overrightarrow{q}, \overrightarrow{\lambda}) = (1 - \overrightarrow{\lambda})U_0(\overrightarrow{q}) + \overrightarrow{\lambda}U_1(\overrightarrow{q}), \tag{19}$$

so-called because the dependence on $\overrightarrow{\lambda}$ is linear. This clearly satisfies the basic requirement that it gives the initial endpoint potential energy $U_0(\overrightarrow{q})$ when $\overrightarrow{\lambda} = 0$ and final endpoint energy $U_1(\overrightarrow{q})$ when $\overrightarrow{\lambda} = 1$.

For many energy terms this is a very good approach, *as long as a repulsive core remains on*. For example, it can be shown that if van der Waals repulsions are left on, then the linear approach is very nearly the optimal path possible for changing, removing, or inserting the electrostatic energy terms, with the alchemical path being within about 10–20% of the minimum possible uncertainty [167] for a fixed amount of simulation time, as well as being nearly optimally efficient for van der Waals attractive terms with repulsion terms turned on

[168]. Although we are not aware of any quantitative tests for dipolar or higher multipole terms, theoretically it should behave equally well for those systems.

However, this approach ends up being terrible for removing or adding repulsive potentials that go to infinity quickly at or near the origin. One way to look at this is to examine how low $\overrightarrow{\lambda}$ values must go to reduce the energy at 0.5σ (the atomic size parameter) down to 1 k_BT, where thermal fluctuations make it possible for other atomic sites to penetrate routinely that deep. Assume we are trying to go from a particle being present, and desire to make it disappear alchemically. If the repulsive terms are of the form $\epsilon\left(\frac{\sigma}{r}\right)^{12}$, and if ϵ is 1 k_BT at the temperature of interest, and we start with the particle present, we may solve for $(1-\overrightarrow{\lambda})(1k_BT)\left(\frac{\sigma}{0.5\sigma}\right)^{12} = 1k_BT$. This yields $\overrightarrow{\lambda} = 1 - 2^{-12} \sim 0.999976$. At this point, we have gone virtually all the way to the end of the transformation, but there is still an impenetrable post in the middle of our simulation! This is not very much like the desired final state of no interactions between the particle and its environment. We can play around with a few ways of modifying this, like simulating many more intermediate states near $\overrightarrow{\lambda} = 1$. However, various analyses have shown that this is not a very good strategy [164, 166, 169–171].

What we need instead is a function that smoothly gets rid of this infinity. A large number of schemes have been tried [164, 168–173], but the most common strategy that appears to be the best practice is to use a "soft-core" potential, of the form:

$$U\left(\overrightarrow{r}_{ij}, \overrightarrow{\lambda}\right) = 4\epsilon_{ij}\overrightarrow{\lambda}\left(\frac{1}{\left(\alpha(1-\overrightarrow{\lambda}) + (r_{ij}/\sigma_{ij})^6\right)^2} - \frac{1}{\alpha(1-\overrightarrow{\lambda}) + (r_{ij}/\sigma_{ij})^6}\right), \qquad (20)$$

where rij is the distance between two particles i and j, ϵ_{ij} and σ_{ij} are the Lennard-Jones parameters corresponding to the interaction between particles i and j, and α is a constant. In particular, α = 0.5 is statistically optimal for the specific functional form shown above. This functional form has exactly the property we are looking for: it recovers the Lennard-Jones potential when $\overrightarrow{\lambda} = 1$, and the at other endpoint ($\overrightarrow{\lambda} = 0$), it is exactly zero for all r_{ij} everywhere, and as $\overrightarrow{\lambda}$ goes to zero, the $\alpha(1-\overrightarrow{\lambda})$ term lowers the infinite energy in the core. There are several different variants of the same functional form [164, 169, 170], but the one given in eq. 20 is easy to understand and implement and fairly numerically stable. This functional form is shown in C and D of Fig. 7.

It has been shown that more complicated forms are not significantly more efficient than eq. 20 [172]. We therefore recommend using the soft-core potential given in eq. 20, unless there is a compelling reason otherwise. Using a similar equation to eq. 20 may be acceptable in most circumstances if that is what is supported in your chosen software. However, if you are inserting or removing entire atomic sites, we heavily recommend against using the linear approach; it will be very difficult to get correct or converged results.

So far in this section, we have discussed optimal ways of disappearing or appearing Lennard-Jones interaction sites and turning on and off electrostatics terms. What about performing both transformations at the same time? We cannot turn off the electrostatics linearly at the same time we turn off the Lennard-Jones terms, as it would leave infinitely large attractive and repulsive electrostatic terms “bare” at small $\vec{\lambda}$, resulting in the simulation crashing. It *is* possible to apply the same soft-core approach to the Coulomb interaction as to the van der Waals interaction, and this is indeed done in a number of implementations. In this case, it is important that the Coulomb interaction is softened as rapidly or more rapidly than the Lennard-Jones interaction to avoid charge penetration issues into the repulsive core, which can be tricky to ensure for multiple types of perturbations simultaneously [174].

A safe but potentially more computationally expensive approach is to perform the transformations in sequence; first, turning off all electrostatics for atoms that must be removed, inserting and removing Lennard-Jones sites (both the insertion and removal can be done simultaneously), and then turning electrostatics for the introduced particles on. Again, if there are no removals or additions to atomic sites, then it is reasonable to change the interactions in the first and third steps linearly.

Other issues, such as whether absolute calculations should retain or remove intramolecular non-bonded interactions through either annihilation [149, 151, 175–177] or decoupling [149, 178], must be considered. Reasonable efficiency can be often obtained with either choice even if some are somewhat better or worse than others, and there is no consensus on which is better in most given situations. Our recommendation is to leave the intramolecular interactions on during the transformation for simplicity if there are no other known issues with this approach. The key feature of the simulation to watch out for is whether the total potential energy, and therefore the intermediate ensembles sampled, changes smoothly from beginning to end. Problems of discontinuous changes of the potential energy can be diagnosed by noticing lack of configuration space overlap between different simulations (see Sec. 8.5).

Relative calculations introduce additional choices, such as the order in which to modify nonbonded interactions. A common process in single topology relative calculations is, as noted above, to first remove electrostatic interactions of any atoms which will be deleted, then modify other non-bonded interactions, then restore electrostatic interactions of any atoms which are being inserted. Although this is a simpler path to understand cognitively and can take advantage of the soft-core potential from Eq. 20, this can lead to more intermediate steps and thus be more computationally expensive. Other schemes, such as simultaneously changing electrostatic and Lennard-Jones interactions with electrostatic soft-core potentials [179], as already discussed above, may be implemented with fewer intermediate states but could require fine-tuning of electrostatic and Lennard-Jones soft-core parameters to avoid numerical instabilities. At the time of writing, there has not been conclusive evidence to suggest the separate or simultaneous approach is in general better than the other, all factors considered, so discretion should be left up to the user as to what is viable from both hardware resources, and what the simulation software supports.

A key additional consideration in choosing the alchemical pathway is the choice of spacing of intermediate states. The spacing depends to some extent on the choice of analysis method, though states should essentially be spaced equidistant in the relevant thermodynamic length [180, 181]. For BAR/MBAR techniques this means that states should be spaced so that the statistical uncertainties between neighboring states be approximately equal [172, 182], where "approximately" is roughly within 30–50% in magnitude. Some schemes to adaptively optimize the spacing of intermediate states based on initial exploratory simulations have been proposed [183]. For molecules changing in dense solvent, then the best path is roughly independent of molecule size and shape, so what works for one molecular transformation is likely to be relatively efficient for another [184].

Some approaches have attempted to find alternative pathways to improve efficiency or find paths of low thermodynamic length [167, 168, 172]. For example, the enveloping distribution sampling (EDS) approach, and its multiple-replica and accelerated variant, works to improve efficiency by creating a single artificial intermediate state which simultaneously samples all end state phase spaces [185–187]. When this can be done, it provides an extremely efficient way to calculate the relative free energy difference between multiple ligands from a single simulation. However, it can often fail whenever the simulation of this intermediate state ends up trapped in configurations characteristic of only one end state. Thus, successful use of EDS can require system-dependent tuning, making it difficult to implement in an automated and reliable way. However, when successful, it can be very efficient.

In our view, there is still some room for further exploration of how to best choose transformation pathways, especially for relative binding calculations or more complex molecules, as most existing studies focus on smaller molecules. As we have stressed, in principle, any pathway that connects the desired end states is rigorously correct, but as discussed above, different paths may differ dramatically in thermodynamic length and therefore efficiency. Additionally, some paths simply may not converge due to issues noted above such as those encountered without soft core potentials. However, the recommendations above are reasonable, reliable, and are likely not that much less efficient than potentially more optimal choices [167, 168, 172], as the real problems with the efficiency of calculating free energies are lack of sampling of slow conformational modes, rather than the lack of efficiency of the transformations.

It is however important to note that different packages also differ in how they handle implementation of alchemical transformations, making it difficult to give rules of thumb concerning specific efficient transformations which work equally well across simulation packages. Thus, we are hesitant to recommend best practices within specific software packages at this point in time, although any good transformation pathway will conform to the guidelines we have outlined above.

7.2.4 Which sampling scheme will work best for my problem?

Though all alchemical simulations must sample from multiple $\vec{\lambda}$ states, different approaches can be used to achieve this. Fig. 8 illustrates the four most common schemes. The simplest approach involves running an independent simulation at each of the predefined $\vec{\lambda}$ values (see Fig. 8A). This type of scheme is currently used for AMBER TI calculations [17] and for Sire as implemented in BioSimSpace [34]. However, if these simulations can be run simultaneously with communication between them, a simple extension allows mixing between these replicas. In this approach, the simulation at each $\vec{\lambda}$ can undergo periodic exchanges with neighboring $\vec{\lambda}$ values. This form of replica exchange, called Hamiltonian replica exchange, is based on ideas developed from Monte Carlo simulations of spin glasses by Swendsen and Wang [188]. With the Metropolis-Hastings acceptance criterion for exchanges, the generated ensemble of all replicas still samples from the Boltzmann distribution for each replica. This approach has been used in many different contexts for molecular simulations [155, 189–191]. The basic idea of the replica exchange scheme is shown in Fig. 8B. It is supported in various software packages that provide alchemical implementations, such as GROMACS [13], GROMOS [192, 193], FEP+ [15], and NAMD [134].

A third approach borrows ideas from simulated tempering [194]. In this scheme a single replica rapidly explores all of $\vec{\lambda}$ space by working out optimal weights that allow switching between different intermediate $\vec{\lambda}$ values, as seen in Fig. 8C. This approach is also referred to as self-adjusted mixture sampling [157, 158, 195] and while promising, has so far only been supported in OpenMM Tools [196] and GROMACS. Although this approach allows multiple states to be simulated in a single simulation, the weights do not always converge to the proper equilibrium distribution, and care must be taken that the final results are converged.

The last approach makes use of non-equilibrium simulations [7]. In this approach, only end state $\vec{\lambda}$ replicas ($\vec{\lambda} = 0, \vec{\lambda} = 1$) are simulated at equilibrium; intermediate information is generated from non-equilibrium simulations that rapidly transition between end states. This approach is available in GROMACS and appears to be coming online in several other packages. A schematic of this approach is shown in Fig. 8D.

Currently, we recommend using Hamiltonian replica exchange type sampling schemes (Fig. 8 B). If these are not available in the code of choice, running independent simulations at different $\vec{\lambda}$ values can be acceptable, especially when configurational sampling is fast (Fig. 8 A). Single replica schemes and non-equilibrium schemes are not as established yet because of potential failure modes, but are very promising for use in the near future.

7.2.5 How long should I run my simulation for and what information should be saved?—

Before launching alchemical free energy calculations it is wise to consider how convergence and completion will be assessed. Different conditions on when to stop alchemical free energy calculations should be determined, and this may require several iterative checks and therefore modifications to the calculation protocol. One useful metric to use for termination is the expected or desired uncertainty of a desired free energy estimate, though care must be exercised should the uncertainty estimate prove unreliable. In particular, if the rate of change in the free energy estimate is significant when this condition is met, the simulation may not be locally converged, and more sampling may be necessary to determine a stable free energy estimate which is no longer changing significantly over time.

However, this is not the only metric which can or should be used, as the uncertainty only captures the information about the sampled phase space, not necessarily the entirety of the phase space. For example, convergence of relative free energy calculations in predictive simulations where the entire phase space is not known in advance, requires sampling the different kinetically stable states [85]. This highlights the importance of choosing the correct thermodynamic path to ensure you sample the required thermodynamic states as discussed in Sec. 7.2.3.

The condition of minimizing the statistical uncertainty of different free energy estimators below a sufficient threshold should be one metric monitored over the simulation. This can be done through the uncertainty estimator built into certain analysis tools such as MBAR, or through more general statistical tools like bootstrap sampling. It should be noted however, that uncertainty estimates have the same limitations as other metrics of convergence, as they are only an uncertainty based on the phase space sampled so far in a simulation, and cannot account for states not sampled, and it is worth considering that they will be an underestimation of the true uncertainty. A target statistical uncertainty should be chosen at the onset of the simulation to avoid excessively long simulations, or falling into the trap of running until the free energy estimate is "good enough," which is subjective and has no defined criteria. This could be a fixed value such as 0.20 kcal/mol, or a functional quantity such as "below 0.5 kcal/mol and 10% of the free energy estimate." The user does not need to monitor this information in real-time and can choose to run simulations for fixed duration (either time or number of samples) and run analysis on the data collected thus far. If more samples are needed, the simulations can be resumed, or, started again in different initial conditions.

Convergence in other alchemical observables should also be monitored to determine if the defined phase space has been sufficiently sampled and enough decorrelated samples have been drawn. These additional observables include, but are not limited to, the variance in $\frac{dU}{d\vec{\lambda}}$ across all $\vec{\lambda}$ values, calculating the variance in free energy using bootstrap analysis, and comparing differences in free energies calculated using different percentages of the simulation in both the forward and reverse directions [43] (see Fig. 9).

Each of these metrics shows some promise for diagnosing when a simulation has a convergence issue beyond simple convergence of uncertainty estimates. Results obtained from calculations with convergence issues should be checked for errors or run for longer before any confidence should be placed in conclusions drawn from their analysis. For example, in relative calculations ligands that share similar binding modes and do not induce large conformational changes when in complex with protein, the need to sample exhaustively to converge estimates in free energy differences is often minimal due to the locality of sampling changes in the molecular topology and shared phase space of the core atoms. However, even subtle induced changes in protein binding configuration will require more sampling or cause local convergence to a free energy estimate that has high error. The confidence a user should have in a free energy estimate is significantly improved when both the uncertainty of the free energy estimate is low, and when other observables have reached a convergence.

The uncertainty in the free energy, for example, can be estimated in multiple ways, e.g. through standard error propagation methods (including MBAR's estimator, which is based on the same principles as standard error propagation), through bootstrap methods, and through multiple independent runs. Independent of how the property is estimated, it is important to remember that results of any free energy analysis are *estimations of the given property*, not the true underlying value of the property itself. These estimators are usually consistent estimators, meaning they will converge to the true answer in the limit of sufficient sampling, not necessarily unbiased ones though. As such, it is a good idea to subject different estimators to the same data to see if they yield either the same estimate (within error and bias), or if they fluctuate wildly. See, for example, the potential of mean force with respect to $\vec{\lambda}$ estimated from a bound simulation of a Tyk2 ligand pair of Wang et al. [15] for both the MBAR and TI estimators, as seen in Fig. 10. This is not a perfect method as some estimators, such as exponential averaging, will converge significantly more slowly, relative to more accurate estimators like MBAR. Therefore, it is a good idea to apply the estimators to different fractions of the data to see if the main estimator of free energy you have chosen is stable.

Each method requires different data from the simulation be collected. If, for instance, the free energy estimator selected is thermodynamic integration, then values of $\frac{dU}{d\vec{\lambda}}$ at uncorrelated data points must be collected. Once you have made a choice of the combination of the type of simulation you will run, which alchemical topology you will simulate, what alchemical path you will simulate along, and what your stopping conditions are, then you are ready to enumerate the information you should capture. Below is a sample of the minimal information you need for a set of common estimators (discussed in more detail in Sec. 8.3):

- Thermodynamic Integration (TI) requires $\frac{\partial u(\vec{q})}{\partial \vec{\lambda}}$.
- Exponential Averaging (EXP) needs *either* $\Delta u_{k,k+1}(\vec{q})$ or $\Delta u_{k,k-1}(\vec{q})$, depending on the direction its being evaluated in.
- Bennett Acceptance Ratio (BAR) needs *both* $\Delta u_{k,k+1}(\vec{q})$ and $\Delta u_{k,k-1}(\vec{q})$.
- Weighted Histogram Analysis Method (WHAM) and Multistate Bennett Acceptance Ratio (MBAR) both need the complete set of $\Delta u_{k,j} \forall j = \{1 \ldots K\}$. WHAM must have this same information binned with some choice of bin width small enough not to affect the results.

The potential derivative required for TI should generally be calculated during the simulation; only under very rare circumstances [167] can it post-processed by a code that does not evaluate the derivatives. Many codes already have options for doing this. If that option is unavailable, you can estimate it through finite difference (if sufficient information is collected), but this will introduce significant error, and is generally not a best practice. The BAR estimator may be a better, and simpler choice at that point as you will have at least the same level of information.

The potential energy differences required for EXP, BAR, MBAR, and WHAM can be calculated either during the simulation or in post-processing. It is recommended to calculate the potential differences in code when possible to avoid extra overhead and possible errors produced by evaluating the energy of the configuration twice, and to reduce the amount of stored information. Although potential energy derivatives must usually be calculated in code there is one condition under which they can be easily computed in post-simulation analysis. If the alchemical path you have chosen is a linear alchemical path, then $\frac{du}{d\overrightarrow{\lambda}} = u_0(\overrightarrow{q}) - u_1(\overrightarrow{q})$, which is the difference between the initial and final states, which are already calculated by the simulation and can be recorded easily without additional computational expense. However, because of the problems with linear paths already discussed in this paper, this simplification is rarely that useful.

Free energy information should generally be saved more frequently than coordinate data, approximately at the rate that uncorrelated samples are produced. The on-disk size of the data for free energy estimation is often significantly smaller than full atomic coordinates, so the information can easily be collected frequently. However, the information should not be collected *every* time step, as most free energy techniques are operated at equilibrium, and need equilibrated *and decorrelated* samples for an unbiased estimate. Samples collected every time step will likely result in most samples being discarded due to the detection of correlation in the time series by decorrelation routines in the analysis. However, if it is computationally cheap and disk space is plentiful, do save often. One may safely assume that the correlation time is greater than 100–200 fs even for relatively simple systems such as small molecules in solvent, so saving no more frequently than every 50–100 steps is recommended. How decorrelation impacts calculations, and how to compute it is discussed in Sec. 8.2.

In general, uncertainties can be assumed to decrease as $1/\sqrt{(N)}$ where N is the number of uncorrelated samples, for all standard free energy calculation methods [197]. However, this carries the notable caveat that such estimates require accurate estimation of the correlation time which, if important motions are slow compared to simulation timescales, can be difficult. Still, this metric provides a good rule of thumb, and as long as conformational transitions are captured by the simulation, increasing the aggregate simulation time by a factor of T will reduce the uncertainty by a factor of approximately $\sqrt{T}$.

7.2.6 Multiple or uncertain binding modes may require considerable care—

In a discovery setting, new ligands can have unknown or at least uncertain binding modes [111, 198–200], complicating binding free energy estimation.

To deal with prospective ligands with unknown binding modes, discovery projects commonly assume that modifications of functional groups on a common scaffold result in a consistent binding mode across all members of a series. This is not necessarily always the case [111], as reviewed elsewhere [199] and in some cases unexpected binding mode changes can be the origin of apparent non-additivity in structure-activity relationships [200]. Binding modes also tend to be particularly variable in the case of fragments, which often may have multiple relevant binding modes [201].

Absolute free energy calculations for dissimilar ligands can have particular challenges because the (potentially incorrect) assumption of consistent binding modes across a series of similar ligands is likely to be even less robust than in the case of relative calculations. This means that researchers performing absolute binding free energy calculations will have to pay particular attention to generating reasonable putative binding modes.

In some cases, it is tempting to simply use docking techniques to generate initial bound structures for starting molecular dynamics simulations. However, timescales for binding mode interconversion are usually slow compared to MD/free energy timescales, meaning that simulations started from different potential binding modes are likely to yield disparate computed binding free energies [47, 85, 149, 202]. Moreover, docking techniques are good at identifying sterically reasonable potential binding modes, but still perform relatively poorly at identifying a single dominant binding mode *a priori*.

It is worth highlighting a recent SAMPL blind challenge on HIV integrase as an illustration of this. Many submissions, using state-of-the-art methods, had difficulty even predicting which *binding site* ligands would bind in—most submissions placed more than half of the ligands into the incorrect binding site—and even given correct binding sites, the binding mode within each site was also quite difficult to predict [133]. The best performing submission for predicting binding modes actually ended up being a human expert (aided by computational tools) with more than 10 years of experience on the particular target [203], rather than a fully automated approach. While free energy calculations on this set had some success, many of the failures actually ended up being cases where the binding mode selected as input for free energy calculations was later found to be incorrect [204], highlighting the importance of these issues.

One approach which has shown some success in identifying accurate binding modes *de novo* is to retain diverse potential binding modes from docking, perform short MD simulations of these to identify distinct stable binding modes, and then consider only these stable modes in subsequent calculations [12, 149, 204–206].

Routes to handle multiple potential binding modes are different depending on whether absolute or relative calculations are selected, unless a method is available to estimate the relative populations of different stable binding modes in advance (e.g. such as the BLUES approach currently in development [47]), in which case this approach could be applied to assist both types of calculations.

Handling multiple potential binding modes within absolute calculations.: Within absolute binding free energy calculations, multiple potential binding modes can be handled by two main strategies: Considering each binding mode separately (a separation of states strategy) or sampling all binding modes within a single simulation [85]. This couples to the choice of restraints selected, as some restraints will allow transitions between binding modes and even binding sites (Sec. 7.1.2), and others do not.

Sampling all potential ligand binding modes within a single free energy calculation is usually impractical without some form of enhanced sampling or at least Hamiltonian replica exchange [153] because barriers for binding mode interconversion result in kinetics which are too slow compared to simulation timescales [47, 85, 149, 202]. Hamiltonian exchange, coupled with appropriate restraints, can allow the ligand to relatively rapidly exchange between potential binding modes when non-interacting, accelerating sampling of binding modes [153]. However, it is not always clear that this is desirable, since this also increases the size of the configuration space which must be sampled even if the binding mode is known.

Separation of states provides a simple though potentially expensive alternative, where each stable binding mode is considered separately with a binding free energy calculation restricted to that binding mode, and then (as long as the binding modes are non-overlapping) the resulting component binding free energies can be combined into a total [85, 149]. This approach necessitates a separate binding free energy calculation for each potential binding mode, however, so it can be computationally quite costly. If relative populations of different stable binding modes were available from some other technique, it could make this separation of states approach considerably more efficient [47, 85].

Handling multiple potential binding modes within relative calculations.: Multiple potential binding modes pose particular problems for relative free energy calculations, as having multiple starting structures for these calculations could yield substantially different calculated relative binding free energies for the same transformation due to kinetic trapping, and, without additional information (specifically, the free energy of binding mode interconversion or, equivalently, the relative populations of different binding modes) it becomes impossible to sort out which of the multiple answers is in fact the correct relative binding free energy.

To deal with this, some practitioners have actually computed relative binding free energies of different binding modes of the same ligand [202]. For example, a perturbation which adds a methyl to an aromatic ring of a larger ligand might yield one result if the methyl points in one direction, and a different value if it points in the other due to slow ring motions [207, 208]. One could compute the free energy of turning off the methyl group in one orientation and turning it back on in the other orientation to obtain the free energy difference between the two potential binding modes. While this approach has precedent, it is relatively difficult to automate at present and requires considerable care.

Overall, this likely means that relative free energy calculations will be susceptible to problems resulting from uncertainty in ligand binding modes until more robust approaches are available to determine dominant binding modes, or the relative populations of different potential binding modes, in advance.

7.3 General simulation setting choices that could affect free energy calculations.

There are many parameter choices that are common to standard MD simulations. Optimal choices may depend on the simulation package and it is best to consult the manual of each package to make the right choices. One particular parameter we want to draw attention to which is often overlooked is the timestep. The choice of integrator can drastically affect the accuracy of a given calculation depending on the timestep [209]. Using a 1.0 fs timestep near the limit of stability without any constraints can lead to non-negligible statistical mechanical errors. However, from various previous free energy studies using 1 fs integration timesteps it is not clear that a significant error would be introduced into a free energy estimate and may warrant some further investigation.

8 Data analysis

Once equilibrium data has been collected from alchemical intermediates, it must be analyzed to produce an estimate of the free energy change (and its associated statistical uncertainty) for each leg of the thermodynamic cycle. While a number of different estimators are available that will give consistent results under optimal circumstances, some approaches are recommended over others due to their robustness and ability to provide information on poor convergence.

8.1 Detecting the boundary between equilibrated and production regions

Much of the infrastructure for analyzing alchemical free energy calculations relies on the concept of asymptotically unbiased estimators, which produce unbiased estimates of the free energy when fed very long simulations [197]. In reality, free energy calculations are often initiated from highly atypical initial conditions (such as a protein-ligand geometry obtained from docking and subjected to a heuristic solvent placement scheme), and simulations are of a finite length dictated by available computational resources and computing demands. As a result, these estimators can produce significantly biased estimates if fed the entirety of simulation data generated without further processing [210].

To minimize this effect, an initial portion of the simulation is often discarded to *equilibration* [41], with the idea of removing the most heavily biased initial portion of simulation data but retaining the unbiased *production* region that represents a stationary Markov chain process sampling from the desired equilibrium target distribution. Because the simulation time required for the atypical initial sampler state to relax toward equilibrium is a property of the specific system being simulated and the specific initial conditions selected, it is simplest to collect data for the whole process and use an automated algorithm to select how much data should be discarded to equilibration in a post-processing step.

A simple approach to automatically partitioning simulation data into equilibration and production regions is described in [210] (illustrated in Fig. 11). Suppose we have a simulation of length T consisting of correlated data. Here, the goal of the post-processing step is to select the equilibration boundary $t_0 \in [0, T]$ so as to *maximize* the number of effectively uncorrelated samples remaining in the production region $N_{[t_0, T]}$, which is defined as

$$N_{[t_0, T]} = \frac{T - t_0}{g_{[t_0, T]}} \qquad (21)$$

where $g_{[t_0, T]}$ is the *statistical inefficiency* of a timeseries a_t, described in more detail below. Conveniently, this procedure also produces the information necessary to decorrelate the simulation data for estimating the free energy differences, a requisite next step in analysis.

This approach is implemented within the MBAR [211] and alchemlyb [212] packages, and is highly recommended for standard practice.

For additional discussion of working with correlated data and autocorrelation analysis, please refer to the work on Best Practices for Quantification of Uncertainty and Sampling Quality in Molecular Simulations [42].

Computing the timeseries for equilibration detection—Typically, the timeseries of note a_t analyzed in automated equilibration detection is the negative logarithm of the probability density $\left(\pi\left(x_t; \overrightarrow{\lambda}\right)\right)$ sampled by the MCMC algorithm (up to an irrelevant additive constant). For simple independent simulations that sample $x_t \sim \pi(x; \overrightarrow{\lambda})$, this is given by the reduced potential

$$a_t \equiv -\ln \pi\left(x_t; \overrightarrow{\lambda}\right) + c = u\left(x_t; \overrightarrow{\lambda}\right). \tag{22}$$

Note that the use of the effective reduced potential is not guaranteed to pick up on all slow relaxation processes that may be coupled to the alchemical free energy, but the simplicity of its computation means it is generally appropriate for most cases.

Cautions in automating equilibration detection—For simulations that are simply not long enough to contain a large number of samples from true equilibrium either because they are very short or contain slow processes, this procedure cannot completely remove the bias. In such cases, this approach simply selects the final portion of the the simulation, which may be contained in a single substate of configurational space, and may itself lead to biased estimates. This situation can be detected if the equilibration boundary t_0 is a significant fraction of the total simulation length T, with a good rule of thumb being that $T \gtrsim 20t_0$. If this is not possible, advanced analysis techniques that assume only local equilibrium (rather than global equilibrium) such as the TRAM estimators [27, 28, 213] may be more appropriate, but are beyond the scope of this paper.

8.2 Decorrelating samples for analysis

Computing the statistical inefficiency—Most estimators require an uncorrelated set of samples from the equilibrium distribution to produce (relatively) unbiased estimates of the free energy difference and its statistical uncertainty. To do this, the production region of the simulation is generally *subsampled* with an interval approximately equal to or greater than the *statistical inefficiency* $g \geq 1$ to produce a set of uncorrelated samples that can be fed to the estimator machinery [210],

$$g \equiv 1 + 2\tau_{\mathrm{eq}} \tag{23}$$

where τ_{eq} is the integrated autocorrelation time, formally defined as

$$\tau_{\mathrm{eq}} \equiv \sum_{t\,=\,1}^{T\,-\,1}\left(1 - \frac{t}{T}\right)C_t, \tag{24}$$

with the discrete-time normalized fluctuation autocorrelation function C_t defined as

$$c_t \equiv \frac{\langle a_n a_{n+t}\rangle - \langle a_n\rangle^2}{\langle a_n^2\rangle - \langle a_n\rangle^2}\,. \tag{25}$$

The basic concept is that τ_{eq} corresponds to the single-exponential decay time for the autocorrelation process that generates samples, so the statistical inefficiency g measures the approximate temporal separation between two effectively uncorrelated samples (where two exponential relaxation times are presumed to be sufficient).

Robust estimation of C_t for $t \sim T$ is difficult due to growth in statistical error, so common estimators of g make use of several additional properties of C_t to provide useful estimates (see *Practical Computation of Statistical Inefficiencies* in [210] for a detailed discussion).

We recommend using the robust statistical inefficiency computation routines available within the MBAR [211] and alchemlyb [212] packages.

Subsampling data to generate uncorrelated samples—Once the statistical inefficiency g has been estimated, it is straightforward to subsample the correlated timeseries simulation data to produce effectively uncorrelated data that can be fed to the free energy estimators. Suppose the correlated timeseries is $\{a_t\}_{t\,=\,1}^{T}$; we can form a new timeseries of $N_{\mathrm{eff}} \approx T/g$ effectively uncorrelated samples by selecting a subset of indices $\{t = \mathrm{round}((n - 1)\ g) \mid n \in \mathrm{range}(1, \ldots, N)\}$ where round(x) denotes rounding to the nearest integer.

If independent simulations are used, the alchemical state $\vec{\lambda}$ may have a significant impact on the correlation time, and these simulations should be subsampled independently using a separate estimate of the statistical inefficiency g for each alchemical state. If coupled simulations are used (such as a Hamiltonian replica exchange simulation), the replicas should undergo equivalent random walks in alchemical space, and the replicas can be subsampled with the same g to generate an equal number of uncorrelated samples at each alchemical state. Conveniently, the approach described above for automated equilibration detection produces an appropriate estimate of g over the production region for automating this process.

Cautions and considerations—Reliable estimation of the statistical inefficiency is difficult, and estimates will not generally be as precise (in a relative error sense) as averages. To ensure there is sufficient data available for reliable decorrelation and estimation of free energy differences, it is recommended that the effective number of uncorrelated samples $N_{\mathrm{eff}} \gtrapprox 50$ if the BAR or MBAR estimators (discussed below in sec 8.3) are used; the number may need to be much higher with alternate estimators.

8.3 Estimators for free energy differences

Free energy differences between two different states differing in the energy function are directly related to the ratio of probabilities of those states. As can be noted, the partition functions in Eq. 5 are simply the total accumulated probabilities for all possible configurations of the system. Virtually all of the ways to estimate this free energy are based in converting this ratio of integrals to something that can be measured in one (or several) simulations.

The Zwanzig relationship (EXP)—The simplest method for calculating free energy differences from simulations is the so-called *Zwanzig relationship* [1], also called one-sided exponential re-weighting (EXP), or simply free energy perturbation, though this final term is sometimes used to encompass all ways of calculating free energy differences.

The (reduced) free energy difference Δf_{01} between an initial state 0 and a final state 1 defined by two different potential energy functions $u_0(\vec{q})$ and $u_1(\vec{q})$ over coordinate space $\vec{q}$ can be calculated as:

$$\Delta f_{01} = -\ln \left\langle e^{-\left(u_1(\vec{q}) - u_0(\vec{q})\right)} \right\rangle_0 = -\ln \left\langle e^{-\Delta u(\vec{q})} \right\rangle_0 \quad (26)$$

and the average is over all samples from the simulation performed with u_0. In the case of NVT (canonical) sampling and assuming the masses do not change, then u is simply $U/k_B T$, and f is $F/k_B T$, but it can be generalized to other ensembles with the proper definition of f and u. Described in words, we take the samples generated during our run with the potential energy function $u_0(\vec{q})$ and calculate what the difference in energy would be if we switched instantaneously to the potential energy function $u_1(\vec{q})$, and average the exponential of the negative energy difference to get the negative of the exponential of the free energy difference. The original distributions, P(u_0) as generated at $\vec{\lambda} = 0$ and P(u_1) would look like those seen in Fig. 8A–C on the right hand side. Reevaluating requires almost no extra code functionality to perform; one need only to save a full precision trajectory, and run an un-modified molecular simulation code using the u_1 in order to calculate the new energies of stored snapshots. The analysis can be written in a line of code. We note that this method is even more general, in that the instantaneous work to change the potential energy function from u_0 to u_1 can be replaced by the non-reversible work W to make the same change beginning from the same equilibrium conditions at either end state [78–80], allowing for non-equilibrium free energy calculations, an alternate approach. We do not detail non-equilibrium transformations here, and refer the reader to more advanced treatments [18, 77, 214–217], as our focus here is on equilibrium free energy techniques.

Although the Zwanzig equation is formally correct as long as the two states considered sample the same phase space volume, which is true for standard molecular models, it has some very important numerical issues that mean that it often performs badly for standard free energy calculations, even for small molecules [197, 218]. One can show that if the standard deviation of the difference $\Delta u(\vec{q}) = u_1(\vec{q}) - u_2(\vec{q})$ over all sampled $\vec{q}$ is large, which in this case, means only several times $k_B T$, then very few samples contribute to the average, and the answer will be both biased and extremely noisy [219]. Essentially, the method is dominated by contributions of rare snapshots [75, 76, 220].

The Bennett Acceptance Ratio (BAR)—If we have the differences in the potential energy sampled from the distribution defined by u_0 to the state defined by u_1, and we also have the differences in potential energies from the distribution sampled by u_1 to the state defined by u_0, we can obtain a significantly improved estimate of the free energy difference compared to that obtained by EXP. This estimate was first derived by Bennett and is hence generally called the Bennett Acceptance Ratio (BAR). It is solved by finding the reduced free energy f_{ij} that satisfied the following implicit equation:

$$\sum_{i=1}^{n_i} \frac{1}{1+\exp\left[\ln\left(\frac{n_i}{n_j}\right)+u_{ij}(\overrightarrow{q})-f_{ij}\right]} = \sum_{i=1}^{n_j} \frac{1}{1+\exp\left[\ln\left(\frac{n_i}{n_j}\right)-u_{ij}(\overrightarrow{q})+f_{ij}\right]}, \tag{27}$$

where n_i and n_j are the number of samples from each state. More recent derivations show that this formula is the maximum likelihood estimate of the free energy difference given sets of samples from the two states [77].

Many studies have demonstrated both the theoretical and practical superiority of BAR over EXP in molecular simulations [197, 218], and BAR converges to EXP in the limit that all samples are from a single state [25, 25, 77]. BAR also requires significantly less overlap between the configurational space of each state to converge than EXP, though some overlap must still exist.

The Bennett acceptance ratio is only defined between two states. Usually, the endpoints of interest in a free energy calculation are sufficiently different that we will need a chain of states that gradually change the potential energy function from u_0 to u_1, as discussed in Sec. 7.2.3. You can carry out a BAR estimate between each pair of states $\Delta f_{1\to N} = \Delta f_{1\to 2} + \Delta f_{2\to 3} + \ldots + \Delta f_{N-1\to N}$.

There is one important thing to note about the uncertainty estimates when summing multiple free energies together to calculate an overall free energy estimate. Although BAR itself gives a free energy estimate that is asymptotically correct and is much less biased than the uncertainty estimate for EXP, the uncertainties in $\Delta f_{i-1\to i}$ and $\Delta f_{i\to i+1}$ are not uncorrelated, because they both involve the energies $u_i(\overrightarrow{q})$. The variances of each of the free energies will *not* propagate as variances usually do (in quadrature) into the variance of the overall free energy. Instead, some other method for propagating the uncertainty, such as bootstrapping [42] must be used.

Thermodynamic integration (TI)—By taking the derivative of the free energy with respect to the variable $\overrightarrow{\lambda}$, we find that:

$$\frac{df}{d\vec{\lambda}} = \frac{d}{d\vec{\lambda}}\left[-\ln \int \frac{\exp - u(\vec{\lambda},\vec{q})}{Z(\vec{\lambda})} d\vec{q}\right] = \left\langle \frac{du(\vec{\lambda},\vec{q})}{d\vec{\lambda}} \right\rangle_{\vec{\lambda}}. \tag{28}$$

And then we can numerically integrate $df/d\vec{\lambda}$ over an alchemical transformation, using a range of different well-established techniques, to obtain:

$$\Delta f = \int_0^1 \left\langle \frac{du(\vec{\lambda},\vec{q})}{d\vec{\lambda}} \right\rangle_{\vec{\lambda}} d\vec{\lambda}. \tag{29}$$

This approach to calculating the free energy is called thermodynamic integration (TI). Averaging over $\left\langle \frac{du}{d\vec{\lambda}} \right\rangle$ requires fewer uncorrelated samples to reach a given level of relative error than averaging $e^{-u(\vec{q})}$, as the distribution of values is usually narrower, with a more Gaussian shape to the distribution. Rather than being limited by overlap, as in the case of BAR and MBAR (see below), we are instead limited by the bias in the numerical quadrature, which must be minimized sufficiently to be beneath the level of statistical noise.

Various numerical integration schemes are possible, but the trapezoid rule provides a simple and robust scheme. All types of numerical integration can be written as:

$$\Delta f \approx \sum_{k=1}^{K} w_k \left\langle \frac{du(\vec{\lambda},\vec{q})}{d\vec{\lambda}} \right\rangle_k,$$

where the weights w_k correspond to a particular choice of numerical integration. Researchers have tried a large number of different integration schemes [221–223]. However, many integration choices require specific choices of $\vec{\lambda}$ to minimize bias, which makes them unsuitable when the intermediates have widely-varying levels of uncertainty. For example, integrating a cubic spline interpolation provided negligible benefits over a simple trapezoid rule [224]. As fitting to higher order polynomials can have numerical instabilities for some energy functions, and because alternate functional forms might only be appropriate with some types of transformations, expertise and experience is required to perform such numerical integration modifications. For starting researchers, we therefore recommend the simple trapezoid rule scheme, as it allows for maximal flexibility in which values of $\vec{\lambda}$ are simulated. In practice, adding 2–3 more intermediate states is typically sufficient to match the performance of these more complicated numerical quadrature schemes. It is also possible to calculate the λ-derivatives at non-simulated states from simulated states in a scheme named extended TI [225] which reduces the integration error.

One drawback of TI is that it requires derivatives with respect to $\vec{\lambda}$ to be calculated directly in the code. Unfortunately, many problems of interest require using pathways (such as the soft-core pathways, for removing repulsive interactions) that are not linear, as we discuss, making this more complex. Still, if the code of interest does compute $\frac{du}{d\overrightarrow{\lambda}}$, then TI is perhaps the simplest method to use, as it involves a very little post-processing effort.

The multistate Bennett acceptance ratio (MBAR)—One can generalize Bennett's logic from two states to multiple states to obtain a free energy estimator that uses energy differences between configurations at all intermediate states to compute free energy differences between all states. MBAR gives a system of implicit equations for the free energies f_i:

$$f_i = -\ln \sum_{n=1}^{N} \frac{\exp\left(-u_i\left(\overrightarrow{q}_n\right)\right)}{\sum_{k=1}^{K} N_k \exp\left(f_k - u_k\left(\overrightarrow{q}_n\right)\right)}, \tag{30}$$

where there are N_k samples from each of K states, with $\sum_k N_k = N$ the total number of samples. Thus, we need to evaluate the energy function u_i for all samples obtained at all states in the transformation. The equations can be solved by a number of different standard routines. We note that there are only $K - 1$ independent equations, so only $K - 1$ of the free energies are independent variables, and one of the f_i must be specified (usually, without loss of generality, setting it to zero).

MBAR is probably the lowest variance asymptotically unbiased estimator of the free energy given the energies of the samples [226], which means that BAR is also the lowest variance estimator for the free energy difference between only 2 states, as it is mathematically exactly the same as MBAR in this case. MBAR also provides an uncertainty estimate, derived from standard error propagation methods for implicit functions, which has been shown to be highly accurate as long as there are sufficient samples at each state [224].

MBAR can also be thought of as the Zwanzig estimator of the free energy to state i where the sampled distribution is the *mixture distribution* of all the other samples thrown together in one "pot", defined by $p_m(\overrightarrow{q}) = N^{-1}\sum_k N_k \exp\left(f_k - u_k \overrightarrow{q}\right)$, which is the weighted average of all the individual normalized probability distributions from the simulations that are performed [227].

Recommendations

- We recommend MBAR if all energy differences are available. It is the lowest variance unbiased free energy estimate given samples from multiple states.
- BAR is essentially just as good as MBAR for highly optimized $\overrightarrow{\lambda}$ intermediates. Specifically, if the $\overrightarrow{\lambda}$s are chosen such that intermediate states have moderate overlap with their neighbors (i.e. between i and $i + 1$ and between i and $i - 1$, they will *not* have significant overlap with their next nearest neighbors $i + 2$ and $i - 2$. Thus MBAR does not actually get significant information from these energy differences, so one might as well not even calculate them, and just perform BAR between nearest neighbors. [224]

- TI usually gives similar values as MBAR implemented with sufficient numbers of intermediates, but quadrature errors that are hard to estimate beforehand can occur if one is not careful. [224]
- WHAM is an approximation to MBAR, and there are no compelling reasons it should be used. If careful, it is not necessarily much worse than the other methods, but it always introduces some degree if binning error.
- Other variants, especially ones that adaptively determine the free energies can be useful in certain circumstances but beyond the scope of a Best Practices article.

8.4 Uncertainty estimation

It is important to consider the variation in your computed free energies from your equilibrium simulations, in order to obtain an estimate of uncertainty of the obtained value for the free energies of interest. A recent Best Practices paper by Grossfield et al., [42] provides substantial detail on how to estimate uncertainties from molecular simulations and is a good starting point for this topic. The uncertainty of a free energy estimate is limited however, as it is only an estimate of the configurations sampled and cannot contain any information from the phase space not sampled during a simulation. As a consequence, the variance in free energy afforded by any estimator will always be an underestimate of the true uncertainty. In general, the quantification of different error metrics depends on both data generation and analysis methods used from the ones discussed above.

The computation of free energies using TI (Sec.n 8.3) is straightforward and the trapezoidal rule is often recommended since it allows unequal spacing of $\vec{\lambda}$ states, which is required to minimize the variance in the free energy estimate, but in principle any good numerical integration method can be used. The determination of regions of high curvature when estimating the integral is helpful to determine regions of phase space where more sampling and/or more $\vec{\lambda}$ states are necessary to obtain the best approximation of the integral. Plotting $\vec{\lambda}$ with respect to the gradients at each of the $\vec{\lambda}$ values can be be a helpful diagnostic. Additionally, computation of the overall variance of TI requires the calculation of the overall variance of integration, rather than each individual $\Delta G_{i,i+1}$ and assuming variances add independently. Therefore, $\mathrm{var}(\Delta \mathrm{f}) = \sum_{i=1}^{K} w_k^2 \mathrm{var}\left(\frac{du}{d\vec{\lambda}}\right)_k$.

For alchemical changes that result in smooth, low curvature sets of $\left\langle \frac{dU}{d\vec{\lambda}} \right\rangle$, a relatively small number of $\vec{\lambda}$ states is necessary for sufficient accuracy and low variance in the free energy estimate. Depending on the difficulty of the perturbation, the bias introduced by discretization of the integral can become large due to increased curvature, and more $\vec{\lambda}$ intermediate states become necessary to reduce error. It is recommended that researchers verify that a sufficient number of states are included such that the free energy is essentially invariant to the number of lambda intermediate states chosen. Good heuristics or measures to assess the ‘difficulty’ of a given perturbation is still an ongoing research topic.

Compared with TI, the MBAR method (Sec. 8.3) discussed above provides uncertainty estimation directly from solving a set of linear equations to compute the variances between all states. The number of states and amount of sampling should be optimized to minimize the uncertainty in the MBAR free energy estimate, while balancing other key considerations such as computational expense.

If possible, it is advisable to analyze the same set of simulations with different estimators, providing an opportunity for synergy. If different estimators agree the free energy estimate is more reliable than if there are differences between methods that are larger than 1 kcal/mol and would indicate poor convergence.

Uncertainty can also be assessed for a particular perturbation by repeating calculations with slight changes in initial configurations, forcefield parameters, and different random seeds in the MD engine. The assessment of variability in free energy calculations due to repeating simulations has been previously reported [11, 16, 162, 224], and large variance in free energies estimated from simulations with different random seeds should be flagged as issues with convergence.

For relative binding free energy calculations, additional sensitivity analysis can be performed by changing the initial configurations of non-core regions of the perturbation topology and determining if this change in configurations results in a large differences in the computed relative free energy, indicating poor sampling of ligand configuration. The proposed changes in configuration are increasingly relevant if no experimental evidence is available to reduce uncertainty in where the changing atoms should be positioned.

In addition to statistical uncertainty and sampling, a variety of other factors can impact results from binding free energy calculations. In addition to the choice of initial configuration, results can depend on the choice of force field for the protein/receptor, water, and small molecule(s), so rerunning calculations with different choices of force field can also be used to assess how sensitive results and conclusions are to these particular choices. Other factors, like system preparation (choice of protonation state, tautomer, counterion presence, salt concentration, etc.) can also substantially impact results [228, 229], so unless modelers are confident they have these factors correct, sensitivity to these choices may also need to be examined.

8.5 Are my simulations any good?

There are different easily measurable indicators that can test how well converged simulations are, and if all alchemical states have been sufficiently sampled for a rigorous analysis. Furthermore, once you have established that individual perturbations are well behaved, there are some tricks to ensure the overall perturbation network gives reliable results.

Convergence of simulations—Fig. 12 illustrates how looking at the convergence of your data may be important. The CB8 host with guest G3 has a longer correlation time than the G6 guest in the octa acid (OA) host. In some cases, slow correlation time may not be expected and therefore not a feature known in advance. To this end, you should always look at all simulation data available and check convergence behaviour for each free energy estimate. Comparing the free energy trajectories as a function of the simulation time in the forward and reverse time direction is a useful convergence test [43]. As shown in Fig. 12, disagreements larger than 1 $k_B T$ in the final part of the forward and reverse trajectories can be useful to detect unconverged results (see also Fig. 9). In this case, one can extend the simulations or try an approach that requires simulations in two separate binding modes where they interconvert at very slow timescales.

Overlap matrix—One way of assessing reliability of the calculations is checking the phase space overlap between neighboring $\vec{\lambda}$-windows [75, 76]. For this purpose, a so-called overlap matrix $\mathcal{O}$ can be used. $\mathcal{O}$ is a $K \times K$ matrix, with K being the number of simulated states, i.e. values of $\vec{\lambda}$. Sufficient overlap is important for reweighting estimators such as BAR or MBAR, but cannot help assess reliability of estimates when using TI. These matrices are graphical representations of the phase space overlap, i.e. the average probability that a sample generated at state $\vec{\lambda}_j$ can be observed at state $\vec{\lambda}_i$. As this probability is computed considering the samples from all states, and not just the adjacent states, the values in each row and column add up to 1. In this analysis, the goal is to ensure every state has overlap with its neighbors in both directions, indicated by off-diagonal elements that are sufficiently larger than zero. For accurate calculations, the matrix should be at least tridiagonal.

Details on the calculation and properties of these matrices can be found elsewhere [43]. In an overlap matrix $\mathcal{O}$, the off-diagonal values ($O_{i,j,i \neq j}$) are negatively correlated with the variance of the free energy difference. Accordingly, the uncertainty of the free energy difference between the states i and j will be smaller when $O_{i,j,i \neq j}$ is larger (and thus the values in the main diagonal ($O_{i,j,i=j}$) are smaller). In order to obtain a reliable estimate of the free energy all neighbouring states must be connected, i.e. there must be sufficient overlap between the samples of these states, such that $O_{i,j,i \neq j} \geq$ threshold). However, due to the mathematical derivation it is difficult to explicitly describe the relation of the overlap matrix and the variance by formulae. Consequently, the threshold has to be derived empirically. It has been proposed that the values of the first off-diagonals (i.e. the diagonals above and below the main diagonal) should at least be 0.03 to obtain a reliable free energy estimate [43]. Smaller values should be considered as a warning sign (see Fig. 13C), as the variance tends to be underestimated in case of poor overlap.

Fig. 13A, B, and C shows examples of good, mediocre, and poor overlap respectively. For Fig. 13 A, the probability to find a sample from state i in its neighbouring state j is about 0.2 for all states adjacent to the main diagonal, and hence the overall connectivity is good. In the case of Fig. 13 B, the overlap is strongly diminishing in the lower right corner, raising concerns regarding the reliability of the free energy estimate obtained. For Fig. 13 C, the state at $\vec{\lambda}$ index = 6 is connected to neither of its neighbouring states. While this does not necessarily imply that the result for this perturbation is wrong, the energy estimate must at least be considered as highly unreliable. In order to overcome the issue of poor overlap in this example, additional sampling should be performed by introducing additional states, i.e. $\vec{\lambda}$ values.

Interestingly, as the variance is inversely correlated with the number of states [43], it can in principle be reduced below any arbitrary threshold with enough simulation time and a large enough number of $\vec{\lambda}$ windows. However, decreasing the variance to a value close to 0 is not feasible, as this approach would significantly increase the calculation time. While variance can be decreased by increasing simulation length, if the overlap between states is known to be poor, increasing the number of $\vec{\lambda}$ values, or adjusting the spacing of those values to better cover regions of poor overlap will likely provide a larger immediate impact. Different approaches are described in Sec. 7 and more details can be found in the literature [230, 231].

Cycle closure error—Relative free energy calculations, which compute the change in free energy on making a change to a molecule (e.g. adding a functional group to a ligand) may provide an additional opportunity for error/consistency checking. Particularly, such calculations are often done to span a graph or tree of free energy calculations [96, 98, 136]. In some cases the free energy change to go between molecules A and B can be obtained via multiple transformation pathways. This allows a type of consistency checking where we assess how much the free energy change for that transformation in practice differs from equivalence.

Significant deviations of agreement from the same transformation by different routes typically indicate insufficient configurational sampling along the lambda schedule of one or more of the transformations involved. This approach may be generalised to sets of connected transformations given the requirement that the sum of free energy changes along edges of a closed cycle should be zero. This analysis is called “cycle closure”. In practice, such thermodynamic cycles do not actually sum to zero, and deviations become increasingly large as the size of the cycle increases owing to propagation of error. Though no firm guidelines have emerged, it may be judicious to perform additional configurational sampling along edges of a network that are involved in cycles closing poorly. This may be done by extending the duration of simulations, or by averaging free energy changes over multiple repeats. The latter approach may yield more reproducible free energy changes, but at the expense of a stronger bias on the estimated free energies due to repeated use of the same input coordinates.

A scheme to reduce cycle closure errors is used in FEP+ whereby calculated free energy changes along the nodes of the network are re-sampled assuming estimates of the calculated free energy change along a node may be obtained from a Gaussian distribution centered on the estimated free energy change and with a standard deviation equal to the estimated standard deviation of the free energy change. The procedure then uses a maximum likelihood method to find new sets of free energy changes that minimize cycle closure errors [98]. An alternative approach computes the free energy change between a target and reference compound as a weighted average over all unique paths in the network, with the weights derived from the propagated uncertainties of each node [16]. Approaches as illustrated by Yang et al. for perturbation map design can also be used to compute relative free energies between target and reference compounds [95].

Reversible binding simulations—An even more stringent test of the correctness of binding free energy calculations is to compare the results to the equilibrium binding constants derived from long timescale reversible binding simulations [62]. For small ligands with millimolar affinities, repeated binding to and unbinding from the protein can occur for a large number of times in a sufficiently long unbiased MD simulation (10–100 μs), and the equilibrium binding constants can be computed from the ratio of bound to unbound fractions of the simulation time. The agreement between the binding free energy calculations and the reversible binding simulations—given the same system preparation and the same force field parameters—will strongly support the correctness of both calculations, as the same results are arrived at by two independent methods, and any discrepancy will suggest some systematic error in one, or both, of the two methods. As part of validation testing of alchemical free energy codes a benchmark set to compare alchemical and direct computation of equilibrium binding constant should become standard in future.

8.6 Common issues to watch out for during analysis

It is important to carefully examine output data for common problems. Some of the most important things to check for are:

- **Sampling of the binding site by the ligand:** Make sure the ligand samples the binding site reasonably tightly for its expected potency and fit, and that it does not depart out of binding site in the coupled end state if it is a moderate to strong binder.
- **Consistency of free energy estimates across different estimators** Significant discrepancies, meaning results that further outside the mutual error estimates than would be plausible statistically, between free energies calculated with different free energy estimators such as TI, BAR, and MBAR. All of these estimators converge to the same results with sufficient sampling. Differences between them indicate poor overlap or errors in processing.
- **Have replicas mixed well?** Poor replica mixing (for replica-exchange) or λ-space sampling for single-replica methods. If the system is not mixing between states, then the states are insufficiently close for mixing, or else there are bottlenecks in the configurational sampling that limit the accuracy.
- **Behaviour of correlation times:** Correlation time that does not vary relatively smoothly as a function of $\vec{\lambda}$. Discontinuities in correlation time with $\vec{\lambda}$ indicate that the system is sampling significantly different configurations with only small changes to the Hamiltonian changes. This usually indicates sampling problems.
- **Dependence of the free energy on initial configuration of the system**. Ensemble average properties should not depend on the starting point.
- **Torsional sampling** Torsions with multiple low-energy minima where some of these minima are visited rarely or not at all. Which torsions have low energy minima can best be found by comparing to the simulation in the solvent. There should be clear physical reasons that simulation in the complex has different torsional distributions that the ligand in the solvent.

- **Free energy dependence on** $\vec{\lambda}$ The free energy difference between states should vary relatively smoothly with $\vec{\lambda}$. If it varies drastically, then either there need to be finer sampling in $\vec{\lambda}$ in this region, or there are sampling problems there.
- **Convergence of free energy** The free energy should clearly converge as a function of simulation time (Fig. 9).
- If using non-equilibrium methods, **is the result independent of the speed at which the non-equilibrium change is performed**? Non-equilibrium methods are in theory independent of the switching time in the limit of good sampling unless the switching time is simply too short.
- **Visualization of data** In general, inspect output data such as energies and visualize the simulation trajectories and assess if they match your expectations. Many issues can be spotted by a straight forward visualization.

8.7 Best practices for reporting data

Following best practices for data generation and their analysis does not mean that data is reported in the optimal way. As a practitioner of alchemical free energy simulations you also should use best practices for reporting and plotting your results. We encourage the following standard set of analyses and ways to represent data.

Statistics to include in reporting data—As with any modelling technique, misuse of statistical analysis can skew the perception of how well models perform in free energy predictions. First, error estimates should always be included on your predictions in whatever form you present your data (scatterplots, barplots, etc; see next paragraph). We recommend performing triplicates of your predictions at minimum, with starting points that are expected to be uncorrelated, to ensure some measure of reliability in your data. This replication may seem excessive, but uncertainty estimates often underestimate the true statistical uncertainty. Where performing multiple replicas of the simulation is not possible, an error estimate from e.g. MBAR can be used, though bearing in mind this is likely an underestimated error.

As alchemical free energy methods are used in drug discovery to quantify and rationalise structure activity relationships (SAR), the models ability to (a) correlate well with experiment and (b) rank-order the molecules by affinity, should both be computed. Conventionally, this means including an R^2 (or Pearson's R), where $R = +1$ means high correlation, $R = 0$ means no correlation, and $R = -1$ means high anti-correlation) and a Kendall τ (with perfect ranking agreement when τ=1 and perfect disagreement when τ=−1) metric in your results. Additionally, practitioners may choose to include a Spearman ρ as well. Brown et al. [232] have provided a useful analysis in terms of upper bounds of expected possible correlations between experiment and computation with a given potency range for the compounds. For example, for potency ranges of 2 log units it would be impossible to get a higher correlation in R than 0.8 because of experimental uncertainties [232]. What often is neglected to include is an error analysis on correlation statistics that arise from the errors of both experimental and computed data. One way to include such error analysis for correlation metrics is using bootstrapping on the datasets. The D3R community challenges follows best practices on their data evaluation with readily available python scripts online [233], based on work by Pat Walters [234]. Other analysis software also provide similar functionality for bootstrapping datasets [235].

Mean unsigned error (MUE, also called mean absolute error/MAE) is another key statistic to include in your results. Even though some models' near-perfect correlation and ranking statistics might suggest excellent accuracy, MUE values can still have errors of multiple kcal/mol, providing important additional insight into performance. Furthermore, MUE allows for unbiased comparisons between predictive models as it is less sensitive to dataset size. Other metrics such as Gaussian Random Affinity Model (GRAM) [236], Predictive Interval (PI) and Relative Absolute Error (RAE), attempt to correct for the inherent potency range of a dataset, which can aid in comparing success between different targets. We recommend further reading on evaluation of computational models [232, 234, 237, 238].

Reporting the results of relative free energy calculations requires care. As shown in Fig. 5, relative free energies can be performed arbitrarily as a forward or a reverse process, and thus relative free energies may be reported as either positively or negatively valued. The consequence of the two possible signs for relative free energies is that correlation statistics (such as Pearson's R and Kendall τ) can be skewed depending on which sign is analysed. The issue of this inconsistency can be circumvented by either plotting all datapoints within a consistent quadrant [90], or by avoiding the use of correlation statistics for assessment of relative free energy calculations and instead measuring accuracy using RMSE and MUE, which are unaffected by choice of sign.

Presenting your data—As essentially all alchemical free energy prediction schemes are regression problems, the preferred type of plot is a scatter plot (see Fig. 14). Most alchemical free energy projects will look at 10–50 ligands. Any study with <10 ligands is more suitable for bar plots (with inclusion of error bars), and is unlikely to provide meaningful statistics. Any study with >50 ligands typically contains multiple protein targets to which alchemical free energies may perform better on some targets than others. Because of this, it is bad practice to place multiple datasets on the same plot as this can suggest high model accuracy even though the individual models perform less well [238].

As we are interested mainly in the linear relationship between the alchemical free energy predictions and the experimentally-determined affinity values, plots should be depicted with the same range on both axes (i.e. $x = y$) with a 1:1 aspect ratio, with units for both experiment and simulation converted to be the same. If this skews the plot to a point where it is difficult to read of information, using the same dimensions, such that e.g. 1 cm is 1 kcal/mol is acceptable. Furthermore, bounds should be depicted for the 1- and 2-kcal/mol confidence regions. These regions can serve as tools to communicate your model performance: any predictions inside the 1 kcal/mol region can be seen as highly reliable, any predictions inside the 2 kcal/mol region should be seen as somewhat reliable, and any predictions outside the confidence regions should be expected to be unreliable and handled as outliers. In a drug discovery context, this type of data depiction may suggest the reliability of alchemical FE predictions in the project, and can give an idea of how trustworthy predictions can be for synthesis ideas. It is also recommended to included experimental error bars in all plots.

An example of a best practice scatter comparison between computed and experimental values is shown in Fig. 14, highlighting outliers, error bars and confidence intervals. The data for this plot is artificially generated for illustration purposes.

9 Conclusions

Alchemical free energy calculations have seen a vast increase in popularity both in academic research as well as pharmaceutical industry applications in structure based drug discovery [37, 39, 239]. Commercial products such as FEP+ and Flare, which provide a convenient user interface make the setup and use of these methods much easier [15, 19], but this convenience comes with less flexibility in terms of choice of simulation protocols. It is also important to understand the current limitations of the methodology to recognise when automated workflow tools can be used effectively for a given protein target and when they are likely to fail still. Prospective prediction challenges such as the Drug Design Data resource grand challenges provide a community driven platform to evaluate different free energy protocols against each other on blinded targets [240, 241]. Such efforts have highlighted that selection of seemingly identical or similar potential energy function or simulation package does not guarantee production of similar free energies owing to differences in simulation protocols. We hope that the best practice guide provides a set of tools that allow a better understanding of how to setup, run, and reliably interpret alchemical free energy calculations.

10 Selection of available software packages

There are many different software solutions available for the setup, running, and analysis of alchemical free energy calculations. These will vary in customizability and ways in which they are ran, e.g. graphical user interface versus command line tool or python script. The following provides a non-exhaustive list of commercial and noncommercial tools available for conducting alchemical free energy calculations.

Simulation software: Commercial

- FEP+ is a tool offered by Schrödinger Inc. under a commercial license. It has an intuitive GUI which makes it easier for non-experts to run alchemical free energy calculations and analyze the results. It runs the DESMOND MD package under the hood and hence parallelizes well on GPUs [15].
- Flare is a commercial structure-based drug design software offered by Cresset. Similar to FEP+ it has an easily accessible graphical user interface and strives to facilitate free energy calculations for non-experts while offering advanced users full control via a Python API. It only runs on GPUs, using CUDA or OpenCL [19]. It is build on top of the open source software packages Sire and BioSimSpace (cf. below).

- The molecular operating environment (MOE) offered by the Chemical Computing Group (CCG) has a tool for performing free energy calculations. It is built on AMBER-TI (cf. below).

All the above tools also provide a convenient setup and analysis suite and are really a one in all product.

Simulation software: Free/low cost academic and Commercial

- CHARMM has a variety of tools developed over the years. The PERT module can be used to define initial and final states and define the intermediate lambda points. FREN and BAR modules can be used to analyze the data after the MD run. Lambda-dynamics-based free energy calculation can be carried out using the BLOCK module.
- AMBER, including its new pmemd.cuda version supports free energy calculations [242].
- GROMOS offers an extensive and flexible molecular dynamics and simulations analysis suites with free energy calculation functionalities including customizable alchemical paths and various sampling protocols [243–245].

Simulation software: Open Source

- PLUMED is a tool which enables the usage of a variety of MD engines. It is designed as a plugin for MD packages such that it analyzes the trajectory on the fly. It also offers a VMD based plugin for the computation of collective variables [246].
- BioSimSpace is a multiscale molecular simulation framework, written to allow computational modellers to quickly prototype and develop new algorithms for molecular simulation and molecular design [34].
- Sire is a multiscale, molecular simulation framework that provides several applications, including SOMD, an MD/MC code for performing FEP calculations via an interface to OpenMM.
- YANK is a tool developed by John Chodera and group on the top of OpenMM MD package. It allows the users to write their inputs in easy-to-use YAML format.
- GROMACS is a molecular simulation package with a significant number of free energy methods implementations. The LiveCOMS GROMACS tutorial includes an example free energy calculation [247].
- PMX, an add-on to GROMACS, offers a mutation free energy calculation module [248].
- Q is MD code for performing FEP calculations using a variety of force fields [249].

Setup tools:

- PMX: Setup of perturbation maps, perturbed topologies and input coordinates for GROMACS simulations at https://github.com/deGrootLab/pmx.
- Lomap/Lomap2: Relative alchemical transformation graph planning for setting up perturbation networks [136].
- CHARMM-GUI is a web-based tool for setting up a variety of MD simulations. It can be used to generate CHARMM scripts for solvation and ligand-binding free energy calculations [250].
- QligFEP offers robust and fast setup of FEP calculations for the software package Q [93].
- ProtoCaller, a setup tool for the automation of Gromacs free energy calculations [251].
- FESetup has been developed primarily to setup calculations in AMBER, GROMACS and SIRE [91].

Analysis tools:

- Alchemlyb: Multipackage free energy analysis https://github.com/alchemistry/alchemlyb [252].
- pymbar: MBAR implementation, but have to roll your own analysis wrapper https://github.com/choderalab/pymbar [26].
- Arsenic: Standardising alchemical free energy analysis https://github.com/openforcefield/Arsenic
- Free Energy Workflows: Sire-specific free energy map analysis using weighted path averages https://github.com/michellab/freenrgworkflows.

Generally, commercial software will offer more complete pipelines in which standalone analysis applications are not necessarily needed; free and open source packages often require manual analysis but allow more flexibility and modification.

11 Alchemical free energy datasets: an overview

The following contains a non-exhaustive summary of alchemical free energy datasets that can serve as a starting point to review approaches or test new implementations. The field is moving towards a more standardised way of generating protein-ligand benchmark datasets and the progress of these efforts can be tracked here: https://github.com/openforcefield/FE-Benchmarks-Best-Practices. Currently lacking an exhaustive set of benchmark datasets, the review by Williams-Noonan et al. [253] contains an overview of recently published alchemical free energy studies. For comparison of FEP+ and Gromacs (using the AMBER99SB-ILDN and GAFF2 force field), cf. the recently published study by Pérez-Benito et al. [90]. An overview of further suggested benchmark sets can be found in the review by Mobley and Gilson [228] or on alchemistry.org [254]. These include cyclodextrins, the Cytochrome C peroxidase (CCP) protein model binding site, thrombin and bromodomains as well as solvation benchmark sets [224]. Please refer to table 1, for a small overview of datasets, what forcefields they used, and what the original study was it came from.

12 Checklist

KNOW WHAT YOU WANT TO SIMULATE

Initial questions you should ask before you set up an alchemical free energy calculation using molecular dynamics simulations

- □ Do I understand the biology, chemistry and physics of my system?
- □ Have I properly prepared my protein and ligand systems?
- □ Does my system contain any structures that require custom parameters?
- □ What simulation protocol will provide the most evidence to verify my hypothesis?
- □ Are the projected computational expense and runtime realistic for my scientific goals?
- □ Will my protocol be reproducible?
- □ Will my statistics be reliable? If not, would more replicates solve the problem?
- □ Can I open-source my data?

PREPARING YOUR SIMULATIONS

Steps to getting started setting up your alchemical free energy calculation

- □ Make sure you know why you have picked your (combination of) force field(s)
- □ Energy minimize your system
- □ Equilibrate your system properly with your choice of thermodynamic ensemble
- □ Check the stability of your system and whether it behaves the way you believe it should

RUNNING ABSOLUTE SIMULATIONS

Steps to running your absolute alchemical free energy calculations

- □ Check your ligands have the same, biologically correct binding pose
- □ Make sure your λ-scheduling is appropriate
- □ Check if your ligands are discharging and decoupling correctly
- □ Set up your restraints correctly
- □ Make sure you subsample the data in your free energy estimation protocol
- □ Apply the appropriate correction terms

RUNNING RELATIVE SIMULATIONS

Steps to running your relative alchemical free energy calculations

□ Check your ligands have the same, biologically correct binding pose

□ Make sure your λ-scheduling is set correctly

□ Make sure your molecular transformations are realistic (1–5 heavy atoms for reliable computations)

□ Generate a perturbation network by your method of choice; check whether you have enough cycle closures to check consistency in the results

□ Check whether dummy atoms were assigned correctly

□ Consider subsampling the data in your free energy estimation protocol

□ Apply the appropriate correction terms

HOW DO I KNOW WHICH SIMULATIONS ARE UNRELIABLE?

Situations suggesting your relative alchemical free energy calculations have not run properly (assuming absence of experimental affinities)

□ Standard error (σ) should not be >1 kcal·mol^{-1}

□ Simulated systems have not converged - trajectories should be manually checked for consistency; other methods such as generating RMSD plots are also recommended

Relative:

□ If you observe hysteresis in perturbations and incorrect cycle closures

□ Energy differences >~15 kcal·mol^{-1} are likely unreliable

Absolute:

□ Energies <~−15 kcal·mol^{-1} are likely unreliable

□ The ligand has not sampled most of the intended region after the decoupling step

□ The ligand is drifting out of the intended region after the decoupling step

WHY ARE THEY NOT RELIABLE?

Suggestions for finding out why your alchemical free energy calculations may not be reliable

□ Check again whether dummy atoms were assigned correctly

□ Inspect the trajectories across the λ-schedule (particularly the endpoints) for problems described in the text

□ Inspect the overlap matrices for lack of overlap

DATA ANALYSIS

Steps to analyzing your output data correctly

□ Make sure you have run enough replicates to ensure statistical reliability (>3)

□ Compute both correlation and ranking coefficients and ranking statistics (e.g. r, ρ, MUE and τ)

□ Include error bars in all your visual analyses

Funding Information

ASJSM and JM acknowledge funding through an EPSRC flag-ship software grant: EP/P022138/1 MK and JM acknowledge funding through Innovate UK by KTP partnership 011120. AR acknowledges partial support from the Tri-Institutional Program in Computational Biology and Medicine and the Sloan Kettering Institute. HEBM acknowledges support from a Molecular Sciences Software Institute Investment Fellowship and Relay Therapeutics. DLM acknowledges support from the National Institutes of Health (R01GM108889, R01GM124270, and R01GM132386), and the National Science Foundation (CHE 1352608). JDC acknowledges support from the National Institutes of Health (NIH P30 CA008748, NIH R01 GM121505, R01 GM132386). A complete funding history for the Chodera lab can be found at http://choderalab.org/funding. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Acronyms

CPU	Central Processing Unit
BAR	Bennett Acceptance Ratio
FEP	Free Energy Perturbation
GPCR	G-Protein Coupled Receptor
GPU	Graphics Processing Unit
MBAR	Multistate Bennett Acceptance Ratio
MCSS	Maximum Common Substructure
MD	Molecular Dynamics
RMSE	Root Mean Square Error
MUE	Mean Unsigned Error
SAR	Structure-Activity Relationships
TI	Thermodynamic Integration

List of Symbols

L* and *R	generic names for ligand and receptor
K_b°	binding constant
$\boldsymbol{c^\circ}$	standard state concentration

U	potential energy
u	reduced (dimensionless) potential describing a thermodynamic state
G	Gibbs free energy (free energy in the isothermal isobaric ensemble), Gibbs function, or free enthalpy, though the most common term Gibbs free energy is used in the text
A	Helmholtz free energy (free energy in the canonical ensemble) or Helmholtz function, with Helmholtz free energy used in the text.
f	reduced (dimensionless) free energy
$\Delta\hat{f}$	estimate from an estimator for the reduced free energy difference between two states
Γ	configurational space accessible by simulations
$\overrightarrow{q}$	vector of a single configuration, i.e. x, y, z coordinates of the simulation system
k_B	Boltzmann constant
Z	partition function
p	pressure
μ	chemical potential (grand canonical ensemble)
T	temperature
$\beta \equiv (k_B T)^{-1}$	inverse thermal energy
$\overrightarrow{\lambda}$	alchemical progress parameter, which may be multidimensional
g	statistical inefficiency
$\mathcal{O}$	overlap matrix
C_t	discrete-time-normalized fluctuation auto-correlation function
τ_{eq}	integrated auto-correlation time
t_0	equilibration time

References

[1]. Zwanzig RW. High-Temperature Equation of State by a Perturbation Method. I. Nonpolar Gases. J Chem Phys. 1954; 22(8):1420–1426. doi: 10.1063/1.1740409.

[2]. Tembre BL, Mc Cammon JA. Ligand-Receptor Interactions. Comput Chem. 1984; 8(4):281–283. doi: 10.1016/00978485(84)85020-2.

[3]. Rustenburg AS, Dancer J, Lin B, Feng JA, Ortwine DF, Mobley DL, Chodera JD. Measuring Experimental Cyclohexane-Water Distribution Coefficients for the SAMPL5 Challenge. J Comput Aided Mol Des. 2016; 30(11):945–958. doi: 10.1007/s10822016-9971-7. [PubMed: 27718028]

[4]. Bosisio S, Mey ASJS, Michel J. Blinded Predictions of Distribution Coefficients in the SAMPL5 Challenge. J Comput Aided Mol Des. 2016; 30(11):1101–1114. doi: 10.1007/s10822-016-9969-1. [PubMed: 27677751]

[5]. Corey RA, Vickery ON, Sansom MSP, Stansfeld PJ. Insights into Membrane Protein-Lipid Interactions from Free Energy Calculations. J Chem Theory Comput. 2019; 15(10):5727–5736. doi: 10.1021/acs.jctc.9b00548. [PubMed: 31476127]

[6]. Hauser K, Negron C, Albanese SK, Ray S, Steinbrecher T, Abel R, Chodera JD, Wang L. Predicting Resistance of Clinical Abl Mutations to Targeted Kinase Inhibitors Using Alchemical Free-Energy Calculations. Commun Biol. 2018; 1(1):70. doi: 10.1038/s42003-018-0075-x. [PubMed: 30159405]

[7]. Aldeghi M, Gapsys V, de Groot BL. Accurate Estimation of Ligand Binding Affinity Changes upon Protein Mutation. ACS Cent Sci. 2018; 4(12):1708–1718. [PubMed: 30648154]

[8]. Seeliger D, De Groot BL. Protein Thermostability Calculations Using Alchemical Free Energy Simulations. Biophys J. 2010; 98(10):2309–2316. doi: 10.1016/j.bpj.2010.01.051. [PubMed: 20483340]

[9]. Gapsys V, Michielssens S, Seeliger D, de Groot BL. Insights from the First Principles Based Large Scale Protein Thermostability Calculations. Biophys J. 2016; 110(3):368a. doi: 10.1016/j.bpj.2015.11.1985.

[10]. Gapsys V, Michielssens S, Seeliger D, de Groot BL. Accurate and Rigorous Prediction of the Changes in Protein Free Energies in a Large-Scale Mutation Scan. Angew Chem Int Ed. 2016; 55(26):7364–7368. doi: 10.1002/anie.201510054.

[11]. Aldeghi M, de Groot BL, Gapsys V. Accurate Calculation of Free Energy Changes upon Amino Acid Mutation. In: Computational Methods in Protein Evolution Springer; 2019.p. 19–47.

[12]. Mobley DL, Graves AP, Chodera JD, McReynolds AC, Shoichet BK, Dill KA. Predicting Absolute Ligand Binding Free Energies to a Simple Model Site. J Mol Biol. 2007; 371(4):1118–1134. doi: 10.1016/j.jmb.2007.06.002. [PubMed: 17599350]

[13]. Aldeghi M, Heifetz A, Bodkin MJ, Knapp S, Biggin PC. Accurate Calculation of the Absolute Free Energy of Binding for Drug Molecules. Chem Sci. 2015; 7(1):207–218. doi: 10.1039/C5SC02678D. [PubMed: 26798447]

[14]. Aldeghi M, Heifetz A, Bodkin MJ, Knapp S, Biggin PC. Predictions of Ligand Selectivity from Absolute Binding Free Energy Calculations. J Am Chem Soc. 2017; 139(2):946–957. doi: 10.1021/jacs.6b11467. [PubMed: 28009512]

[15]. Wang L, Wu Y, Deng Y, Kim B, Pierce L, Krilov G, Lupyan D, Robinson S, Dahlgren MK, Greenwood J, Romero DL, Masse C, Knight JL, Steinbrecher T, Beuming T, Damm W, Harder E, Sherman W, Brewer M, Wester R, 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(7):2695–2703. doi: 10.1021/ja512751q. [PubMed: 25625324]

[16]. Mey ASJS, Juárez-Jiménez J, Hennessy A, Michel J. Blinded Predictions of Binding Modes and Energies of HSP90-α Ligands for the 2015 D3R Grand Challenge. Bioorg Med Chem. 2016; 24(20):4890–4899. doi: 10.1016/j.bmc.2016.07.044. [PubMed: 27485604]

[17]. Song LF, Lee TS, Zhu C, York DM, Merz KM. Using AMBER18 for Relative Free Energy Calculations. J Chem Inf Model. 2019; 59(7):3128–3135. doi:10.1021/acs.jcim.9b00105. [PubMed: 31244091]

[18]. Gapsys V, Pérez-Benito L, Aldeghi M, Seeliger D, van Vlijmen H, Tresadern G, de Groot BL. Large Scale Relative Protein Ligand Binding Affinities Using Non-Equilibrium Alchemy. Chem Sci. 2020; doi: 10.1039/C9SC03754C.

[19]. Kuhn M, Firth-Clark S, Tosco P, Mey ASJS, Mackey M, Michel J. Assessment of Binding Affinity via Alchemical Free-Energy Calculations. J Chem Inf Model. 2020; 60(6):3120–3130. doi: 10.1021/acs.jcim.0c00165. [PubMed: 32437145]

[20]. Kirkwood JG. Statistical Mechanics of Fluid Mixtures. J Chem Phys. 1935; 3:300–313. doi: 10.1063/1.1749657.

[21]. Jorgensen WL, Ravimohan C. Monte Carlo Simulation of Differences in Free Energies of Hydration. J Chem Phys. 1985; 83(6):3050–3054. doi: 10.1063/1.449208.

[22]. Kollman PA. Free Energy Calculations: Applications to Chemical and Biochemical Phenomena. Chem Rev. 1993; 7:2395–2417. doi: 10.1021/cr00023a004.

[23]. Wong CF, McCammon JA. Dynamics and Design of Enzymes and Inhibitors. J Am Chem Soc. 1986; 108(13):3830–3832. doi: 10.1021/ja00273a048.

[24]. Merz KM, Kollman PA. Free Energy Perturbation Simulations of the Inhibition of Thermolysin: Prediction of the Free Energy of Binding of a New Inhibitor. J Am Chem Soc. 1989; 111(15):5649–5658. doi: 10.1021/ja00197a022.

[25]. Bennett CH. Efficient Estimation of Free Energy Differences from Monte Carlo Data. J Comput Phys. 1976; 22:245–268. doi: 10.1016/0021-9991(76)90078-4.

[26]. Shirts MR, Chodera JD. Statistically Optimal Analysis of Samples from Multiple Equilibrium States. J Chem Phys. 2008; 129(12):124105. doi: 10.1063/1.2978177. [PubMed: 19045004]

[27]. Wu H, Paul F, Wehmeyer C, Noé F. Multiensemble Markov Models of Molecular Thermodynamics and Kinetics. Proc Natl Acad Sci. 2016; 113(23):E3221–E3230. doi: 10.1073/pnas.1525092113. [PubMed: 27226302]

[28]. Mey ASJS, Wu H, Noé F. xTRAM: Estimating Equilibrium Expectations from Time-Correlated Simulation Data at Multiple Thermodynamic States. Phys Rev X. 2014; 4(4):041018. doi: 10.1103/PhysRevX.4.041018.

[29]. Wu H, Mey ASJS, Rosta E, Noé F. Statistically Optimal Analysis of State-Discretized Trajectory Data from Multiple Thermodynamic States. J Chem Phys. 2014; 141(21):214106. doi: 10.1063/1.4902240. [PubMed: 25481128]

[30]. Shirts MR, Pitera JW, Swope WC, Pande VS. Extremely Precise Free Energy Calculations of Amino Acid Side Chain Analogs: Comparison of Common Molecular Mechanics Force Fields for Proteins. J Chem Phys. 2003; 119(11):5740–5761. doi: 10.1063/1.1587119.

[31]. Shirts MR, Pande VS. Solvation Free Energies of Amino Acid Side Chains for Common Molecular Mechanics Water Models. J Chem Phys. 2005; 122:134508. doi: 10.1063/1.1877132. [PubMed: 15847482]

[32]. van der Spoel D, Lindahl E, Hess B, Groenhof G, Mark AE, Berendsen HJC. GROMACS: Fast, Flexible, and Free. J Comput Chem. 2005; 26:1701–1718. doi: 10.1002/jcc.20291. [PubMed: 16211538]

[33]. Mermelstein DJ, Lin C, Nelson G, Kretsch R, McCammon JA, Walker RC. Fast and Flexible Gpu Accelerated Binding Free Energy Calculations within the Amber Molecular Dynamics Package. J Comput Chem. 2018; 39(19):1354–1358. doi: 10.1002/jcc.25187. [PubMed: 29532496]

[34]. 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(43):1831. doi: 10.21105/joss.01831.

[35]. Riniker S, Christ CD, Hansen HS, Hünenberger PH, Oostenbrink C, Steiner D, van Gunsteren WF. Calculation of Relative Free Energies for Ligand-Protein Binding, Solvation, and Conformational Transitions Using the GROMOS Software. J Phys Chem B. 2011; 115(46):13570–13577. doi: 10.1021/jp204303a. [PubMed: 22039957]

[36]. Fratev F, Sirimulla S. An Improved Free Energy Perturbation FEP+ Sampling Protocol for Flexible Ligand-Binding Domains. chemrxivorg. 2019; doi: 10.26434/chemrxiv.6204167.v2.

[37]. Schindler CEM, Baumann H, Blum A, Böse D, Buchstaller HP, Burgdorf L, Cappel D, Chekler E, Czodrowski P, Dorsch D, Eguida MKI, Follows B, Fuchß T, Grädler U, Gunera J, Johnson T, Jorand Lebrun C, Karra S, Klein M, Knehans T, et al. Large-Scale Assessment of Binding Free Energy Calculations in Active Drug Discovery Projects. J Chem Inf Model. 2020; 60(11):5457–5474. doi: 10.1021/acs.jcim.0c00900. [PubMed: 32813975]

[38]. Cournia Z, Allen B, Sherman W. Relative Binding Free Energy Calculations in Drug Discovery: Recent Advances and Practical Considerations. J Chem Inf Model. 2017; 57(12):2911–2937. doi: 10.1021/acs.jcim.7b00564. [PubMed: 29243483]

[39]. Sherborne B, Shanmugasundaram V, Cheng AC, Christ CD, DesJarlais RL, Duca JS, Lewis RA, Loughney DA, Manas ES, McGaughey GB, Peishoff CE, van Vlijmen H. Collaborating to Improve the Use of Free-Energy and Other Quantitative Methods in Drug Discovery. J Comput Aided Mol Des. 2016; 30(12):1139–1141. doi: 10.1007/s10822-016-9996-y. [PubMed: 28013427]

[40]. Salomon-Ferrer R, Götz AW, Poole D, Le Grand S, Walker RC. Routine Microsecond Molecular Dynamics Simulations with AMBER on GPUs. 2. Explicit Solvent Particle Mesh Ewald. J Chem Theory Comput. 2013; 9(9):3878–3888. doi: 10.1021/ct400314y. [PubMed: 26592383]

[41]. Braun E, Gilmer J, Mayes HB, Mobley DL, Monroe JI, Prasad S, Zuckerman DM. Best Practices for Foundations in Molecular Simulations [Article v1.0]. LiveCoMS. 2019; 1(1). doi: 10.33011/livecoms.1.1.5957.

[42]. Grossfield A, Patrone PN, Roe DR, Schultz AJ, Siderius D, Zuckerman DM. Best Practices for Quantification of Uncertainty and Sampling Quality in Molecular Simulations [Article v1.0]. LiveCoMS. 2018; 1(1):5067. doi: 10.33011/livecoms.1.1.5067. [PubMed: 30533602]

[43]. Klimovich PV, Shirts MR, Mobley DL. Guidelines for the Analysis of Free Energy Calculations. J Comput Aided Mol Des. 2015; 29(5):397–411. doi: 10.1007/s10822-015-9840-9. [PubMed: 25808134]

[44]. Shirts MR. Best Practices in Free Energy Calculations for Drug Design. In: Baron R, editor. Computational Drug Discovery and Design Methods in Molecular Biology, New York, NY: Springer; 2012.p. 425–467. doi: 10.1007/978-1-61779-465-0_26.

[45]. Grinter SZ, Zou X. Challenges, Applications, and Recent Advances of Protein-Ligand Docking in Structure-Based Drug Design. Molecules. 2014; 19(7):10150–10176. doi: 10.3390/molecules190710150. [PubMed: 25019558]

[46]. Lameira J, Bonatto V, Cianni L, Rocho FdR, Leitão A, Montanari CA. Predicting the Affinity of Halogenated Reversible Covalent Inhibitors through Relative Binding Free Energy. Phys Chem Chem Phys. 2019; 21(44):24723–24730. doi: 10.1039/C9CP04820K. [PubMed: 31680132]

[47]. Gill SC, Lim NM, Grinaway PB, Rustenburg AS, Fass J, Ross GA, Chodera JD, Mobley DL. Binding Modes of Ligands Using Enhanced Sampling (BLUES): Rapid Decorrelation of Ligand Binding Modes via Nonequilibrium Candidate Monte Carlo. J Phys Chem B. 2018; doi: 10.1021/acs.jpcb.7b11820.

[48]. Genheden S, Ryde U. The MM/PBSA and MM/GBSA Methods to Estimate Ligand-Binding Affinities. Expert Opin Drug Dis. 2015; 10(5):449–461. doi: 10.1517/17460441.2015.1032936.

[49]. Gutiérrez-de-Terán H, Åqvist J. Linear Interaction Energy: Method and Applications in Drug Design. In: Baron R, editor. Computational Drug Discovery and Design Methods in Molecular Biology, New York, NY: Springer New York; 2012.p. 305–323. doi: 10.1007/978-1-61779-465-0_20.

[50]. Heinzelmann G, Henriksen NM, Gilson MK. Attach-Pull-Release Calculations of Ligand Binding and Conformational Changes on the First BRD4 Bromodomain. J Chem Theory Comput. 2017; 13(7):3260–3275. doi: 10.1021/acs.jctc.7b00275. [PubMed: 28564537]

[51]. Loeffler HH, Bosisio S, Duarte Ramos Matos G, Suh D, Roux B, Mobley DL, Michel J. Reproducibility of Free Energy Calculations across Different Molecular Simulation Software Packages. J Chem Theory Comput. 2018; 14(11):5567–5582. doi: 10.1021/acs.jctc.8b00544. [PubMed: 30289712]

[52]. Vassetti D, Pagliai M, Procacci P. Assessment of GAFF2 and OPLS-AA General Force Fields in Combination with the Water Models TIP3P, SPCE, and OPC3 for the Solvation Free Energy of Druglike Organic Molecules. J Chem Theory Comput. 2019; 15(3):1983–1995. doi: 10.1021/acs.jctc.8b01039. [PubMed: 30694667]

[53]. Lopes PEM, Guvench O, MacKerell AD. Current Status of Protein Force Fields for Molecular Dynamics. Methods Mol Biol. 2015; 1215:47–71. doi: 10.1007/978-1-4939-1465-4_3. [PubMed: 25330958]

[54]. Beierlein FR, Michel J, Essex JW. A Simple QM/MM Approach for Capturing Polarization Effects in Protein-Ligand Binding Free Energy Calculations. J Phys Chem B. 2011; 115(17):4911–4926. doi: 10.1021/jp109054j. [PubMed: 21476567]

[55]. Dybeck EC, König G, Brooks BR, Shirts MR. Comparison of Methods To Reweight from Classical Molecular Simulations to QM/MM Potentials. J Chem Theory Comput. 2016; 12(4):1466–1480. doi: 10.1021/acs.jctc.5b01188. [PubMed: 26928941]

[56]. Cave-Ayland C, Skylaris CK, Essex JW. Direct Validation of the Single Step Classical to Quantum Free Energy Perturbation. J Phys Chem B. 2015; 119(3):1017–1025. doi: 10.1021/jp506459v. [PubMed: 25238649]

[57]. Rufa DA, Macdonald HEB, Fass J, Wieder M, Grinaway PB, Roitberg AE, Isayev O, Chodera JD. Towards Chemical Accuracy for Alchemical Free Energy Calculations with Hybrid Physics-Based Machine Learning / Molecular Mechanics Potentials. bioRxiv. 2020; p. 2020.07.29.227959. doi: 10.1101/2020.07.29.227959.

[58]. Scheen J, Wu W, Mey ASJS, Tosco P, Mackey M, Michel J. Hybrid Alchemical Free Energy/Machine-Learning Methodology for the Computation of Hydration Free Energies. J Chem Inf Model. 2020; doi: 10.1021/acs.jcim.0c00600.

[59]. Cole DJ, Mones L, Csányi G. A Machine Learning Based Intramolecular Potential for a Flexible Organic Molecule. Faraday Discuss. 2020; doi: 10.1039/D0FD00028K.

[60]. Gilson MK, Given JA, Bush BL, McCammon JA. A StatisticalThermodynamic Basis for Computation of Binding Affinities: A Critical Review. Biophys J. 1997; 72(3):1047–1069. doi: 10.1016/S0006-3495(97)78756-3. [PubMed: 9138555]

[61]. Jong DHD, Schäfer LV, Vries AHD, Marrink SJ, Berendsen HJC, Grubmüller H. Determining Equilibrium Constants for Dimerization Reactions from Molecular Dynamics Simulations. J Comput Chem. 2011; 32(9):1919–1928. doi: 10.1002/jcc.21776. [PubMed: 21469160]

[62]. Pan AC, Xu H, Palpant T, Shaw DE. Quantitative Characterization of the Binding and Unbinding of Millimolar Drug Fragments with Molecular Dynamics Simulations. J Chem Theory Comput. 2017; 13(7):3372–3377. doi: 10.1021/acs.jctc.7b00172. [PubMed: 28582625]

[63]. Teo I, Mayne CG, Schulten K, Lelièvre T. Adaptive Multilevel Splitting Method for Molecular Dynamics Calculation of Benzamidine-Trypsin Dissociation Time. J Chem Theory Comput. 2016; 12(6):2983–2989. doi: 10.1021/acs.jctc.6b00277. [PubMed: 27159059]

[64]. Votapka LW, Jagger BR, Heyneman AL, Amaro RE. SEEKR: Simulation Enabled Estimation of Kinetic Rates, A Computational Tool to Estimate Molecular Kinetics and Its Application to Trypsin–Benzamidine Binding. J Phys Chem B. 2017; 121(15):3597–3606. doi: 10.1021/acs.jpcb.6b09388. [PubMed: 28191969]

[65]. Doerr S, De Fabritiis G. On-the-Fly Learning and Sampling of Ligand Binding by High-Throughput Molecular Simulations. J Chem Theory Comput. 2014; 10(5):2064–2069. doi: 10.1021/ct400919u. [PubMed: 26580533]

[66]. Plattner N, Noé F. Protein Conformational Plasticity and Complex Ligand-Binding Kinetics Explored by Atomistic Simulations and Markov Models. Nat Commun. 2015; 6(1):1–10. doi: 10.1038/ncomms8653.

[67]. Dixon T, Lotz SD, Dickson A. Predicting Ligand Binding Affinity Using On- and off-Rates for the SAMPL6 SAMPLing Challenge. J Comput Aided Mol Des. 2018; 32(10):1001–1012. doi: 10.1007/s10822-018-0149-3. [PubMed: 30141102]

[68]. Basavapathruni A, Jin L, Daigle SR, Majer CRA, Therkelsen CA, Wigle TJ, Kuntz KW, Chesworth R, Pollock RM, Scott MP, Moyer MP, Richon VM, Copeland RA, Olhava EJ. Conformational Adaptation Drives Potent, Selective and Durable Inhibition of the Human Protein Methyltransferase DOT1L. Chemical Biology & Drug Design. 2012; 80(6):971–980. doi: 10.1111/cbdd.12050. [PubMed: 22978415]

[69]. Hyre DE, Trong IL, Merritt EA, Eccleston JF, Green NM, Stenkamp RE, Stayton PS. Cooperative Hydrogen Bond Interactions in the Streptavidin–Biotin System. Protein Sci. 2006; 15(3):459–467. doi: 10.1110/ps.051970306. [PubMed: 16452627]

[70]. Eastman P, Swails J, Chodera JD, McGibbon RT, Zhao Y, Beauchamp KA, Wang LP, Simmonett AC, Harrigan MP, Stern CD, Wiewiora RP, Brooks BR, Pande VS. OpenMM 7: Rapid

Development of High Performance Algorithms for Molecular Dynamics. PLoS Comput Biol. 2017; 13(7):e1005659. doi: 10.1371/journal.pcbi.1005659. [PubMed: 28746339]

[71]. Kutzner C, Páll S, Fechner M, Esztermann A, de Groot BL, Grubmüller H. More Bang for Your Buck: Improved Use of GPU Nodes for GROMACS 2018. J Comput Chem. 2019; 40(27):2418–2431. doi: 10.1002/jcc.26011. [PubMed: 31260119]

[72]. Woo HJ, Roux B. Calculation of Absolute Protein–Ligand Binding Free Energy from Computer Simulations. Proc Natl Acad Sci. 2005; 102(19):6825–6830. doi: 10.1073/pnas.0409005102. [PubMed: 15867154]

[73]. Velez-Vega C, Gilson MK. Overcoming Dissipation in the Calculation of Standard Binding Free Energies by Ligand Extraction. J Comput Chem. 2013; 34(27):2360–2371. doi: 10.1002/jcc.23398. [PubMed: 24038118]

[74]. Limongelli V, Bonomi M, Parrinello M. Funnel Metadynamics as Accurate Binding Free-Energy Method. Proc Natl Acad Sci. 2013; 110(16):6358–6363. doi: 10.1073/pnas.1303186110. [PubMed: 23553839]

[75]. Wu D, Kofke DA. Phase-Space Overlap Measures. II. Design and Implementation of Staging Methods for Free-Energy Calculations. J Chem Phys. 2005; 123(8):084109. doi: 10.1063/1.2011391. [PubMed: 16164284]

[76]. Wu D, Kofke DA. Phase-Space Overlap Measures. I. FailSafe Bias Detection in Free Energies Calculated by Molecular Simulation. J Chem Phys. 2005; 123(5):054103. doi: 10.1063/1.1992483. [PubMed: 16108627]

[77]. Shirts MR, Bair E, Hooker G, Pande VS. Equilibrium Free Energies from Nonequilibrium Measurements Using Maximum-Likelihood Methods. Phys Rev Lett. 2003; 91(14):140601. doi: 10.1103/PhysRevLett.91.140601. [PubMed: 14611511]

[78]. Jarzynski C Nonequilibrium Equality for Free Energy Differences. Phys Rev Lett. 1997; 78(14):2690–2693. doi: 10.1103/PhysRevLett.78.2690.

[79]. Jarzynski C Equilibrium Free Energies from Nonequilibrium Processes. Act Phys Pol B. 1998; 29(6):1609–1622.

[80]. Crooks GE. Path-Ensemble Averages in Systems Driven Far from Equilibrium. Phys Rev E. 2000; 61(3):2361–2366. doi: 10.1103/PhysRevE.61.2361.

[81]. Harder E, Damm W, Maple J, Wu C, Reboul M, Xiang JY, Wang L, Lupyan D, Dahlgren MK, Knight JL, Kaus JW, Cerutti DS, Krilov G, Jorgensen WL, Abel R, Friesner RA. OPLS3: A Force Field Providing Broad Coverage of Drug-like Small Molecules and Proteins. J Chem Theory Comput. 2016; 12(1):281–296. doi: 10.1021/acs.jctc.5b00864. [PubMed: 26584231]

[82]. Keränen H, Pérez-Benito L, Ciordia M, Delgado F, Steinbrecher TB, Oehlrich D, van Vlijmen HWT, Trabanco AA, Tresadern G. Acylguanidine Beta Secretase 1 Inhibitors: A Combined Experimental and Free Energy Perturbation Study. J Chem Theory Comput. 2017; 13(3):1439–1453. doi: 10.1021/acs.jctc.6b01141. [PubMed: 28103438]

[83]. Rizzi A, Jensen T, Slochower DR, Aldeghi M, Gapsys V, Ntekoumes D, Bosisio S, Papadourakis M, Henriksen NM, De Groot BL, et al. The SAMPL6 SAMPLing challenge: Assessing the reliability and efficiency of binding free energy calculations. BioRxiv. 2019; p. 795005.

[84]. Schindler T, Bornmann W, Pellicena P, Miller WT, Clarkson B, Kuriyan J. Structural Mechanism for STI-571 Inhibition of Abelson Tyrosine Kinase. Science. 2000; 289(5486):1938–1942. doi: 10.1126/science.289.5486.1938. [PubMed: 10988075]

[85]. Mobley DL, Klimovich PV. Perspective: Alchemical Free Energy Calculations for Drug Discovery. J Chem Phys. 2012; 137(23). doi: 10.1063/1.4769292.

[86]. Brown S, Shirts M, Mobley D. Free Energy Calculations in Structure-Based Drug Design. In:; 2010.p. 61–86.

[87]. Kuhn B, Tichý M, Wang L, Robinson S, Martin RE, Kuglstatter A, Benz J, Giroud M, Schirmeister T, Abel R, Diederich F, Hert J. Prospective Evaluation of Free Energy Calculations for the Prioritization of Cathepsin L Inhibitors. J Med Chem. 2017; 60(6):2485–2497. doi: 10.1021/acs.jmedchem.6b01881. [PubMed: 28287264]

[88]. Pérez-Benito L, Keränen H, van Vlijmen H, Tresadern G. Predicting Binding Free Energies of PDE2 Inhibitors. The Difficulties of Protein Conformation. Sci Rep. 2018; 8(1):1–10. doi: 10.1038/s41598-018-23039-5. [PubMed: 29311619]

[89]. Warren GL, Do TD, Kelley BP, Nicholls A, Warren SD. Essential Considerations for Using Protein–Ligand Structures in Drug Discovery. Drug Discov. 2012; 17(23):1270–1281. doi: 10.1016/j.drudis.2012.06.011.

[90]. Pérez-Benito L, Casajuana-Martin N, Jiménez-Rosés M, van Vlijmen H, Tresadern G. Predicting Activity Cliffs with Free-Energy Perturbation. J Chem Theory Comput. 2019; 15(3):1884–1895. doi: 10.1021/acs.jctc.8b01290. [PubMed: 30776226]

[91]. Loeffler HH, Michel J, Woods C. FESetup: Automating Setup for Alchemical Free Energy Simulations. J Chem Inf Model. 2015; 55(12):2485–2490. doi: 10.1021/acs.jcim.5b00368. [PubMed: 26544598]

[92]. Gapsys V, Michielssens S, Seeliger D, de Groot BL. Pmx: Automated Protein Structure and Topology Generation for Alchemical Perturbations. J Comput Chem. 2015; 36(5):348–354. doi: 10.1002/jcc.23804. [PubMed: 25487359]

[93]. Jespers W, Esguerra M, Åqvist J, Gutiérrez-de-Terán H. QligFEP: An Automated Workflow for Small Molecule Free Energy Calculations in Q. J Cheminformatics. 2019; 11(1):26. doi: 10.1186/s13321-019-0348-5.

[94]. Konze KD, Bos PH, Dahlgren MK, Leswing K, Tubert-Brohman I, Bortolato A, Robbason B, Abel R, Bhat S. ReactionBased Enumeration, Active Learning, and Free Energy Calculations To Rapidly Explore Synthetically Tractable Chemical Space and Optimize Potency of Cyclin-Dependent Kinase 2 Inhibitors. J Chem Inf Model. 2019; 59(9):3782–3793. doi: 10.1021/acs.jcim.9b00367. [PubMed: 31404495]

[95]. Yang Q, Burchett W, Steeno GS, Liu S, Yang M, Mobley DL, Hou X. Optimal Designs for Pairwise Calculation: An Application to Free Energy Perturbation in Minimizing Prediction Variability. J Comput Chem. 2020; 41(3):247–257. doi: 10.1002/jcc.26095. [PubMed: 31721260]

[96]. Xu H. Optimal Measurement Network of Pairwise Differences. J Chem Inf Model. 2019; 59(11):4720–4728. doi: 10.1021/acs.jcim.9b00528. [PubMed: 31613620]

[97]. Ciordia M, Pérez-Benito L, Delgado F, Trabanco AA, Tresadern G. Application of Free Energy Perturbation for the Design of BACE1 Inhibitors. J Chem Inf Model. 2016; 56(9):1856–1871. doi: 10.1021/acs.jcim.6b00220. [PubMed: 27500414]

[98]. Wang L, Deng Y, Knight JL, Wu Y, Kim B, Sherman W, Shelley JC, Lin T, Abel R. Modeling Local Structural Rearrangements Using FEP/REST: Application to Relative Binding Affinity Predictions of CDK2 Inhibitors. J Chem Theory Comput. 2013; 9(2):1282–1293. doi: 10.1021/ct300911a. [PubMed: 26588769]

[99]. Robert A, Lingle W, David LM, Richard AF. A Critical Review of Validation, Blind Testing, and Real- World Use of Alchemical Protein-Ligand Binding Free Energy Calculations. Current Topics in Medicinal Chemistry. 2017; 17(23):2577–2585. [PubMed: 28413950]

[100]. Cappel D, Hall ML, Lenselink EB, Beuming T, Qi J, Bradner J, Sherman W. Relative Binding Free Energy Calculations Applied to Protein Homology Models. J Chem Inf Model. 2016; 56(12):2388–2400. doi: 10.1021/acs.jcim.6b00362. [PubMed: 28024402]

[101]. Deflorian F, Perez-Benito L, Lenselink EB, Congreve M, van Vlijmen HWT, Mason JS, de Graaf C, Tresadern G. Accurate Prediction of GPCR Ligand Binding Affinity with Free Energy Perturbation. J Chem Inf Model. 2020; doi: 10.1021/acs.jcim.0c00449.

[102]. Lenselink EB, Louvel J, Forti AF, van Veldhoven JPD, de Vries H, Mulder-Krieger T, McRobb FM, Negri A, Goose J, Abel R, van Vlijmen HWT, Wang L, Harder E, Sherman W, IJzerman AP, Beuming T. Predicting Binding Affinities for GPCR Ligands Using Free-Energy Perturbation. ACS Omega. 2016; 1(2):293–304. doi: 10.1021/acsomega.6b00086. [PubMed: 30023478]

[103]. 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(12):6346–6358. doi: 10.1021/acs.jctc.8b00825. [PubMed: 30375870]

[104]. Wang L, Deng Y, Wu Y, Kim B, LeBard DN, Wandschneider D, Beachy M, Friesner RA, Abel R. Accurate Modeling of Scaffold Hopping Transformations in Drug Discovery. J Chem Theory Comput. 2017; 13(1):42–54. doi: 10.1021/acs.jctc.6b00991. [PubMed: 27933808]

[105]. Albanese SK, Chodera JD, Volkamer A, Keng S, Abel R, Wang L. Is Structure Based Drug Design Ready for Selectivity Optimization? bioRxiv. 2020; p. 2020.07.02.185132. doi: 10.1101/2020.07.02.185132.

[106]. Mondal S, Tresadern G, Greenwood J, Kim B, Kaus J, Wirtala M, Steinbrecher T, Wang L, Masse C, Farid R, Abel R. A Free Energy Perturbation Approach to Estimate the Intrinsic Solubilities of Drug-like Small Molecules.. 2019; doi: 10.26434/chemrxiv.10263077.v1.

[107]. Berman H, Henrick K, Nakamura H. Announcing the World-wide Protein Data Bank. Nat Struct Mol Biol. 2003; 10(12):980–980. doi: 10.1038/nsb1203-980.

[108]. Wacker D, Fenalti G, Brown MA, Katritch V, Abagyan R, Cherezov V, Stevens RC. Conserved Binding Mode of Human *B*2 Adrenergic Receptor Inverse Agonists and Antagonist Revealed by X-Ray Crystallography. J Am Chem Soc. 2010; 132(33):11443–11445. doi: 10.1021/ja105108q. [PubMed: 20669948]

[109]. Brandt T, Holzmann N, Muley L, Khayat M, Wegscheid-Gerlach C, Baum B, Heine A, Hangauer D, Klebe G. Congeneric but Still Distinct: How Closely Related Trypsin Ligands Exhibit Different Thermodynamic and Structural Properties. J Mol Biol. 2011; 405(5):1170–1187. doi: 10.1016/j.jmb.2010.11.038. [PubMed: 21111747]

[110]. Nazaré M, Will DW, Matter H, Schreuder H, Ritter K, Urmann M, Essrich M, Bauer A, Wagner M, Czech J, Lorenz M, Laux V, Wehner V. Probing the Subpockets of Factor Xa Reveals Two Binding Modes for Inhibitors Based on a 2-Carboxyindole Scaffold: A Study Combining Structure-Activity Relationship and X-Ray Crystallography. J Med Chem. 2005; 48(14):4511–4525. doi: 10.1021/jm0490540. [PubMed: 15999990]

[111]. Kaus JW, Harder E, Lin T, Abel R, McCammon JA, Wang L. How To Deal with Multiple Binding Poses in Alchemical Relative Protein-Ligand Binding Free Energy Calculations. J Chem Theory Comput. 2015; 11:2670. doi: 10.1021/acs.jctc.5b00214. [PubMed: 26085821]

[112]. Michel J, Essex JW. Prediction of Protein–Ligand Binding Affinity by Free Energy Simulations: Assumptions, Pitfalls and Expectations. J Comput Aided Mol Des. 2010; 24(8):639–658. doi: 10.1007/s10822-010-9363-3. [PubMed: 20509041]

[113]. Wang L, Berne BJ, Friesner RA. Ligand Binding to ProteinBinding Pockets with Wet and Dry Regions. Proc Natl Acad Sci. 2011; 108(4):1326–1330. doi: 10.1073/pnas.1016793108. [PubMed: 21205906]

[114]. Bruce Macdonald HE, Cave-Ayland C, Ross GA, Essex JW. Ligand Binding Free Energies with Adaptive Water Networks: Two-Dimensional Grand Canonical Alchemical Perturbations. J Chem Theory Comput. 2018; 14(12):6586–6597. doi: 10.1021/acs.jctc.8b00614. [PubMed: 30451501]

[115]. Ross GA, Bodnarchuk MS, Essex JW. Water Sites, Networks, And Free Energies with Grand Canonical Monte Carlo. J Am Chem Soc. 2015; 137(47):14930–14943. doi: 10.1021/jacs.5b07940. [PubMed: 26509924]

[116]. Michel J, Tirado-Rives J, Jorgensen WL. Energetics of Displacing Water Molecules from Protein Binding Sites: Consequences for Ligand Optimization. J Am Chem Soc. 2009; 131(42):15403–15411. doi: 10.1021/ja906058w. [PubMed: 19778066]

[117]. Anandakrishnan R, Aguilar B, Onufriev AV. H++ 3.0: Automating pK Prediction and the Preparation of Biomolecular Structures for Atomistic Molecular Modeling and Simulations. Nucleic Acids Res. 2012; 40(Web Server issue):W537–W541. doi: 10.1093/nar/gks375. [PubMed: 22570416]

[118]. Søndergaard CR, Olsson MHM, Rostkowski M, Jensen JH. Improved Treatment of Ligands and Coupling Effects in Empirical Calculation and Rationalization of pKa Values. J Chem Theory Comput. 2011; 7(7):2284–2295. doi: 10.1021/ct200133y. [PubMed: 26606496]

[119]. Jurrus E, Engel D, Star K, Monson K, Brandi J, Felberg LE, Brookes DH, Wilson L, Chen J, Liles K, Chun M, Li P, Gohara DW, Dolinsky T, Konecny R, Koes DR, Nielsen JE, Head-Gordon T, Geng W, Krasny R, et al. Improvements to the APBS Biomolecular Solvation Software Suite. Protein Sci. 2018; 27(1):112–128. doi: 10.1002/pro.3280. [PubMed: 28836357]

[120]. Schrödinger Release 2020–2: Maestro,; 2020. Schrödinger, LLC, New York, NY,.

[121]. Olsson MHM, Søndergaard CR, Rostkowski M, Jensen JH. PROPKA3: Consistent Treatment of Internal and Surface Residues in Empirical pKa Predictions. J Chem Theory Comput. 2011; 7(2):525–537. doi: 10.1021/ct100578z. [PubMed: 26596171]

[122]. Işık M, Levorse D, Rustenburg AS, Ndukwe IE, Wang H, Wang X, Reibarkh M, Martin GE, Makarov AA, Mobley DL, Rhodes T, Chodera JD. pKa Measurements for the SAMPL6

Prediction Challenge for a Set of Kinase Inhibitor-like Fragments. J Comput Aided Mol Des. 2018; 32(10):1117–1138. doi: 10.1007/s10822-018-0168-0. [PubMed: 30406372]

[123]. Onufriev AV, Alexov E. Protonation and pK Changes in Protein-Ligand Binding. Q Rev Biophys. 2013; 46(2):181–209. doi: 10.1017/S0033583513000024. [PubMed: 23889892]

[124]. Kim MO, Blachly PG, McCammon JA. Conformational Dynamics and Binding Free Energies of Inhibitors of BACE-1: From the Perspective of Protonation Equilibria. PLOS Computational Biology. 2015; 11(10):e1004341. doi: 10.1371/journal.pcbi.1004341. [PubMed: 26506513]

[125]. Evoli S, Mobley DL, Guzzi R, Rizzuti B. Multiple Binding Modes of Ibuprofen in Human Serum Albumin Identified by Absolute Binding Free Energy Calculations. Phys Chem Chem Phys. 2016; 18(47):32358–32368. doi: 10.1039/C6CP05680F. [PubMed: 27854368]

[126]. Ruiz-Carmona S, Alvarez-Garcia D, Foloppe N, Garmendia-Doval AB, Juhos S, Schmidtke P, Barril X, Hubbard RE, Morley SD. rDock: A Fast, Versatile and Open Source Program for Docking Ligands to Proteins and Nucleic Acids. PLoS Comput Biol. 2014; 10(4):e1003571. doi: 10.1371/journal.pcbi.1003571. [PubMed: 24722481]

[127]. Trott O, Olson AJ. AutoDock Vina: Improving the Speed and Accuracy of Docking with a New Scoring Function, Efficient Optimization, and Multithreading. J Comput Chem. 2010; 31(2):455–461. doi: 10.1002/jcc.21334. [PubMed: 19499576]

[128]. Friesner RA, Banks JL, Murphy RB, Halgren TA, Klicic JJ, Mainz DT, Repasky MP, Knoll EH, Shelley M, Perry JK, Shaw DE, Francis P, Shenkin PS. Glide: A New Approach for Rapid, Accurate Docking and Scoring. 1. Method and Assessment of Docking Accuracy. J Med Chem. 2004; 47(7):1739–1749. doi: 10.1021/jm0306430. [PubMed: 15027865]

[129]. O'Boyle NM, Banck M, James CA, Morley C, Vandermeersch T, Hutchison GR. Open Babel: An Open Chemical Toolbox. J Cheminformatics. 2011; 3(1):33. doi: 10.1186/1758-2946-3-33.

[130]. Klimovich PV, Mobley DL. Predicting Hydration Free Energies Using All-Atom Molecular Dynamics Simulations and Multiple Starting Conformations. J Comput Aided Mol Des. 2010; 24(4):307–316. doi: 10.1007/s10822-010-9343-7. [PubMed: 20372973]

[131]. Lim VT, Bayly CI, Fusti-Molnar L, Mobley DL. Assessing the Conformational Equilibrium of Carboxylic Acid via Quantum Mechanical and Molecular Dynamics Studies on Acetic Acid. J Chem Inf Model. 2019; 59(5):1957–1964. doi: 10.1021/acs.jcim.8b00835. [PubMed: 30742770]

[132]. Boresch S, Karplus M. The Role of Bonded Terms in Free Energy Simulations. 2. Calculation of Their Influence on Free Energy Differences of Solvation. J Phys Chem A. 1999; 103(1):119–136. doi: 10.1021/jp981629f.

[133]. Mobley DL, Liu S, Lim NM, Wymer KL, Perryman AL, Forli S, Deng N, Su J, Branson K, Olson AJ. Blind Prediction of HIV Integrase Binding from the SAMPL4 Challenge. J Comput Aided Mol Des. 2014; 28(4):327–345. doi: 10.1007/s10822-014-9723-5. [PubMed: 24595873]

[134]. Jiang W, Chipot C, Roux B. Computing Relative Binding Affinity of Ligands to Receptor: An Effective Hybrid Single-Dual-Topology Free-Energy Perturbation Approach in NAMD. J Chem Inf Model. 2019; 59(9):3794–3802. doi: 10.1021/acs.jcim.9b00362. [PubMed: 31411473]

[135]. Rocklin GJ, Mobley DL, Dill KA. Separated Topologies—A Method for Relative Binding Free Energy Calculations Using Orientational Restraints. J Chem Phys. 2013; 138(8):085104. doi: 10.1063/1.4792251. [PubMed: 23464180]

[136]. Liu S, Wu Y, Lin T, Abel R, Redmann JP, Summa CM, Jaber VR, Lim NM, Mobley DL. Lead Optimization Mapper: Automating Free Energy Calculations for Lead Optimization. J Comput Aided Mol Des. 2013; 27(9):755–770. doi: 10.1007/s10822-0139678-y. [PubMed: 24072356]

[137]. RDkit, Rdkit: Open Source Chem Informatics software; 2019. http://www.rdkit.org, [Online; accessed 9. Dec. 2019].

[138]. Kawabata T, Nakamura H. 3D Flexible Alignment Using 2D Maximum Common Substructure: Dependence of Prediction Accuracy on Target-Reference Chemical Similarity. J Chem Inf Model. 2014; 54(7):1850–1863. doi: 10.1021/ci500006d. [PubMed: 24895842]

[139]. Raymond JW, Willett P. Maximum Common Subgraph Isomorphism Algorithms for the Matching of Chemical Structures. J Comput Aided Mol Des. 2002; 16(7):521–533. doi: 10.1023/A:1021271615909. [PubMed: 12510884]

[140]. Klabunde T, Giegerich C, Evers A. MARS: Computing Three-Dimensional Alignments for Multiple Ligands Using Pairwise Similarities. J Chem Inf Model. 2012; 52(8):2022–2030. doi: 10.1021/ci3000369. [PubMed: 22794356]

[141]. Jones G, Gao Y, Sage CR. Elucidating Molecular Overlays from Pairwise Alignments Using a Genetic Algorithm. J Chem Inf Model. 2009; 49(7):1847–1855. doi: 10.1021/ci900109n. [PubMed: 19537722]

[142]. Liu S, Wang L, Mobley DL. Is Ring Breaking Feasible in Relative Binding Free Energy Calculations? J Chem Inf Model. 2015; 55(4):727–735. doi: 10.1021/acs.jcim.5b00057. [PubMed: 25835054]

[143]. Clark AJ, Negron C, Hauser K, Sun M, Wang L, Abel R, Friesner RA. Relative Binding Affinity Prediction of Charge-Changing Sequence Mutations with FEP in Protein–Protein Interfaces. J Mol Biol. 2019; 431(7):1481–1493. doi: 10.1016/j.jmb.2019.02.003. [PubMed: 30776430]

[144]. Kräutler V, van Gunsteren WF, Hünenberger PH. A Fast SHAKE Algorithm to Solve Distance Constraint Equations for Small Molecules in Molecular Dynamics Simulations. J Comput Chem. 2001; 22(5):501–508. doi: 10.1002/1096987X(20010415)22:5<501::AID-JCC1021>3.0.CO;2-V.

[145]. Pearlman DA. Determining the Contributions of Constraints in Free Energy Calculations: Development, Characterization, and Recommendations. J Chem Phys. 1993; 98(11):8946–8957. doi: 10.1063/1.464453.

[146]. Straatsma TP, Zacharias M, McCammon JA. Holonomic Constraint Contributions to Free Energy Differences from Thermodynamic Integration Molecular Dynamics Simulations. Chem Phys Lett. 1992; 196(3):297–302. doi: 10.1016/00092614(92)85971-C.

[147]. Pearlman DA, Kollman PA. The Overlooked Bond-stretching Contribution in Free Energy Perturbation Calculations. J Chem Phys. 1991; 94(6):4532–4545. doi: 10.1063/1.460608.

[148]. van Gunsteren WF, Weiner PK. Computer Simulation of Biomolecular Systems: Theoretical and Experimental Applications Volume 1, vol. 1 of Computer Simulations of Biomolecular Systems Springer Netherlands; 1989.

[149]. Mobley DL, Chodera JD, Dill KA. On the Use of Orientational Restraints and Symmetry Corrections in Alchemical Free Energy Calculations. J Chem Phys. 2006; 125:084902. doi: 10.1063/1.2221683. [PubMed: 16965052]

[150]. Chen J, Brooks CL. Can Molecular Dynamics Simulations Provide High-Resolution Refinement of Protein Structure? Proteins: Struct, Funct, Bioinf. 2007; 67(4):922–930. doi: 10.1002/prot.21345.

[151]. Boresch S, Tettinger F, Leitgeb M, Karplus M. Absolute Binding Free Energies: A Quantitative Approach for Their Calculation. J Phys Chem B. 2003; 107(35):9535–9551. doi: 10.1021/jp0217839.

[152]. Leitgeb M, Schröder C, Boresch S. Alchemical Free Energy Calculations and Multiple Conformational Substates. J Chem Phys. 2005; 122(8):084109. doi: 10.1063/1.1850900.

[153]. Wang K, Chodera JD, Yang Y, Shirts MR. Identifying Ligand Binding Sites and Poses Using GPU-Accelerated Hamiltonian Replica Exchange Molecular Dynamics. J Comput Aided Mol Des. 2013; 27(12):989–1007. doi: 10.1007/s10822-013-9689-8. [PubMed: 24297454]

[154]. Georgiou C, McNae I, Wear M, Ioannidis H, Michel J, Walkinshaw M. Pushing the Limits of Detection of Weak Binding Using Fragment-Based Drug Discovery: Identification of New Cyclophilin Binders. J Mol Biol. 2017; 429(16):2556–2570. doi: 10.1016/j.jmb.2017.06.016. [PubMed: 28673552]

[155]. Sugita Y, Kitao A, Okamoto Y. Multidimensional Replica-Exchange Method for Free-Energy Calculations. J Chem Phys. 2000; 113(15):6042–6051. doi: 10.1063/1.1308516.

[156]. Chodera JD, Shirts MR. Replica Exchange and Expanded Ensemble Simulations as Gibbs Sampling: Simple Improvements for Enhanced Mixing. J Chem Phys. 2011; 135(19):194110. doi: 10.1063/1.3660669. [PubMed: 22112069]

[157]. Lyubartsev AP, Martsinovski AA, Shevkunov SV, Vorontsov-Velyaminov PN. New Approach to Monte Carlo Calculation of the Free Energy: Method of Expanded Ensembles. J Chem Phys. 1992; 96(3):1776–1783. doi: 10.1063/1.462133.

[158]. Li H, Fajer M, Yang W. Simulated Scaling Method for Localized Enhanced Sampling and Simultaneous “Alchemical” Free Energy Simulations: A General Method for

Molecular Mechanical, Quantum Mechanical, and Quantum Mechanical/Molecular Mechanical Simulations. J Chem Phys. 2007; 126(2):024106. doi: 10.1063/1.2424700. [PubMed: 17228942]

[159]. Lin YL, Aleksandrov A, Simonson T, Roux B. An Overview of Electrostatic Free Energy Computations for Solutions and Proteins. J Chem Theory Comput. 2014; 10(7):2690–2709. doi: 10.1021/ct500195p. [PubMed: 26586504]

[160]. 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(10):986–999. doi: 10.1002/jcc.26143. [PubMed: 31930547]

[161]. Rocklin GJ, Mobley DL, Dill KA, Hünenberger PH. Calculating the Binding Free Energies of Charged Species Based on Explicit-Solvent Simulations Employing Lattice-Sum Methods: An Accurate Correction Scheme for Electrostatic FiniteSize Effects. J Chem Phys. 2013; 139(18):184103. doi: 10.1063/1.4826261. [PubMed: 24320250]

[162]. Mey ASJS Jiménez JJ, Michel J. Impact of Domain Knowledge on Blinded Predictions of Binding Energies by Alchemical Free Energy Calculations. J Comput Aided Mol Des. 2018; 32(1):199–210. doi: 10.1007/s10822-017-0083-9. [PubMed: 29134431]

[163]. Gapsys V, Michielssens S, Peters JH, de Groot BL, Leonov H. Calculation of Binding Free Energies. In: Kukol A, editor. Molecular Modeling of Proteins Methods in Molecular Biology, New York, NY: Springer; 2015.p. 173–209. doi: 10.1007/978-1-49391465-4_9.

[164]. Beutler TC, Mark AE, van Schaik RC, Gerber PR, van Gunsteren WF. Avoiding Singularities and Numerical Instabilities in Free Energy Calculations Based on Molecular Simulations. Chem Phys Lett. 1994; 222:529–539. doi: 10.1016/00092614(94)00397-1.

[165]. Beutler TC, van Gunsteren WF. Molecular Dynamics Free Energy Calculation in Four Dimensions. J Chem Phys. 1994; 101(2):1417–1422. doi: 10.1063/1.467765.

[166]. Gapsys V, Seeliger D, de Groot BL. New Soft-Core Potential Function for Molecular Dynamics Based Alchemical Free Energy Calculations. J Chem Theory Comput. 2012; 8(7):2373–2382. doi: 10.1021/ct300220p. [PubMed: 26588970]

[167]. Naden LN, Shirts MR. Linear Basis Function Approach to Efficient Alchemical Free Energy Calculations. 2. Inserting and Deleting Particles with Coulombic Interactions. J Chem Theory Comput. 2015; 11(6):2536–2549. doi: 10.1021/ct501047e. [PubMed: 26575553]

[168]. Naden LN, Pham TT, Shirts MR. Linear Basis Function Approach to Efficient Alchemical Free Energy Calculations. 1. Removal of Uncharged Atomic Sites. J Chem Theory Comput. 2014; 10(3):1128–1149. doi: 10.1021/ct4009188. [PubMed: 26580188]

[169]. Pham TT, Shirts MR. Identifying Low Variance Pathways for Free Energy Calculations of Molecular Transformations in Solution Phase. J Chem Phys. 2011; 135(3):034114. doi: 10.1063/1.3607597. [PubMed: 21786994]

[170]. Zacharias M, Straatsma TP, McCammon JA. Separation-Shifted Scaling, a New Scaling Method for Lennard-Jones Interactions in Thermodynamic Integration. J Phys Chem. 1994; 100(12):9025–9031. doi: 10.1063/1.466707.

[171]. Blondel A. Ensemble Variance in Free Energy Calculations by Thermodynamic Integration: Theory, Optimal Alchemical Path, and Practical Solutions. J Comput Chem. 2004; 25(7):985–993. doi: 10.1002/jcc.20025. [PubMed: 15027110]

[172]. Pham TT, Shirts MR. Optimal Pairwise and Non-Pairwise Alchemical Pathways for Free Energy Calculations of Molecular Transformation in Solution Phase. J Chem Phys. 2012; 136(12):124120. doi: 10.1063/1.3697833. [PubMed: 22462848]

[173]. Donnini S, Mark AE, Juffer AH, Villa A. Incorporating the Effect of Ionic Strength in Free Energy Calculations Using Explicit Ions. Journal of Computational Chemistry. 2005; 26(2):115–122. doi: 10.1002/jcc.20156. [PubMed: 15584080]

[174]. Steinbrecher T, Joung I, Case DA. Soft-Core Potentials in Thermodynamic Integration: Comparing One- and Two-Step Transformations. J Comput Chem. 2011; 32(15):3253–3263. doi: 10.1002/jcc.21909. [PubMed: 21953558]

[175]. Hermans J, Wang L. Inclusion of Loss of Translational and Rotational Freedom in Theoretical Estimates of Free Energies of Binding. Application to a Complex of Benzene and Mutant T4 Lysozyme. J Am Chem Soc. 1997; 119(11):2707–2714. doi: 10.1021/ja963568+.

[176]. Mann G, Hermans J. Modeling Protein-Small Molecule Interactions: Structure and Thermodynamics of Noble Gases Binding in a Cavity in Mutant Phage T4 Lysozyme L99A11Edited by B. Honig. J Mol Biol. 2000; 302(4):979–989. doi: 10.1006/jmbi.2000.4064. [PubMed: 10993736]

[177]. Wang J, Deng Y, Roux B. Absolute Binding Free Energy Calculations Using Molecular Dynamics Simulations with Restraining Potentials. Biophysical Journal. 2006; 91(8):2798–2814. doi: 10.1529/biophysj.106.084301. [PubMed: 16844742]

[178]. Fujitani H, Tanida Y, Ito M, Jayachandran G, Snow CD, Shirts MR, Sorin EJ, Pande VS. Direct Calculation of the Binding Free Energies of FKBP Ligands. J Chem Phys. 2005; 123(8):084108. doi: 10.1063/1.1999637. [PubMed: 16164283]

[179]. Steinbrecher T, Mobley DL, Case DA. Nonlinear Scaling Schemes for Lennard-Jones Interactions in Free Energy Calculations. J Chem Phys. 2007; 127(21). doi: 10.1063/1.2799191.

[180]. Crooks GE. Measuring Thermodynamic Length. Phys Rev Lett. 2007; 99(10):100602. doi: 10.1103/PhysRevLett.99.100602. [PubMed: 17930381]

[181]. Sivak DA, Crooks GE. Thermodynamic Metrics and Optimal Paths. Phys Rev Lett. 2012; 108(19):190602. doi: 10.1103/PhysRevLett.108.190602. [PubMed: 23003019]

[182]. Shenfeld DK, Xu H, Eastwood MP, Dror RO, Shaw DE. Minimizing Thermodynamic Length to Select Intermediate States for Free-Energy Calculations and Replica-Exchange Simulations. Phys Rev E. 2009; 80(4):046705. doi: 10.1103/PhysRevE.80.046705.

[183]. Hayes RL, Armacost KA, Vilseck JZ, Brooks CL. Adaptive Landscape Flattening Accelerates Sampling of Alchemical Space in Multisite λ Dynamics. J Phys Chem B. 2017; 121(15):3626–3635. doi: 10.1021/acs.jpcb.6b09656. [PubMed: 28112940]

[184]. Monroe JI, Shirts MR. Converging Free Energies of Binding in Cucurbit[7]Uril and Octa-Acid Host–Guest Systems from SAMPL4 Using Expanded Ensemble Simulations. J Comput Aided Mol Des. 2014; 28(4):401–415. doi: 10.1007/s10822-0149716-4. [PubMed: 24610238]

[185]. Perthold JW, Oostenbrink C. Accelerated Enveloping Distribution Sampling: Enabling Sampling of Multiple End States While Preserving Local Energy Minima. J Phys Chem B. 2018; 122(19):5030–5037. doi: 10.1021/acs.jpcb.8b02725. [PubMed: 29669415]

[186]. Sidler D, Cristòfol-Clough M, Riniker S. Efficient Round-Trip Time Optimization for Replica-Exchange Enveloping Distribution Sampling (RE-EDS). J Chem Theory Comput. 2017; 13(6):3020–3030. doi: 10.1021/acs.jctc.7b00286. [PubMed: 28510459]

[187]. Christ CD, van Gunsteren WF. Enveloping Distribution Sampling: A Method to Calculate Free Energy Differences from a Single Simulation. J Chem Phys. 2007; 126(18):184110. doi: 10.1063/1.2730508. [PubMed: 17508795]

[188]. Swendsen RH, Wang JS. Replica Monte Carlo Simulation of Spin-Glasses. Phys Rev Lett. 1986; 57(21):2607–2609. doi: 10.1103/PhysRevLett.57.2607. [PubMed: 10033814]

[189]. Sugita Y, Okamoto Y. Replica-Exchange Molecular Dynamics Method for Protein Folding. Chem Phys Lett. 1999; 314(1):141–151. doi: 10.1016/S0009-2614(99)01123-9.

[190]. Woods CJ, Essex JW, King MA. The Development of Replica-Exchange-Based Free-Energy Methods. J Phys Chem B. 2003; 107(49):13703–13710. doi: 10.1021/jp0356620.

[191]. Jiang W, Roux B. Free Energy Perturbation Hamiltonian Replica-Exchange Molecular Dynamics (FEP/H-REMD) for Absolute Ligand Binding Free Energy Calculations. J Chem Theory Comput. 2010; 6(9):2559–2565. doi: 10.1021/ct1001768. [PubMed: 21857813]

[192]. Hritz J, Oostenbrink C. Hamiltonian Replica Exchange Molecular Dynamics Using Soft-Core Interactions. J Chem Phys. 2008; 128(14):144121. doi: 10.1063/1.2888998. [PubMed: 18412437]

[193]. Hritz J, Oostenbrink C. Optimization of Replica Exchange Molecular Dynamics by Fast Mimicking. J Chem Phys. 2007; 127(20):204104. doi: 10.1063/1.2790427. [PubMed: 18052416]

[194]. Marinari E, Parisi G. Simulated Tempering: A New Monte Carlo Scheme. Europhys Lett. 1992; 19(6):451–458. doi: 10.1209/0295-5075/19/6/002.

[195]. Tan Z Optimally Adjusted Mixture Sampling and Locally Weighted Histogram Analysis. J Comput Graph Stat. 2017; 26(1):54–65. doi: 10.1080/10618600.2015.1113975.

[196]. Rizzi A, Chodera J, Naden L, Beauchamp K, Grinaway P, Fass J, adw62, Rustenburg B, Ross GA, Krämer A, Macdonald HB, Swenson DWH, Simmonett A, hb0402, ajsilveira, Choderalab/

Openmmtools: 0.19.0 - Multiple Alchemical Regions; 2019. doi: 10.5281/zenodo.3532826. Zenodo.

[197]. Shirts MR, Pande VS. Comparison of Efficiency and Bias of Free Energies Computed by Exponential Averaging, the Bennett Acceptance Ratio, and Thermodynamic Integration. J Chem Phys. 2005; 122:144107. doi: 10.1063/1.1873592. [PubMed: 15847516]

[198]. Plount Price ML, Jorgensen WL. Analysis of Binding Affinities for Celecoxib Analogues with COX-1 and COX-2 from Combined Docking and Monte Carlo Simulations and Insight into the COX-2/COX-1 Selectivity. J Am Chem Soc. 2000; 122(39):9455–9466. doi: 10.1021/ja001018c.

[199]. Mobley DL, Dill KA. Binding of Small-Molecule Ligands to Proteins: “What You See” Is Not Always “What You Get”. Structure. 2009; 17(4):489–498. doi: 10.1016/j.str.2009.02.010. [PubMed: 19368882]

[200]. Calabrò G, Woods CJ, Powlesland F, Mey ASJS, Mulholland AJ, Michel J. Elucidation of Nonadditive Effects in Protein–Ligand Binding Energies: Thrombin as a Case Study. J Phys Chem B. 2016; 120(24):5340–5350. doi: 10.1021/acs.jpcb.6b03296. [PubMed: 27248478]

[201]. Steinbrecher TB, Dahlgren M, Cappel D, Lin T, Wang L, Krilov G, Abel R, Friesner R, Sherman W. Accurate Binding Free Energy Predictions in Fragment Optimization. J Chem Inf Model. 2015; 55(11):2411–2420. doi: 10.1021/acs.jcim.5b00538. [PubMed: 26457994]

[202]. Palma PN, Bonifácio MJ, Loureiro AI, Soares-da-Silva P. Computation of the Binding Affinities of Catechol-O-Methyltransferase Inhibitors: Multisubstate Relative Free Energy Calculations. J Comput Chem. 2012; 33(9):970–986. doi: 10.1002/jcc.22926. [PubMed: 22278964]

[203]. Voet ARD, Kumar A, Berenger F, Zhang KYJ. Combining in Silico and in Cerebro Approaches for Virtual Screening and Pose Prediction in SAMPL4. J Comput Aided Mol Des. 2014; 28(4):363–373. doi: 10.1007/s10822-013-9702-2. [PubMed: 24446075]

[204]. Gallicchio E, Deng N, He P, Wickstrom L, Perryman AL, Santiago DN, Forli S, Olson AJ, Levy RM. Virtual Screening of Integrase Inhibitors by Large Scale Binding Free Energy Calculations: The SAMPL4 Challenge. J Comput Aided Mol Des. 2014; 28(4):475–490. doi: 10.1007/s10822-014-9711-9. [PubMed: 24504704]

[205]. Rocklin GJ, Boyce SE, Fischer M, Fish I, Mobley DL, Shoichet BK, Dill KA. Blind Prediction of Charged Ligand Binding Affinities in a Model Binding Site. J Mol Biol. 2013; 425(22):4569–4583. doi: 10.1016/j.jmb.2013.07.030. [PubMed: 23896298]

[206]. Boyce SE, Mobley DL, Rocklin GJ, Graves AP, Dill KA, Shoichet BK. Predicting Ligand Binding Affinity with Alchemical Free Energy Methods in a Polar Model Binding Site. J Mol Biol. 2009; 394(4):747–763. doi: 10.1016/j.jmb.2009.09.049. [PubMed: 19782087]

[207]. Lincoff J, Sasmal S, Head-Gordon T. Comparing Generalized Ensemble Methods for Sampling of Systems with Many Degrees of Freedom. J Chem Phys. 2016; 145(17):174107. doi: 10.1063/1.4965439. [PubMed: 27825215]

[208]. Sasmal S, Gill SC, Lim NM, Mobley DL. Sampling Conformational Changes of Bound Ligands Using Nonequilibrium Candidate Monte Carlo and Molecular Dynamics. J Chem Theory Comput. 2020; 16(3):1854–1865. doi: 10.1021/acs.jctc.9b01066. [PubMed: 32058713]

[209]. Leimkuhler B, Matthews C. Efficient Molecular Dynamics Using Geodesic Integration and Solvent–Solute Splitting. Proc R Soc A. 2016; 472(2189):20160138. doi: 10.1098/rspa.2016.0138. [PubMed: 27279779]

[210]. Chodera JD. A Simple Method for Automated Equilibration Detection in Molecular Simulations. J Chem Theory Comput. 2016; 12(4):1799–1805. doi: 10.1021/acs.jctc.5b00784. [PubMed: 26771390]

[211]. Beauchamp K, Chodera J, Naden L, Shirts M, Martiniani S, Stern C, McGibbon RT, Gowers R, Dotson D, Choderalab/Pymbar: Critical Bugfix Release; 2019. doi: 10.5281/zenodo.3559263. Zenodo.

[212]. Dotson D, Beckstein O, Wille D, Kenney I, shuail, trje3733, Lee H, Lim V, brycestx, Barhaghi MS, Alchemistry/Alchemlyb: 0.3.0; 2019. doi: 10.5281/zenodo.3361016. Zenodo.

[213]. Nüske F, Wu H, Prinz JH, Wehmeyer C, Clementi C, Noé F. Markov State Models from Short Non-Equilibrium Simulations—Analysis and Correction of Estimation Bias. J Chem Phys. 2017; 146(9):094104. doi: 10.1063/1.4976518.

[214]. Maragakis P, Ritort F, Bustamante C, Karplus M, Crooks GE. Bayesian Estimates of Free Energies from Nonequilibrium Work Data in the Presence of Instrument Noise. J Chem Phys. 2008; 129(2):024102–8. doi: 10.1063/1.2937892. [PubMed: 18624511]

[215]. Oberhofer H, Dellago C, Geissler PL. Biased Sampling of Nonequilibrium Trajectories: Can Fast Switching Simulations Outperform Conventional Free Energy Calculation Methods? J Phys Chem B. 2005; 109:6902–6915. doi: 10.1021/jp044556a. [PubMed: 16851777]

[216]. Procacci P. Unbiased Free Energy Estimates in Fast Nonequilibrium Transformations Using Gaussian Mixtures. J Chem Phys. 2015; 142(15):154117. doi: 10.1063/1.4918558. [PubMed: 25903876]

[217]. Ytreberg FM, Zuckerman DM. Single-Ensemble Nonequilibrium Path-Sampling Estimates of Free Energy Differences. J Chem Phys. 2004; 120(23):10876–9. doi: 10.1063/1.1760511. [PubMed: 15268117]

[218]. Lu ND, Singh JK, Kofke DA. Appropriate Methods to Combine Forward and Reverse Free-Energy Perturbation Averages. J Chem Phys. 2003; 118(7):2977–2984. doi: 10.1063/1.1537241.

[219]. Lelièvre T, Rousset M, Stoltz G. Free Energy Computations. IMPERIAL COLLEGE PRESS; 2010. doi: 10.1142/p579.

[220]. Jarzynski C. Rare Events and the Convergence of Exponentially Averaged Work Values. Phys Rev E. 2006; 73:046105. doi: 10.1103/PhysRevE.73.046105.

[221]. Resat H, Mezei M. Studies on Free Energy Calculations. I. Thermodynamic Integration Using a Polynomial Path. J Chem Phys. 1993; 99(8):6052–6061. doi: 10.1063/1.465902.

[222]. Jorge M, Garrido N, Queimada A, Economou I, Macedo E. Effect of the Integration Method on the Accuracy and Computational Efficiency of Free Energy Calculations Using Thermodynamic Integration. J Chem Theory Comput. 2010; 6(4):1018–1027. doi: 10.1021/ct900661c. [PubMed: 20467461]

[223]. Shyu C, Ytreberg FM. Reducing the Bias and Uncertainty of Free Energy Estimates by Using Regression to Fit Thermodynamic Integration Data. J Comput Chem. 2009; 30(14):2297–2304. doi: 10.1002/jcc.21231. [PubMed: 19266482]

[224]. Paliwal H, Shirts MR. A Benchmark Test Set for Alchemical Free Energy Transformations and Its Use to Quantify Error in Common Free Energy Methods. J Chem Theory Comput. 2011; 7(12):4115–4134. doi: 10.1021/ct2003995. [PubMed: 26598357]

[225]. de Ruiter A, Oostenbrink C. Extended Thermodynamic Integration: Efficient Prediction of Lambda Derivatives at Nonsimulated Points. J Chem Theory Comput. 2016; 12(9):4476–4486. doi: 10.1021/acs.jctc.6b00458. [PubMed: 27494138]

[226]. Tan Z. On a Likelihood Approach for Monte Carlo Integration. J Am Stat Soc. 2004; 99(468):1027–1036. doi: 10.1198/016214504000001664.

[227]. Shirts MR. Reweighting from the Mixture Distribution as a Better Way to Describe the Multistate Bennett Acceptance Ratio. arXiv:170400891 [cond-mat]. 2017;.

[228]. Mobley DL, Gilson MK. Predicting Binding Free Energies: Frontiers and Benchmarks. Annu Rev Biophys. 2017; 46(1):531–558. doi: 10.1146/annurev-biophys-070816-033654. [PubMed: 28399632]

[229]. Mobley DL, Heinzelmann G, Henriksen NM, Gilson MK. Predicting Binding Free Energies: Frontiers and Benchmarks (a Perpetual Review). Annu Rev Biophys. 2017; 46:531. doi: 10.1146/annurev-biophys-070816-033654. [PubMed: 28399632]

[230]. Dakka J, Farkas-Pall K, Turilli M, Wright DW, Coveney PV, Jha S. Concurrent and Adaptive Extreme Scale Binding Free Energy Calculations. arXiv:180101174 [cs]. 2018; doi: 10.1109/eScience.2018.00034.

[231]. Hahn DF, Hünenberger PH. Alchemical Free-Energy Calculations by Multiple-Replica λ-Dynamics: The Conveyor Belt Thermodynamic Integration Scheme. J Chem Theory Comput. 2019; 15(4):2392–2419. doi: 10.1021/acs.jctc.8b00782. [PubMed: 30821973]

[232]. Brown SP, Muchmore SW, Hajduk PJ. Healthy Skepticism: Assessing Realistic Model Performance. Drug Discov. 2009; 14(7):420–427. doi: 10.1016/j.drudis.2009.01.012.

[233]. Drugdata/Metk; 2018. Drug Design Data Resource.

[234]. Walters WP. What Are Our Models Really Telling Us? A Practical Tutorial on Avoiding Common Mistakes When Building Predictive Models. In: Chemoinformatics for Drug Discovery John Wiley & Sons, Ltd; 2013.p. 1–31. doi: 10.1002/9781118742785.ch1.

[235]. Antonia M, Michellab/Freenrgworkflows; 2019. michellab.

[236]. Cui G, Graves AP, Manas ES. GRAM: A True Null Model for Relative Binding Affinity Predictions. J Chem Inf Model. 2020; 60(1):11–16. doi: 10.1021/acs.jcim.9b00939. [PubMed: 31874032]

[237]. Jain AN, Nicholls A. Recommendations for Evaluation of Computational Methods. J Comput Aided Mol Des. 2008; 22(3):133–139. doi: 10.1007/s10822-008-9196-5. [PubMed: 18338228]

[238]. Walter P, Some Thoughts on Evaluating Predictive Models;. http://practicalcheminformatics.blogspot.com/2019/02/somethoughts-on-evaluating-predictive.html.

[239]. Wagner V, Jantz L, Briem H, Sommer K, Rarey M, Christ CD. Computational Macrocyclization: From de Novo Macrocycle Generation to Binding Affinity Estimation. ChemMedChem. 2017; 12(22):1866–1872. doi: 10.1002/cmdc.201700478. [PubMed: 28977738]

[240]. Gaieb Z, Liu S, Gathiaka S, Chiu M, Yang H, Shao C, Feher VA, Walters WP, Kuhn B, Rudolph MG, Burley SK, Gilson MK, Amaro RE. D3R Grand Challenge 2: Blind Prediction of Protein–Ligand Poses, Affinity Rankings, and Relative Binding Free Energies. J Comput Aided Mol Des. 2018; 32(1):1–20. doi: 10.1007/s10822-017-0088-4. [PubMed: 29204945]

[241]. Gaieb Z, Parks CD, Chiu M, Yang H, Shao C, Walters WP, Lambert MH, Nevins N, Bembenek SD, Ameriks MK, Mirzadegan T, Burley SK, Amaro RE, Gilson MK. D3R Grand Challenge 3: Blind Prediction of Protein–Ligand Poses and Affinity Rankings. J Comput Aided Mol Des. 2019; 33(1):1–18. doi: 10.1007/s10822-018-0180-4. [PubMed: 30632055]

[242]. Salomon-Ferrer R, Case DA, Walker RC. An Overview of the Amber Biomolecular Simulation Package. WIREs Comput Mol Sci. 2013; 3(2):198–210. doi: 10.1002/wcms.1121.

[243]. Schmid N, Christ CD, Christen M, Eichenberger AP, van Gunsteren WF. Architecture, Implementation and Parallelisation of the GROMOS Software for Biomolecular Simulation. Comput Phys Commun. 2012; 183(4):890–903. doi: 10.1016/j.cpc.2011.12.014.

[244]. Kunz APE, Allison JR, Geerke DP, Horta BAC, Hünenberger PH, Riniker S, Schmid N, van Gunsteren WF. New Functionalities in the GROMOS Biomolecular Simulation Software. Journal of Computational Chemistry. 2012; 33(3):340–353. doi: 10.1002/jcc.21954. [PubMed: 22076815]

[245]. Eichenberger AP, Allison JR, Dolenc J, Geerke DP, Horta BAC, Meier K, Oostenbrink C, Schmid N, Steiner D, Wang D, van Gunsteren WF. GROMOS++ Software for the Analysis of Biomolecular Simulation Trajectories. J Chem Theory Comput. 2011; 7(10):3379–3390. doi: 10.1021/ct2003622. [PubMed: 26598168]

[246]. Bonomi M, Bussi G, Camilloni C, Tribello GA, Banáš P, Barducci A, Bernetti M, Bolhuis PG, Bottaro S, Branduardi D, Capelli R, Carloni P, Ceriotti M, Cesari A, Chen H, Chen W, Colizzi F, De S, De La Pierre M, Donadio D, et al. Promoting Transparency and Reproducibility in Enhanced Molecular Simulations. Nat Methods. 2019; 16(8):670–673. doi: 10.1038/s41592-019-0506-8. [PubMed: 31363226]

[247]. Lemkul J. From Proteins to Perturbed Hamiltonians: A Suite of Tutorials for the GROMACS-2018 Molecular Simulation Package [Article v1.0]. LiveCoMS. 2018; 1(1):5068–. doi: 10.33011/livecoms.1.1.5068.

[248]. Abraham MJ, Murtola T, Schulz R, Páll S, Smith JC, Hess B, Lindahl E. GROMACS: High Performance Molecular Simulations through Multi-Level Parallelism from Laptops to Supercomputers. SoftwareX. 2015; 1–2:19–25. doi: 10.1016/j.softx.2015.06.001.

[249]. {\AAquist} J, Kamerlin SCL, Bauer P, Marelius J, Q6: A Comprehensive Toolkit for Empirical Valence Bond and Related Free Energy Calculations; 2017. doi: 10.5281/zenodo.1002739. Zenodo.

[250]. Jo S, Kim T, Iyer VG, Im W. CHARMM-GUI: A Web-Based Graphical User Interface for CHARMM. J Comput Chem. 2008; 29(11):1859–1865. doi: 10.1002/jcc.20945. [PubMed: 18351591]

[251]. Suruzhon M, Senapathi T, Bodnarchuk MS, Viner R, Wall ID, Barnett CB, Naidoo KJ, Essex JW. ProtoCaller: Robust Automation of Binding Free Energy Calculations. J Chem Inf Model. 2020; 60(4):1917–1921. doi: 10.1021/acs.jcim.9b01158. [PubMed: 32092258]

[252]. Dotson D, Beckstein O, Wille D, Kenney I, shuail, trje3733, Lee H, Lim V, Allen B, Barhaghi MS, Alchemistry/Alchemlyb: 0.3.1; 2020. doi: 10.5281/zenodo.3610564. Zenodo.

[253]. Williams-Noonan BJ, Yuriev E, Chalmers DK. Free Energy Methods in Drug Design: Prospects of “Alchemical Perturbation” in Medicinal Chemistry. J Med Chem. 2018; 61(3):638–649. doi: 10.1021/acs.jmedchem.7b00681. [PubMed: 28745501]

[254]. Alchemistry.org;. Accessed: 2019-08-01. http://www.alchemistry.org/wiki/Test_System_Repository.

[255]. D3R Grand Challenges;. Accessed: 2019-08-01. https://drugdesigndata.org/about/grand-challenge.

[256]. Gathiaka S, Liu S, Chiu M, Yang H, Stuckey JA, Kang YN, Delproposto J, Kubish G, Dunbar JB, Carlson HA, Burley SK, Walters WP, Amaro RE, Feher VA, Gilson MK. D3R Grand Challenge 2015: Evaluation of Protein–Ligand Pose and Affinity Predictions. J Comput Aided Mol Des. 2016; 30(9):651–668. doi: 10.1007/s10822-016-9946-8. [PubMed: 27696240]

[257]. SAMPL Challenges;. Accessed: 2019-08-01. https://samplchallenges.github.io/.

[258]. Rizzi A, Murkli S, McNeill JN, Yao W, Sullivan M, Gilson MK, Chiu MW, Isaacs L, Gibb BC, Mobley DL, Chodera JD. Overview of the SAMPL6 Host–Guest Binding Affinity Prediction Challenge. J Comput Aided Mol Des. 2018; 32(10):937–963. doi: 10.1007/s10822-018-0170-6. [PubMed: 30415285]

[259]. Yin J, Henriksen NM, Slochower DR, Shirts MR, Chiu MW, Mobley DL, Gilson MK. Overview of the SAMPL5 Host–Guest Challenge: Are We Doing Better? J Comput Aided Mol Des. 2017; 31(1):1–19. doi: 10.1007/s10822-016-9974-4. [PubMed: 27658802]

[260]. Muddana HS, Fenley AT, Mobley DL, Gilson MK. The SAMPL4 Host–Guest Blind Prediction Challenge: An Overview. J Comput Aided Mol Des. 2014; 28(4):305–317. doi: 10.1007/s10822014-9735-1. [PubMed: 24599514]

[261]. Roos K, Wu C, Damm W, Reboul M, Stevenson JM, Lu C, Dahlgren MK, Mondal S, Chen W, Wang L, Abel R, Friesner RA, Harder ED. OPLS3e: Extending Force Field Coverage for Drug-Like Small Molecules. J Chem Theory Comput. 2019; 15(3):1863–1874. doi: 10.1021/acs.jctc.8b01026. [PubMed: 30768902]

[262]. Yu HS, Deng Y, Wu Y, Sindhikara D, Rask AR, Kimura T, Abel R, Wang L. Accurate and Reliable Prediction of the Binding Affinities of Macrocycles to Their Protein Targets. J Chem Theory Comput. 2017; 13(12):6290–6300. doi: 10.1021/acs.jctc.7b00885. [PubMed: 29120625]

[263]. MCompChem, fep-benchmark; 2019. https://github.com/MCompChem/fep-benchmark, [Online; accessed 9. Dec. 2019].

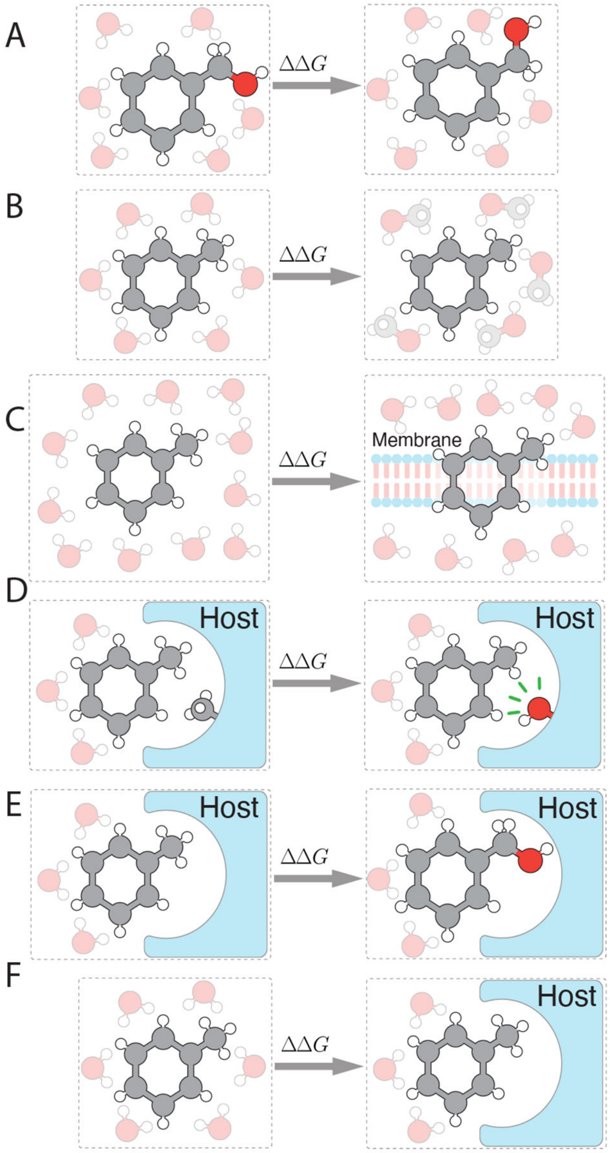


Figure 1. Illustration of common types of free energies differences that can be calculated using alchemical free energy methods.

A: Change in free energy due to a conformational change of the molecule across a high barrier. **B**: Partition coefficient such as log *P* or log *D* depend on a change in free energy between different phases; here, as an example the partition coefficient between methanol and water is shown. **C**: Free energy difference associated with the insertion of a molecule into a membrane. **D**: Effect of mutations of protein or host residues on free energies of binding. **E**: Relative free energy of binding of one molecule with respect to another, here toluene and benzyl alcohol, **F**: absolute free energies of binding of a small molecule to a host (e.g. protein).

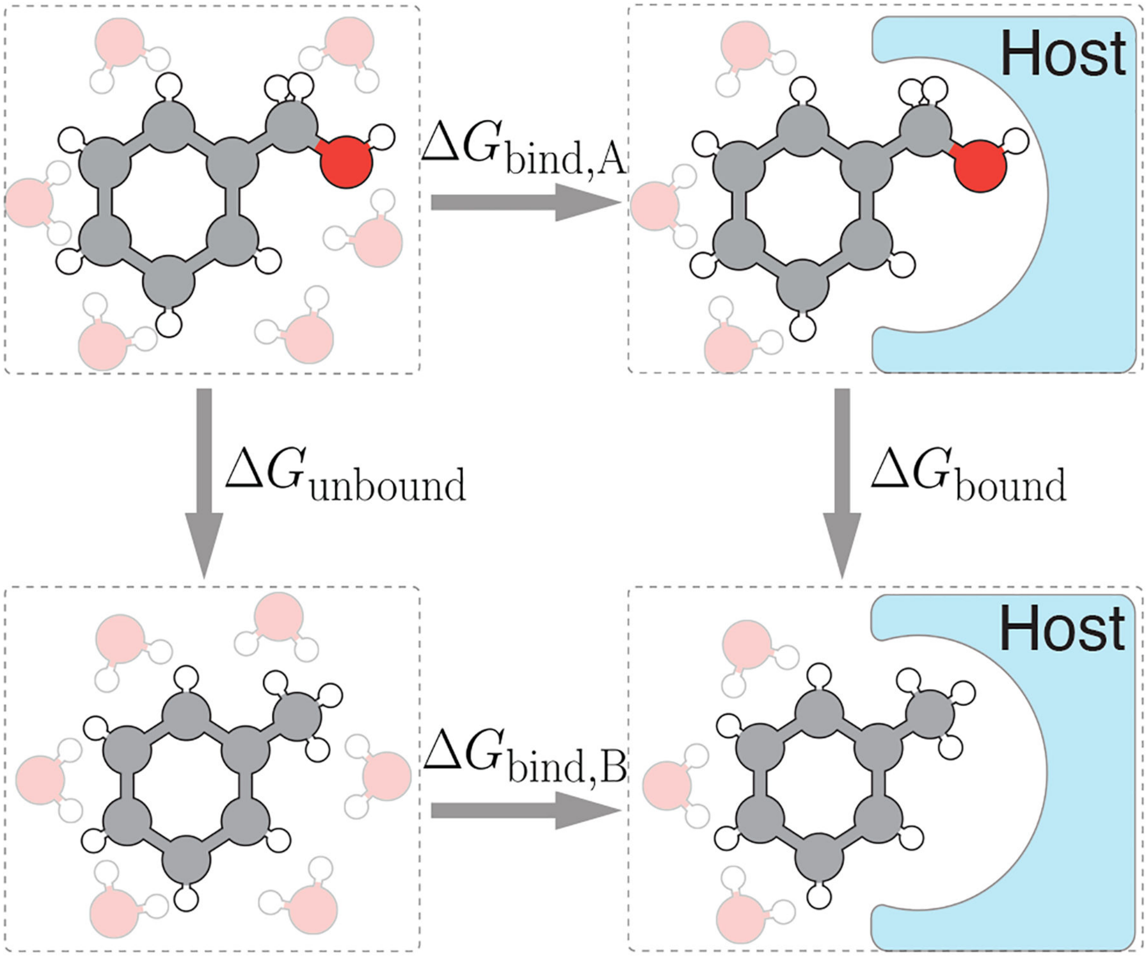

Host
$\Delta G_{\text{bind,A}}$
$\Delta G_{\text{unbound}}$
ΔG_{bound}
$\Delta G_{\text{bind,B}}$
Host

Figure 2. Thermodynamic cycle for computing the relative free energy of binding ($\Delta\Delta G$) between two related small molecules to a supramolecular host or a rigid receptor.

The relative binding free energy difference between two small molecules, $\Delta\Delta G_{\text{bind},A \to B} \equiv \Delta G_{\text{bind},B} - \Delta G_{\text{bind},A}$—here benzyl alcohol (top) to toluene (bottom)—can be computed as a difference between two alchemical transformations, $\Delta G_{\text{bound}} - \Delta G_{\text{solvated}}$, where ΔG_{bound} represents the free energy change of transforming $A \to B$ in complex, i.e. bound to a host molecule, and $\Delta G_{\text{unbound}}$ the free energy change of transforming $A \to B$ in solvent, typically water.

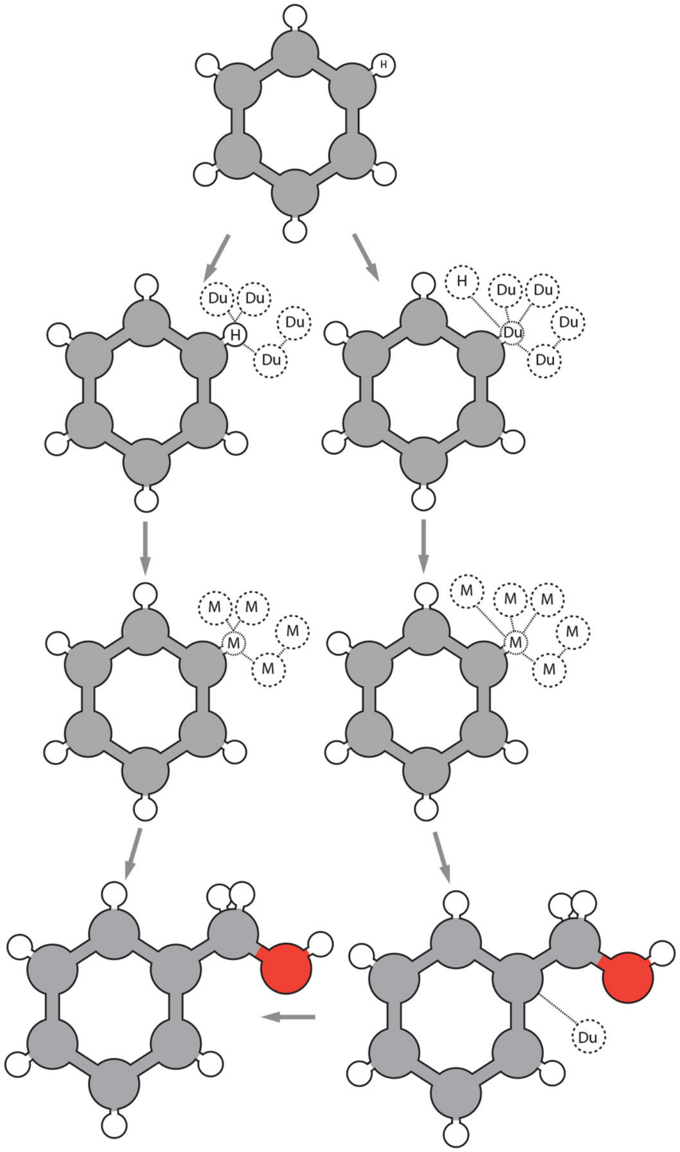


Figure 3. Two common topologies for alchemical calculations: single and dual topology.

Left: A single topology converts from one type of atom to an other. Dummy atoms (Du) are used when there is no corresponding maximum common substructure match between the two molecules for certain atoms, using a soft-core interaction to improve overlap between the dummy atoms and the “real” atoms. **Right**: The dual topology does not convert one species to another, but only converts between Du atoms and an interacting species, but usually uses soft-core potentials for this. The ‘mixed’ intermediate atoms (M) are used in both dual and single topology approaches. Only the way the transformation occurs and the end states differ. Following the arrow along the left and right illustrate the differences. Figure adapted from http://www.alchemistry.org/wiki/Constructing_a_Pathway_of_Intermediate_States

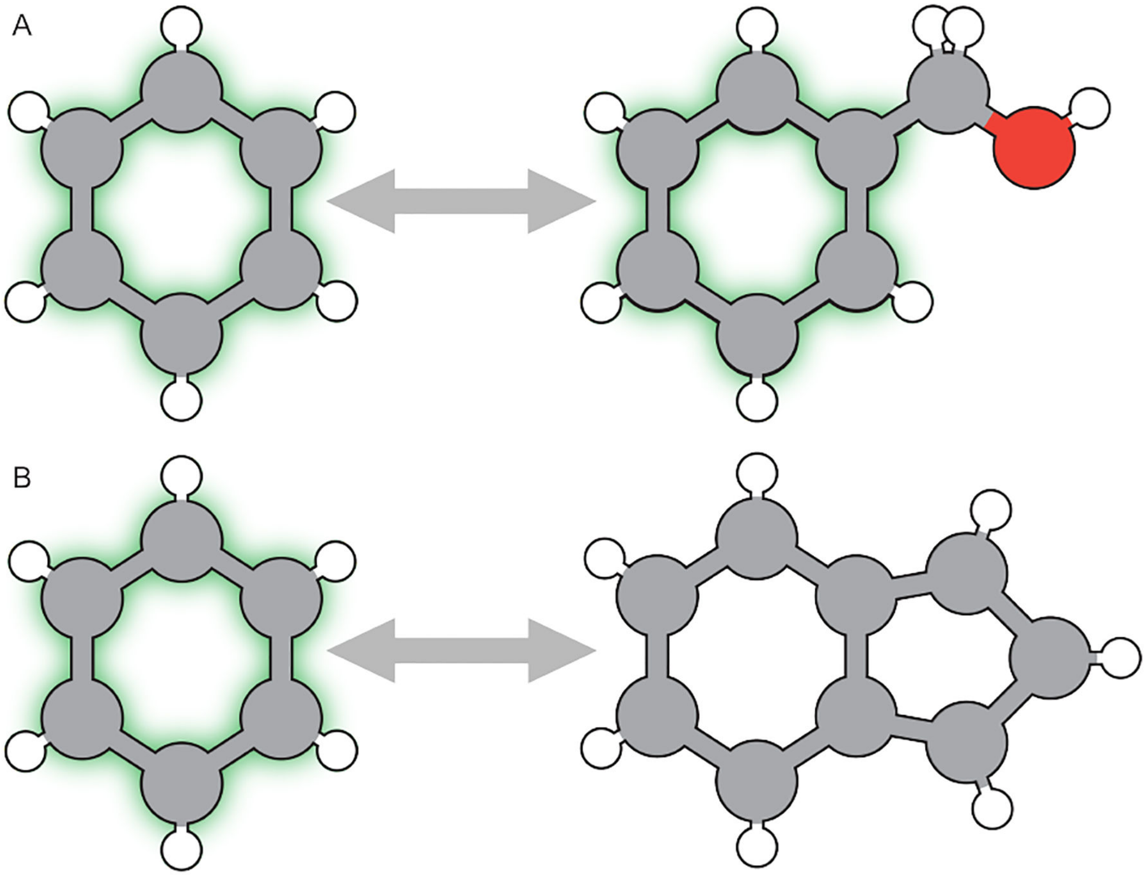


Figure 4. Illustration of maximum common substructure matches

MCSS is shown in green for when (**A**) a restrictive MCSS match is used and in (**B**) ring breaking is allowed, meaning there is no MCSS match between the two compounds.

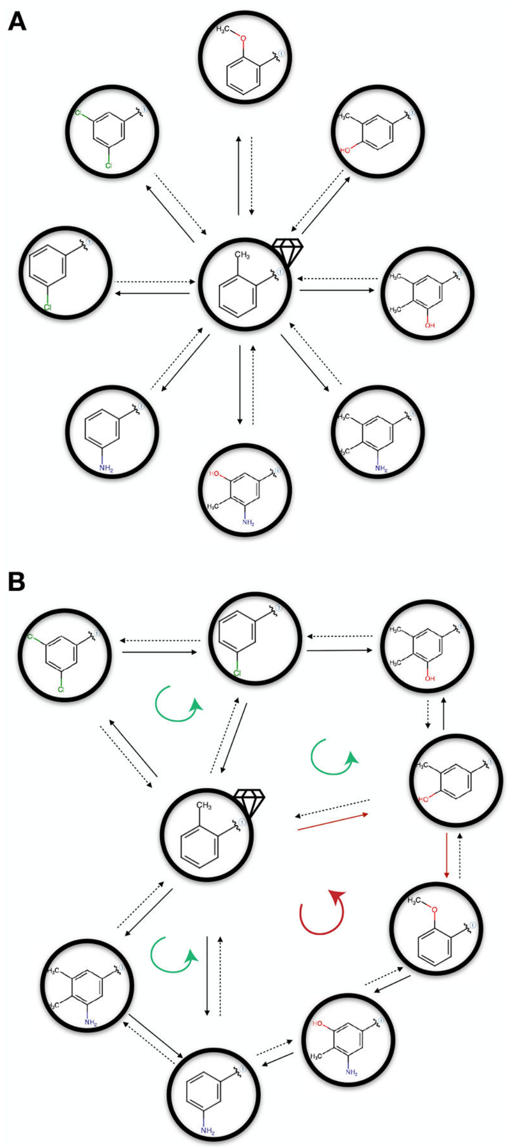


Figure 5. Examples of perturbation networks

(**A**) Star shaped network with the crystal structure in the center. (**B**) Network with cycle closures (see more on this in Sec. 8.5). Arrows indicate the direction of the perturbation. Fully converged binding free energy calculations yield binding free energy changes which sum to zero around any closed cycle. However, in practice errors may not sum to zero around closed cycles, providing a way to look for potential sampling problems. Here in (B), green cycles indicate cycles with hypothetically good cycle closure, red those with poor cycle closure. The red arrow indicates a poorly converged simulation that would give rise to bad cycle closures. The diamond indicates the use of a crystallographic binding mode.

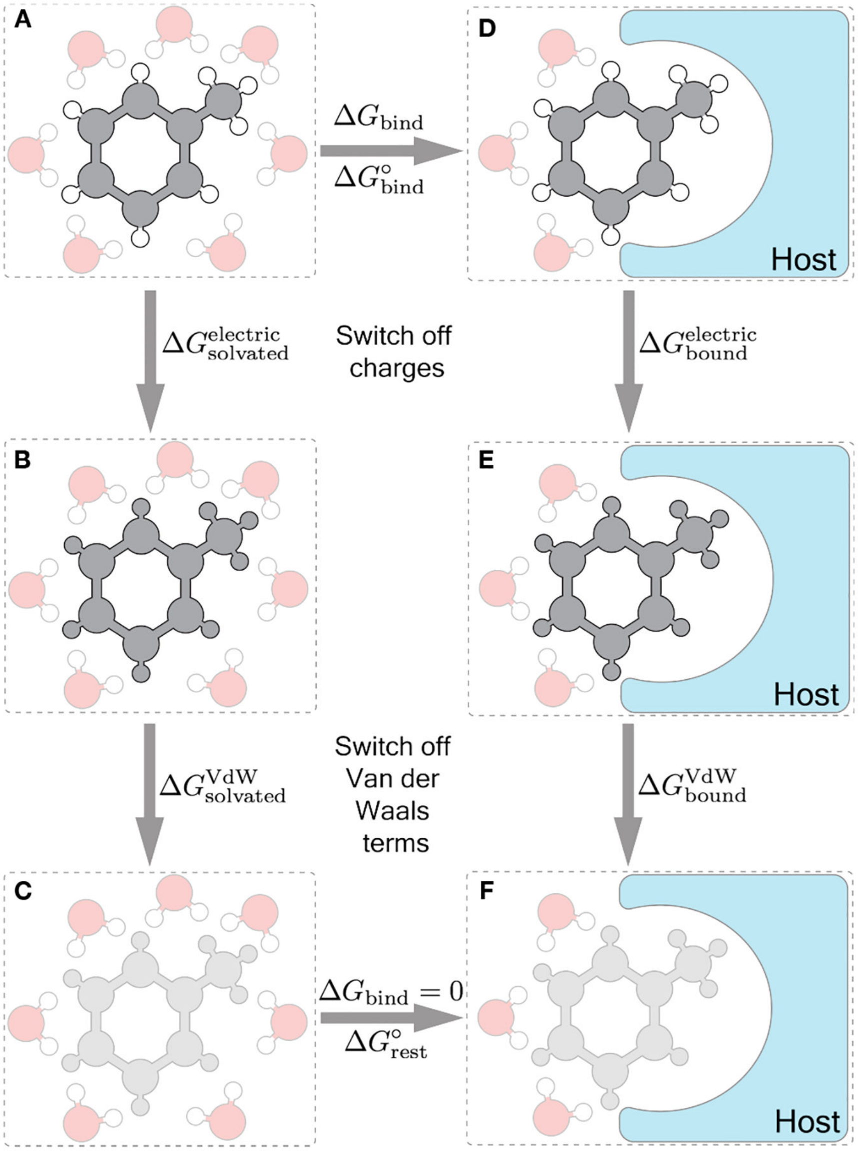

A
D
ΔG_{bind}
$\Delta G^{\circ}_{\mathrm{bind}}$
Host
$\Delta G^{\mathrm{electric}}_{\mathrm{solvated}}$
Switch off charges
$\Delta G^{\mathrm{electric}}_{\mathrm{bound}}$
B
E
Host
$\Delta G^{\mathrm{VdW}}_{\mathrm{solvated}}$
Switch off Van der Waals terms
$\Delta G^{\mathrm{VdW}}_{\mathrm{bound}}$
C
F
$\Delta G_{\mathrm{bind}} = 0$
$\Delta G^{\circ}_{\mathrm{rest}}$
Host

Figure 6. Thermodynamic cycle required for an absolute free energy calculation – absolute free energy of binding example

The fully interacting ligand in water (**A**), has its charges turned off to pass to (**B**) followed by turning of van der Waals terms, resulting in a non-interacting ligand in water in (**C**). Restraints are used on the fully interacting ligand in the binding site of a protein or host molecule (**D**). The next step is to turn off the charges again (**E**) followed by the van der Waals interactions resulting in a non-interacting complex state (**F**). Free energyes can be computed as $\Delta G_{bind} = \left(\Delta G^{\mathrm{elec}}_{\mathrm{solv}} + \Delta G^{\mathrm{VdW}}_{\mathrm{solv}}\right) - \left(\Delta G^{\mathrm{elec}}_{\mathrm{bound}} + \Delta G^{\mathrm{VdW}}_{\mathrm{bound}}\right)$.

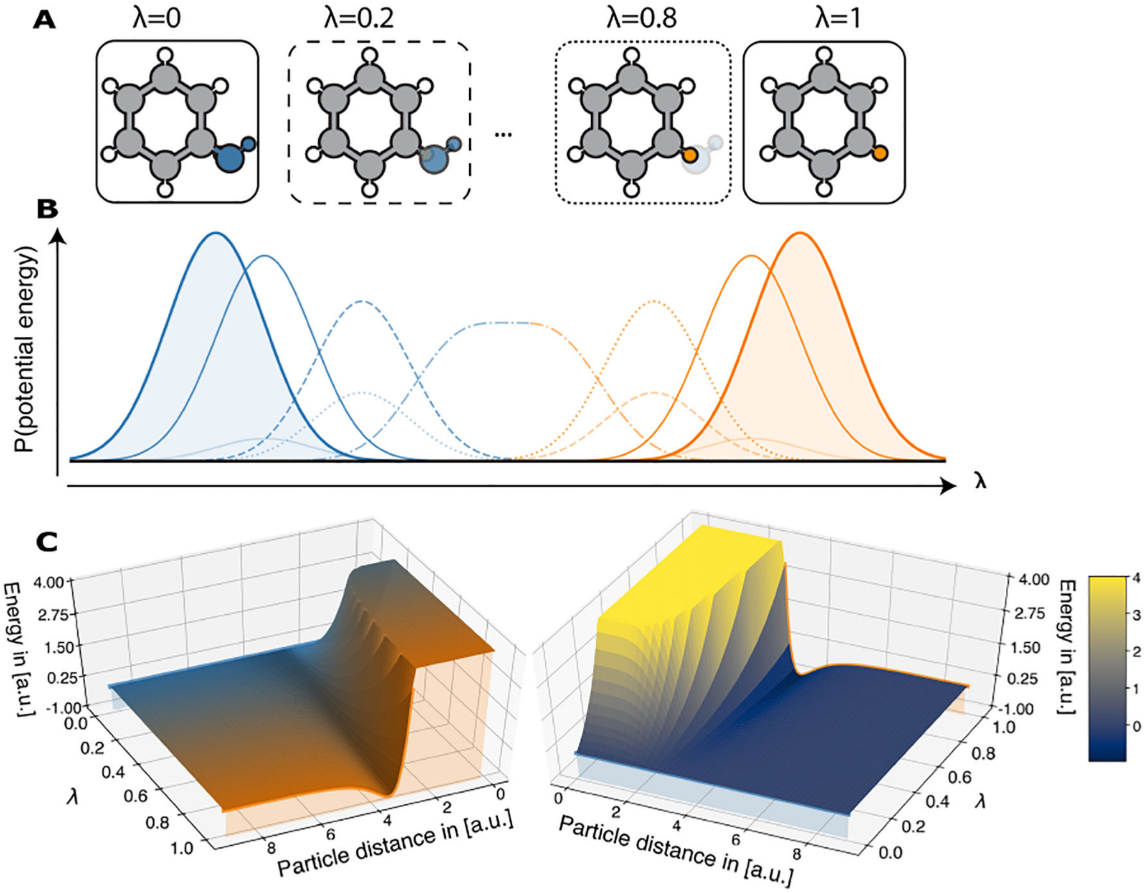


Figure 7.

Alchemical intermediates are created by making the potential energy depend on an additional variable $\vec{\lambda}$ that interpolates between the chemical endpoints. In (**A**), at $\vec{\lambda} = 0$ the molecule is a fully interacting phenol and at $\vec{\lambda} = 1$, a fully interacting benzene. (**B**) shows an illustration of the probability distribution of the potential energies as the switching function takes values of $\vec{\lambda} = 0$ to $\vec{\lambda} = 1$. Intermediates states are required for a sufficient overlap in potential energies to estimate a free energy difference between $\vec{\lambda} = 0$ and $\vec{\lambda} = 1$. Soft-core potentials provide one of the most efficient families of intermediate pathways, with a $\vec{\lambda}$ dependence. In (**C**) the potential energy surface is coloured according to $\vec{\lambda}$ with blue being $\vec{\lambda} = 0$ and $\vec{\lambda} = 1$ orange. In (**D**) the potential is coloured according to the potential energy. Note how as $\vec{\lambda}$ approaches 0, the energy smoothly approaches zero at all *r*, a necessary requirement for efficient and stable calculations.

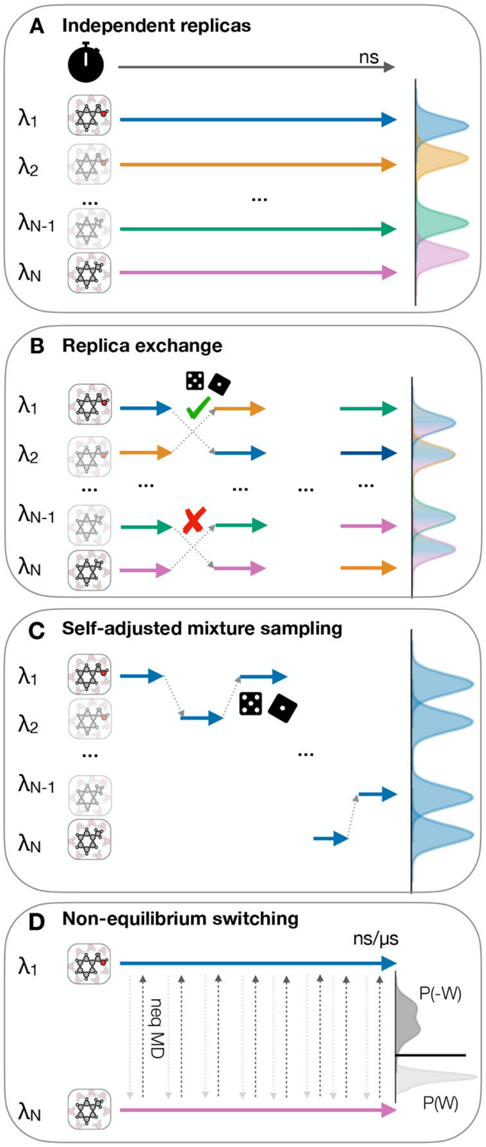


Figure 8. Four most common sampling strategies.

(**A**): Multiple replicas in parallel at different lambda states. Each arrow symbolises an independent $\vec{\lambda}$ simulation. (**B**): Hamiltonian replica exchange scheme. Each arrow represents a short simulation interval before an exchange through Metropolis Hastings acceptance (dice) is attempted. A tick means an accepted exchange, a cross a rejected exchange. (**C**): Single replica scheme sampling from all $\vec{\lambda}$ states. After a short simulation time symbolised by the arrow, the lambdastate is attempted to change until all N lambda states will be sampled. (**D**): Non-equilibrium sampling scheme, where two equilibrium simulations at the end states are run as indicated by the blue and pink arrow. Non-equilibrium simulations are attempted at intervals to switch between the two end states.

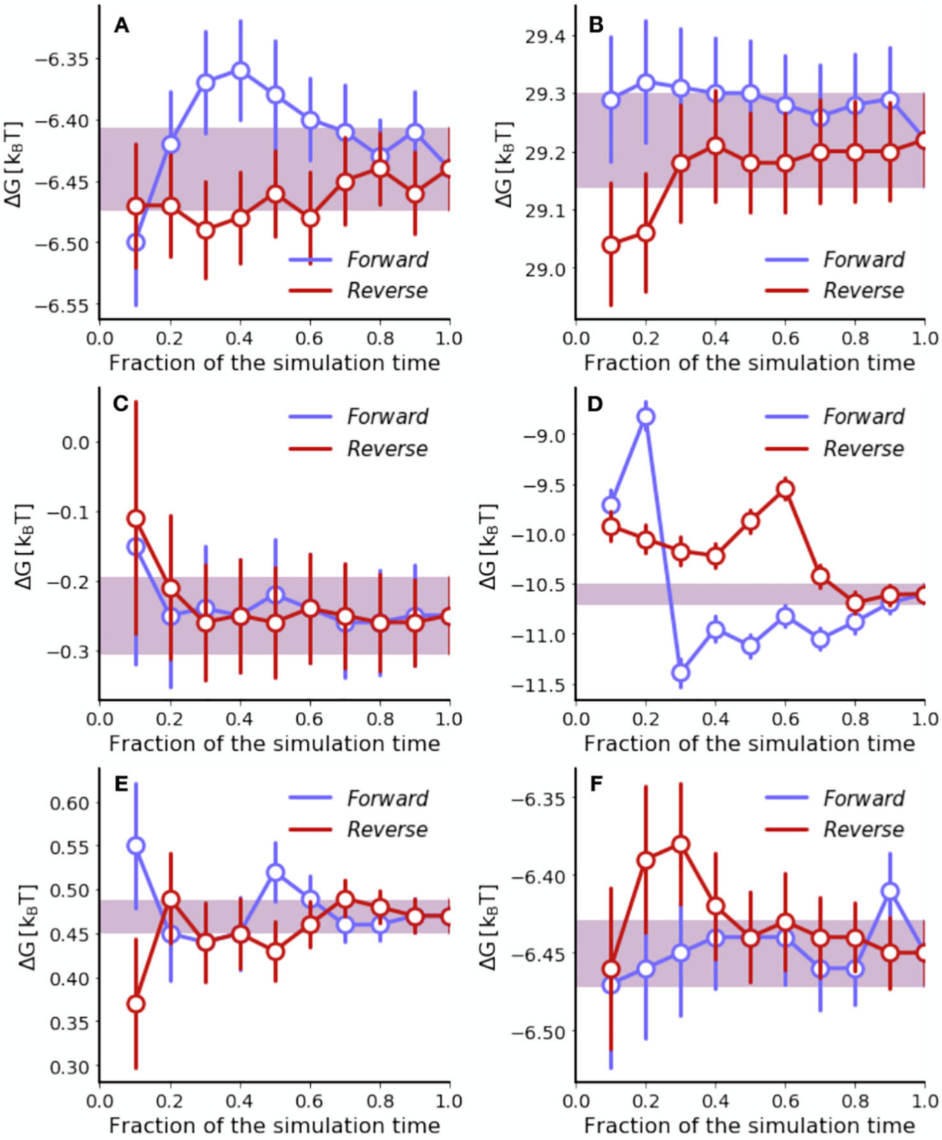


Figure 9. Free energy (in k_BT) for two different relative binding free energy perturbations.
Each plot shows the estimated free energy change using a varying fraction of total simulation time (up to 5 ns total). Subplots (**A**), (**C**), and (**E**) show a three step protocol for a perturbation involving 3 perturbed atoms, while (**B**), (**D**), and (**F**) shows the same protocol for a perturbation involving 10 perturbed atoms. The first step of the protocol is the decharging then removing van der Waals interactions and then recharging. The difference in energy between the forward (blue) and reverse (red) free energy calculations at the midpoint of the simulation time gives an indication of the overall convergence of the simulation, with differences over 1 k_BT indicating poor convergence.

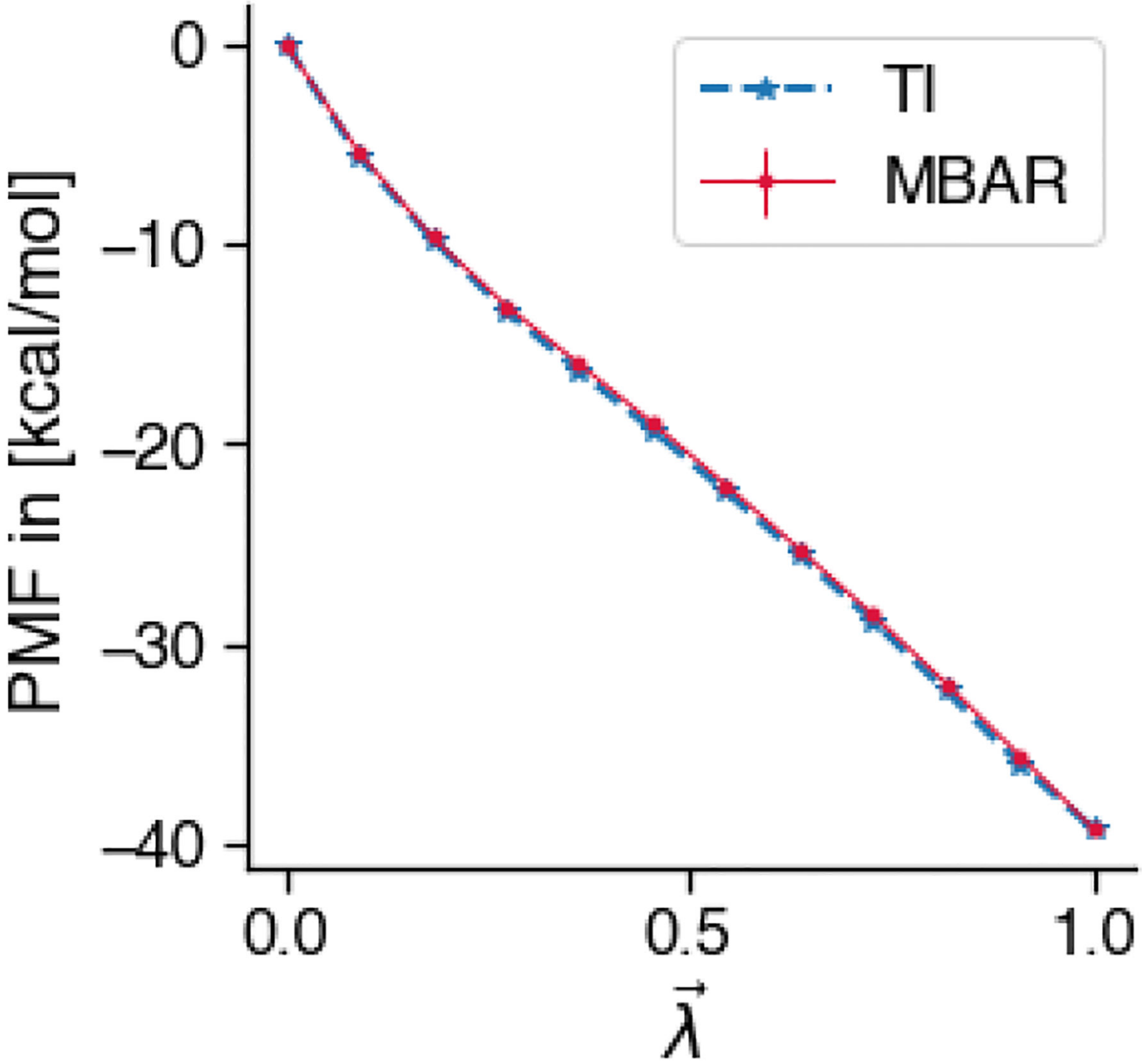


Figure 10. Potential of mean force with respect to $\overrightarrow{\lambda}$ for TI and MBAR

The estimated PMF for a bound calculation of a Tyk2 ligand pair of Wang et al. [15] with respect to $\overrightarrow{\lambda}$ estimated from TI and MBAR and showing agreement within errorbars.

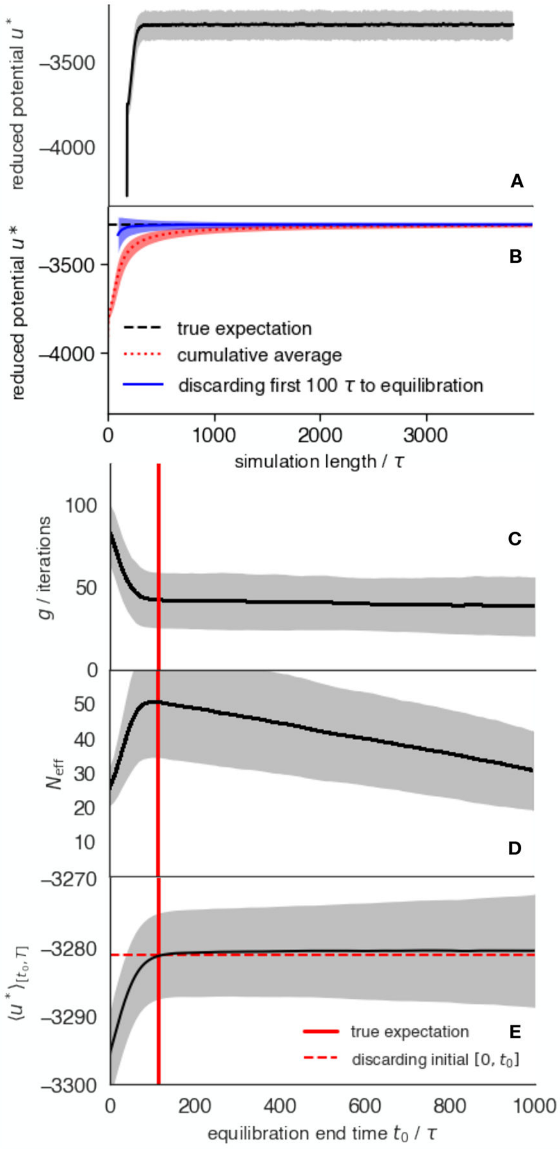


Figure 11. Automatic partitioning into equilibration and production regions.

(**A**) The average (black line) standard deviation (shaded region) of the reduced potential *u** over many independent replicate simulations started from the same initial conditions show a significant initial transient change before relaxing to the true average potential energy (**B**). A cumulative average (red) of the entire simulation data demonstrates simulation bias not seen when initial simulation data is omitted (blue). Using an automated approach to detect equilibration of the boundary t_0 using statistical inefficiency g (**C**) for an effective simulation interval (**D**). (**E**) The optimal equilibration boundary t_0 is selected to maximize the number of uncorrelated samples. *Figure adapted from* [210].

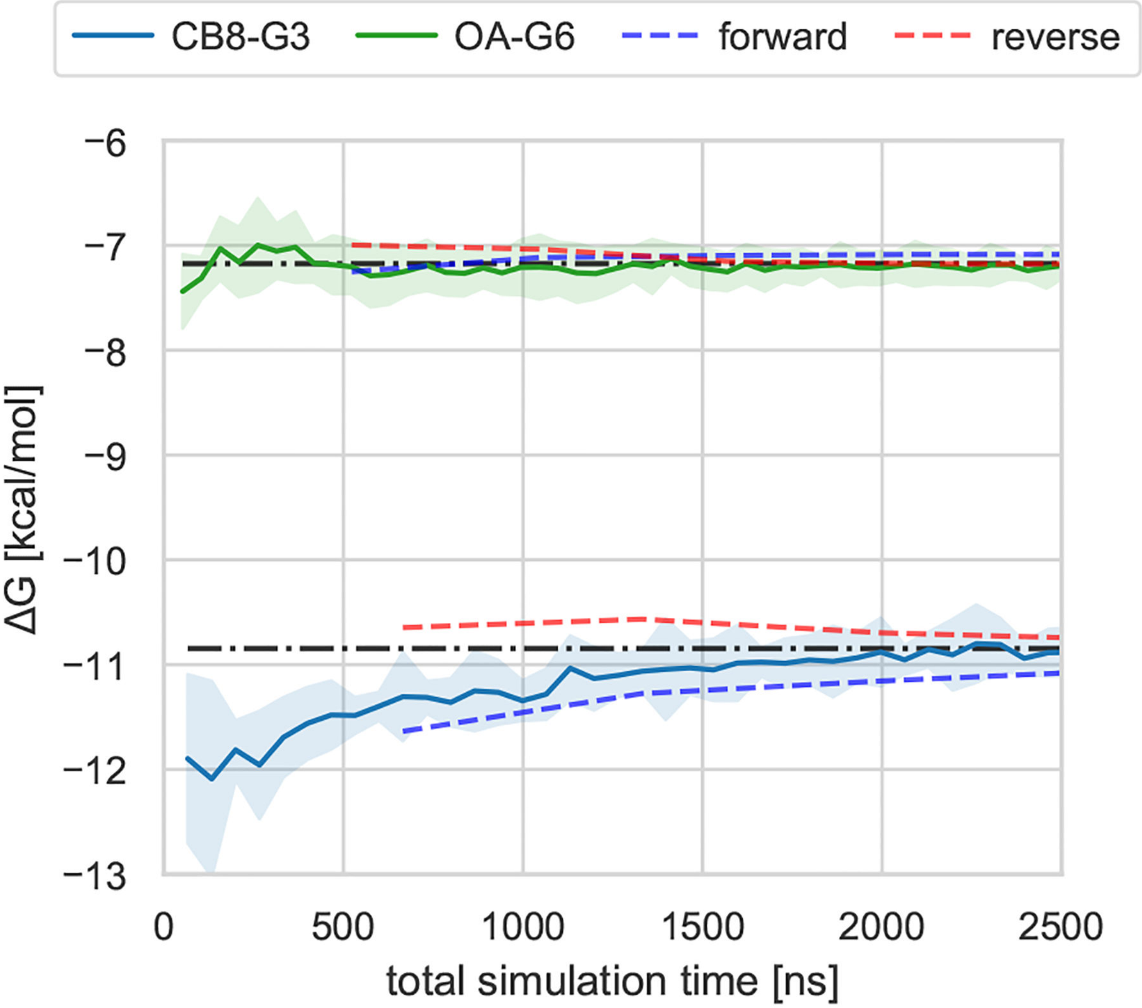
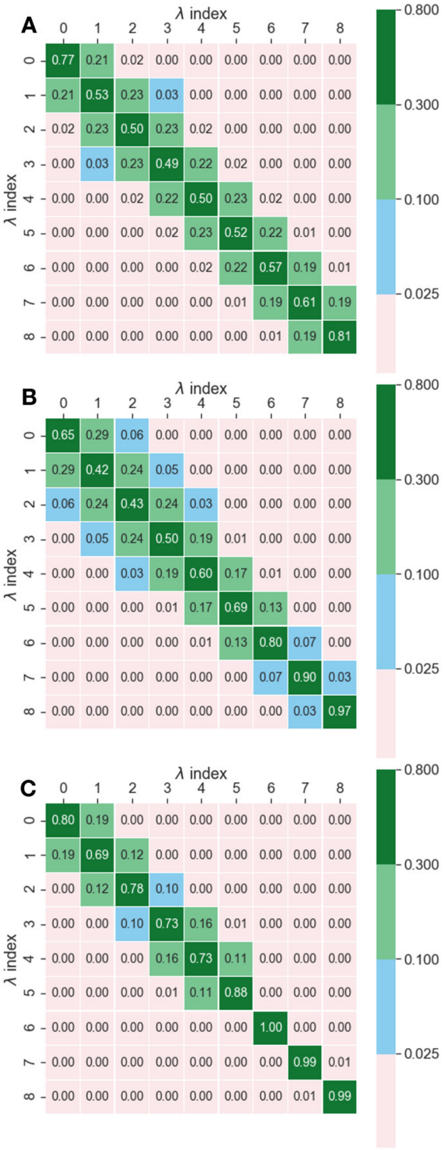


Figure 12.

Average binding free energy of 5 replicate Hamiltonian replica exchange calculations as a function of total simulation time (i.e. the sum of the simulation time of all replicas) for the two host-guest systems CB8-G3 and OA-G6. Shaded areas represent 95% confidence intervals around the mean computed from the 5 replicates data. The horizontal dash-dot lines show the final binding free energy prediction of the two calculations after a total of 5230 ns for OA-G6 and 6650 ns for CB8-G3. Dashed lines are the free energy trajectories computed in the forward (blue) and reverse (red) time direction for a single replicate calculation. Longer correlation times in CB8-G3 cause the calculation to converge more slowly. The original data used to generate the plot can be found at https://github.com/samplchallenges/SAMPL6/blob/master/host_guest/Analysis/SAMPLing/Data/reference_free_energies.csv.



A

λ index	0	1	2	3	4	5	6	7	8
0	0.77	0.21	0.02	0.00	0.00	0.00	0.00	0.00	0.00
1	0.21	0.53	0.23	0.03	0.00	0.00	0.00	0.00	0.00
2	0.02	0.23	0.50	0.23	0.02	0.00	0.00	0.00	0.00
3	0.00	0.03	0.23	0.49	0.22	0.02	0.00	0.00	0.00
4	0.00	0.00	0.02	0.22	0.50	0.23	0.02	0.00	0.00
5	0.00	0.00	0.00	0.02	0.23	0.52	0.22	0.01	0.00
6	0.00	0.00	0.00	0.00	0.02	0.22	0.57	0.19	0.01
7	0.00	0.00	0.00	0.00	0.00	0.01	0.19	0.61	0.19
8	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.19	0.81

B

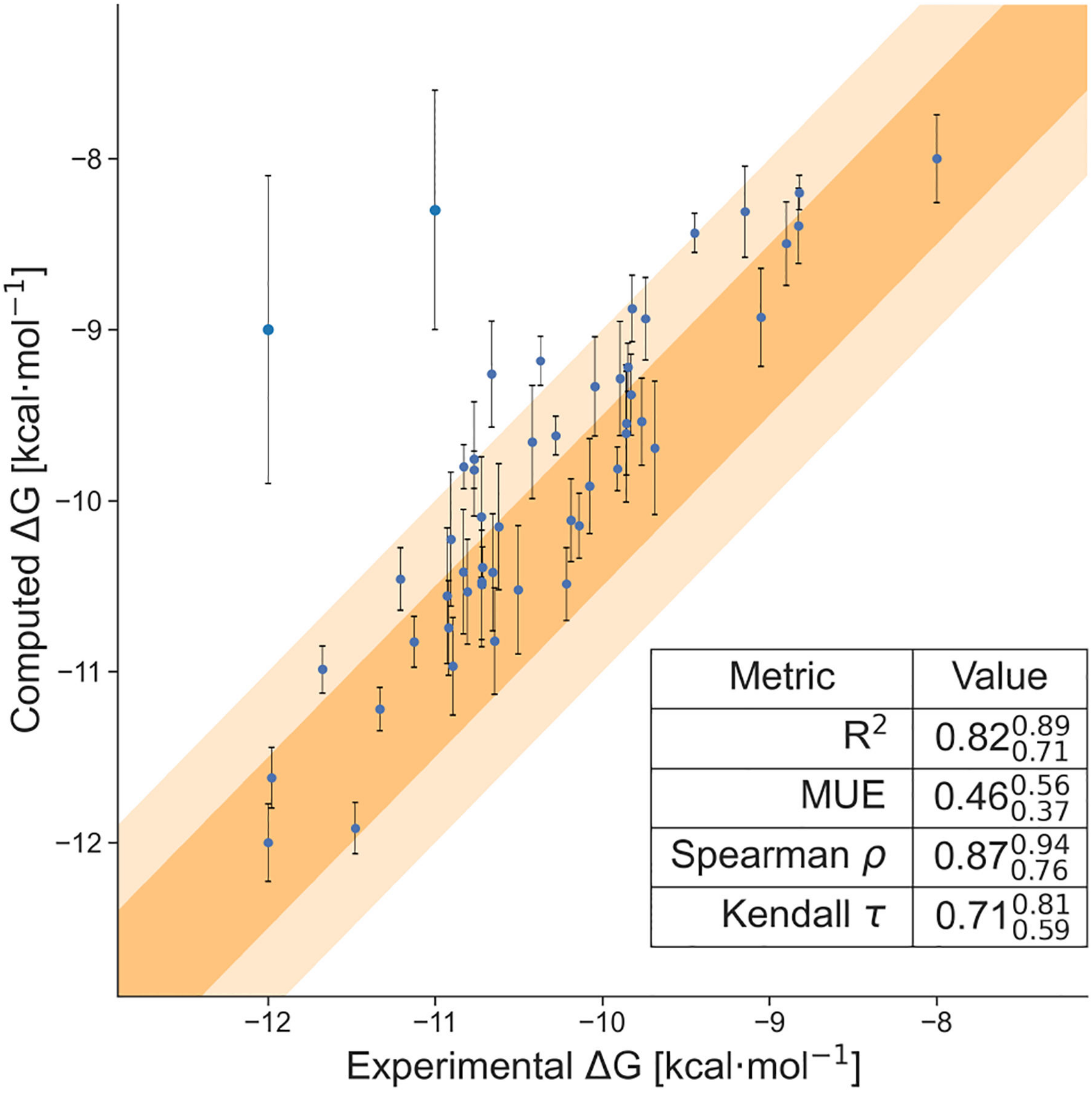
λ index	0	1	2	3	4	5	6	7	8
0	0.65	0.29	0.06	0.00	0.00	0.00	0.00	0.00	0.00
1	0.29	0.42	0.24	0.05	0.00	0.00	0.00	0.00	0.00
2	0.06	0.24	0.43	0.24	0.03	0.00	0.00	0.00	0.00
3	0.00	0.05	0.24	0.50	0.19	0.01	0.00	0.00	0.00
4	0.00	0.00	0.03	0.19	0.60	0.17	0.01	0.00	0.00
5	0.00	0.00	0.00	0.01	0.17	0.69	0.13	0.00	0.00
6	0.00	0.00	0.00	0.00	0.01	0.13	0.80	0.07	0.00
7	0.00	0.00	0.00	0.00	0.00	0.00	0.07	0.90	0.03
8	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.03	0.97

C

λ index	0	1	2	3	4	5	6	7	8
0	0.80	0.19	0.00	0.00	0.00	0.00	0.00	0.00	0.00
1	0.19	0.69	0.12	0.00	0.00	0.00	0.00	0.00	0.00
2	0.00	0.12	0.78	0.10	0.00	0.00	0.00	0.00	0.00
3	0.00	0.00	0.10	0.73	0.16	0.01	0.00	0.00	0.00
4	0.00	0.00	0.00	0.16	0.73	0.11	0.00	0.00	0.00
5	0.00	0.00	0.00	0.01	0.11	0.88	0.00	0.00	0.00
6	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00
7	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.99	0.01
8	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.99

Figure 13. Overlap matrices:

Visualising overlap matrices can help with assessing the quality of simulation data. (**A**) shows good overlap with all first off-diagonal entries well above 0.03, the suggested threshold, (**B**) is an example of mediocre overlap with good overlap at lower $\vec{\lambda}$ values and poor overlap at high $\vec{\lambda}$ values. (**C**) shows poor overlap resulting in disconnected simulations with unreliable MBAR estimates.



Metric	Value
R^2	$0.82_{0.71}^{0.89}$
MUE	$0.46_{0.37}^{0.56}$
Spearman ρ	$0.87_{0.76}^{0.94}$
Kendall τ	$0.71_{0.59}^{0.81}$

Figure 14. An example of recommended practices for graphing alchemical free energy predictions.

This figure shows the relation between predicted and experimentally-determined Gibbs free energy in kcal/mol with standard errors as error bars. The dark and light-orange regions depict the 1- and 2-kcal/mol confidence bounds. Statistical metrics for the data are reported, with 95% confidence intervals determined by bootstrapping analysis. Extra care should be taken when investigating potential outliers further.

Table 1.

Selection of example datasets

Publication	Targets	Ligands	Force Field
D3R Grand Challenges [255]			
GC3 [241]	6	266	various
GC2 [240]	1	102	various
GC2015 [256]	2	215	various
SAMPL Challenges [257]			
SAMPL6 [258]	3	21	various
SAMPL5 [259]	3	22	various
SAMPL4 [260]	2	23	various
Schrödinger Datasets			
FEP+ Dataset [15]	8	199	OPLS2.1
FEP+ Dataset [81]	8	199	OPLS3
FEP+ Dataset [261]	8	199	OPLS3e
FEP+ Dataset [17]	8	199	GAFF 1.8
FEP+ Dataset [18]	8	199	various
FEP+ Dataset [19]	8	199	GAFF2.1
Fragments [201]	8	96	OPLS2.1
Scaffold Hopping [104]	6	21	OPLS3
Scaffold Hopping [19]	6	21	GAFF2.1
Macrocycles [262]	7	33	OPLS3
Further Suggested Datasets			
Cucurbit[7]uril (CB7) [228]	1	15	NA
Deep cavity cavitand [228]	2	19	NA
T4 Lysozyme [228]	2	20	NA
Merck set [263]	5	169	OPSL3