

UC Irvine

UC Irvine Previously Published Works

Title

Addressing Intersite Coupling Unlocks Large Combinatorial Chemical Spaces for Alchemical Free Energy Methods

Permalink

<https://escholarship.org/uc/item/1zf827wf>

Journal

Journal of Chemical Theory and Computation, 18(4)

ISSN

1549-9618

Authors

Hayes, Ryan L
Vilseck, Jonah Z
Brooks, Charles L

Publication Date

2022-04-12

DOI

10.1021/acs.jctc.1c00948

Peer reviewed



Published in final edited form as:

J Chem Theory Comput. 2022 April 12; 18(4): 2114–2123. doi:10.1021/acs.jctc.1c00948.

Addressing Intersite Coupling Unlocks Large Combinatorial Chemical Spaces for Alchemical Free Energy Methods

Ryan L. Hayes^{†,‡}, Jonah Z. Vilseck^{¶,§}, Charles L. Brooks III^{†,‡}

[†]Department of Chemistry, University of Michigan, Ann Arbor, Michigan 48109, United States

[‡]Biophysics Program, University of Michigan, Ann Arbor, Michigan 48109, United States

[¶]Department of Biochemistry and Molecular Biology, Indiana University School of Medicine, Indianapolis, Indiana 46202, United States

[§]Center for Computational Biology and Bioinformatics, Indiana University School of Medicine, Indianapolis, Indiana 46202, United States

Abstract

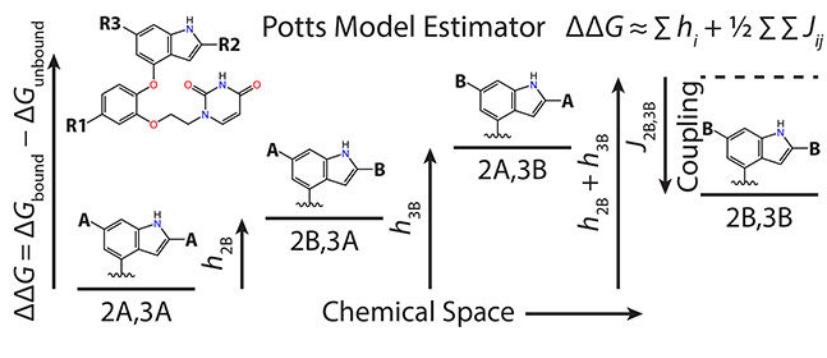
Alchemical free energy methods are playing a growing role in molecular design, both for computer-aided drug design of small molecules and for computational protein design. Multisite λ dynamics (MS λ D) is a uniquely scalable alchemical free energy method that enables more efficient exploration of combinatorial alchemical spaces encountered in molecular design, but simulations have typically been limited to a few hundred ligands or sequences. Here we focus on coupling between sites to enable scaling to larger alchemical spaces. We first discuss updates to the biasing potentials that facilitate MS λ D sampling to include coupling terms and show that this can provide more thorough sampling of alchemical states. We then harness coupling between sites by developing a new free energy estimator based on the Potts models underlying direct coupling analysis, a method for predicting contacts from sequence coevolution, and find it yields more accurate free energies than previous estimators. The sampling requirements of the Potts model estimator scale with the square of the number of sites, a substantial improvement over the exponential scaling of the standard estimator. This opens up exploration of much larger alchemical spaces with MS λ D for molecular design.

Graphical Abstract

brooksc1@umich.edu, Phone: (734) 647-6682.

Supporting Information Available

Supporting Information contains details of the various ALF methods, a discussion of FPL, details for system setup, results highlighting the superiority of PME electrostatics over force switching, the specific flattening protocols used to obtain results, more detailed analysis of the number of sequences sampled with varying ALF methods, calculations of estimator accuracy relative to experiment and Potts model coupling magnitude and uncertainty, and reference 66. Scripts used for landscape flattening are available for download at <https://github.com/RyanLeeHayes/ALF/releases/tag/ALF-3.1>. System input files are available for download at <https://github.com/RyanLeeHayes/PublicationScripts/blob/main/2021Coupling.tgz>.



Introduction

Alchemical free energy methods are an exciting class of molecular simulation techniques that allow calculation of relative free energies for problems including protein-ligand binding,¹⁻⁴ protein folding,⁵⁻⁸ pH-driven protonation events,⁹⁻¹⁴ host-guest binding,^{15,16} and small molecule solvation.^{17,18} Among these diverse problems, computing binding free energies for computer-aided drug design (CADD)¹⁹⁻²¹ and computing changes in stability for computational protein design^{22,23} are becoming increasingly relevant for molecular design.

In order to maximize the impact of alchemical free energy methods on molecular design projects, it is essential to be able to explore large ligand chemical spaces and large protein sequence spaces. The accuracy of alchemical methods for CADD has been well established in retrospective studies when the experimental result is known,^{1,2} proven in prospective studies when it is not,¹⁹ and streamlined for commercial use.^{20,24,25} For alchemical methods to contribute to industrial scale CADD, it is necessary to computationally screen an order of magnitude more ligands than can be synthesized, which equates to thousands of candidate molecules.^{26,27} Alchemical methods are less developed in the context of protein design, but there are reports of impressive accuracy and early successes.^{5-8,22,23} Protein design often involves optimization at dozens of amino acid positions,²⁸ which requires searching through truly astronomical numbers of sequences. Beyond protein design, there is substantial experimental interest in these combinatorial sequence spaces for insight into epistasis and protein evolution.²⁹⁻³¹

Among alchemical free energy methods, multisite λ dynamics (MS λ D) is uniquely well suited for exploring large combinatorial alchemical spaces relevant in CADD and computational protein design. While many free energy methods like free energy perturbation (FEP),³² thermodynamic integration (TI),³³ and non-equilibrium fast growth TI³⁴⁻³⁶ require a set of simulations to compare a single pair of ligands in chemical space or a pair of sequences in sequence space, a single set of MS λ D simulations allows comparison of combinatorial chemical and sequence spaces arising from several perturbations at several sites.^{37,38} Consequently, MS λ D has previously been used to explore spaces of hundreds of ligands^{3,20} or sequences.⁷

While MS λ D is ideally suited to explore large combinatorial spaces, two technical limitations related to coupling between sites have prevented exploration of spaces larger

than several hundred ligands or sequences. In this work, coupling refers to interactions between sites (pairwise or higher order) that prevent the free energy of sequences or ligands from being broken down into an independent contribution from each site. The first technical limitation is that converged free energy estimates by MS λ D require many spontaneous transitions between ligands or sequences within the alchemical space, so an adaptive landscape flattening (ALF) algorithm is used to optimize a biasing potential that flattens barriers in alchemical space;^{7,39} however, ALF currently treats each site as independent. As the number of sites increases, couplings between sites focus excessive sampling on some favorable combinations of perturbations, while preventing sampling of other mutually exclusive pairs of perturbations. Second, MS λ D utilizes a histogram-based free energy estimator that includes all higher order couplings and determines the relative free energy of a ligand or sequence from the fraction of time the alchemical coordinates are within a threshold of that particular ligand or sequence. As the alchemical space grows, more of the time is spent in irrelevant alchemical intermediates rather than the alchemical endpoints corresponding to ligands or sequences. Even if measures are taken to minimize time spent in alchemical intermediates, the time is split between more states, so sampling requirements grow at least linearly with the number of ligands or sequences considered, or exponentially with the number of sites.

Consequently, new techniques for ALF and free energy estimation are required to enable sampling of much larger alchemical spaces. Therefore, we include additional bias terms in the bias optimized by ALF to overcome the sampling difficulties caused by pairwise coupling between sites. We then harness these pairwise couplings by introducing a Potts model free energy estimator that has been repurposed from predictions of protein contacts with direct coupling analysis.^{40–42} The Potts model estimator includes only single site terms and pairwise couplings between sites, and sacrifices higher order couplings that tend to be small and noisy, in order to substantially reduce sampling requirements. We demonstrate that including coupling bias terms in ALF can significantly improve the quality of the biasing potential in repeated production runs of a previously studied protein perturbation system, T4 lysozyme,⁷ as quantified by a factor of 2 to 5 decrease in the number of sequences that are not sampled. We further show that the Potts model estimator gives superior results to the histogram-based estimator and a recently proposed independent site estimator²⁰ on several multisite T4 lysozyme systems⁷ and on two multisite drug binding systems.^{3,20} Finally, we find that in contrast to the exponential scaling of the histogram-based estimator, the Potts model estimator sampling requirements scale with the square of the number of sites or better, depending on the choice of convergence criteria. These advances will enable exploration of much larger chemical and sequence spaces with MS λ D and have already found use in an ongoing study of a space of 32768 sequences arising from mutations at 15 sites in ribonuclease H.

Theoretical Methods

Basics of MS λ D

Alchemical free energy methods all make use of thermodynamic cycles like the two shown in Figure 1. Taking ligand binding as an example, the relative free energy of binding can be

expressed as the difference of two physical processes (horizontal arrows) or two alchemical processes (vertical arrows):

$$\begin{aligned} \Delta \Delta G_{\text{binding}}(\text{L1} \rightarrow \text{L2}) &= \Delta G_{\text{binding}}(\text{L2}) - \Delta G_{\text{binding}}(\text{L1}) \\ &= \Delta G_{\text{bound}}(\text{L1} \rightarrow \text{L2}) - \Delta G_{\text{unbound}}(\text{L1} \rightarrow \text{L2}) \end{aligned} \quad (1)$$

Because physical processes such as binding or folding are much slower than the time scales accessible with molecular dynamics simulation, alchemical free energy methods evaluate the free energies of the two vertical processes, typically by introducing a coupling parameter λ into the potential energy function U .

In MS λ D, the single dimensional coupling parameter can be generalized to a higher dimensional alchemical space allowing combinatorial permutations of perturbations at M different sites

$$U = U_0 + \sum_{s=1}^M \sum_{i=1}^{N_s} \lambda_{si} U_{si} + \sum_{s=1}^M \sum_{t=s+1}^M \sum_{i=1}^{N_s} \sum_{j=1}^{N_t} \lambda_{si} \lambda_{tj} U_{si,tj} + U_{\text{bias}}(\vec{\lambda}) \quad (2)$$

where λ_{si} is the λ coefficient for site s , substituent i , N_s is the number of substituents at site s , U_0 are interactions of environment atoms not involved in the perturbation, U_{si} are interactions of atoms at site s substituent i among themselves and with the environment, and $U_{si,tj}$ are interactions between atoms at two different sites s and t . Substituent bonded terms are often not scaled by λ , and in the present work, bonds, angles, and improper dihedrals of substituents appear in U_0 and only proper dihedrals of substituents are included in U_{si} . (See Reference 8 for a discussion of when this is necessary or appropriate.) Equation 2 has been generalized to include particle mesh Ewald electrostatics;^{12,43,44} generalizations for implicit solvent or polarizable forcefields should be possible, but have not been described. The alchemical λ degrees of freedom fluctuate according to implicit constraints,

$$\lambda_{si} = \frac{\exp(c \sin(\theta_{si}))}{\sum_j^{N_s} \exp(c \sin(\theta_{sj}))} \quad (3)$$

where the θ_{si} degrees of freedom are propagated analogously to the spatial degrees of freedom, and c is typically set to 5.5.⁴⁵ Consequently, the alchemical degrees of freedom can become trapped in favorable regions of alchemical space, so a bias U_{bias} that is a function of the $\vec{\lambda}$ state of the system is added to the potential and tuned to flatten the landscape and prevent trapping.

Adaptive Landscape Flattening of Coupled Biases

Adaptive Landscape Flattening (ALF) was introduced to tune the biasing potential in an automated fashion to optimize sampling,³⁹ and was subsequently updated to utilize a linearized least squares approach that allows ALF to be easily extended to new biasing potentials.⁷ Briefly, several free energy profiles are computed as a function of alchemical

coordinates, and precomputed reference profiles for ideal flat landscapes are subtracted. For each bin in each profile, the square of the deviation from the average value of the profile is added to a penalty function. Consequently, the penalty function is minimized for a flat landscape when all bins in each profile have the same free energy. A flat landscape accelerates convergence of free energy estimates both through barrierless transitions between end states and through even sampling of end states.

To optimize the bias, the linear dependence of the profiles on the bias parameters is computed, and the penalty function is minimized with respect to the bias parameters based on this linear approximation, (see Supporting Information for mathematical details). Previously, ALF treated sites as completely independent in two ways. First, the biasing potentials in Equations 4–7 below only included single site terms for which $s = t$. Second, in bias optimization, the derivatives representing the change in profiles at one site with respect to biases at another site were approximated as zero. We denote this previous approach as fully independent ALF. Several sequences failed to sample during simulations of multisite mutants in a previous study of T4 lysozyme,⁷ and further analysis revealed this was partially due to coupling between sites, where free energy differences between substituents at one site depend on the substituent at another site.

To capture coupling between pairs of sites, several modifications to ALF are tested. As a first alternative, the original bias with no intersite coupling terms is used, but additional free energy profiles that are functions of alchemical coordinates at multiple sites may be included. We further account for coupling during bias optimization by explicitly computing and including the derivatives describing the change in profiles at one site with respect to changes in bias parameters at another site, which we call coupling aware ALF. This may result in identification of better bias parameters, but cannot actually flatten the coupling between sites to give a landscape in which free energy differences at one site are independent of substituents other sites.

To flatten coupling between sites requires biases that are functions of alchemical coordinates at multiple sites, so the biasing potentials are generalized to the following forms

$$V_{\text{Fixed}} = \sum_s^M \sum_i^{N_s} \phi_{si} \lambda_{si} \quad (4)$$

$$V_{\text{Quad}} = \sum_s^M \sum_i^{N_s} \sum_t^M \sum_j^{N_t} \psi_{si,tj} \lambda_{si} \lambda_{tj} \quad (5)$$

$$V_{\text{Skew}} = \sum_s^M \sum_i^{N_s} \sum_t^M \sum_j^{N_t} \chi_{si,tj} \lambda_{tj} (1 - \exp(-\lambda_{si} / \sigma)) \quad (6)$$

$$V_{\text{End}} = \sum_s^M \sum_i^{N_s} \sum_t^M \sum_j^{N_t} \omega_{si,tj} \lambda_{tj} \lambda_{si} / (\alpha + \lambda_{si}) \quad (7)$$

where $\sigma = 0.18$, $\alpha = 0.017$, and $si = tj$ terms are omitted. Previous work has shown the shape of various components of the free energy profiles, and thus of the biases needed to flatten them, is relatively consistent, and only the amplitudes, ϕ , ψ , χ , and ω , change.^{7,39} The V_{Fixed} term can be tuned to ensure that each substituent at a site is sampled equally,^{39,46} while the V_{Quad} can remove the bulk of the barriers to alchemical transitions between substituents.³⁹ The forms of the two remaining terms were determined by heuristic fitting of free energy barriers. The V_{End} term is needed to account for deep traps near alchemical endpoints that are partially due to the free energy cost of displacing solvent as a substituent turns on;³⁹ this term removed the need to use an implicit constraint parameter of $c = 2.5$ to avoid becoming trapped at the endpoints that was observed in Reference 47. While ALF did not initially utilize V_{Skew} ,³⁹ its was subsequently found to give improved fits to free energy profiles when using soft-core interactions.⁷

While the intersite ψ terms are essential for capturing coupling, the intersite χ and ω terms tend to be small, but can drift to large values as a group to cover deficiencies in the bias that ought to be corrected by adjusting the intrasite χ and ω terms. Two different approaches were taken to keep the intersite χ and ω terms small. First, the regularization term in the ALF algorithm was modified to restrain these terms to remain close to zero (see Supporting Information), which we call ψ, χ, ω coupling ALF. Alternatively, since the computational cost of the least squares approach for ALF scales like the number of biases, we set $\chi_{si,tj} = \omega_{si,tj} = 0$ for $s \neq t$, which cuts the number of intersite biases by a factor of five, and which we call only ψ coupling ALF. For the systems considered here, ALF is fast, and MS λ D simulations, whose computational cost has negligible dependence on the number of biases,⁴⁸ are rate limiting. However, for larger alchemical systems, ALF may become rate limiting, and utilizing only ψ coupling ALF may improve generalizability to these larger alchemical spaces.

Free Energy Estimators and Potts Models

Historically, free energy differences are estimated from MS λ D simulations by binning states above a λ_c cutoff, but in this work, three estimators are considered: the original histogram-based estimator, the Potts model estimator, and the independent site estimator.

The histogram-based estimator counts the number of times each ligand is sampled, Boltzmann inverts the populations, and corrects for the biasing potentials.^{7,38,46} A ligand is considered to be sampled during a frame if λ for each substituent is greater than the λ_c cutoff, usually chosen to be 0.99. λ_c is chosen to be close to 1 to minimize errors due to the finite width of the bin.^{49,50} The relative free energy of an alchemical state u is then given by

$$\Delta G_H(u) = -U_{\text{bias}}(\vec{\lambda}_u) - k_B T \ln \left(\int dt \prod_{s=1}^M \Theta(\lambda_{su_s}(t) - \lambda_c) \right) \quad (8)$$

up to an arbitrary additive constant, where $U_{\text{bias}}(\vec{\lambda}_u)$ is the value of the biasing potential at the position in alchemical space $\vec{\lambda}_u$ corresponding to sequence or ligand u , u_s is the particular substituent present in a sequence or ligand at site s , $\lambda_{su_s}(t)$ is the trajectory that alchemical coordinate takes during a simulation, Θ is the heaviside function, and integration over time is approximated by a discrete sum of samples from the simulation.

As mentioned in the introduction, the sampling requirements of the histogram-based free energy estimator scale exponentially with the number of perturbation sites. In exchange for this high cost, the estimator can in principle be used to calculate all higher order couplings between sites. Exponential scaling occurs as the product of two factors. First the number of sequences or ligands scales exponentially with the number of sites and each state must be visited repeatedly to estimate its free energy. Second, the fraction of the time the system spends in proximity to any of the ligands is called the fraction physical ligand (FPL) and determines how much of the simulation is useful for estimating free energies (see Supporting Information for a deeper discussion of FPL). As the number of sites increases, the FPL decreases exponentially with the number of sites, because it is more likely that at least one site will be in an alchemical intermediate state. Raising the FPL has the effect of decreasing the base of the exponential scaling of the estimator and can substantially improve scaling.

Several approaches have previously been used to increase the FPL, including increasing the c constant in the implicit constraints,^{3,45} adding a small barrier between alchemical endpoints,³ or variable biasing replica exchange,⁷ but these techniques decrease numerical stability, slow convergence by lowering transition rates, or increase the computational cost, respectively. The implicit constraints affect FPL and numerical stability by controlling how sharply λ and the potential energy change as a function of θ . The default value for c is 5.5; a lower value of 2.5 has been used to prevent trapping before Equation 7,⁴⁷ but does not approach alchemical endpoints closely enough to give accurate results; higher values up to 15.5 were found to improve FPL, while further increases degraded results.³ Adding a small barrier between alchemical endpoints increases FPL, but lowers transition rates, and a previous study balanced these effects by increasing all intrasite ψ terms by 2 kcal/mol and all intrasite ω terms by 0.5 kcal/mol to give a 1 kcal/mol barrier.³ Alternatively, in biasing potential replica exchange, some replicas have higher barriers to increase FPL, while others have lower barriers to encourage transitions, but the computational cost is proportional to the number of replicas.⁷ In this work, some simulations use standard parameters ($c = 5.5$ and no barrier) to test the ability of the estimators to contend with low FPL, while others follow the endpoint focused parameter choices in Reference 3 ($c = 15.5$ and a 1 kcal/mol barrier) to maintain reasonable FPL.

Because the histogram-based estimator requires undesirable accommodations to maintain a high FPL, and does not scale well to large alchemical spaces even if these accommodations are made, an alternative estimator is desirable. The Potts model is borrowed from DCA as a simplifying approximation.^{41,42,51} The Potts model neglects third and higher order coupling between sites by assuming the free energy only depends on the substituent u_s at site s due to one-body terms called “fields” (h_s) and two-body terms capturing interactions between sites s and t called “couplings” (J_{st}):

$$U_{\text{Potts}}(u) = \sum_{s=1}^M h_s(u_s) + \frac{1}{2} \sum_s^M \sum_{t \neq s}^M J_{st}(u_s, u_t) \quad (9)$$

The probability of observing any sequence is then

$$P(u) = \frac{1}{Z} \exp(-\beta U_{\text{Potts}}(u)) \quad (10)$$

where $Z = \sum_u \exp(-\beta U_{\text{Potts}}(u))$, and $\beta = 1/k_B T$. In DCA, $P(u)$ is fit to the distribution of extant sequences in a multiple sequence alignment, and T corresponds to some selection temperature.^{52,53} In MS λ D, $P(u)$ is fit to the distribution of alchemical states observed during the simulation, and T is the temperature the simulation was run at. An additional state u_s is included for each site s for the alchemical intermediates, when no λ_{si} is greater than λ_c . For the alchemical spaces considered in this work, $U_{\text{Potts}}(u)$ is determined using log-likelihood maximization with Broyden, Fletcher, Goldfarb, and Shanno quasi-Newton optimization, where the log-likelihood is given by

$$L(h, J) = \log Z - \sum_s^M \sum_i^{N_s+1} f_{si} h_s(i) - \frac{1}{2} \sum_s^M \sum_t^M \sum_i^{N_s+1} \sum_j^{N_t+1} f_{si,tj} J_{st}(i, j) + \frac{k_h}{2} \sum_h h_s(i)^2 + \frac{k_J}{2} \sum_J J_{st}(i, j)^2 \quad (11)$$

$$f_{si} = \int dt \Theta(\lambda_{si}(t) - \lambda_c) \quad (12)$$

$$f_{si,tj} = \int dt \Theta(\lambda_{si}(t) - \lambda_c) \Theta(\lambda_{tj}(t) - \lambda_c) \quad (13)$$

where f_{si} and $f_{si,tj}$ are the first and second moments, or the probability of observing a particular perturbation or pair of perturbations with λ greater than λ_c during the simulation, the sum on i runs to $N_s + 1$ due to the inclusion of a state for the alchemical intermediates, and the final two terms are included for regularization with $k_h = k_J = 10^{-6} \beta$. For alchemical spaces exceeding roughly a billion ligands or sequences, calculation of Z is not practical, and pseudolikelihood maximization is required instead.^{42,54} Free energies can be calculated from $h_s(i)$ and $J_{st}(i, j)$ using Equation 9 and correcting for the biasing potentials, with uncertainties estimated by bootstrapping between independent simulations.

$$\Delta G_P(u) = -U_{\text{bias}}\left(\vec{\lambda}_u\right) + U_{\text{Potts}}(u) \quad (14)$$

Finally, a recent MS λ D study noted that treating each site as independent, evaluating the free energies for each site with the histogram-based estimator, and then summing the free energies at each site gave superior results to the full histogram-based estimator.²⁰

$$\Delta G_I(u) = -U_{\text{bias}}(\vec{\lambda}_u) - k_B T \ln \left(\prod_{s=1}^M \int dt \Theta(\lambda_{su_s}(t) - \lambda_c) \right) \quad (15)$$

This approach goes further than the Potts model estimator, and also sacrifices pairwise two-body couplings between sites, retaining only one-body terms. The success of the independent site estimator relative to the histogram-based estimator is likely due to decreased noise, and merits comparison with the Potts model estimator.

Systems

Three systems were used to test ALF with intersite biases and the various estimators. These systems were chosen because they were the largest alchemical spaces previously studied with MS λ D, which allowed the study to focus on difficulties sampling and estimating free energy in large alchemical spaces within familiar systems.

Large couplings were first observed in the previously studied multisite systems from T4 lysozyme (T4L) with the natural disulfide removed (C54T/C97A),⁷ so these systems were explored first. T4L calculations were run with 3 mutating sites with $2 \times 2 \times 2 = 8$ sequences, 4 mutating sites with $2 \times 2 \times 3 \times 2 = 24$ sequences, and 5 mutating sites with $3 \times 5 \times 4 \times 2 \times 2 = 240$ sequences (Supporting Information Figure S1 and Table S3). Simulations were run starting from the PDB 1L63 crystal structure,⁵⁵ and compared with experimental data compiled in Reference 56.

To assess the usefulness of these techniques for CADD, two ligand binding systems were also considered: a system of $8 \times 8 \times 8 = 512$ HIV reverse transcriptase (HIV-RT) indole based inhibitors that had been studied previously,³ and a new set of $5 \times 7 \times 7 = 245$ p38 inhibitors similar to a smaller set of inhibitors studied previously²⁰ (see Supporting Information). Initial coordinates for the HIV-RT simulations were obtained from the PDB 4MFB crystal structure⁵⁷ and the full system was truncated, as done previously, to focus sampling on the non-nucleoside binding site.^{3,58} Experimental data was taken from reference 57. The p38 protein system was built from the PDB 3FLY crystal structure,⁵⁹ mirroring what has been done previously in the field.¹ Experimental data was taken from reference 59. In both ligand binding systems, the ligands were built in Chimera,⁶⁰ parametrized with the MATCH atom-typing tool to obtain CGenFF small molecule parameters,⁶¹ and modeled as a multiple topology model for use with MS λ D.

Simulation Details

System setup and simulation details have been described previously.^{3,7} Preliminary simulations using variable biasing potential replica exchange⁷ were performed with both force switching (FSWITCH)⁶² and particle mesh Ewald (PME) electrostatics^{43,44} using the MS λ D PME formalism derived by Shen and coworkers,¹² which scales substituent charges by λ . These simulations clearly demonstrated that PME results improved with

longer sampling times while FSWITCH results degraded with longer sampling times, so FSWITCH was abandoned, and PME was used exclusively for the remaining simulations (see Supporting Information). Rather than use biasing potential replica exchange^{3,47} or variable biasing potential replica exchange⁷ to improve sampling, we instead chose to use an implicit constraint c parameter of 15.5 and apply a small barrier to the biasing potentials as described above. Simulations were run using the newly developed BLaDE MSAD engine⁴⁸ in CHARMM,^{63,64} because this is roughly 5-6 times faster than the DOMDEC module of CHARMM.⁶⁵ Flattening protocols are described in detail in the Supporting Information.

Results

Coupling in Adaptive Landscape Flattening

In order to demonstrate the benefits of including coupling terms in the adaptive landscape flattening algorithm, the folded side of the 5 site T4L system was flattened with various ALF methods. A limited amount of sampling, 5 independent trials of 20 ns each, was utilized so that not all sequences would be visited, and differences between the number of sequences visited would highlight differences in the quality of the biasing potentials. The thoroughness of sampling was quantified by the number of sequences sampled in any of the 5 independent trials, which is required for histogram-based free energy estimates (Table 1). The average number sampled per trial and the number sampled by all trials, which are less indicative of sampling quality, were also computed (Supporting Information Table S6). After initial flattening, production runs of 5×20 ns were performed 20 times with bias optimization after each production run, and statistics (mean and standard error of mean) were collected from the last 15 production runs to allow the biases some time to converge (Table 1). We note that the number of states visited reflects more traditional measures of bias potential quality such as evenness of sampling all sequences and the rate of transition into less favorable sequence states.

The results in Table 1 reveal that ALF methods that optimize coupling terms in the biases give significantly superior results to ALF methods that do not. For endpoint focused systems, subtracting the numbers in the Table 1 from 240 indicates methods without coupling failed to sample 14.7 ± 1.6 or 14.2 ± 1.6 sequences in any trial, while methods with coupling only failed to sample 6.0 ± 1.7 or 3.3 ± 0.6 sequences. For the two methods that do not include coupling terms, fully independent ALF, which analyzes each site independently, gave statistically indistinguishable sampling with coupling aware ALF, which optimizes all sites simultaneously without any intersite coupling. For the two methods that do include coupling terms, optimizing ψ , χ , and ω terms seemed to perform better with the standard parameters, while only optimizing ψ seemed to perform better with the endpoint focused parameters, but these differences may be due to statistical variation.

Two factors account for the differences between the two ALF methods including bias potential coupling terms and the two that lack it. First, the coupling biases enable more even sampling between all sequences, and second, because the sampling is more even, the biases for subsequent runs can be estimated better, which results in further improvements in sampling in subsequent production runs.

While the ALF methods which include coupling biasing terms give better sampling, in the subsequent sections evaluating free energy estimators, we utilize coupling aware ALF, which includes no coupling terms, in order to be able to compare the independent site estimator along with the Potts model and histogram-based estimators.

Free Energy Estimator Accuracy

To evaluate the three different free energy estimators, five independent 100 ns (5×100 ns) simulations of the T4L systems were compared against long 5×500 ns reference simulations, and 12×50 ns ligand binding simulations were compared against 12×500 ns reference simulations (Table 2). Each estimator was applied to the same simulation data, and both endpoint focused simulations with $c = 15.5$ and a 1 kcal/mol barrier and standard simulations with $c = 5.5$ and no barrier were examined as described above. These parameters were chosen to access performance of each estimator under both ideal conditions with high FPL and standard parameter conditions with low FPL. Accuracy was assessed with the centered root mean squared error (cRMSE) of $(\langle x_i^2 \rangle - \langle x_i \rangle^2)^{1/2}$, the mean unsigned error (MUE) of $|\langle x_i \rangle - 0|$, and the Pearson correlation (R), where x_i is the free energy of a particular sequence or ligand i in a simulation with a particular estimator minus the reference free energy of that sequence or ligand in either a longer reference simulation or in experiment. Reference simulations were run with endpoint focused conditions and analyzed with the histogram-based estimator to preserve any higher order coupling and are an ideal point of comparison because they give an approximation of the correct answer for the force field. It is worth noting that these exhaustively sampled reference calculations were primarily possible because of the development of BLaDE, a significantly faster MSAD engine.⁴⁸ Table 2 reveals that the Potts model estimator gives improved results over both the independent site estimator and the histogram-based estimator. The most accurate estimator is highlighted for each metric, and for the endpoint focused simulations, the Potts model estimator is best in 8 out of 15 cases, while for the standard simulations, the Potts model estimator is best in 13 out of 15 cases.

Agreement with the force field answer is the mark of a successful free energy estimator; however, because the reference simulations may not be fully converged to the force field answer, comparison with experiment is also a relevant test. Interestingly, when comparing to experimental data rather than a reference calculation, the Potts model estimator still maintains an advantage over the histogram-based estimator, but is comparable to the independent site estimator (Supporting Information Table S7). The Potts model estimator performs better on the protein mutations while the independent site estimator performs better on the ligand perturbations. This may be due to smaller and more noisy couplings for the ligand perturbations, especially in p38. Supporting Information Table S8 shows p38 ligand perturbations have smaller root mean square couplings in the reference simulations, and their larger root mean square coupling in the shorter production simulations may be a source of error. This source of error could be anticipated from the root mean square uncertainty of the coupling parameters obtained from bootstrapping: p38 clearly exhibits the largest absolute uncertainty in the couplings, and the largest relative uncertainty at roughly half the root mean square coupling strength during production (Supporting Information Table S9). Other possible explanations beyond noise in the couplings include that the

experimental data points for the ligand include few (HIV-RT) or no (p38) perturbations at multiple sites that could highlight deficiencies in the independent site estimator and that amino acid parameters are better calibrated than ligand parameters, so the force field correct answer may deviate further from experiment for the ligand systems.

Overall, the Potts model estimator clearly outperforms the histogram-based estimator. In the smallest alchemical space, the T4L 3 site system, the histogram-based estimator gives marginally better results because of the thorough sampling of the small higher order couplings, but the improvements are small. In larger systems, the Potts model estimator gives better results because of decreased noise, and can still give robust free energy estimates when the histogram-based estimator fails because of insufficient sampling of the alchemical endpoints. The Potts model estimator also produces results in closer agreement with our estimate of the force field answer than the independent site estimator, which is the true test of a free energy estimator; however, the closer agreement of the independent site estimator with experimental results in ligand binding systems deserves further attention, as it may highlight deficiencies in our estimate of the force field answer.

Scaling of Potts Model Estimator

The Potts model estimator clearly gives improved results over the original histogram-based estimator at a reduced computational cost, but confidently applying this estimator to larger alchemical spaces requires an understanding of its scaling behavior. Therefore, we developed a simple numerical test system to assess scaling of the Potts model estimator.

In the test system, each site has $N_s = 2$ substituents, each substituent end state is equally probable, sites are uncoupled, and the ratio between substituent end states and alchemical intermediates is chosen to match that of a flat landscape with $c = 5.5$ and $\lambda_c = 0.99$. This reflects a molecular system with negligible three-site couplings for which ALF has run long enough to obtain converged biasing potentials that effectively flatten the alchemical landscape, including pairwise couplings. Monte Carlo sampling is used to generate between 1000 and 64000 statistically independent samples for this ideal system; the amount of simulation time this corresponds to in a molecular system varies depending on the time scale of relaxation processes, but reveals the effects of increased sampling. Potts model estimates of the fields, couplings, and free energies are made with varying numbers of sites and varying numbers of samples, and the uncertainty of these values is determined from the standard deviation of fields, couplings, or free energies. By symmetry, all fields, couplings, or free energies for a particular number of sites and samples should only exhibit statistical variation around the same mean value. This mean value may differ from zero because alchemical intermediates have their own fields and couplings which are different from those of alchemical end states, but should be the same for all end states. Thus, the standard deviation of the free energy values for different states gives an estimate of the root mean square error or the uncertainty, and these terms will subsequently be used interchangeably.

Plotting the standard deviation as a function of sampling reveals that it decays like the square root of the amount of sampling as expected (Figure 3A). (By the central limit theorem, fluctuations in the average of uncorrelated measurements scale like $N^{-1/2}$ while fluctuations in the sum of uncorrelated measurements scale like $N^{+1/2}$, where N is

the number of measurements.) Strikingly, the standard deviations of both the fields and couplings are constant as the number of sites increases (Figure 3B). This indicates that regardless of the number of sites in a molecular system, as long as ALF is converged and the relaxation time scales are the same, pairwise couplings between sites can be estimated with the same precision. The standard deviation of the free energy grows linearly with the number of sites (Figure 3B) as a direct consequence of the constant standard deviation of the couplings. The number of couplings contributing to any free energy value is proportional to M^2 , so the uncertainty in their sum is proportional to M , the number of sites. Similar arguments using the central limit theorem can show that the amount of sampling required to determine a single model parameter, the free energy difference of a point mutant, or the free energy difference of two arbitrary alchemical endpoints to a desired level of uncertainty scales like 1 , M , or M^2 , respectively. This represents a substantial improvement over $\exp(M)$ scaling for the histogram-based estimator.

These scaling estimates are made for systems with $N_s = 2$ substituents. Sampling requirements are expected to scale like $N_s N_t / F_s F_t$ for couplings between sites s and t , where F_s is the single-site FPL in Supporting Information Table S1, and like $\sum_s^M \sum_{t>s}^M N_s N_t / F_s F_t$ for overall free energy estimates if all sites have the same relaxation time scale. This is because roughly the same number of samples of a pair of substituents is expected to be required for a desired accuracy in the couplings, but the samples will be split between more pairs of substituents and alchemical intermediates. This may further indicate why the Potts model performed better in the protein systems than in the ligand systems: the number of substituents in the protein systems ranged from 2 to 5, while the number of substituents in the ligand systems ranged from 5 to 8.

In systems where three body couplings are important, one could generalize the Potts model to include three body terms. The uncertainty in the three body terms would be expected to be independent of the number of sites like the one body fields and the two body couplings, albeit with a larger prefactor, and since the number of three body terms included in any free energy estimate is proportional to the number of sites cubed, the uncertainty in free energy would be proportional to $M^{3/2}$ and the sampling requirements would be proportional to M^3 .

Discussion

The objective of this study has been to enable studies of larger alchemical spaces for applications like CADD and computational protein design. Energetic coupling between sites has been at the core of this effort, both in adding coupling terms to the biasing potential to allow more even exploration of alchemical space, and in terms of a more scalable Potts model free energy estimator that includes significant pairwise couplings while ignoring typically negligible higher order couplings that bring higher levels of noise and computational expense to the histogram-based estimator. It is natural that couplings should be fundamentally connected to large scale studies with MS λ D because the computational advantage of MS λ D depends on the level of coupling between sites. If all sites are independent, MS λ D provides a modest computational advantage because they can all be explored simultaneously, but the same information could be obtained from FEP calculations treating each site independently. In contrast, if sites are coupled, then the computational

advantage of MS λ D is much greater, because obtaining the same information with FEP would require running a much larger set of simulations to observe all pairs of perturbations for pairwise couplings, and all combinations of perturbations for higher order couplings.

Pairwise couplings were observed for the T4L systems, which consisted of buried mutations in close contact, and in bound ligand simulations, where perturbations can shift a ligand within the binding pocket and create or release strain at other sites. Pairwise couplings were notably smaller in unbound ligand simulations (Table S8), suggesting coupling may play a lesser role in solvation free energies. Smaller pairwise couplings are also more likely for widely separated protein surface mutations. These considerations together with the success of the Potts estimator suggest that the pairwise couplings are necessary and sufficient for high accuracy in most cases.

Consequently, for systems with two or three sites where the user has reason to expect couplings, as well as in systems with more sites where even small couplings are likely to build up, it is advisable to use an ALF method with coupling terms in the bias potential, unless one intends to use the independent site estimator. In larger alchemical spaces, including only ψ coupling will provide a performance advantage over ψ , χ , and ω coupling.

The Potts model estimator provides several advantages over the histogram-based estimator for systems with three or more sites, (they are identical for two site systems), by improving the scaling of the sampling requirements as a function of the number of sites and relaxing the need to focus sampling on the alchemical endpoints. Previously, with the histogram-based estimator, attaining sufficient FPL for systems with several sites was a major concern, and while increasing FPL can still be mildly helpful with the Potts model estimator due to the weak predicted dependence of sampling requirements on FPL, a more relevant metric for the quality of the results is the uncertainty in the Potts model couplings. In rare cases where the amount of sampling is large and large three body couplings are expected, the histogram-based estimator may still be useful. The Potts model estimator also provides several advantages over the independent site estimator by including pairwise couplings and allowing estimation of the uncertainty of the couplings with bootstrapping. While the error introduced by neglecting these couplings is unknown with the independent site estimator, the noise introduced by including them can be estimated with Potts model estimator, and if the uncertainties in the couplings are too large, one can fall back on the independent site estimator. In this case it may also be possible to increase the regularization constant k_J of the couplings in the Potts model estimator to reduce the magnitude of the couplings and the noise they contribute to free energy estimates.

These developments enable new questions to be asked that require exploration of larger alchemical spaces. An ongoing study in our group of the combinatorial sequence space arising from 15 concurrent mutations in ribonuclease H has already been mentioned, and other studies of protein design and protein epistasis involving dozens of mutations become possible. In the context of CADD, experimental studies rarely go beyond three perturbation sites, but these methods can reveal synergies between sites that may be overlooked by optimizing one site at a time as is typically done.⁵⁹ Furthermore, studies of drug resistance

involving perturbations to both the binding pocket and the ligand become much more practical.

Conclusions

In this work we have provided an updated ALF framework and biasing potentials to overcome couplings between sites to enable more efficient sampling. We have also harnessed the couplings to develop a new Potts model free energy estimator that is more accurate than previous free energy estimators. The scaling requirements of the Potts model estimator are proportional to the square of the number of sites, which opens up much larger alchemical spaces useful for CADD and computational protein design.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgement

We gratefully acknowledge funding from the NIH (GM130587 and GM37554) and the NSF (CHE 1506273).

References

- (1). Wang L; Wu Y; Deng Y; Kim B; Pierce L; Krilov G; Lupyan D; Robinson S; Dahlgren MK; Greenwood J 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. *Journal of the American Chemical Society* 2015, 137, 2695–2703. [PubMed: 25625324]
- (2). 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. *Chemical Science* 2020, 11, 1140–1152.
- (3). Vilseck JZ; Armacost KA; Hayes RL; Goh GB; Brooks CL III Predicting Binding Free Energies in a Large Combinatorial Chemical Space Using Multisite λ Dynamics. *Journal of Physical Chemistry Letters* 2018, 9, 3328–3332. [PubMed: 29847134]
- (4). Vilseck JZ; Sohail N; Hayes RL; Brooks CL III Overcoming Challenging Substituent Perturbations with Multisite λ -Dynamics: A Case Study Targeting β -Secretase 1. *Journal of Physical Chemistry Letters* 2019, 10, 4875–4880. [PubMed: 31386370]
- (5). Seeliger D; de Groot BL Protein Thermostability Calculations Using Alchemical Free Energy Simulations. *Biophysical Journal* 2010, 98, 2309–2316. [PubMed: 20483340]
- (6). Steinbrecher T; Zhu C; Wang L; Abel R; Negron C; Pearlman D; Feyfant E; Duan J; Sherman W Predicting the Effect of Amino Acid Single-Point Mutations on Protein Stability: Large-Scale Validation of MD-Based Relative Free Energy Calculations. *Journal of Molecular Biology* 2017, 429, 948–963. [PubMed: 27964946]
- (7). Hayes RL; Vilseck JZ; Brooks CL III Approaching Protein Design with Multisite λ Dynamics: Accurate and Scalable Mutational Folding Free Energies in T4 Lysozyme. *Protein Science* 2018, 27, 1910–1922. [PubMed: 30175503]
- (8). Hayes RL; Brooks CL III, A Strategy for Proline and Glycine Mutations to Proteins with Alchemical Free Energy Calculations. *Journal of Computational Chemistry* 2021, 42, 1088–1094. [PubMed: 33844328]
- (9). Donnini S; Tegeler F; Groenhof G; Grubmüller H Constant pH Molecular Dynamics in Explicit Solvent with λ -Dynamics. *Journal of Chemical Theory and Computation* 2011, 7, 1962–1978. [PubMed: 21687785]

- (10). Wallace JA; Shen JK Charge-Leveling and Proper Treatment of Long-Range Electrostatics in All-Atom Molecular Dynamics at Constant pH. *Journal of Chemical Physics* 2012, 137, 184105. [PubMed: 23163362]
- (11). Goh GB; Hulbert BS; Zhou H; Brooks CL III, Constant pH Molecular Dynamics of Proteins in Explicit Solvent with Proton Tautomerism. *Proteins: Structure, Function, and Bioinformatics* 2014, 82, 1319–1331.
- (12). Huang Y; Chen W; Wallace JA; Shen J All-Atom Continuous Constant pH Molecular Dynamics with Particle Mesh Ewald and Titratable Water. *Journal of Chemical Theory and Computation* 2016, 12, 5411–5421. [PubMed: 27709966]
- (13). Ellis CR; Tsai C-C; Hou X; Shen J Constant pH Molecular Dynamics Reveals pH-Modulated Binding of Two Small-Molecule BACE1 Inhibitors. *Journal of Physical Chemistry Letters* 2016, 7, 944–949. [PubMed: 26905811]
- (14). Hu Y; Sherborne B; Lee T-S; Case DA; York DM; Guo Z The Importance of Protonation and Tautomerization in Relative Binding Affinity Prediction: A Comparison of AMBER TI and Schrödinger FEP. *Journal of Computer-Aided Molecular Design* 2016, 30, 533–539. [PubMed: 27480697]
- (15). Paul TJ; Vilseck JZ; Hayes RL; Brooks CL III, Exploring pH Dependent Host/Guest Binding Affinities. *Journal of Physical Chemistry B* 20 20, 124, 6520–6528.
- (16). Shi Y; Laury ML; Wang Z; Ponder JW AMOEBA Binding Free Energies for the SAMPL7 TrimerTrip Host-Guest Challenge. *Journal of Computer-Aided Molecular Design* 2021, 35, 79–93. [PubMed: 33140208]
- (17). Guthrie JP A Blind Challenge for Computational Solvation Free Energies: Introduction and Overview. *Journal of Physical Chemistry B* 2009, 113, 4501–4507. [PubMed: 19338360]
- (18). Mobley DL; Guthrie JP FreeSolv: A Database of Experimental and Calculated Hydration Free Energies, with Input Files. *Journal of Computer-Aided Molecular Design* 2014, 28, 711–720. [PubMed: 24928188]
- (19). Abel R; Wang L; Harder ED; Berne BJ; Friesner RA Advancing Drug Discovery through Enhanced Free Energy Calculations. *Accounts of Chemical Research* 2017, 50, 1625–1632. [PubMed: 28677954]
- (20). Raman EP; Paul TJ; Hayes RL; Brooks CL III Automated, Accurate, and Scalable Relative Protein-Ligand Binding Free Energy Calculations using Lambda Dynamics. *Journal of Chemical Theory and Computation* 2020, 16, 7895–7914. [PubMed: 33201701]
- (21). Parks CD; Gaieb Z; Chiu M; Yang H; Shao C; Walters WP; Jansen JM; McGaughey G; Lewis RA; Bembenek SD et al. D3R Grand Challenge 4: Blind Prediction of Protein-Ligand Poses, Affinity Rankings, and Relative Binding Free Energies. *Journal of Computer-Aided Molecular Design* 2020, 34, 99–119. [PubMed: 31974851]
- (22). 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. *Angewandte Chemie* 2016, 55, 7364–7368. [PubMed: 27122231]
- (23). Duan J; Lupyan D; Wang L Improving the Accuracy of Protein Thermostability Predictions for Single Point Mutations. *Biophysical Journal* 2020, 119, 115–127. [PubMed: 32533939]
- (24). Wang L; Chambers J; Abel R *Biomolecular Simulations; Methods in Molecular Biology*; 2019; Vol. 2022; pp 201–232.
- (25). Jespers W; Esguerra M; Åqvist J; de Terán HG QligFEP: An Automated Workflow for Small Molecule Free Energy Calculations in Q. *Journal of Cheminformatics* 2019, 11, 26. [PubMed: 30941533]
- (26). Abel R; Manas ES; Friesner RA; Farid RS; Wang L Modeling the Value of Predictive Affinity Scoring in Preclinical Drug Discovery. *Current Opinion in Structural Biology* 2018, 52, 103–110. [PubMed: 30321805]
- (27). Schindler CEM; Baumann H; Blum A; Bose D; Buchstaller H-P; Burgdorf L; Cappel D; Chekler E; Czodrowski P; Dorsch D et al. Large-Scale Assessment of Binding Free Energy Calculations in Active Drug Discovery Projects. *Journal of Chemical Information and Modeling* 2020, 60, 5457–5474. [PubMed: 32813975]

- (28). Bornscheuer UT; Huisman GW; Kazlauskas RJ; Lutz S; Moore JC; Robins K Engineering the third wave of biocatalysis. *Nature* 2012, 485, 185–194. [PubMed: 22575958]
- (29). Olson CA; Wu NC; Sun R A Comprehensive Biophysical Description of Pairwise Epistasis throughout an Entire Protein Domain. *Current Biology* 2014, 24, 2643–2651. [PubMed: 25455030]
- (30). Starr TN; Picton LK; Thornton JW Alternative Evolutionary Histories in the Sequence Space of an Ancient Protein. *Nature Year*, 549, 409–413.
- (31). Wu NC; Dai L; Olson CA; Lloyd-Smith JO; Sun R Adaptation in Protein Fitness Landscapes Is Facilitated by Indirect Paths. *eLife* 2016, 5, e16965. [PubMed: 27391790]
- (32). Zwanzig RW High-Temperature Equation of State by a Perturbation Method. I. Nonpolar Gases. *Journal of Chemical Physics* 1954, 22, 1420–1426.
- (33). Straatsma TP; Berendsen HJC Free Energy of Ionic Hydration: Analysis of a Thermodynamic Integration Technique to Evaluate Free Energy Differences by Molecular Dynamics Simulations. *Journal of Chemical Physics* 1988, 89, 5876–5886.
- (34). Crooks GE Path-Ensemble Averages in Systems Driven Far from Equilibrium. *Physical Review E* 2000, 61, 2361.
- (35). Shirts MR; Bair E; Hooker G; Pande VS Equilibrium Free Energies from Nonequilibrium Measurements Using Maximum-Likelihood Methods. *Physical Review Letters* 2003, 91, 140601. [PubMed: 14611511]
- (36). Goette M; Grubmüller H Accuracy and Convergence of Free Energy Differences Calculated from Nonequilibrium Switching Processes. *Journal of Computational Chemistry* 2009, 30, 447–456. [PubMed: 18677708]
- (37). Kong X; Brooks CL III λ -Dynamics: A New Approach to Free Energy Calculations. *Journal of Chemical Physics* 1996, 105, 2414–2423.
- (38). Knight JL; Brooks CL III Multisite λ Dynamics for Simulated Structure-Activity Relationship Studies. *Journal of Chemical Theory and Computation* 2011, 7, 2728–2739. [PubMed: 22125476]
- (39). Hayes RL; Armacost KA; Vilseck JZ; Brooks CL III, Adaptive Landscape Flattening Accelerates Sampling of Alchemical Space in Multisite λ Dynamics. *Journal of Physical Chemistry B* 2017, 121, 3626–3635. [PubMed: 28112940]
- (40). Schug A; Weigt M; Hoch JA; Onuchic JN; Hwa T; Szurmant H Chapter 3 - Computational Modeling of Phosphotransfer Complexes in Two-Component Signaling. *Methods in Enzymology* 2010, 471, 43–58. [PubMed: 20946841]
- (41). Morcos F; Pagnani A; Lunt B; Bertolino A; Marks DS; Sander C; Zecchina R; Onuchic JN; Hwa T; Weigt M Direct-coupling analysis of residue coevolution captures native contacts across many protein families. *Proceedings of the National Academy of Sciences of the United States of America* 2011, 108, E1293–E1301. [PubMed: 22106262]
- (42). Ekeberg M; Lovkvist C; Lan Y; Weigt M; Aurell E Improved contact prediction in proteins: Using pseudolikelihoods to infer Potts models. *Physical Review E* 2013, 87, 012707.
- (43). Darden T; York D; Pedersen L Particle Mesh Ewald: An $N \cdot \log(N)$ Method for Ewald Sums in Large Systems. *Journal of Chemical Physics* 1993, 98, 10089–10092.
- (44). Essmann U; Perera L; Berkowitz ML; Darden T; Lee H; Pedersen LG A Smooth Particle Mesh Ewald Method. *Journal of Chemical Physics* 1995, 103, 8577–8593.
- (45). Knight JL; Brooks CL III Applying Efficient Implicit Nongeometric Constraints in Alchemical Free Energy Simulations. *Journal of Computational Chemistry* 2011, 32, 3423–3432. [PubMed: 21919014]
- (46). Guo Z; Brooks CL III; Kong X Efficient and Flexible Algorithm for Free Energy Calculations Using the λ -Dynamics Approach. *Journal of Physical Chemistry B* 1998, 102, 2032–2036.
- (47). Armacost KA; Goh GB; Brooks CL III Biasing Potential Replica Exchange Multisite λ -Dynamics for Efficient Free Energy Calculations. *Journal of Chemical Theory and Computation* 2015, 11, 1267–1277. [PubMed: 26579773]
- (48). Hayes RL; Buckner J; Brooks CL III BLADE: A Basic Lambda Dynamics Engine for GPU Accelerated Molecular Dynamics Free Energy Calculations. Submitted to *Journal of Chemical Theory and Computation* 2021, -, -.

- (49). Ding X; Vilseck JZ; Hayes RL; Brooks CL III Gibbs Sampler-Based λ -Dynamics and Rao-Blackwell Estimator for Alchemical Free Energy Calculation. *Journal of Chemical Theory and Computation* 2017, 13, 2501–2510. [PubMed: 28510433]
- (50). Vilseck JZ; Ding X; Hayes RL; Brooks CL III, Generalizing the Discrete Gibbs Sampler-based λ -Dynamics Approach for Multisite Sampling of Many Ligands. *Journal of Chemical Theory and Computation* 2021, 17, 3895–3907. [PubMed: 34101448]
- (51). Schug A; Weigt M; Onuchic JN; Hwa T; Szurmant H High-Resolution Protein Complexes from Integrating Genomic Information with Molecular Simulation. *Proceedings of the National Academy of Sciences of the United States of America* 2009, 106, 22124–22129. [PubMed: 20018738]
- (52). Pande V; Grosberg A; Tanaka T Statistical Mechanics of Simple Models of Protein Folding and Design. *Biophysical Journal* 1997, 73, 3192–3210. [PubMed: 9414231]
- (53). Morcos F; Schafer NP; Cheng RR; Onuchic JN; Wolynes PG Coevolutionary Information, Protein Folding Landscapes, and the Thermodynamics of Natural Selection. *Proceedings of the National Academy of Sciences of the United States of America* 2014, 111, 12408–12413. [PubMed: 25114242]
- (54). Ekeberg M; Hartonen T; Aurell E Fast pseudolikelihood maximization for direct-coupling analysis of protein structure from many homologous amino-acid sequences. *Journal of Computational Physics* 2014, 276, 341–356.
- (55). Nicholson H; Anderson DE; Pin SD; Matthews BW Analysis of the Interaction Between Charged Side Chains and the α -Helix Dipole Using Designed Thermostable Mutants of Phage T4 Lysozyme. *Biochemistry* 1991, 30, 9816–9828. [PubMed: 1911773]
- (56). Baase WA; Liu L; Tronrud DE; Matthews BW Lessons from the Lysozyme of Phage T4. *Protein Science* 2010, 19, 631–641. [PubMed: 20095051]
- (57). Lee W-G; Gallardo-Macias R; Frey KM; Spasov KA; Bollini M; Anderson KS; Jorgensen WL Picomolar Inhibitors of HIV Reverse Transcriptase Featuring Bicyclic Replacement of a Cyanovinylphenyl Group. *Journal of the American Chemical Society* 2013, 135, 16705–16713. [PubMed: 24151856]
- (58). Brooks CL III; Brünger A; Karplus M Active Site Dynamics in Protein Molecules: A Stochastic Boundary Molecular-Dynamics Approach. *Biopolymers* 1985, 24, 843–865. [PubMed: 2410050]
- (59). Goldstein DM; Soth M; Gabriel T; Dewdney N; Kuglstatter A; Arzeno H; Chen J; Bingenheimer W; Dalrymple SA; Dunn J et al. Discovery of 6-(2,4-Difluorophenoxy)-2-[3-hydroxy-1-(2-hydroxyethyl)propylamino]-8-methyl-8H-pyrido[2,3-d]pyrimidin-7-one (Pamapimod) and 6-(2,4-Difluorophenoxy)-8-methyl-2-(tetrahydro-2H-pyran-4-ylamino)pyrido[2,3-d]pyrimidin-7(8H)-one (R1487) as Orally Bioavailable and Highly Selective Inhibitors of p38 α Mitogen-Activated Protein Kinase. *Journal of Medicinal Chemistry* 2011, 54, 2255–2265. [PubMed: 21375264]
- (60). Pettersen EF; Goddard TD; Huang CC; Couch GS; Greenblatt DM; Meng EC; Ferrin TE UCSF Chimera-A Visualization System for Exploratory Research and Analysis. *Journal of Computational Chemistry* 2004, 25, 1605–1612. [PubMed: 15264254]
- (61). Yesselman JD; Price DJ; Knight JL; Brooks CL III MATCH: An Atom-Typing Toolset for Molecular Mechanics Force Fields. *Journal of Computational Chemistry* 2011, 33, P189–202.
- (62). Steinbach PJ; Brooks BR New Spherical-Cutoff Methods for Long-Range Forces in Macromolecular Simulation. *Journal of Computational Chemistry* 1994, 15, 667–683.
- (63). Brooks BR; Bruccoleri RE; Olafson BD; States DJ; Swaminathan S; Karplus M CHARMM: A Program for Macromolecular Energy, Minimization, and Dynamics Calculations. *Journal of Computational Chemistry* 1983, 4, 187–217.
- (64). Brooks BR; Brooks CL III; Mackerell AD Jr.; Nilsson L; Petrella RJ; Roux B; Won Y; Archontis G; Bartels C; Boresch S et al. CHARMM: The Biomolecular Simulation Program. *Journal of Computational Chemistry* 2009, 30, 1545–1614. [PubMed: 19444816]
- (65). Hynninen A-P; Crowley MF New Faster CHARMM Molecular Dynamics Engine. *Journal of Computational Chemistry* 2014, 35, 406–413. [PubMed: 24302199]

- (66). Kumar S; Rosenberg JM; Bouzida D; Swendsen RH; Kollman PA The Weighted Histogram Analysis Method for Free-Energy Calculations on Biomolecules. I. The Method. *Journal of Computational Chemistry* 1992, 13, 1011–1021.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

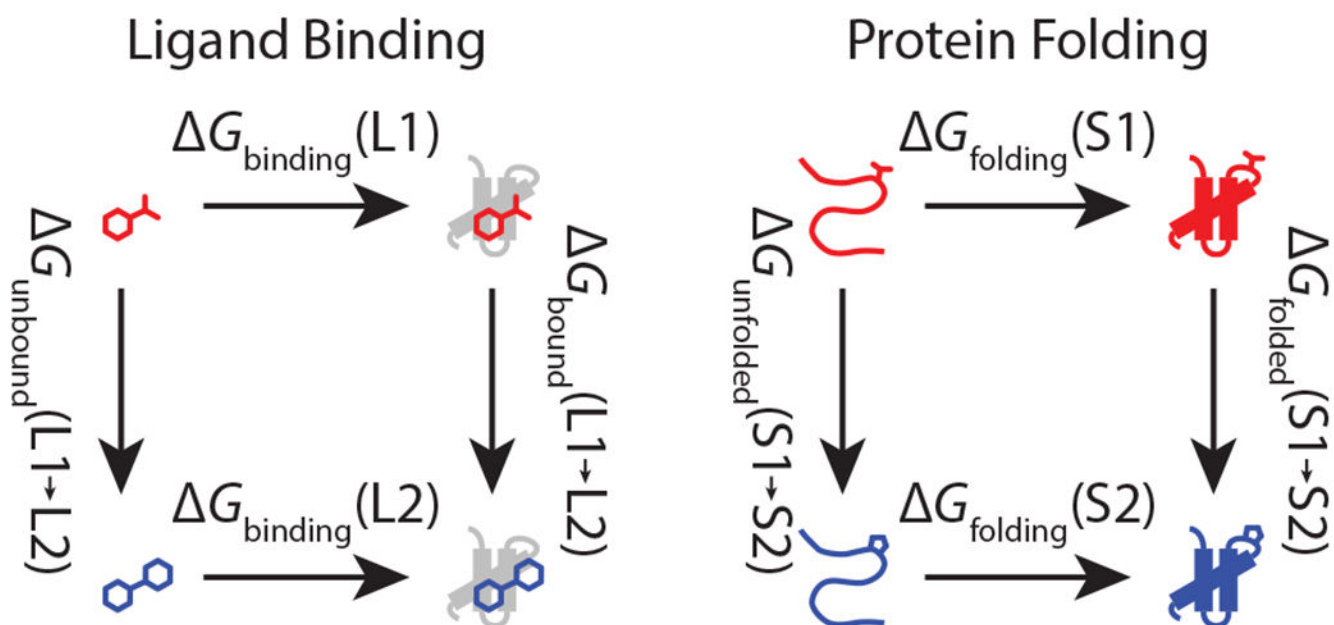


Figure 1: Alchemical free energy simulations for ligand binding and protein folding use similar free energy cycles. Because the horizontal physical processes (G_{binding} and G_{folding}) converge slowly, alchemical free energy methods take the difference of the two vertical alchemical processes, perturbing from L1 to L2 or S1 to S2 in each physical ensemble.

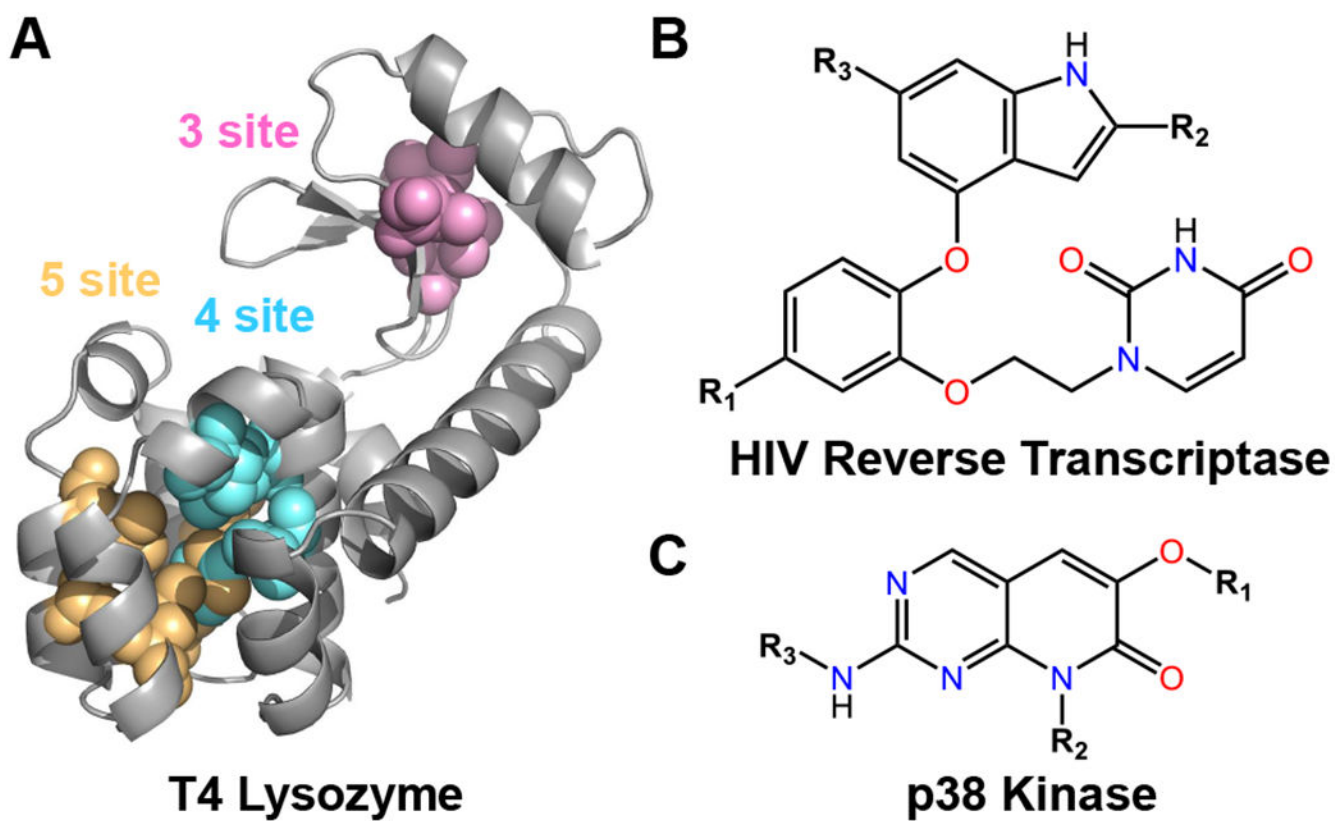
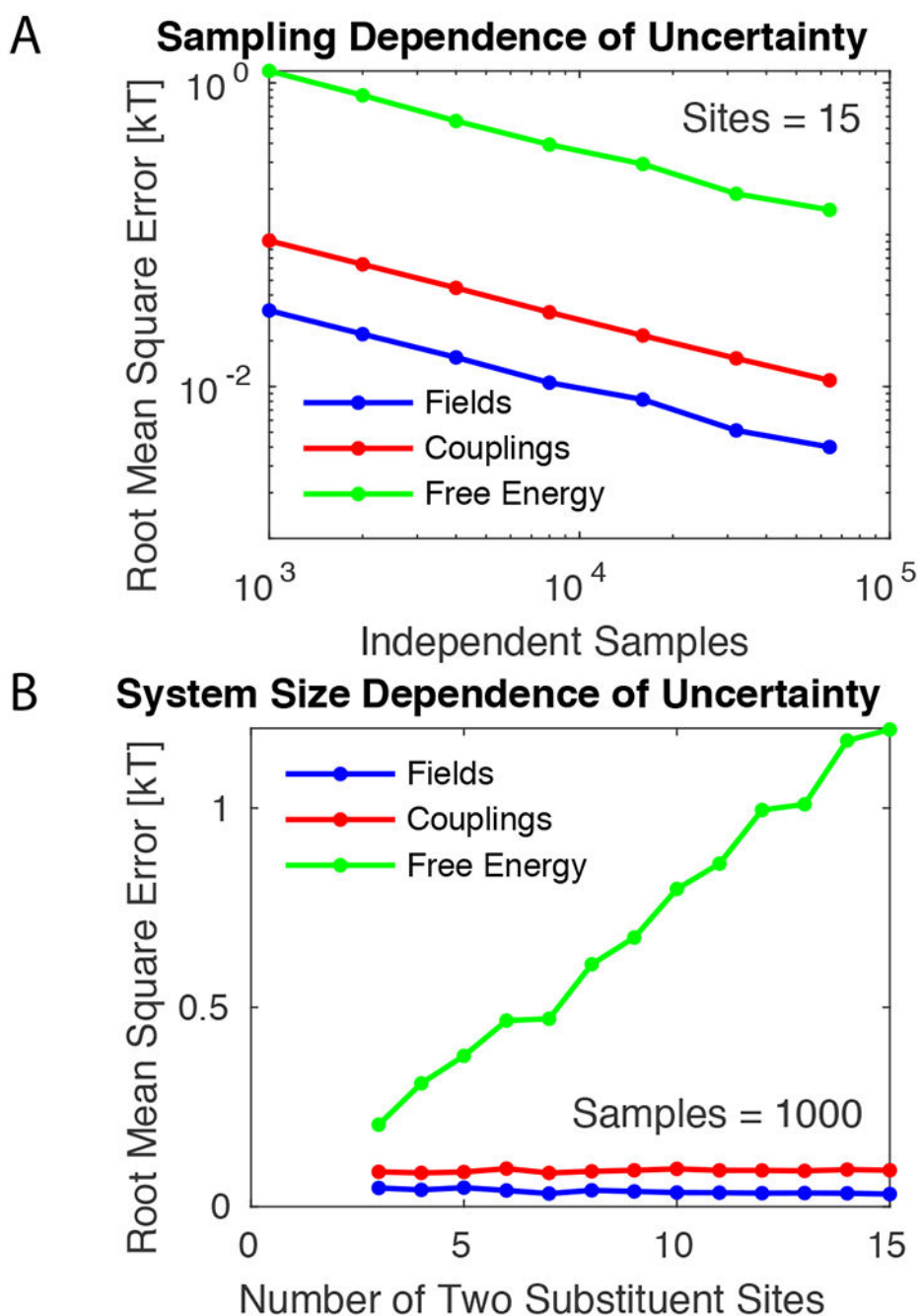


Figure 2:
The three systems used in the present study. (A) T4 lysozyme protein perturbations,⁷ (B) HIV reverse transcriptase indole based inhibitors,³ and (C) p38 ligands.²⁰

**Figure 3:**

(A) Root mean square error as a function of the number of independent samples, a proxy for the total amount of sampling time. The straight line and slope on a log-log plot indicate the error decays as the square root of the number of samples. (B) Root mean square error as a function of the number of sites. The error for the fields and couplings are both independent of the number of sites, while the error for free energy estimates grows linearly with the number of sites.

Table 1:

Number of Sequences Sampled in Any of 5 Trials of 20 ns out of 240 Sequences in T4L 5 Site System

	Fully Independent	Coupling Aware	ψ, χ, ω Coupling	Only ψ Coupling
Endpoint Focused ^a	225.3 ± 1.6	225.8 ± 1.6	234.0 ± 1.7	236.7 ± 0.6
Standard Parameters ^b	151.0 ± 2.6	154.9 ± 3.2	199.6 ± 4.5	182.2 ± 4.0

^a
c = 15.5 and 1 kcal/mol barrier^b
c = 5.5 and 0 kcal/mol barrier

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 2:

Free Energy Estimator Accuracy Relative to Reference Calculation

Estimator ^a	<i>c</i> = 15.5			<i>c</i> = 5.5		
	1 kcal/mol barrier			No barrier		
	cRMSE	MUE	R	cRMSE	MUE	R
T4L 3 Site System						
Independent site	0.41	0.73	0.972	0.43	0.94	0.971
Potts model	0.22	0.38	0.994	0.07	0.07	0.999
Histogram-based	0.23	0.36	0.994	0.07	0.08	0.999
T4L 4 Site System						
Independent site	0.53	0.59	0.923	0.70	0.64	0.896
Potts model	0.48	0.41	0.922	0.44	0.43	0.948
Histogram-based	0.51	0.41	0.922	0.46	0.51	0.940
T4L 5 Site System						
Independent site	0.68	0.88	0.892	0.92	1.44	0.789
Potts model	0.59	0.46	0.924	0.55	0.57	0.933
Histogram-based	-	-	-	-	-	-
HIV-RT Inhibitors						
Independent site	0.48	0.97	0.974	0.50	1.02	0.973
Potts model	0.30	0.26	0.990	0.37	0.29	0.985
Histogram-based	0.31	0.25	0.989	-	-	-
p38 Inhibitors						
Independent site	0.36	0.87	0.958	0.45	0.90	0.945
Potts model	0.42	0.59	0.952	0.52	0.61	0.935
Histogram-based	0.46	0.54	0.942	-	-	-

^aThe best estimator for each system and metric is highlighted in bold. If insufficient sampling is available to estimate free energies for all alchemical states with a particular estimator, a dash is displayed.