Achieving Rapid Free Energy Estimates from λ-Dynamics with Bias Updated Gibbs Sampling

Michael Robo
Indiana University School of Medicine

Ryan Hayes
Univeristy of Michigan

Xinqiang Ding
Massachesetts Institute of Technology   https://orcid.org/0000-0002-4598-8732

Brian Pulawski
Indiana University School of Medicine

Jonah Vilseck (✉ jvilseck@iu.edu)
Indiana University School of Medicine

Article

Keywords:

Additional Declarations: There is **NO** Competing Interest.

Abstract

Relative binding free energy (RBFE) calculations have become an integral computational tool for lead optimization in drug design. However, commonly used methods, such as free energy perturbation (FEP), thermodynamic integration (TI), or multistate Bennett acceptance ratio (MBAR), have an exceedingly high computational cost associated with the need to explore many pairwise calculations. To reduce this cost and accelerate molecular design workflows, we present a new Gibbs Sampler λ-Dynamics method that uses dynamic biases to continuously drive the sampling of multiple discrete λ states, and therefore multiple ligand analogues, collectively within a single simulation. As a result, many relative free energy differences can be rapidly determined without compromising accuracy in the computed results. We refer to this method as λ-Dynamics with Bias Updated Gibbs Sampling (LaDyBUGS). For three benchmark systems, errors in computed LaDyBUGS free energy differences relative to experiment are near 1.0 kcal/mol or less, and agreement with TI/MBAR is within statistical noise (approximately 0.4 kcal/mol or less). Large efficiency gains of 18- to 66-fold improvements are observed with LaDyBUGS compared to TI/MBAR for modeling just 5–10 perturbations at a single site of substitution off a central ligand core. We expect that the rapid and efficient determination of RBFEs with LaDyBUGS will enable larger chemical spaces to be more readily explored at reduced costs and structure-based drug design to be accelerated.

full text

Relative binding free energy (RBFE) calculations have emerged as a promising tool for the lead optimization of small molecule pharmaceuticals.[1−3] In an RBFE calculation, a small molecule bound to a protein target is alchemically transformed into a different small molecule, such as an analog formed by modifying one or more functional groups of the lead compound. The relative difference in free energies of binding ($\Delta\Delta G_{bind}$) between the two molecules can then be calculated using a thermodynamic cycle (Figure 1).[4] Compared to methods such as molecular docking, RBFE calculations have shown significantly improved correlation between computed and experimental binding affinities, with errors of roughly 1 kcal/mol or less for the state-of-the-art.[5−8] Although too high to eliminate the need for experiment entirely, this degree of accuracy is low enough to separate compounds with stronger versus weaker binding affinities and efficiently prioritize molecules for experimental investigation.[9,10] Using stochastic simulations, Mobley and Klimovich quantified the effect that this computational prioritization can have on a drug discovery project; they estimate that RBFE calculations with an average of 1.0 kcal/mol of error to experiment can improve the odds of identifying a 10-fold potency boost by a factor of 5.[11] When optimizing lead compounds for other drug-like properties, RBFE calculations can also be used to filter out compound modifications that might negatively affect potency.[7]

While many methods of RBFE calculations exist, the most commonly used methods are free energy perturbation (FEP), thermodynamic integration (TI), and multistate Bennett acceptance ratio (MBAR).[12−15] With these methods, an alchemical coupling parameter, called λ, is used to alchemically transform one molecule into another. To ensure sufficient phase space overlap between λ states and achieve

convergence in computed free energy differences, many intermediate discrete λ states are also defined (typically 10-20) that span a range of λ values between 0 and 1, the two molecule end states of interest.[4,7,9] Molecular dynamics (MD) simulations are performed at each of these discrete λ states, with λ values held constant for the duration of the simulation, and the MD trajectories are then postprocessed to calculate the final free energy difference. Though effective, FEP and TI calculations require considerable computational resources, are exclusively pairwise, and inherently unable to evaluate more than one RBFE at a time. For example, in a typical FEP/TI experiment,[5,7,8,16] 11 discrete λ states may be used to model a single perturbation, requiring 11 MD trajectories of 5-20 ns per λ window to be run, which amounts to a total of 55-220 ns simulation time for a single RBFE result.[6,16] Longer simulations or additional windows may be added for more challenging perturbations, such as ring additions or polar-to-non-polar transformations.[17] Further compounding computational costs, recommended best practices for investigating sets of multiple ligands with FEP/TI necessitates the use of redundant calculations to provide improved accuracy around closed perturbation cycles.[18-20] Although recent adoption of running MD simulations on graphical processing units (GPUs) has accelerated computational throughput and facilitated routine employment of RBFE calculations on large high-performance computing (HPC) resources for drug discovery,[21-23] costs per RBFE calculation remains considerably high.

Driven by the high cost of pairwise RBFE calculations, many groups have investigated alternative methods to perform RBFE calculations with an aim of achieving comparable accuracy with lower computational costs per computed RBFE. Non-equilibrium switching free energy calculations have seen renewed interest of late.[24–27] Mostly run in a pairwise manner, these calculations require only ca. 20-40 ns per transformation and are highly parallelizable,[24,27] making them good candidates for HPC or cloud computing. A variety of expanded ensemble methods have also grown in popularity.[28-32] λ-dynamics (λD),[33,34] enveloping distribution sampling,[35–37] and λ-local elevation umbrella sampling (λ-LEUS) methods,[38–40] to name a few, have all sought to calculate free energy differences between multiple thermodynamic states within a single calculation to increase efficiency through improved scalability.

In a conventional λD simulation, λ is treated as a continuous parameter and its value can change dynamically in conjunction with the coordinates of an MD simulation, using extended Lagrangian methods.[34,41] Sampling of multiple ligand end states or multisite sampling of many substituents at 2 or more sites of substitution are both feasible with multisite λ-dynamics (MSλD) via holonomic constraints.[34,42] Hence, multiple RBFEs can be computed from a single λD simulation, lending large efficiency gains over conventional approaches. Recent benchmarks have shown that single-site perturbations can be performed with cost savings in the range of 3-5.4 times better than TI/MBAR.[6] Advantageously, sampling with λD also produces a strongly connected graph between all ligand end states without the need for redundant calculations or cycle closure connections (Figure 2).[18,20] In recent years a variety of new developments have been introduced to expand the utility of λD for drug discovery, including an Adaptive Landscape Flattening (ALF) algorithm for automated bias determination,[41,43] a Potts model-based estimator for computing free energy differences and intersite couplings,[44] an

accelerated GPU engine,[45] and an alternative λ sampling strategy using Gibbs sampling, a Markov chain Monte Carlo algorithm.[46,47] This work builds upon this latter development of discrete Gibbs sampler λ-dynamics (d-GSλD).[47]

To sample multiple ligands collectively within a single λD simulation, free energy barriers between ligand end states in λ-space must first be flattened. This can be accomplished by identifying and incorporating a variety of biasing potentials into a λD simulation.[41] These biases flatten intermediate free energy barriers and ensure ligand end states have equivalent free energies to facilitate rapid transitioning between end states. Though effective, a non-negligible amount of simulation time must be devoted to determining these biases prior to production sampling, e.g., recent multisite λD simulations have taken 20-40 ns for bias determination.[6,44,48,49] With the advent of d-GSλD, discrete λ states can also be used to propagate alchemical transformations while maintaining the sampling of all λ states within a single simulation. The use of discrete λ states in d-GSλD is advantageous because conventional λD biasing potentials can be simplified from a functional form into a single scalar value per discrete λ state. This reduces the average amount of time needed to identify biases to 5-10 ns.[47] Furthermore, highly accurate free energy estimates can be obtained from d-GSλD with the MBAR free energy estimator.[15,47,50] Nonetheless, the computational cost of identifying biases for GSλD, or λD in general, reduces the efficiency and cost advantages of these methods. Thus, this work was motivated by an effort to eliminate these costs and accelerate RBFE calculations by removing the need to identify biases prior to production sampling. If such "biasing runs" could be avoided, we estimate that λD-based methods could screen hundreds of compound analogs at a fraction of the cost of FEP/TI methods for drug discovery.

In this report, we describe the use of continuous bias updates in conjunction with discrete Gibbs sampler λ-dynamics to achieve rapid and accurate RBFE estimates. We refer to this new method as λ-Dynamics with Bias Updated Gibbs Sampling (LaDyBUGS). In contrast to the static biases used with d-GSλD, and which were determined with a Wang Landau-like algorithm,[47] LaDyBUGS uses an aggressive dynamic bias that changes and continuously drives the system to sample different λ states. This avoids the need to run separate simulations for bias determination prior to production sampling and continually focuses sampling towards the least visited λ states to provide exceptionally smooth sampling of all λ states. FastMBAR, a GPU implementation of MBAR, is used for rapid free energy determination and on-the-fly bias refinement.[50] For three protein-ligand benchmark systems we observe large efficiency gains of 18–66-fold improvements with LaDyBUGS compared to TI/MBAR without compromising accuracy in the predicted $\Delta\Delta G_{bind}$ results. In the following sections, we describe the workflow of LaDyBUGS, and we evaluate it in terms of accuracy compared to experiment and efficiency compared to TI/MBAR, as implemented in OpenMM.[51]

Methods

λ-Dynamics with Bias Updated Gibbs Sampling is an advancement of the previously described discrete Gibbs sampler λ-dynamics approach.[47] Because much of the framework of LaDyBUGS builds on top of

d-GSλD, we quickly review d-GSλD before describing the new workflow for LaDyBUGS.

Discrete-Gibbs Sampler λ-Dynamics. To investigate alchemical transformations of a chemical system, d-GSλD samples the joint distribution of atomic coordinates, X, and alchemical states, λ, (P(X, λ)).[47] With Gibbs sampling this is accomplished via indirect sampling of two related conditional distributions, P(X|λ) and P(λ|X), which are formed by freezing a subset of variables, λ and X respectively. A single Gibbs sampler (GS) step consists of sequential sampling of P(X|λ) and P(λ|X) to yield X and λ at time t (X_t, $λ_t$).[32,52,53] Molecular dynamics can be used to sample P(X|λ), the coordinate space of the chemical system, while $λ_{t+1}$ can be chosen using a pseudorandom number generator to sample P(λ|X) (described in more detail below). In contrast to the use of continuous λ variables used in most λD-based techniques, the use of discrete λ states in d-GSλD was advantageous for several reasons. Notably it allowed for soft-core potentials to be used in conjunction with the alchemical perturbations and it facilitated the exploration of perturbations at multiple sites around a central ligand core, tasks not feasible using a continuous λ variable with GSλD.[46] Though no unique solution exists for defining discrete λ states between multiple ligand end states, a representation of λ states along connective edges between ligands provides a strongly connected map for sampling multiple ligands simultaneously and has yielded good free energy results in prior benchmarks (Fig. 2).[47] We note that to sample multiple ligands simultaneously, a single λ state *i* ($λ^i$) consists of a vector of substituent-specific λ variables that scale the interactions of each alchemical functional group individually.[47] Like most λD-based methods, all λ values within a single $λ^i$ state must sum to 1.0 to prevent more than 1 ligand from interacting with the rest of the chemical system at one time.[34,41,42] Furthermore, as mentioned earlier, biases are necessary to reduce free energy barriers in λ space and facilitate transitions between λ states at equilibrium. Discrete λ states simplified these biases into a single scalar quantity per state.[47] Prior to production sampling, static biases for each λ state were identified with a Wang-Landau-like algorithm with ~ 5–10 ns of sampling.[54,55] Production sampling for a preset amount of GS steps then ensued, followed by a FastMBAR[50] calculation to compute all final relative free energy differences.

λ-Dynamics with Bias Updated Gibbs Sampling. In this newly reported method, LaDyBUGS uses Gibbs sampling with discrete λ states to sample multiple ligand end states collectively within a single simulation. However, in an endeavor to accelerate d-GSλD and achieve rapid free energy results, Gibbs sampling is performed with dynamic biases, rather than static biases, to drive the exploration of many λ states without prior bias determination. Figure 3 describes the workflow of LaDyBUGS. Following initialization and minimization of a chemical system, the atomic coordinates and the alchemical states of the system are alternatively sampled.

While P(X|λ) continues to be sampled with MD, P(λ|X) is more involved. With LaDyBUGS, the conditional distribution P(λ|X) can be described as a multinomial distribution (Eq. 1):

$$P\left(\lambda^i|X\right) = \frac{\exp\left(-\beta\left[V_{SS}\left(X,\lambda^i\right) + V_{MS}\left(X,\lambda^i\right) + E^i\right]\right)}{\sum_{k=1}^{K}\exp\left(-\beta\left[V_{SS}\left(X,\lambda^k\right) + V_{MS}\left(X,\lambda^k\right) + E^k\right]\right)}$$

1

k represents the total number of λ states and E^i is a scalar bias added to each λ^i state. The single-site V_{SS} and multi-site V_{MS} potentials, necessary for investigating multisite perturbations of substituents i and j at sites s and t, are defined by equations 2 and 3 respectively:

$$V_{SS}\left(X = (x_0, \{x\}), \lambda\right) = \sum_{s=1}^{M}\sum_{i=1}^{N_s}\lambda_{s,i}\left(V\left(x_0, x_{s,i}\right) + V\left(x_{s,i}\right)\right)$$

2

$$V_{MS}\left(X = (x_0, \{x\}), \lambda\right) = \sum_{s=1}^{M}\sum_{i=1}^{N_s}\sum_{t=s+1}^{M}\sum_{j=1}^{N_t}\lambda_{s,i}\lambda_{t,j}V\left(x_{s,i}, x_{t,j}\right)$$

3

where X comprises atomic coordinates for both environment (x_0) and alchemical components ($x_{s,i}$), {x} represents the set of all $x_{s,i}$, and $\lambda_{s,i}$ are the substituent-specific λ variables. However, if systems with only single site modifications are studied, V_{MS} equals zero and can be ignored. The conditional distribution P(λ|X) at time t is thus formed by first calculating the potential energy of the system at coordinates X_t for each alchemical state λ^i and normalizing to form a Boltzmann distribution. A new λ_t^i state can then be selected by randomly choosing a new state proportional to its probability with a pseudorandom number generator. Finally, prior to the end of each GS step, the biases for all λ states are updated (as described below). As shown in Fig. 3, Gibbs sampling is an iterative process that is performed repeatedly for a preset amount time, usually quantified as an amount of cumulative MD sampling. At designated break points, Gibbs sampling is halted and FastMBAR is called to compute relative free energy differences (ΔG^i) for each λ^i state compared to the reference state, λ^1. Advantageously, the MBAR input, i.e., equilibrium energies of all λ states at every sampled configuration of the system (X), are calculated and saved on-the-fly when P(λ|X) is sampled, thus no trajectory postprocessing is necessary with LaDyBUGS. In addition, though an MBAR calculation is traditionally performed only at the conclusion of an alchemical free energy calculation, LaDyBUGS also uses MBAR free energy results at intermediate stages of an on-going simulation to update and refine the E^i biases for the next set of Gibbs sampler steps. As shown below, this helps provide smooth and even sampling of all λ states in a LaDyBUGS simulation.

Choice of Bias Function. In LaDyBUGS, the E^i biases are changed at the end of each Gibbs sampler step and intermediate FastMBAR calculations are performed regularly throughout a simulation to provide

additional bias refinement. When a LaDyBUGS simulation is initiated, a relatively aggressive biasing scheme is used to ensure every λ state is sampled prior to running FastMBAR for the first time. For example, in this work we used a flat external bias of 100 kcal/mol, which is added to each λ^i state every time that state is sampled. While any flat bias value would work, in principle, a large bias ($\geq$ 10 kcal/mol) ensures rapid sampling of all λ states within a short amount of GS sampling at the onset of a LaDyBUGS simulation. Prior to the first iteration of running FastMBAR, the total bias on λ^i is $E^i = 100 * N_i$, where N_i is the number of times λ^i was sampled. At time $t = u$ updates, Gibbs sampling is stopped and a FastMBAR calculation is performed to estimate the free energy differences of each λ^i state up to that point in time ($\Delta G_{t=u}^i$). At this stage of a LaDyBUGS simulation, the E^i biases are replaced with the negative value of the MBAR results ($-\Delta G_{t=u}^i$) and an additional exponential bias[39] is used to penalize each λ^i state based on the number of times λ^i is sampled compared to the least-sampled state ($min\left[N\left(\lambda\right)\right]$) (Eq. 4). After each GS step, the biases are updated with Eq. 4 to reflect the new number of counts per λ^i state, but the $-\Delta G_{t=u}^i$ component remains unchanged until the next FastMBAR calculation. Through this continuous changing of the biases, complete and smooth sampling of all λ states can be achieved (see Results/Discussion sections). In the supplementary information we present a mathematical proof that, assuming the MD simulation used for sampling from P(X|λ) in each GS step reaches equilibrium, the value of the scalar bias used during Gibbs sampling has no effect on the FastMBAR calculation, facilitating the use of unbiased equilibrium energies in FastMBAR for free energy estimation.

$$E^i = -\Delta G_{t=u}^i + 2^{N_i - min[N(\lambda)]}$$

4

Computational Details

Our goal in benchmarking our new method is to demonstrate that it provides comparable accuracy to classical methods for RBFE calculation with significant improvements in efficiency and cost savings. To that end, we selected three literature examples to benchmark LaDyBUGS, DNA Ligase,[56] Major Urinary Protein 1 (MUP1),[57] and c-Met kinase (c-Met),[58] which have been featured in previous benchmarking studies of FEP + and non-equilibrium switching.[5,24,57] In total, binding free energies were calculated for 24 different ligands: 6 for MUP1, 7 for DNA ligase, and 11 for c-Met (Fig. 4). To avoid the complications of charge-changing perturbations,[5,59,60] only neutral ligands were used in this study.

Benchmark System Details. LaDyBUGS has been implemented in OpenMM and all simulations were run using the CUDA platform.[51] CHARMM-based force field parameters were used to represent different components of the chemical systems. CHARMM36 was used for all protein atoms.[61−63] Small molecule ligand atoms were parameterized with ParamChem/CGenFF atom types[64−66] and partial atomic charges from the MATCH atom parameterization tool.[67] The TIP3P water model was used to represent water.[68] Initial protein complex starting coordinates were taken from PDBIDs 4CC5,[56] 1I06,[69] and 4R1Y[58] for DNA

Ligase, MUP1, and c-Met, respectively. Protonation states of titratable residues at a pH of 7.0 were determined with the assistance of MolProbity[70] and ProPKa.[71] Protein systems were prepared and solvated using the CHARMM-GUI webserver[72] and cubic water boxes were constructed with a 10 Å buffer between solute atoms and box edges. Enough ions to neutralize the system and create a 0.1 M NaCl solution were added. Small molecule structure files for MUP1 and DNA ligase ligands were constructed manually using UCSF Chimera.[73] Published structure files were used as initial coordinates for the c-Met compounds.[5] Alchemical functional groups were created as multiple topology models, with explicit atoms for every unique functional group, using the *msld_py_prep* utility.[74] Unbound ligand cubic solvent boxes were constructed with the *convpdb.pl* tool from the MMTSB toolset,[75] with a 12 Å buffer between solute atoms and box edges. Starting psf topology and pdb coordinate files were generated with the CHARMM molecular simulation package prior to running LaDyBUGS.[76,77] A series of discrete λ states were created to track alchemical transformations along connective edges between ligand end states, following the procedure used for d-GSλD.[47] In OpenMM, the CHARMM generated psf and pdb files were loaded in with the *CharmmPsfFile* and *CharmmParameterSet* classes. A nonbonded lookup table was generated to handle CHARMM's NBFIX nonbonded parameter exceptions, and custom non-bonded forces were written to facilitate λ scaling of all alchemical functional groups. These custom nonbonded forces included CHARMM's force switching and λD-based soft-core potentials.[41,78] All LaDyBUGS simulations were performed at 25°C and 1 atm in the isothermal-isobaric ensemble. In OpenMM, this was accomplished with a Monte Carlo barostat[79,80] and a Langevin integrator[81] with a friction coefficient of 10 ps^{-1}. An integration time step of 2 fs was used, facilitated by constraining all hydrogen to heavy atom bond lengths with the SHAKE algorithm.[82] Periodic boundary conditions were employed, and force switching was used to gradually smooth nonbonded forces to zero between 10 and 12 Å.[78] During a LaDyBUGS simulation, trajectory frames were saved at the end of a GS step, if an alchemical end state was sampled. VMD[83] and PyMOL[84] were used to visualize and analyze simulation trajectories.

LaDyBUGS Free Energy Calculations. Ligands in the three test systems were grouped together as follows: 6 MUP1 ligands were sampled collectively, 7 DNA ligase ligands were sampled collectively, and 11 c-Met ligands were grouped into two sets of 6 ligands each, including a common reference compound to connect the datasets (Fig. 4). A symmetric lambda spacing ($\Delta\lambda$) of 0.1 was used for all transformations. For ligands with a single site of alchemical perturbation, the number of total λ states (N_λ) scales quadratically with the number of ligands analyzed (N_s), as shown in Eq. 5. As a result, 141 λ states were used in LaDyBUGS calculations analyzing 6 ligands, and 196 λ states were used to evaluate 7 ligands collectively. Though this work only investigates single-site perturbations, multisite perturbations would be equally feasible as well, as demonstrated with d-GSλD.[47]

$$N_\lambda = N_s + \frac{N_s(N_s - 1)}{2}\left(\frac{1}{\Delta\lambda} - 1\right)$$

5

For each LaDyBUGS calculation, the chemical system was subjected to 1000 steps of energy minimization at a random fixed λ state, followed by 5000 steps of MD equilibration to briefly relax the system. The workflow in Fig. 3 was then followed, with iterative sampling of $P(X|\lambda)$ and $P(\lambda|X)$ conditional distributions. A 100 time steps (200 fs)-long MD trajectory was used to sample $P(X|\lambda)$, and biases were updated after every $P(\lambda|X)$ sample was taken. After 1000 Gibbs sampler steps, a FastMBAR calculation was performed, and the $\Delta G_{t=u}$ results were used to update the biases according to Eq. 4. Gibbs sampling with bias updates then resumed. LaDyBUGS simulations were run for 15 ns each, during which FastMBAR was called 75 times for bias refinement (every 1000 GS steps). Simulations were run in triplicate for a total of 45 ns of simulation time expended per compound group. FastMBAR was used to collate data from all duplicate runs to yield the final free energy results, and free energy differences were bootstrapped to provide an estimate of precision. To further account for the variability in the relative free energies and bootstrapped errors, this last FastMBAR calculation was performed 10 times and the average value of the relative free energy and bootstrapped error estimates were used as the final results. To investigate the effects of running LaDyBUGS for shorter or longer, simulations were also run for 5 ns and 25 ns each, respectively. Computed relative free energy differences ($\Delta\Delta G_{\text{comp}}$) were converted into absolute free energy differences (ΔG_{comp}) for comparison to experiment (ΔG_{expt}) with Eq. 6.[16,49]

$$\Delta G_{comp} = \Delta\Delta G_{comp} - \left(\frac{\sum \Delta\Delta G_{comp}}{n} - \frac{\sum \Delta G_{expt}}{n} \right)$$

6

TI/MBAR Free Energy Calculations. For each chemical system, pairwise perturbations were run between a reference ligand, highlighted with a gray box in Fig. 4, and all other ligand analogues. This perturbation approach has sometimes been called a "star map" (Fig. 2); redundant calculations for cycle closure were not performed to maximize TI/MBAR efficiency.[18−20] Alchemical transformations were accomplished over 11 λ windows with a $\Delta\lambda$ schedule of 0.1. For each λ window, the chemical system was subjected to 1000 steps of energy minimization and 5000 steps of MD equilibration. MD simulations were run for 5 ns per λ window, and configurations were saved every 100 time steps (200 fs) for a subsequent FastMBAR analysis. Similar to LaDyBUGS, each calculation was run in triplicate for a total of 165 ns of sampling per pairwise perturbation. Configurations from all λ windows and duplicates were pooled together and supplied to FastMBAR to estimate a final relative free energy difference. We refer to these results as "TI/MBAR (5 ns/window)". Bootstrapping and duplicate FastMBAR calculations were performed to account for variability in the free energy differences and bootstrapped errors, and averaged values were used as the final results. Relative binding free energies were again converted into absolute binding affinities (ΔG_{bind}) for comparison to LaDyBUGS and experiment. To investigate the effects of running TI/MBAR for longer, λ window simulations were also extended and sampled for 15 ns each, referred to as "TI/MBAR (15 ns/window)". These longer simulations required 495 ns of total sampling for a single pairwise perturbation.

Results And Discussion

Benchmarking Results. In this work, we assess the accuracy and efficiency of the introduced LaDyBUGS method compared to experiment and a community standard alchemical free energy method (TI/MBAR). Free energies of binding were calculated for 24 ligands bound to one of three benchmark protein systems: DNA ligase, MUP1, or c-Met. Figure 5 plots the correlation between experiment and computed ΔG_{bind} with LaDyBUGS (15 ns) and TI/MBAR (5 ns/window); all data points are reported in Table S1 of the Supplementary Information. Root mean square error (RMSE) and Kendall τ scores[85] were computed for each test system individually and for the combined dataset. Using these metrics, we see a uniform improvement in both RMSE and Kendall τ with LaDyBUGS relative to TI/MBAR. For all 24 ligands, the LaDyBUGS RMSE was 0.89 kcal/mol and the Kendall τ was 0.67. For every test case, the calculated LaDyBUGS RMSE was about or below 1.0 kcal/mol, a typical goal and state-of-the-art for predictive accuracy in free energy calculations for drug discovery.[5−7, 11,16,17,24] It is important to note that accuracy is dependent on both correct force field representation of a chemical system and thorough configurational sampling with a given free energy method.[86] The larger RMSE of 1.03 kcal/mol and reduced Kendall τ of 0.59 from TI/MBAR (5 ns/window), which used the same force field parameters as LaDyBUGS, suggests LaDyBUGS is providing improved sampling proficiency over TI/MBAR for the same benchmark systems. This seems extraordinary, since LaDyBUGS used 19.2 times less sampling than TI/MBAR (5 ns/window) (Fig. 6).

Because both LaDyBUGS and TI/MBAR calculations used the same force field parameters, we can also compare the precision of their ΔG_{bind} predictions (Fig. 6). The two computational methods agree well with each other, with an overall RMSE of 0.43 kcal/mol. Considering that most protein-ligand ΔG_{bind} calculations have computed uncertainties between 0.3−0.5 kcal/mol,[16,48,49] and that LaDyBUGS bootstrapped errors ranged from 0.08−0.48 kcal/mol (Table S1), these results suggest good agreement exists between these free energy methods. To explore the effect of sampling time on the ΔG_{bind} results, we also compared these methods with less sampling per LaDyBUGS (5 ns) simulation and more sampling per TI/MBAR calculation (15 ns/window). In Fig. 7, agreement between LaDyBUGS simulations with only 5 ns of sampling compared to TI/MBAR (5 ns/window) remains high with a RMSE of 0.40 kcal/mol and a Kendall τ of 0.83. As expected with a reduction in sampling, the mean bootstrapping error for LaDyBUGS increased from 0.21 to 0.37 kcal/mol, although the RMSE between LaDyBUGS (5 ns) and experiment remains about the same at 0.92 kcal/mol. This level of agreement seems remarkable considering LaDyBUGS (5 ns) used 57.8 times less sampling on average than TI/MBAR (5 ns/window). For the DNA ligase test case, LaDyBUGS used 66 times perturbations in all benchmark systems, for the protein-bound edge of the thermodynamic cycle less sampling than TI/MBAR. Thus, from a total of 60 ns expended to sample 24 alchemical (Fig. 1), LaDyBUGS (5 ns) provided ΔG_{bind} predictions with errors at or below 1.0 kcal/mol compared to experiment. On a per-ligand basis, LaDyBUGS spent only an average of 2.5 ns of protein-ligand sampling. These results highlight that significant cost savings are achievable with LaDyBUGS without compromising accuracy in the computed ΔG_{bind} results. In contrast, TI/MBAR required almost 3.5 μs of total protein-ligand sampling (165 ns/ligand), without employing commonly used redundant calculations for cycle closure and hysteresis error reduction.[18−20] To test LaDyBUGS

convergence, 25 ns simulations were also run. No large deviations were observed and the RMSE between 15 ns LaDyBUGS and 25 ns LaDyBUGS simulations was 0.11 kcal/mol, well within statistical noise (Table S1). The RMSE of 0.23 kcal/mol between 5 ns and 25 ns LaDyBUGS results was slightly larger but still within noise, suggesting a high degree of convergence even with a minimal amount of LaDyBUGS sampling. Figure 7 also shows the effects of extending TI/MBAR sampling to 15 ns/window for the DNA ligase and c-Met systems. Large improvements are observed in comparison to TI/MBAR (5 ns/window). The RMSE to experiment improves to 0.93 kcal/mol, and the RMSE to LaDyBUGS decreases to 0.27 kcal/mol. The strong agreement between short 5 ns runs of TI/MBAR and LaDyBUGS, as well as between the longer 15 ns runs of TI/MBAR and LaDyBUGS, suggests LaDyBUGS is able of deliver comparable accuracy as TI/MBAR with significant cost savings in terms of sampling (18–66 times less simulation time required). Improved efficiency with LaDyBUGS directly stems from its ability to investigate several alchemical perturbations collectively within a single simulation, without the need to break up transformations into separate λ windows spread across multiple separate simulations (Fig. 2).

Uniformity of λ Sampling with LaDyBUGS. One potential issue associated with expanded ensemble free energy methods is a difficulty in achieving sampling smoothness of all λ states, e.g., avoiding becoming stuck primarily sampling one or several states too often and neglecting to sample all other λ states.[41,87] This problem has been observed in an expanded ensemble investigation that used a Wang-Landau algorithm to propagate λ switching.[28] In conventional λD simulations, including d-GSλD, static biases are added to a simulation to reduce free energy barriers between λ states and facilitate transitions between λ states.[41,47] In most situations these biases work well, all λ states are evenly sampled, and reliable free energy predictions are obtained. But burn-in time is required to first identify appropriate biases for these methods, which decreases their overall efficiency. Furthermore, if the protein-ligand system experiences a rare event during production sampling, such as a conformational change associated with a slow degree of freedom, static biases may be mismatched to the free energy surface of the new conformation and thus unable to facilitate continuous λ sampling; the simulation would become trapped, requiring new biases to be identified and sampling to be restarted. However, the dynamic biases used in LaDyBUGS continuously propagate the sampling of many λ states without prior burn-in time for bias identification and allow for conformational plasticity of the chemical system without getting trapped in a small number of λ states. Biases from Eq. 4 rely solely on the number of times each λ state has been sampled and on-the-fly FastMBAR free energy estimates; thus, LaDyBUGS can provide incredibly smooth λ sampling throughout an entire simulation. Figure 8 shows the difference between the minimum and maximum number of times a λ state was sampled as a function of time, referred to as "counts", averaged across all protein simulations used for benchmarking. The difference does not exceed a count of 4, even though λ states are sampled more than 500–800 times by the end of each simulation. This level of sampling smoothness ensures that LaDyBUGS does not become trapped in λ sampling and provides rapid transitions between multiple ligand end states for accurate free energy calculation with FastMBAR.

LaDyBUGS Samples a Mixture Distribution of λ States. Smooth transitions between states are also facilitated by strong energetic overlap between neighboring λ states. In our benchmark studies, c-Met

group 1 consists of different 5-membered heterocycles while group 2 contains a mixture of carbamate and aryl substituents (Fig. 4). As shown in Fig. 9 for two example c-Met group 1 and 2 perturbations, a uniform $\Delta\lambda$ schedule provides good energetic overlap between both similar (c-Met Group 1) and dissimilar (c-Met Group 2) transformations. This enables facile transitions to adjacent λ states when sampling the $P(\lambda|X)$ conditional distribution (Fig. 3). As shown in Fig. 9, most transitions occur to + 1 or + 2 states away, although large jumps (> 4 states) are sometimes observed. The degree of overlap between λ states affects the transition distance traveled, with higher overlap facilitating larger jumps (see also Table S2). The mean transition distance traveled for c-Met group 1 is 2.47 states, but it is smaller at 1.64 states for c-Met group 2 which has less overlap between adjacent states (Fig. 9). Fortuitously, transitions between energetically similar and adjacent λ states enables the chemical system to quickly relax and equilibrate during the brief 200 ps MD simulation whenever a new λ state is sampled. Therefore, we assume the MD configuration drawn from $P(X|\lambda)$ in each GS step represents an equilibrium sample. By constant sampling of different λ states and atomic coordinates, LaDyBUGS can efficiently sample a mixture distribution of λ states within a single simulation. Pairing free energy determination with the MBAR algorithm is natural then, because MBAR pools and reweights samples as if they originated from a mixture distribution.[15,88] Our proof in the SI demonstrates that samples drawn from the same λ state with different external biases can be treated as coming from the same state. We then use FastMBAR to obtain equilibrium free energy results from a LaDyBUGS simulation, under the stated assumptions of the proof. Furthermore, because sampling of the $P(\lambda|X)$ conditional distribution requires energies to be calculated for every λ state at every sampled $P(X|\lambda)$ configuration, no postprocessing of LaDyBUGS trajectories is required to run MBAR; all necessary information is generated on-the-fly and available at the conclusion of a LaDyBUGS simulation.

Software Implementation. LaDyBUGS has been implemented in OpenMM[51] and all LaDyBUGS scripts are available for download on the Vilseck-Lab GitHub page. One advantage of using OpenMM for LaDyBUGS is the ability to use force groups to partition the interactions of different components of an alchemical system and thus enable λ state-dependent energies to be evaluated without recalculating the energy of the entire chemical system. This feature speeds up the sampling of $P(\lambda|X)$ which requires λ-dependent energies to be calculated for every λ-state at every $P(X|\lambda)$ configuration. Consequently, we find that sampling a group of 6 ligands collectively with 141 λ states is only marginally slower than performing a standard pairwise perturbation of 11 λ states with LaDyBUGS. For example, on a NVIDIA 2080 TI graphics processing unit (GPU), 6 duplicate 5 ns LaDyBUGS c-Met group 1 cmet_9 → cmet_10 pairwise perturbations each took ca. 10.85 hours to run. Similarly, 6 duplicate 5 ns simulations of all 6 c-Met group1 ligands sampled collectively took ca. 11.01 hours each. Thus, the combined 6-ligand calculation was only ~ 1.5% slower, highlighting the effectiveness and cost-savings of sampling multiple ligands simultaneously with LaDyBUGS. With our current implementation of LaDyBUGS in OpenMM and using an assumption of sampling 6 perturbations per LaDyBUGS simulation, we estimate that ca. 4−13 compound perturbations can be investigated per day per 1 GPU with LaDyBUGS using a range of 15 ns to 5 ns of sampling per calculation, respectively. On a modest cluster of 25 GPUs, this readily scales to 100−325 perturbations per day! Hence, rapid high-throughput screening of hundreds of lead compound analogues

with highly accurate free energy predictions is obtainable with LaDyBUGS within a day using minimal computational resources.

Work is on-going to further optimize our implementation of LaDyBUGS in OpenMM as well as incorporate it into other software suites, including CHARMM. If a program lacks the ability to partition energetic interactions via a "force group"-like algorithm, $P(\lambda|X)$ may be sampled by calculating the energy of the entire chemical system; all non-alchemical environment-to-environment interactions should cancel out when λ state-dependent energies are compared. Though some wall-time slowdown may be expected to occur as a consequence of running a larger energy evaluation, we anticipate that LaDyBUGS would still provide highly efficient results, nonetheless. Incorporating LaDyBUGS into CHARMM, or other programs, could provide additional benefits too. For example, the *CustomNonbondedForce* class in OpenMM makes it challenging to use particle mesh Ewald (PME) methods with LaDyBUGS. However, a λD-based PME approach is already available in CHARMM and BLaDE for running MSλD simulations,[45,76,77,89] and this could be utilized with LaDyBUGS in CHARMM to facilitate the inclusion of long-range electrostatic interactions in future calculations.

Multisite sampling. Finally, we emphasize that the efficiency gains for LaDyBUGS reported in this work used only single site perturbations, where substituent group modifications occurred at only one site off a central ligand core. Multisite perturbations, with functional group substitutions occurring at multiple sites around a ligand core, could also be accomplished by using the previously described d-GSλD approach for creating two- or three-dimensional λ states.[47] Such LaDyBUGS simulations may need longer total sampling to obtain converged results due to the increased number of λ states required for multisite sampling, but this has not been tested yet. Instead, this work focused on single site perturbations to match structure-activity relationship strategies typically pursued experimentally by changing one component of a lead compound at a time.[16,56–58] In this manner, LaDyBUGS seems especially adept at exploring incremental changes to a lead compound. Future investigations will reveal the applicability of LaDyBUGS to tackling larger or more challenging perturbations or for molecular decoupling to compute absolute free energies of binding directly.

Conclusions

Alchemical free energy methods such as FEP and TI have played pivotal roles in the lead optimization phase of drug design,[1–3,5–11] yet they require large computational costs to explore many tens to hundreds of alchemical perturbations. λD-based methods have shown improved scalability and efficiency in exploring large chemical spaces with reduced costs.[6,33,34,44,46–49] Hence, the object of this study was to investigate new approaches to further accelerate λD-based free energy methods by eliminating burn-in time commonly expended to identify static biases prior to production sampling. In this work, we have described the λ-Dynamics with Bias Updated Gibbs Sampling method, which is a Gibbs sampler-based λ-dynamics approach. To eliminate time spent for bias identification, LaDyBUGS uses continuous bias updating to rigorously drive the sampling of multiple λ states, and consequently multiple different

ligands, simultaneously within a single simulation. This results in very even and complete sampling of all λ states and significant efficiency gains, compared to TI/MBAR. Evaluated against three experimental benchmarks, LaDyBUGS root-mean-square errors of computed ΔG_{bind} compared to experiment were less than 1 kcal/mol on average with only 5 to 15 ns of sampling per simulation. LaDyBUGS RMSEs were lower than the corresponding error with TI/MBAR in all test cases notwithstanding the use of only ~ 2–5% of the total amount of TI sampling. From these results, we estimate that highly accurate ΔG_{bind} estimates can be obtained with only ~ 2.5–5 ns of LaDyBUGS sampling per ligand. From timing benchmarks of LaDyBUGS implemented in OpenMM, we estimate that ca. 4–13 perturbations can be examined per day per GPU with LaDyBUGS, depending on the length of sampling. Using a modest amount of GPU resources (with as few as 25 GPUs), this can easily scale to hundreds of compounds examined within a day. We envision that the rapid delivery of ΔG_{bind} predictions via LaDyBUGS could thus be used to screen hundreds of compound analogs with minimal computational costs, accelerating computer-aided drug discovery at an incredible pace.

Declarations

ASSOCIATED CONTENT

AUTHOR INFORMATION

Corresponding Author

*E-mail: jvilseck@iu.edu. Address: Department of Biochemistry and Molecular Biology, Center for Computational Biology and Bioinformatics, Indiana University School of Medicine, Indianapolis, Indiana, 46202, United States.

ORCID

Michael Robo: 0000-0003-1356-246X.

Ryan Hayes: 0000-0003-1052-6391.

Xinqiang Ding: 0000-0002-4598-8732.

Jonah Z. Vilseck: 0000-0001-7076-8996.

Notes

Scripts to setup and run LaDyBUGS simulations in OpenMM can be found online at github.com/Vilseck-Lab/LaDyBUGS for download.

References

1. Jorgensen, W. L. The Many Roles of Computation in Drug Discovery. Science **303**, 1813–1818 (2004).

2. Song, L. F. & Merz, K. M. Evolution of Alchemical Free Energy Methods in Drug Discovery. J. Chem. Inf. Model. **60**, 5308–5318 (2020).

3. Chodera, J. D. *et al.* Alchemical free energy methods for drug discovery: progress and challenges. Curr. Opin. Struct. Biol. **21**, 150–160 (2011).

4. Kollman, Peter. Free energy calculations: Applications to chemical and biochemical phenomena. Chem. Rev. **93**, 2395–2417 (1993).

5. Schindler, C. E. M. *et al.* Large-Scale Assessment of Binding Free Energy Calculations in Active Drug Discovery Projects. J. Chem. Inf. Model. **60**, 5457–5474 (2020).

6. Raman, E. P., Paul, T. J., Hayes, R. L. & Brooks, C. L. Automated, Accurate, and Scalable Relative Protein-Ligand Binding Free-Energy Calculations Using Lambda Dynamics. J. Chem. Theory Comput. **16**, 7895–7914 (2020).

7. Abel, R., Wang, L., Harder, E. D., Berne, B. J. & Friesner, R. A. Advancing Drug Discovery through Enhanced Free Energy Calculations. Acc. Chem. Res. **50**, 1625–1632 (2017).

8. Lee, T.-S. *et al.* Alchemical Binding Free Energy Calculations in AMBER20: Advances and Best Practices for Drug Discovery. J. Chem. Inf. Model. **60**, 5595–5623 (2020).

9. Cournia, Z., Allen, B. & Sherman, W. Relative Binding Free Energy Calculations in Drug Discovery: Recent Advances and Practical Considerations. J. Chem. Inf. Model. **57**, 2911–2937 (2017).

10. Abel, R., Manas, E. S., Friesner, R. A., Farid, R. S. & Wang, L. Modeling the value of predictive affinity scoring in preclinical drug discovery. Curr. Opin. Struct. Biol. **52**, 103–110 (2018).

11. Mobley, D. L. & Klimovich, P. V. Perspective: Alchemical free energy calculations for drug discovery. J. Chem. Phys. **137**, 230901 (2012).

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

13. Wang, L., Berne, B. J. & Friesner, R. A. On achieving high accuracy and reliability in the calculation of relative protein–ligand binding affinities. *Proc. Natl. Acad. Sci.* **109**, 1937–1942 (2012).

14. Kirkwood, J. G. Statistical Mechanics of Fluid Mixtures. J. Chem. Phys. **3**, 300–313 (1935).

15. Shirts, M. R. & Chodera, J. D. Statistically optimal analysis of samples from multiple equilibrium states. J. Chem. Phys. **129**, 124105–124105 (2008).

16. Wang, L. *et al.* Accurate and Reliable Prediction of Relative Ligand Binding Potency in Prospective Drug Discovery by Way of a Modern Free-Energy Calculation Protocol and Force Field. J. Am. Chem. Soc. **137**, 2695–2703 (2015).

17. Kuhn, M. *et al.* Assessment of Binding Affinity via Alchemical Free-Energy Calculations. J. Chem. Inf. Model. **60**, 3120–3130 (2020).

18. Liu, S. *et al.* Lead optimization mapper: automating free energy calculations for lead optimization. J. Comput. Aided Mol. Des. **27**, 755–770 (2013).

19. Yang, Q., Burchett, W., Steeno, G. S., Liu, S., Yang, M., Mobley, D. L. & Hou, X. Optimal designs for pairwise calculation: An application to free energy perturbation in minimizing prediction variability. J. Comp. Chem. **41**, 247–257 (2020).

20. Wang, L., Deng, Y., Knight, J. L., Wu, Y., Kim, B., Sherman, W., Shelley, J. C., Lin, T. and Abel, R. Modeling Local Structural Rearrangements Using FEP/REST: Application to Relative Binding Affinity Predictions of CDK2 Inhibitors. J. Chem. Theory Comput. **9**, 1282–1293 (2013).

21. Chen, H., Maia, J. D. C., Radak, B. K., Hardy, D. J., Cai, W., Chipot, C. & Tajkhorshid, E. Boosting Free-Energy Perturbation Calculations with GPU-Accelerated NAMD. J. Chem. Inf. Model. **60**, 5301–5307 (2020).

22. Hynninen, A. P. & Crowley, M. F. New Faster CHARMM Molecular Dynamics Engine. J. Comput. Chem. **35**, 406–413 (2014).

23. Kutzner, C., Pall, S., Fechner, M, Esztermann, A., de Groot, B. L. & Grubmueller, H. More bang for your buck: Improved use of GPU nodes for GROMACS 2018. J. Comp. Chem. **40**, 2418–2431 (2019).

24. Gapsys, V. *et al.* Large scale relative protein ligand binding affinities using non-equilibrium alchemy. Chem. Sci. **11**, 1140–1152 (2020).

25. Khalak, Y., Tresadern, G., de Groot, B. L. & Gapsys, V. Non-equilibrium approach for binding free energies in cyclodextrins in SAMPL7: force fields and software. J. Comput. Aided Mol. Des. **35**, 49–61 (2021).

26. Baumann, H. M., Gapsys, V., de Groot, B. L. & Mobley, D. L. Challenges Encountered Applying Equilibrium and Nonequilibrium Binding Free Energy Calculations. J. Phys. Chem. B **125**, 4241–4261 (2021).

27. Gapsys, V. *et al.* Pre-Exascale Computing of Protein-Ligand Binding Free Energies with Open Source Software for Drug Design. J. Chem. Inf. Model. **62**, 1172–1177 (2022).

28. Zhang, S., Hahn, D. F., Shirts, M. R. & Voelz, V. A. Expanded Ensemble Methods Can be Used to Accurately Predict Protein-Ligand Relative Binding Free Energies. J. Chem. Theory Comput. **17**,
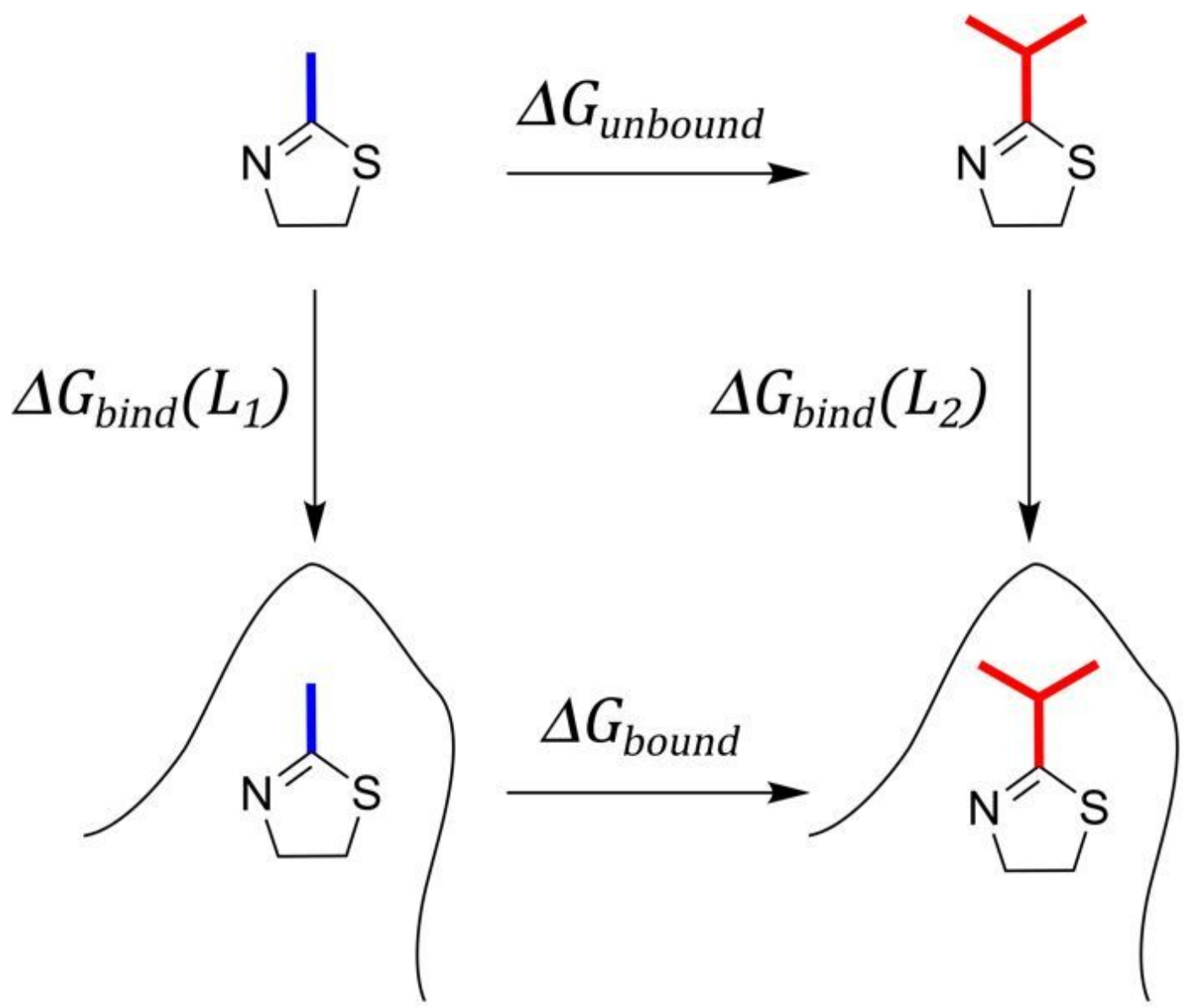
6536–6547 (2021).

29. Lyubartsev, A., Martsinovski, A., Shevkunov, S. & Vorontsov-Velyaminov, P. New Approach to Monte Carlo Calculation of the Free Energy: Method of Expanded Ensembles. J. Chem. Phys. **96**, 1776–1783 (1992).

30. Tan, Z. Optimally Adjusted Mixture Sampling and Locally Weighted Histogram Analysis. J. Comput. Graph. Stat. **26**, 54–65 (2017).

31. Lindahl, V., Lidmar, J. & Hess, B. Accelerated Weight Histogram Method for Exploring Free Energy Landscapes. J. Chem. Phys. **141**, No. 044110 (2014).

32. Chodera, J. D. & Shirts, M. R. Replica exchange and expanded ensemble simulations as Gibbs sampling: Simple improvements for enhanced mixing. J. Chem. Phys. **135**, 194110 (2011).

33. Kong, X. & Brooks, C. L. λ-dynamics: A new approach to free energy calculations. J. Chem. Phys. **105**, 2414–2423 (1996).

34. Knight, J. L. & Brooks, C. L. Multisite λ Dynamics for Simulated Structure–Activity Relationship Studies. J. Chem. Theory Comput. **7**, 2728–2739 (2011).

35. Christ, C. D. & van Gunsteren, W. F. Enveloping distribution sampling: A method to calculate free energy differences from a single simulation. J. Chem. Phys. **126**, 184110 (2007).

36. Christ, C. D. & van Gunsteren, W. F. Simple, Efficient, and Reliable Computation of Multiple Free Energy Differences from a Single Simulation: A Reference Hamiltonian Parameter Update Scheme for Enveloping Distribution Sampling (EDS). J. Chem. Theory Comput. **5**, 276–286 (2009).

37. Perthold, J. W. & Oostenbrink, C. Accelerated Enveloping Distribution Sampling: Enabling Sampling of Multiple End States while Preserving Local Energy Minima. J. Phys. Chem. B **122**, 5030–5037 (2018).

38. Bieler, N. S. & Hünenberger, P. H. Communication: Estimating the initial biasing potential for λ-local-elevation umbrella-sampling (λ-LEUS) simulations via slow growth. J. Chem. Phys. **141**, 201101 (2014).

39. Bieler, N. S., Häuselmann, R. & Hünenberger, P. H. Local Elevation Umbrella Sampling Applied to the Calculation of Alchemical Free-Energy Changes via λ-Dynamics: The λ-LEUS Scheme. J. Chem. Theory Comput. **10**, 3006–3022 (2014).

40. Bieler, N. S., Tschopp, J. P. & Hünenberger, P. H. Multistate λ-Local-Elevation Umbrella-Sampling (MS-λ-LEUS): Method and Application to the Complexation of Cations by Crown Ethers. J. Chem. Theory Comput. **11**, 2575–2588 (2015).

41. Hayes, R. L., Armacost, K. A., Vilseck, J. Z. & Brooks, C. L. Adaptive Landscape Flattening Accelerates Sampling of Alchemical Space in Multisite λ Dynamics. J. Phys. Chem. B **121**, 3626–3635 (2017).

42. Knight, J. L. & Brooks, C. L. III Applying Efficient Implicit Constraints in Alchemical Free Energy Simulations. J. Comput. Chem. **32**, 3423–3432 (2011).

43. Hayes, R. L., Vilseck, J. Z. & Brooks, C. L. III. Approaching protein design with multisite λ dynamics: Accurate and scalable mutational folding free energies in T4 lysozyme. Prot. Sci. **27**, 1910–1922

(2018).

44. Hayes, R. L., Vilseck, J. Z. & Brooks, C. L. Addressing Intersite Coupling Unlocks Large Combinatorial Chemical Spaces for Alchemical Free Energy Methods. J. Chem. Theory Comput. ASAP Article, (2022) DOI: 10.1021/acs.jctc.1c00948.

45. Hayes, R. L., Buckner, J. & Brooks, C. L. BLaDE: A Basic Lambda Dynamics Engine for GPU-Accelerated Molecular Dynamics Free Energy Calculations. J. Chem. Theory Comput. **17**, 6799−6807 (2021).

46. Ding, X., Vilseck, J. Z., Hayes, R. L. & Brooks, C. L. Gibbs Sampler-Based λ-Dynamics and Rao−Blackwell Estimator for Alchemical Free Energy Calculation. J. Chem. Theory Comput. **13**, 2501−2510 (2017).

47. Vilseck, J. Z., Ding, X., Hayes, R. L. & Brooks, C. L. Generalizing the Discrete Gibbs Sampler-Based λ-Dynamics Approach for Multisite Sampling of Many Ligands. J Chem Theory Comput 13 (2021).

48. Vilseck, J. Z., Armacost, K. A., Hayes, R. L., Goh, G. B. & Brooks, C. L. Predicting Binding Free Energies in a Large Combinatorial Chemical Space Using Multisite λ Dynamics. J. Phys. Chem. Lett. **9**, 3328−3332 (2018).

49. Vilseck, J. Z., Sohail, N., Hayes, R. L. & Brooks, C. L. Overcoming Challenging Substituent Perturbations with Multisite λ-Dynamics: A Case Study Targeting β-Secretase 1. J. Phys. Chem. Lett. **10**, 4875−4880 (2019).

50. Ding, X., Vilseck, J. Z. & Brooks, C. L. Fast Solver for Large Scale Multistate Bennett Acceptance Ratio Equations. J. Chem. Theory Comput. **15**, 799−799 (2019).

51. Eastman, P. *et al.* OpenMM 7: Rapid development of high performance algorithms for molecular dynamics. *PLOS Comput. Biol.* **13**, e1005659 (2017).

52. Geman, S. & Geman, D. Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images. IEEE Trans. Pattern Anal. Mach. Intell. **PAMI-6**, 721−741 (1984).

53. Smith, A. F. & Roberts, G. O. Bayesian computation via the Gibbs sampler and related Markov chain Monte Carlo methods. J. R. Stat. Soc. B **55**, 3−23 (1993).

54. Wang, F. & Landau, D. P. Efficient, Multiple-Range Random Walk Algorithm to Calculate the Density of States. Phys. Rev. Lett. **86**, 2050−2053 (2001).

55. Belardinelli, R. E. & Pereyra, V. D. Wang-Landau algorithm: A theoretical analysis of the saturation of the error. J. Chem. Phys. **127**, 184105 (2007).

56. Howard, S. *et al.* Fragment-Based Discovery of 6-Azaindazoles as Inhibitors of Bacterial DNA Ligase. ACS Med. Chem. Lett. **4**, 1208−1212 (2013).

57. Steinbrecher, T. B. *et al.* Accurate Binding Free Energy Predictions in Fragment Optimization. J. Chem. Inf. Model. **55**, 2411−2420 (2015).

58. Dorsch, D. *et al.* Identification and optimization of pyridazinones as potent and selective c-Met kinase inhibitors. Bioorg. Med. Chem. Lett. **25**, 1597−1602 (2015).

59. 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. **14**, 6346–6358 (2018).

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

61. Best, R. B., Mittal, J., Feig, M. & MacKerell, A. D. Inclusion of Many-Body Effects in the Additive CHARMM Protein CMAP Potential Results in Enhanced Cooperativity of α-Helix and β-Hairpin Formation. Biophys. J. **103**, 1045–1051 (2012).

62. Best, R. B., Zhu, X., Shim, J., Lopes, P. E. M., Mittal, J., Feig, M. & MacKerell, A. D. Optimization of the Additive CHARMM All-Atom Protein Force Field Targeting Improved Sampling of the Backbone ϕ, ψ and Side-Chain χ1 and χ2 Dihedral Angles. J. Chem. Theory Comput. **9**, 3257–3273 (2012).

63. Huang, J., Rauscher, S., Nawrocki, G., et al. CHARMM36m: an improved force field for folded and intrinsically disordered proteins. Nat Methods **14**, 71–73 (2017).

64. Vanommeslaeghe, K. *et al.* CHARMM general force field: A force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields. *J. Comput. Chem.* **31**, NA-NA (2009).

65. Vanommeslaeghe, K. & MacKerell, A. D. Automation of the CHARMM General Force Field (CGenFF) I: Bond Perception and Atom Typing. J. Chem. Inf. Model. **52**, 3144–3154 (2012).

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

67. Yesselman, J. D., Price, D. J., Knight, J. L. & Brooks, C. L. MATCH: An atom-typing toolset for molecular mechanics force fields. J. Comput. Chem. **33**, 189–202 (2012).

68. Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W. & Klein, M. L. Comparison of simple potential functions for simulating liquid water. J. Chem. Phys. **79**, 926–935 (1983).

69. Timm, D. E., Baker, L. J., Mueller, H., Zidek, L. & Novotny, M. V. Structural basis of pheromone binding to mouse major urinary protein (MUP-I). Prot. Sci. **10**, 997–1004 (2001).

70. Williams, C. J. *et al.* MolProbity: More and better reference data for improved all-atom structure validation. Protein Sci. **27**, 293–315 (2018).

71. Søndergaard, C. R., Olsson, M. H. M., Rostkowski, M. & Jensen, J. H. Improved Treatment of Ligands and Coupling Effects in Empirical Calculation and Rationalization of pKa Values. J. Chem. Theory Comput. **7**, 2284–2295 (2011).

72. Jo, S., Kim, T., Iyer, V. G. & Im, W. CHARMM-GUI: A web-based graphical user interface for CHARMM. J. Comput. Chem. **29**, 1859–1865 (2008).

73. Pettersen, E. F. *et al.* UCSF Chimera—A visualization system for exploratory research and analysis. J. Comput. Chem. **25**, 1605–1612 (2004).

74. Vilseck, J. Z., Cervantes, L. F., Hayes, R. L. & Brooks, C. L. Optimizing Multisite λ-Dynamics Throughput with Charge Renormalization. J. Chem. Inf. Model. **62**, 1479–1488 (2022).

75. Feig, M., Karanicolas, J. & Brooks, C. L. MMTSB Tool Set: enhanced sampling and multiscale modeling methods for applications in structural biology. Conform. Sampl. **22**, 377–395 (2004).

76. Brooks, B. R. *et al.* CHARMM: The biomolecular simulation program. J. Comput. Chem. **30**, 1545–1614 (2009).

77. Brooks, B. R. *et al.* CHARMM: A program for macromolecular energy, minimization, and dynamics calculations. J. Comput. Chem. **4**, 187–217 (1983).

78. Steinbach, P. J. & Brooks, B. R. New spherical-cutoff methods for long-range forces in macromolecular simulation. J. Comput. Chem. **15**, 667–683 (1994).

79. Chow, K.-H. & Ferguson, D. M. Isothermal-isobaric molecular dynamics simulations with Monte Carlo volume sampling. Comput. Phys. Commun. **91**, 283–289 (1995).

80. Åqvist, J., Wennerström, P., Nervall, M., Bjelic, S. & Brandsdal, B. O. Molecular dynamics simulations of water and biomolecules with a Monte Carlo constant pressure algorithm. Chem. Phys. Lett. **384**, 288–294 (2004).

81. Feller, S. E., Zhang, Y., Pastor, R. W. & Brooks, B. R. Constant pressure molecular dynamics simulation: The Langevin piston method. J. Chem. Phys. **103**, 4613–4621 (1995).

82. Ryckaert, J.-P., Ciccotti, G. & Berendsen, H. J. C. Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. J. Comput. Phys. **23**, 327–341 (1977).

83. Humphrey, W., Dalke, A. & Schulten, K. VMD: Visual molecular dynamics. J. Mol. Graph. **14**, 33–38 (1996).

84. Schrödinger, LLC. The PyMOL Molecular Graphics System, Version 1.8. (2015).

85. Kendall, M. G. A New Measure of Rank Correlation. Biometrika **30**, 81–93 (1938).

86. Mobley, D. L. Let's get honest about sampling. J. Comput. Aided Mol. Des. **26**, 93–95 (2012).

87. König, G., Ries, B., Hünenberger, P. H. & Riniker, S. Efficient Alchemical Intermediate States in Free Energy Calculations Using λ-Enveloping Distribution Sampling. J. Chem. Theory Comput. **17**, 58-5-5815 (2021).

88. Shirts, M. R. Reweighting from the mixture distribution as a better way to describe the Multistate Bennett Acceptance Ratio. *ArXiv170400891 Cond-Mat* (2017).

89. Huang, Y., Chen, W., Wallace, J. A. & Shen, J. All-atom continuous constant pH molecular dynamics with parti- cle mesh Ewald and titratable water. J. Chem. Theory Comput. **12**, 5411–5421 (2016).

Figures

$$\Delta\Delta G_{bind}(L_1 \to L_2) = \Delta G_{bind}(L_2) - \Delta G_{bind}(L_1)$$
$$= \Delta G_{bound} - \Delta G_{unbound}$$

Figure 1

Thermodynamic cycle for computing relative binding free energies.

Strongly Connected
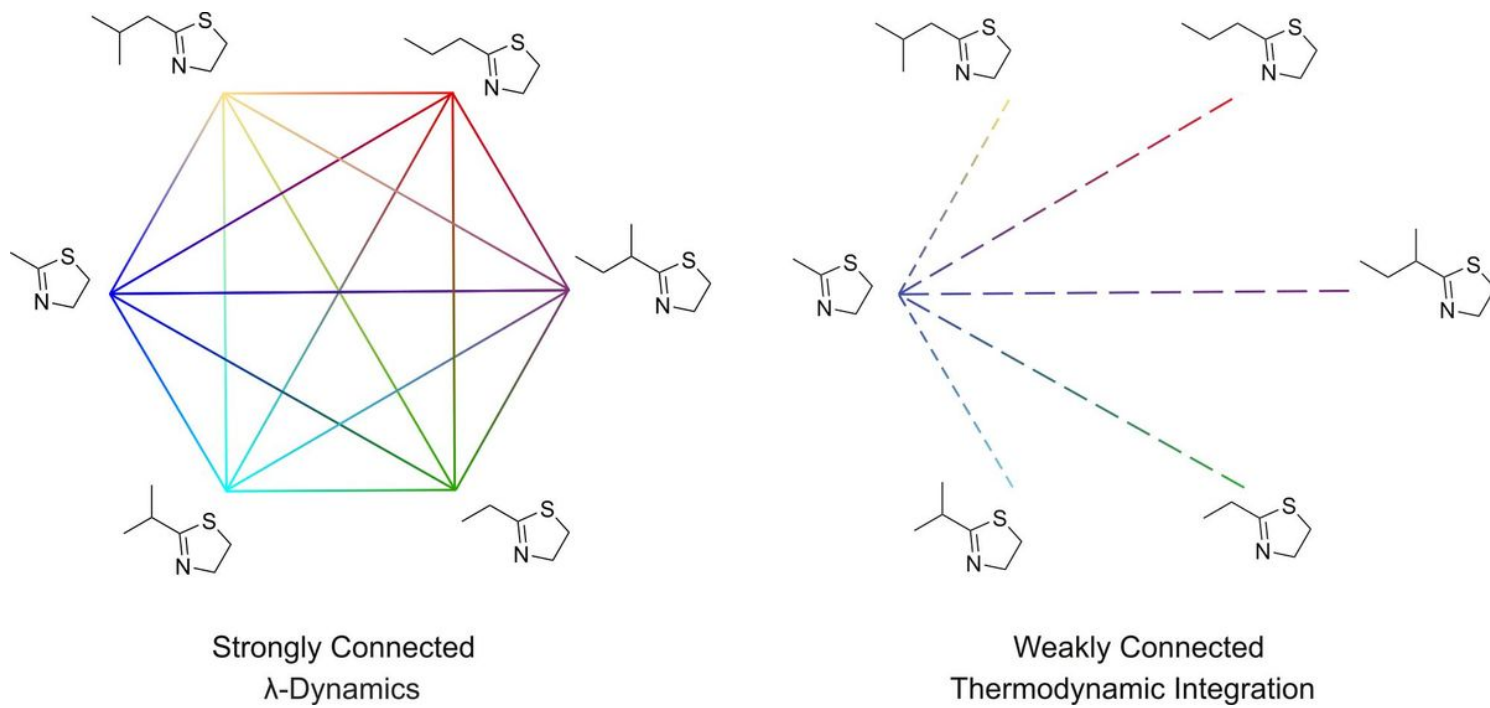λ-Dynamics

Weakly Connected
Thermodynamic Integration

Figure 2

λD-based methods sample a strongly connected graph of ligand end-states. Efficiency gains obtained with λD-based methods over traditional TI or FEP free energy methods originate from two key sources: (1) all physical and intermediate λ states are sampled within a single simulation, represented as a solid line, and (2) multiple ligands can be sampled simultaneously. In contrast, without redundant calculations for cycle closure, TI or FEP methods sample a weakly connected graph via pairwise perturbations (a "star map") and require many intermediate simulations to be run, represented as dashed lines with each dash representing a separate simulation.
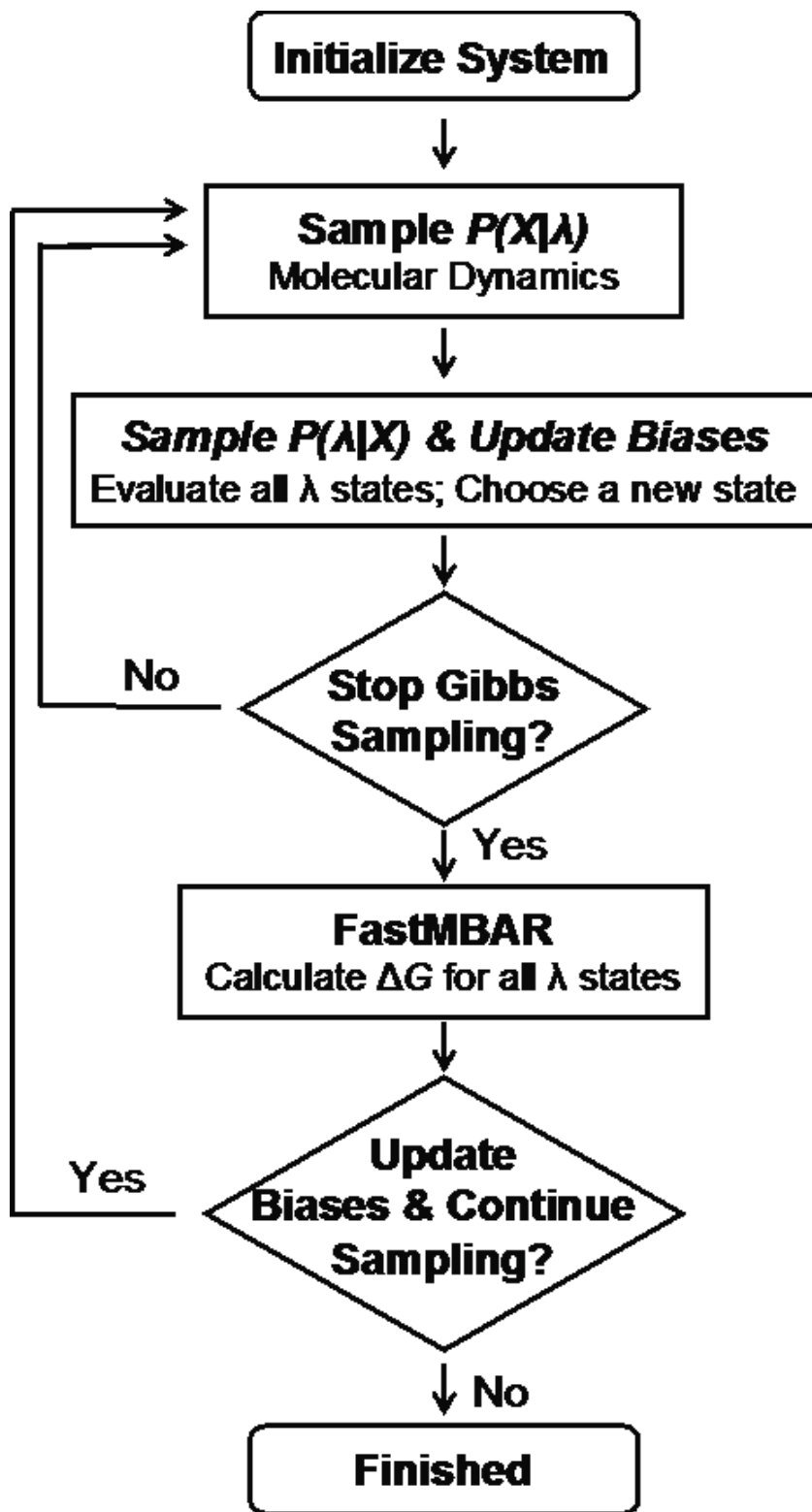
Figure 3

The LaDyBUGS workflow. Gibbs sampling is used to sample atomic coordinates and λ states of an alchemical system; a dynamic bias is continually updated to ensure continuous sampling of all λ states. Free energy differences and periodic bias refinements are computed with FastMBAR.
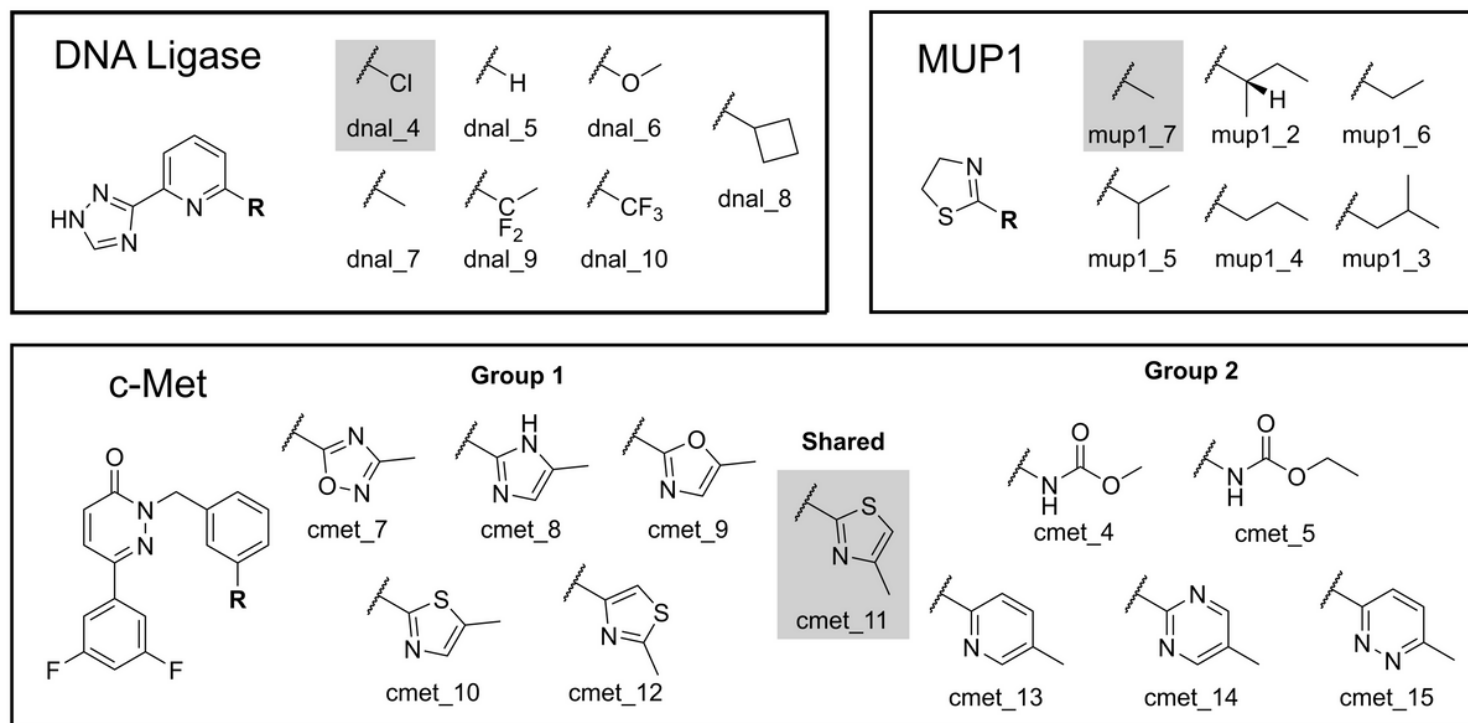
Figure 4

Three benchmark systems with 24 total ligands were used to evaluate and compare LaDyBUGS and TI/MBAR. LaDyBUGS calculations evaluated several ligands within a single simulation, and compounds were grouped as shown above. For the TI/MBAR calculations, a star map of pairwise perturbations was utilized (Figure 2). All relative free energy differences were calculated with respect to the gray-boxed reference compound for each system. Ligand numbering was kept consistent with the original experimental report (DNA ligase[56] and c-Met[58]) or a previous benchmark study (MUP1[57]).
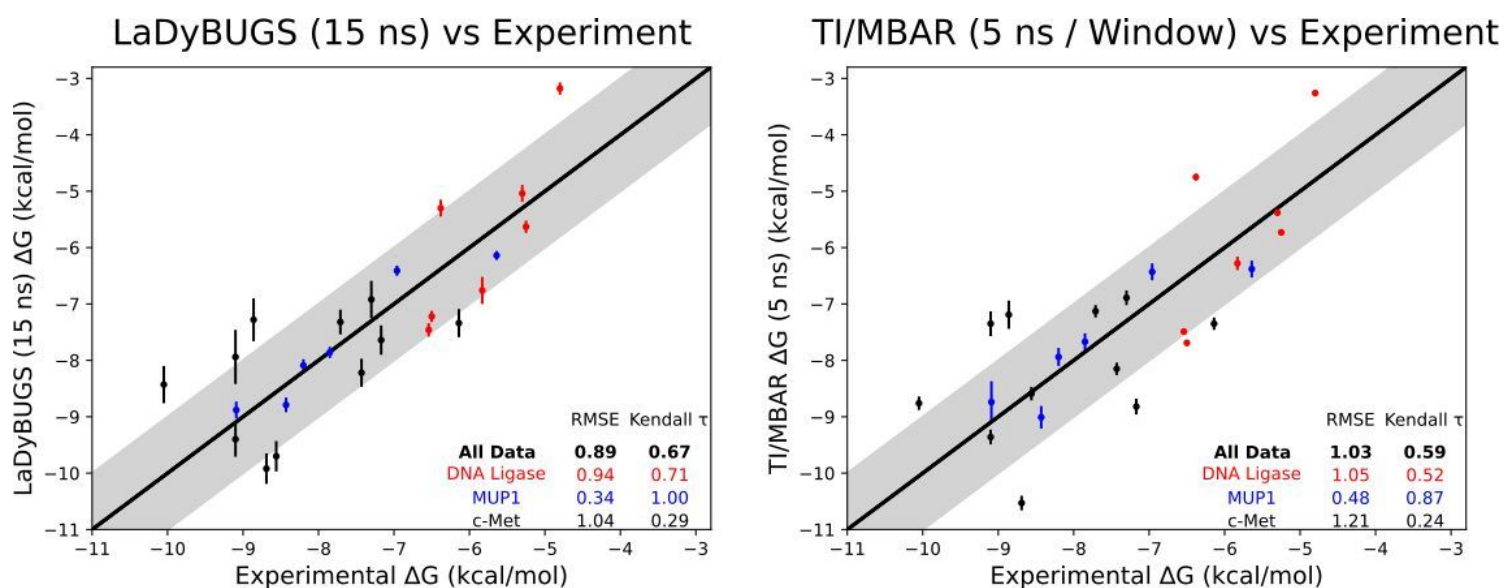

Figure 5

Computed LaDyBUGS and TI/MBAR ΔG_{bind} compared to experiment. The center black line represents perfect one-to-one correlation; the shaded gray area represents an error of ± 1 kcal/mol. Root-mean-square errors and Kendall τ statistics are reported.

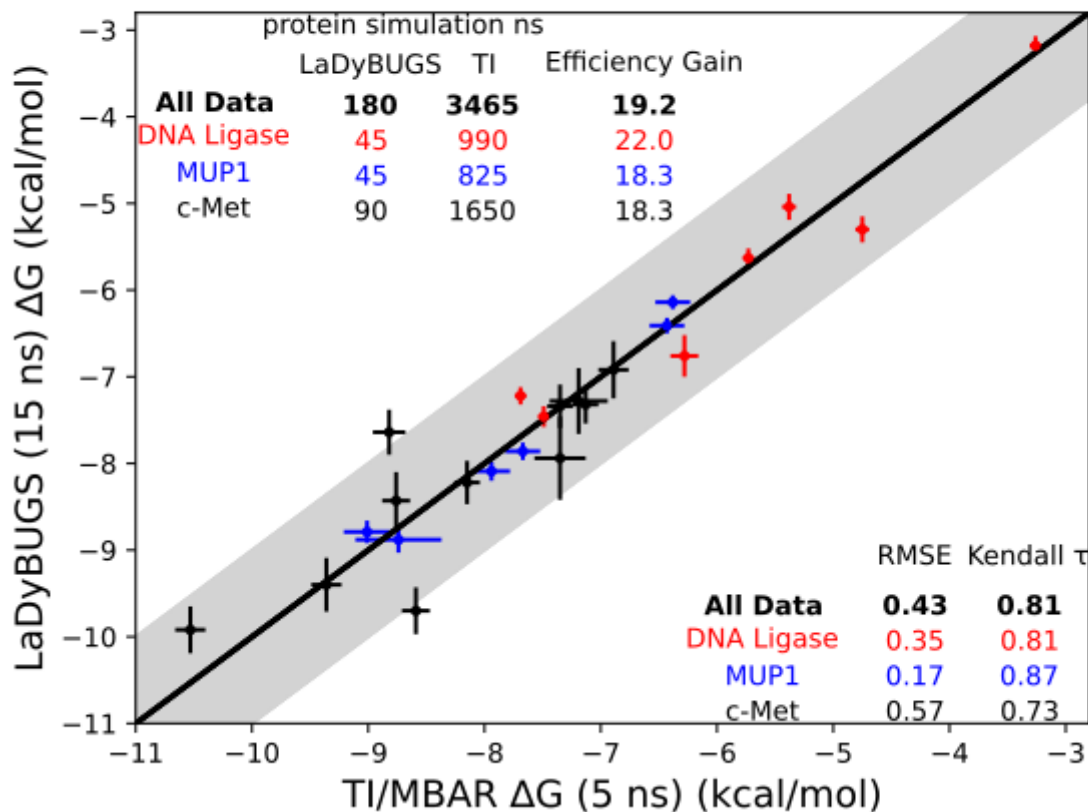## LaDyBUGS (15 ns) vs TI/MBAR (5 ns / Window)

protein simulation ns	LaDyBUGS	TI	Efficiency Gain
All Data	**180**	**3465**	**19.2**
DNA Ligase	45	990	22.0
MUP1	45	825	18.3
c-Met	90	1650	18.3

	RMSE	Kendall τ
All Data	**0.43**	**0.81**
DNA Ligase	0.35	0.81
MUP1	0.17	0.87
c-Met	0.57	0.73

Figure 6

Correlation between LaDyBUGS and TI/MBAR computed ΔG_{bind} results. The center black line represents y=x; the shaded gray area represents an error of ± 1 kcal/mol. Root-mean-square errors, Kendall τ statistics, total amount of sampling, and efficiency gain of LaDyBUGS over TI/MBAR in terms of sampling only is reported.
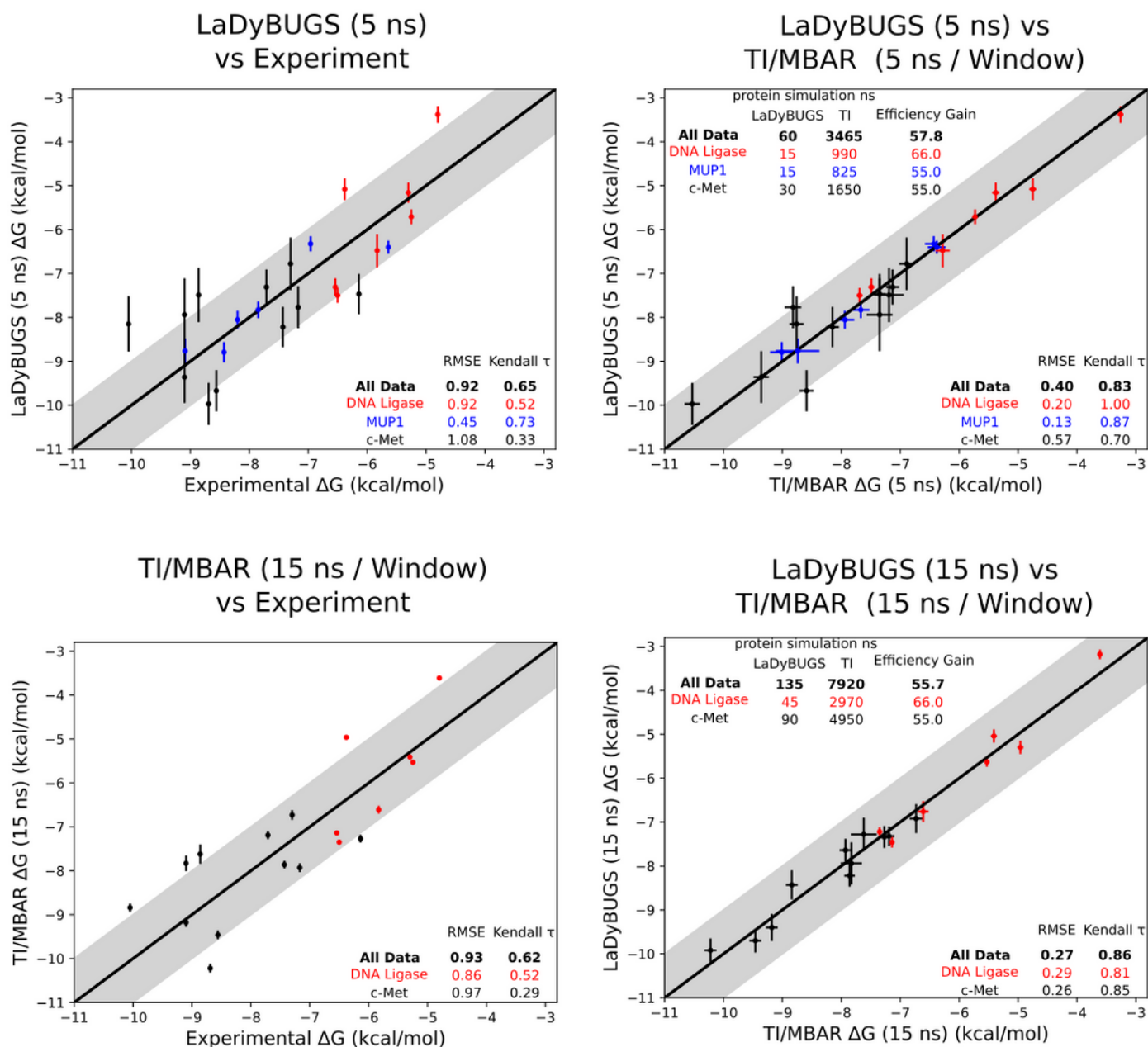
Figure 7

(Top) Correlation between 5 ns LaDyBUGS simulations with experiment (left) and TI/MBAR (right). (Bottom) Correlation between TI/MAR 15 ns/window runs compared to experiment (left) and 15 ns LaDyBUGS results (right), for DNA Ligase and c-Met data sets only. The center black line represents ideal one-to-one agreement; the shaded gray area represents an error of ± 1 kcal/mol. Root-mean-square errors, Kendall τ statistics, total amount of sampling, and efficiency gain of LaDyBUGS over TI/MBAR in terms of sampling only is reported.
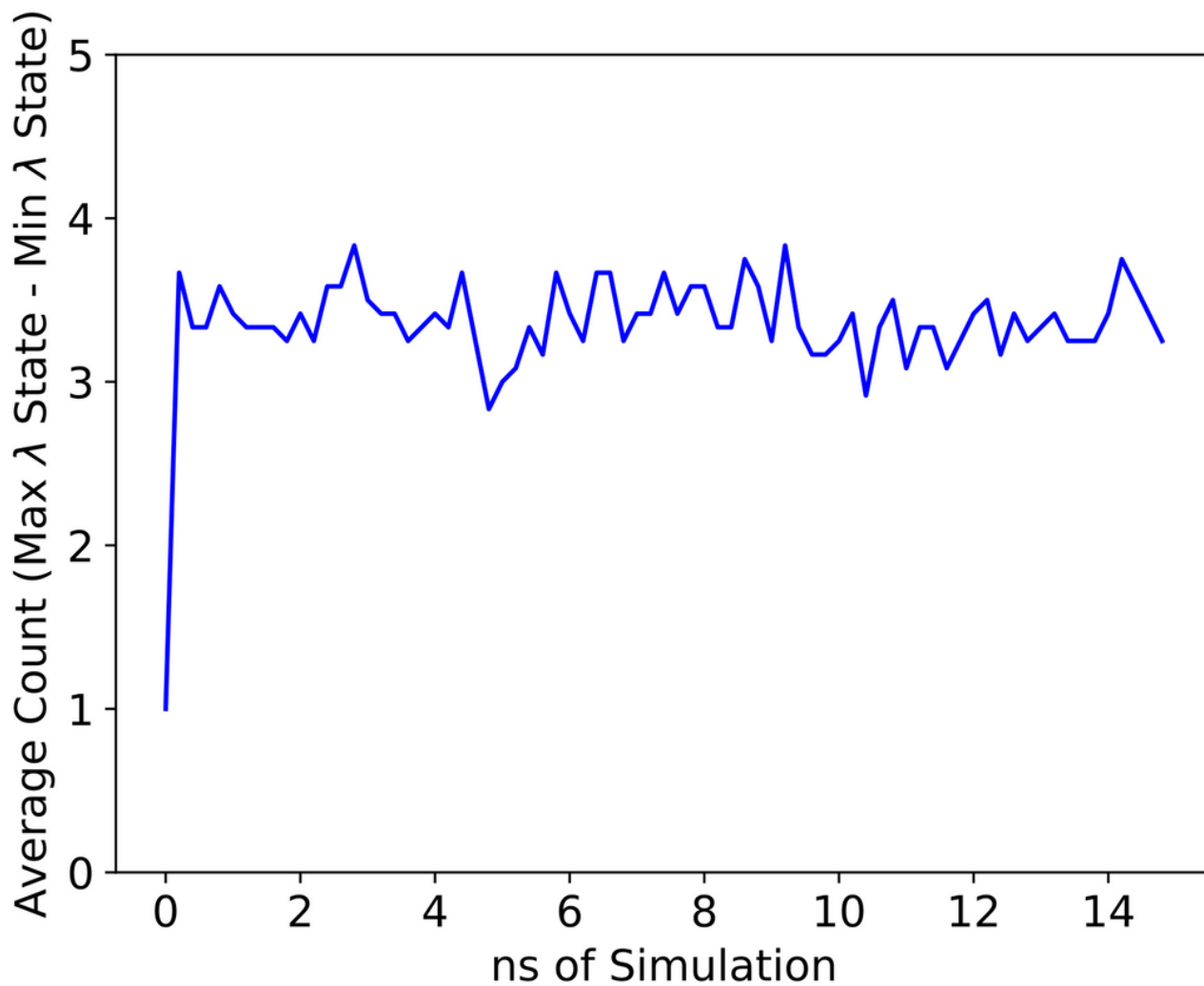
Figure 8

The average difference between the maximum and minimum number of times a λ state was sampled (referred to as "counts") as a function of time. Averages were calculated using all LaDyBUGS benchmark simulations.
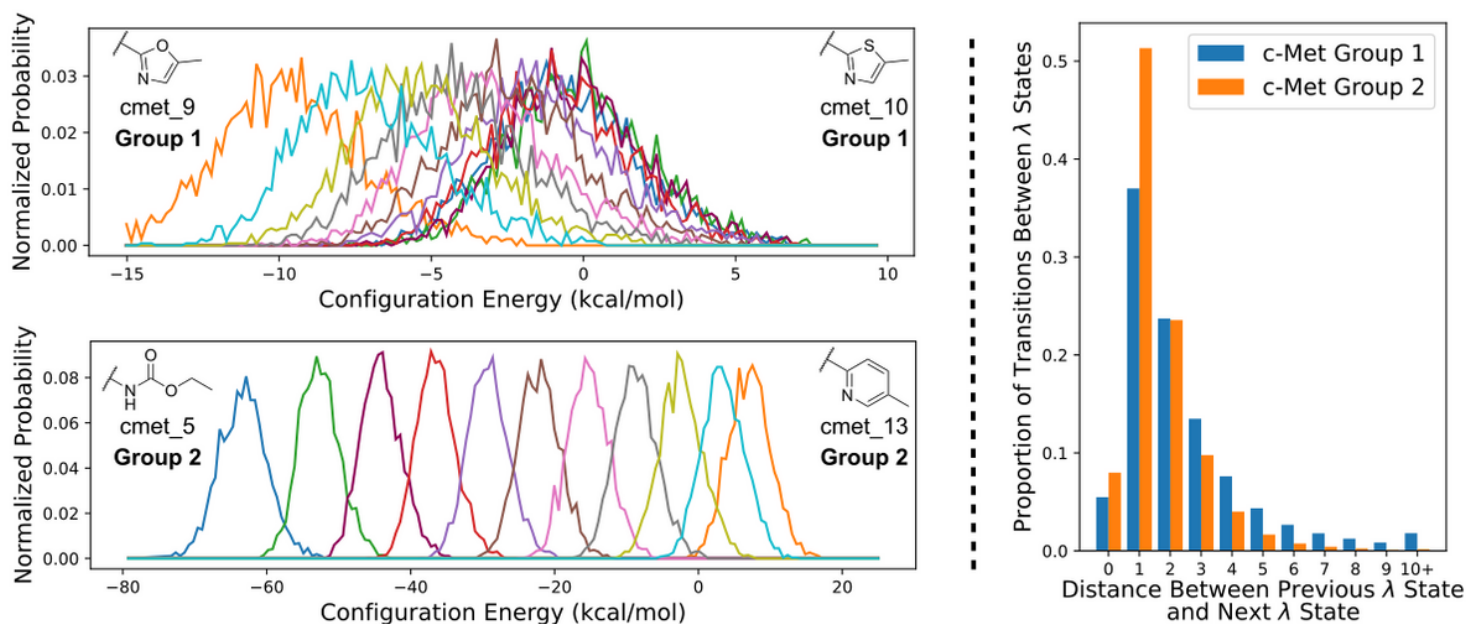
Figure 9

(Left) Normalized LaDyBUGS configuration energy distributions between adjacent λ states in two example c-Met group 1 and a c-Met group 2 perturbations. (Right) Normalized probabilities of the transition distances between λ states sampled in all c-Met group 1 and c-Met group 2 LaDyBUGS simulations. The greater the degree of overlap between adjacent λ states, the more probable long-range λ transitions become.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- floatimage10.jpeg
- ladybugsSIsubmitted.docx