

Expanded ensemble methods can be used to accurately predict protein-ligand relative binding free energies

Si Zhang,[†] David F. Hahn,[‡] Michael R. Shirts,^{*,¶} and Vincent A. Voelz^{*,†}

[†]*Department of Chemistry, Temple University, Philadelphia, PA 19122, USA*

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

[¶]*Department of Chemical and Biological Engineering, University of Colorado Boulder, Boulder, CO 80309*

E-mail: michael.shirts@colorado.edu; voelz@temple.edu

Abstract

Alchemical free energy methods have become indispensable in computational drug discovery for their ability to calculate highly accurate estimates of protein-ligand affinities. Expanded ensemble (EE) methods, which involve single simulations visiting all of the alchemical intermediates, have some key advantages for alchemical free energy calculation. However, there have been relatively few examples published in the literature of using expanded ensemble simulations for free energies of protein-ligand binding. In this paper, as a test of expanded ensemble methods, we computed relative binding free energies using the Open Force Field Initiative force field (codename “Parsley”) for twenty-four pairs of Tyk2 inhibitors derived from a congeneric series of 16 compounds. The EE predictions agree well with the experimental values (RMSE of 0.94 ± 0.13 kcal mol⁻¹ and MUE of 0.75 ± 0.12 kcal mol⁻¹). We find that while increasing the

number of alchemical intermediates can improve the phase space overlap, faster convergence can be obtained with fewer intermediates, as long as the acceptance rates are sufficient. We find that convergence can be improved using more aggressive updating of the biases, and that estimates can be improved by performing multiple independent EE calculations. This work demonstrates that EE is a viable option for alchemical free energy calculation. We discuss the implications of these findings for rational drug design, as well as future directions for improvement.

Introduction

Over the last decade, alchemical free energy methods have increased in accuracy and computational efficiency to become the dominant modeling approach for computing high-quality estimates of ligand binding free energy.¹⁻⁴ Particularly popular have been relative binding free energy methods, which can typically deliver accuracies within 1 kcal mol⁻¹.⁵⁻⁷ Unlike absolute free energy methods, which require the complete decoupling of a molecule’s non-bonded interactions, relative free energy methods only require alchemically transforming the set of atoms that differ between two molecules, resulting in better phase space overlap between alchemical intermediates. This efficient approach is especially useful for structure-based computational lead optimization, where one frequently makes small changes to a single scaffold, and many tools are now widely available to automate and perform these calculations.⁸⁻¹³

The expanded ensemble (EE) method, similar to simulated tempering, is an algorithm in which multiple thermodynamic ensembles are adaptively sampled in the same simulation.¹⁴ Specific variants of this adaptive approach include self-adjusted mixture sampling (SAMS)^{15,16} and the accelerated weight histogram method.¹⁷ Most other free energy approaches rely on multiple parallel or coupled simulations performed for each alchemical intermediate, as in Hamiltonian replica exchange,^{18,19} Therefore, the ability to sample multiple ensembles in a single simulation replica is attractive for many applications, such as

enabling large-scale virtual screening on distributed parallel computing platforms.^{20–22} EE methods, however, have not been widely adopted for estimating ligand binding free energies *per se*, although much work has been done towards improving adaptive expanded ensemble estimates for calculating free energies from perturbed Hamiltonians.^{23–26}

Of the handful of studies that have previously used EE for ligand binding, all have focused on the problem of computing *absolute* binding free energies (ABFE).^{22,27,28} In the SAMPL4 host-guest challenge, Monroe et al. showed that EE yields estimates comparable to other methods, in a system where molecular flexibility and multiple binding modalities are important.²⁷ Rizzi et al. report similarly results in the recent SAMPL6 host-guest challenge.²⁸ To our knowledge, no group has published an example of *relative* binding free energy (RBFE) estimates predicted using EE. In theory, expanded ensemble simulations are particularly well-suited for absolute binding free energies, as the ligand is able to freely bind and unbind, sampling different binding modes. However, even with relative binding free energies, the available phase space changes between ligands, meaning there is likely to be some sampling advantage even here. Given the better thermodynamic overlap of relative vs. absolute methods, we hypothesized that EE calculations would perform well at this task.

To see if this was the case, we examined the performance of EE in estimating RBFE for a set of twenty-four pairs of Tyk2 inhibitors (Figure 1) selected from a congeneric series of 16 compounds designed by Liang et al.^{29,30} Below, we present our methodology and results, which suggest that EE can predict RBFEs accurately and efficiently, given appropriately chosen protocols. We explore the sensitivity of prediction accuracy and adaptive convergence to a number of different parameters and analysis choices, and discuss some possible improvements of the algorithm for the future.

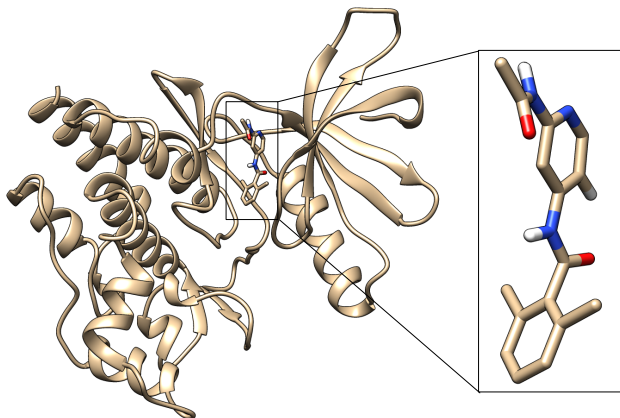


Figure 1: Illustration of the structure of Tyk2 (PDBid: 4GIH) used for all calculations. Tyk2 is shown in complex with ligand **ejm_31** in Table 1.²⁹

Methods

The expanded ensemble method

Consider a set of N thermodynamic ensembles parameterized by λ , where λ ranges between 0 and 1. For a set of λ_i values indexed by $i = 1, \dots, N$, each ensemble is defined by a reduced potential energy function $u_i(x) = \beta U(x|\lambda_i)$, where $\beta = (k_B T)^{-1}$. The goal of the expanded ensemble (EE) method is to use Monte Carlo sampling to perform a random walk in λ -space, throughout a single simulation, where all thermodynamic states defined by different values of λ are uniformly sampled, or alternately, satisfy some other desired distribution as a function of λ . This is achieved through the use of configuration-independent bias potentials \tilde{f}_i , that modify each potential energy function as $u'_i(x) = u_i(x) - \tilde{f}_i$. The values of \tilde{f}_i are adjusted on-the-fly, and adaptively refined to achieve, in this case, equal visitation of each of the N states. If the states are all visited equally, then the difference in the biases $\Delta\tilde{f}_{ij} = \tilde{f}_j - \tilde{f}_i$ are equal to the true difference in free energies Δf_{ij} . The value $\Delta f_{1N} = f_N - f_1$ is the change in the free energy over the entire alchemical transformation.

In this study, to adaptively refine estimates of \tilde{f}_i , the Wang-Landau (WL) flat-histogram algorithm is used.³¹ This algorithm periodically evaluates a histogram of values h_i storing the number of times state i is visited. At each iteration t , the histogram and the free energy

estimates \tilde{f}_i are updated:

$$h_i^{(t+1)} \leftarrow h_i^{(t)} + 1 \quad (1)$$

$$\tilde{f}_i^{(t+1)} \leftarrow \tilde{f}_i^{(t)} - \delta \quad (2)$$

The quantity δ is called in this study Wang-Landau (WL) increment, and it is initially set to a large value (for example, $10 k_B T$). This has the effect of penalizing visits to state i so that subsequent MC moves to other states will more likely be accepted. When the histogram is sufficiently flat, the WL increment is scaled by a factor $\alpha < 1$, and the histogram counts h_i are reset to zero. The histogram is sufficiently flat if the ratio N_{ratio} of all histogram values h_i to the mean value $\bar{h} = \frac{1}{N} \sum_i h_i$ is sufficiently close to 1. This is determined by ensuring that both $N_{\text{ratio}} > \eta$ and $1/N_{\text{ratio}} > \eta$ for all values of i , for some value of η (for example, 0.8) which we will call the Wang-Landau (WL) ratio.³²⁻³⁴

Relative binding free energy calculations

In RBFE calculations, two separate EE simulations are needed to predict relative free energy of binding, $\Delta\Delta G$ for a pair of ligands L and L*. In one simulation, the ligand L is alchemically transformed to L* in aqueous solution to obtain ΔG_L . In the other simulation, a receptor-bound ligand RL is transformed to RL* to obtain ΔG_{RL} . The RBFE can then be obtained as $\Delta\Delta G = \Delta G_{RL} - \Delta G_L$ (Figure 2).

Preparation of structures and force field parameters for Tyk2 and its inhibitors

Structure preparation and initial coordinates. The Tyk2 dataset is part of the Schrödinger’s “JACS set”⁵ and was used in several RBFE benchmark studies.^{5,6,35,36} The receptor model is based on the X-ray structure of PDB ID 4GIH (resolution 2.00 Å, OpenEye Iridium score 0.5, highly trustworthy).²⁹ The preparation and initial coordinates of the protein are the same as in Gapsys et al.⁶, where it was prepared with the Protein Prepara-

tion Wizard³⁷ using default settings: missing atoms, sidechains, and loops were modelled, protein protonation states were assigned with PROPKA at pH 7.0, the hydrogen bonding network was optimized. To relieve local clashes, a restrained minimization was performed with a 0.5 Å heavy-atom RMSD displacement cut-off. The inhibitors were modelled in their neutral form according to their protonation state at pH 7. For the construction of their initial coordinates, the coordinates of the crystallographic ligand **ejm_46** (see Table 1) were used and all other ligands were flexibly aligned to the reference ligand for improving the 3D-overlay of their Bemis-Murcko scaffolds.

Generation of Force Field Parameters. The receptor was parameterized using the AMBER ff99sb*ILDN force field parameters^{38–40} using the GROMACS gmx tool `pdb2gmx` and the AMBER ff99sb*ILDN parameter files available in the pmx distribution.⁴¹

The ligands were parameterized and prepared using a workflow⁴² based on the Open Force Field toolkit,⁴³ the pmx toolkit^{41,44} and GROMACS gmx program suite. The parameter set employed was the Open Force Field version 1.0.0 (codenamed “Parsley”).⁴⁵ Hybrid structures and topologies for the ligand pairs were generated using pmx^{41,44} following a single topology approach. The workflow establishes a mapping between atoms of the two ligands based on the maximum common substructure and conformational alignment. Polar hydrogens are not mapped to each other to decrease the influence of the perturbation on the hydrogen bond network. The mass of atoms in the non-interacting state (“dummies”) was set to 12u. Covalent force field parameters between atoms in the dummy state are not changed during the perturbation. The mappings are illustrated in the Supplementary Information Figure S1.

Simulation Boxes. The following system preparation steps were performed with GROMACS gmx tools `editconf`, `solvate`, and `genion`. The ligand (L) or the ligand-bound receptor (RL) were placed into rhombic dodecahedral boxes which fits the solute plus 1.5 nm buffer distance to the box walls. They were filled with TIP3P explicit water solvent

model. Sodium and chloride ions were added to neutralize the charge of the bare receptor ($-3e$) and to account for physiological ionic strength (150 mM NaCl). Solvated ligand systems (L) contained approximately 6000 (between 5630 to 6521) total comprising one (hybrid) ligand molecule, approximately 2000 (between 1861 and 2155) TIP3P solvent molecules, and approximately 5 sodium and 5 chloride ions. The diameter of the boxes (shortest distance between opposite box walls) was 4.4 nm. Solvated ligand-bound receptor systems (RL) comprised approximately 62k (62272–62295) total atoms of the receptor, one (hybrid) ligand molecule, approximately 19k TIP3P solvent molecules, 59 sodium ions and 56 chloride ions. The RL boxes had a diameter of 9.6 nm.

Molecular simulation protocol

Simulations were performed on the Owlsnest and CB2RR High-Performance Computing clusters at Temple University, and TACC Stampede (XSEDE). Molecular dynamics production runs were performed using the expanded ensemble functionality of GROMACS 5.1.4⁴⁶ (GROMACS/EE). Solvated systems (L and RL) were energy minimized using GROMACS with 50,000 steepest descent steps. Equilibration was performed Verlet integration was performed in the NPT ensemble at 300K using a 1 fs timestep. For each system, L or RL, 2.5 ns NPT equilibrium simulation was generated. The pressure was kept at 1 bar using the Parrinello-Rahman barostat. Long-range electrostatic interactions were handled by Particle Mesh Ewald (PME).

EE perturbation calculations were performed at 300 K in the NVT ensemble using Verlet integration with a 1 fs timestep and a velocity-rescaling thermostat. Long-range electrostatics were modeled using PME, and long-range dispersion correction was used.

GROMACS/EE parameters. The Metropolized Gibbs algorithm³² implemented in GROMACS/EE was used to perform Monte Carlo sampling, with moves proposed every 500 time steps. This method proposes moves from the current state i to all states $j \neq i$ with Gibbs

probability $\exp(-u_j)/\sum_k \exp(-u_k)$, and with a rejection step to satisfy detailed balance. Metropolized Gibbs sampling has been proven to enhance the mixing rate.³² The initial WL increment was set to $10.0 k_B T$. The WL free energy estimates \tilde{f}_i were updated every Monte Carlo move. Two values were explored for the WL scaling factor α : 0.8 (default) and 0.5. Modification of weights was set to be discontinued when the WL increment reached a value of $10^{-5} k_B T$, with the assumption that the weights are sufficiently close to equilibrium at that point, however, none of the simulations in this study reach this limit.

Data for free energy estimation was collected after the WL increment fell below $0.01 k_B T$. We chose this threshold based on the empirical observation that free energy estimates begin to converge beyond this value (see Results). Final free energy estimates were made by averaging the values of $\Delta\tilde{f}_{1N} = \tilde{f}_N - \tilde{f}_1$ collected after the WL increment fell below $0.01 k_B T$.⁴⁷ Our default protocol was to collect an aggregate simulation time of 400 ns for each solvated ligand transformation ($L \rightarrow L^*$), and 100 ns for each receptor-bound complex ($RL \rightarrow RL^*$). We chose these trajectory length to achieve $\sim 80\%$ of the sampling with a WL increment below $0.01 k_B T$.

The modified potential used for each alchemical intermediate $i = 1, \dots, N$ is $U_{\lambda_i}(x) = U_0(x) + (1 - \lambda_i)U_L(x|\lambda_i) + \lambda_i U_{L^*}(x|\lambda_i)$, where λ_i ranges from $\lambda_1 = 0$ to $\lambda_N = 1$. The $U_0(x)$ term represents potential energy terms not coupled to the ligand, while $U_L(x|\lambda_i)$ and $U_{L^*}(x|\lambda_i)$ are potentials energy terms coupled to ligands L and L^* , respectively, which depend on λ_i through the use of soft-core potentials (`sc-alpha` = 0.5, `sc-power` = 1, `sc-sigma` = 0.3).

Several numbers of alchemical intermediates were chosen for comparison: 21, 54 and 109. Although using smaller numbers of intermediates is more efficient if the overlap between states is moderate (see Results), our default protocol used 109 intermediates, chosen to avoid possible convergence problems due to poor thermodynamic overlap, which are difficult to predict *a priori*. We elaborate on this issue in the Results and Discussion sections. Simulations with 21 intermediates had λ_i values spaced linearly in the interval $[0, 1]$ in increments

of 0.05. Simulations with 54 intermediates had 14, 4, 15, 6 and 15 λ_i values linearly spaced in the intervals $[0, 0.1]$, $[0.1, 0.3]$, $[0.3, 0.6]$, $[0.6, 0.9]$, and $[0.9, 1.0]$, respectively. Simulations with 109 intermediates had 5, 98, and 6 λ_i values linearly spaced in the intervals $[0, 0.01]$, $[0.01, 0.99]$, and $[0.99, 1.0]$, respectively (Figure S2). These schedules were arrived at by trial and error, i.e. in previous tests they were found to result in reasonable acceptance. For all schedules used in this study, we find traversal to all intermediates for RL and L. Examples of λ -trajectories for L and RL simulations are shown in Figures S3 and S4, respectively. Empirical transition matrices for ligand-only (L) simulations are shown in Figure S5. While we did not attempt to further optimize the schedule of λ_i values, we note that several good algorithms exist for this purpose.⁴⁸

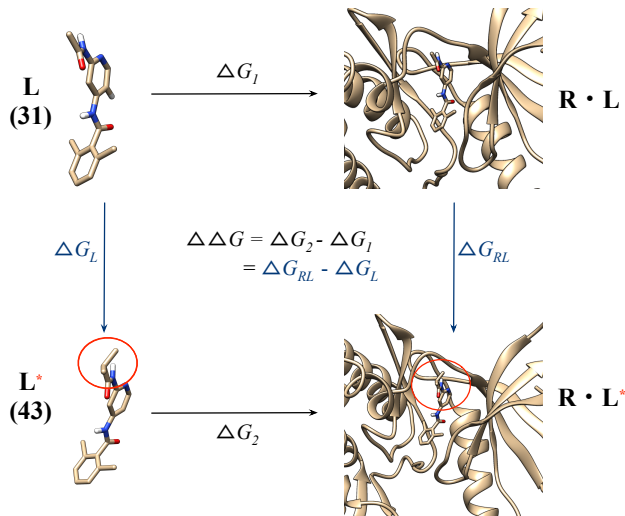
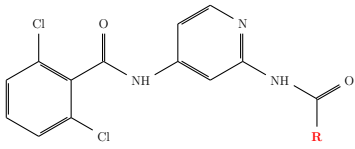


Figure 2: Thermodynamic cycle for relative binding free energy (RBFE) calculation. The RBFE difference between two similar ligands $L \rightarrow L^*$ (**ejm_31** and **ejm_43** in Table 1, respectively) is $\Delta\Delta G = \Delta G_2 - \Delta G_1$, which can be computed as $\Delta G_{RL} - \Delta G_L$.

Table 1: Twenty-four pairs of Tyk2 inhibitors. The numbering of the inhibitors corresponds to the numbering in References.^{29,30} The naming convention “ejm” denotes they are from Reference²⁹ and “jmc” denoted they are from Reference.³⁰



#	R	#	R*
ejm_31	CH ₃	ejm_43	CH(CH ₃) ₂
		ejm_45	CH ₂ –cyclopropyl
		ejm_46	Cyclopropyl
		ejm_48	Cyclopentyl
		jmc_28	methylcyclopropyl
ejm_42	CH ₂ CH ₃	ejm_48	Cyclopentyl
		ejm_54	NHCH ₂ CH ₃
		ejm_55	OCH ₃
ejm_43	CH(CH ₃) ₂	ejm_55	OCH ₃
ejm_44	C(CH ₃) ₃	ejm_42	CH ₂ CH ₃
		ejm_55	OCH ₃
ejm_45	CH ₂ –cyclopropyl	ejm_42	CH ₂ CH ₃
ejm_47	Cyclobutyl	ejm_31	CH ₃
		ejm_55	OCH ₃
ejm_49	Phenyl	ejm_31	CH ₃
		ejm_50	CH ₂ OH
ejm_50	CH ₂ OH	ejm_42	CH ₂ CH ₃
ejm_55	OCH ₃	ejm_54	NHCH ₂ CH ₃
jmc_23	fluorocyclopropyl	ejm_46	Cyclopropyl
		ejm_55	OCH ₃
		jmc_27	chlorocyclopropyl
		jmc_30	cyanocyclopropyl
jmc_28	methylcyclopropyl	jmc_27	chlorocyclopropyl
		jmc_30	cyanocyclopropyl

Results

EE convergence times vary with WL increment scaling and numbers of alchemical intermediates.

As an illustrative example, we show how the EE algorithm converges for L and RL calculations of the ejm_31 → ejm_45 alchemical transformation (Figure 3).

Inspection of the instantaneous estimates of free energy estimates for ΔG_L and ΔG_{RL} over time show occasional sudden deviations and corrections that occur as the ligand undergoes slow conformational changes. In the example (Figure 3), we can trace this behavior to slow interconversion between torsional states, which temporarily “switches” the system to a new effective free energy landscape (Figure S6). This behavior is observed in all simulations, regardless of the number of alchemical intermediates. Although enhanced sampling over these slow sampling barriers is not specifically addressed by the EE algorithm, we nevertheless observe adequate convergence of RBFE calculations for all of the systems considered in this study. Using fewer alchemical intermediates can generally converge EE estimates faster, because the flat-histogram criteria can be achieved in a smaller amount of simulation time.

Another parameter that affects convergence is the WL scaling factor. We performed tests on seven different alchemical transformations (the first seven pairs of inhibitors in Table 1), using two different values for the scaling factor: 0.5 (more aggressive), versus 0.8 (the GROMACS default). These tests used 400 ns simulation data used for each $L \rightarrow L^*$ transformation, and 150 ns for each $RL \rightarrow RL^*$ transformation. The only exception was **ejm_31** (RL) \rightarrow **jmc_28** (RL*), which required ~ 190 ns to converge.

The results show that while the accuracy of the predicted $\Delta\Delta G$ values using a scaling factor of 0.5 is statistically indistinguishable from results using 0.8 (Figure 4), using a scaling factor of 0.5 decreases the convergence time (the time it takes for the WL increment to reach $0.01\ kT$) by about a factor of 2, with ligand-only simulations converging within an average time of 52 ns. Interestingly, despite the more complex protein environment, receptor-ligand simulations converged within an shorter average time of 36 ns (Figure 4B and 4D). These results suggest —at least for these RBFE calculations and with many intermediates —that more aggressive WL scaling may be more efficient.

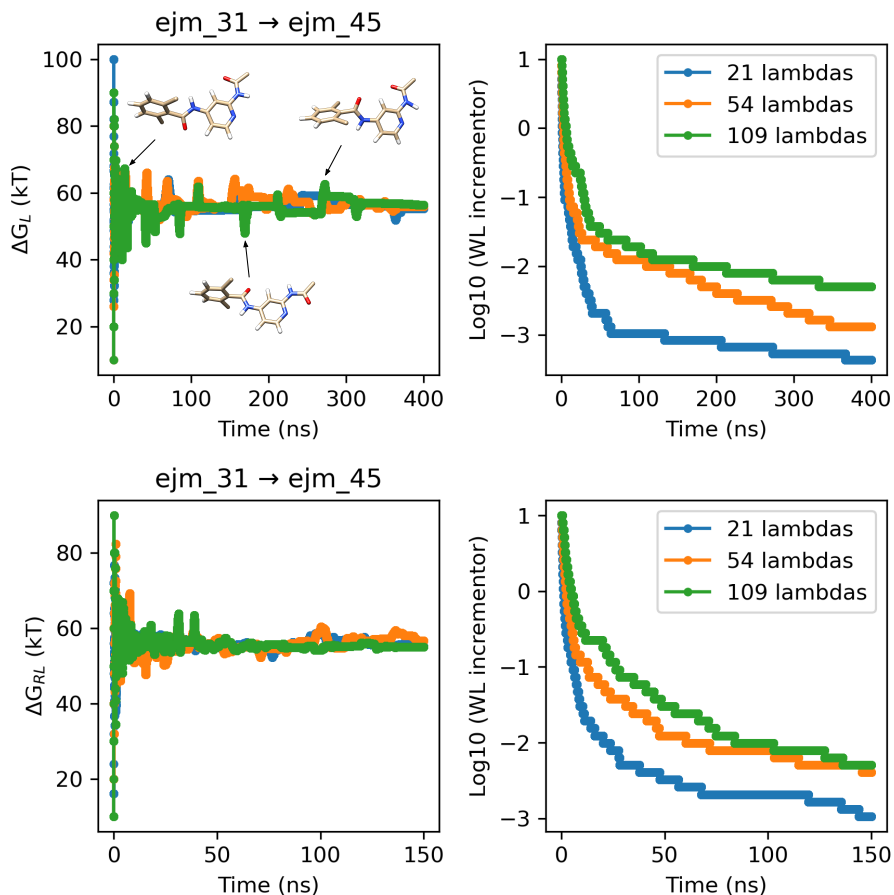


Figure 3: Convergence of EE free energy estimates vary with different numbers of alchemical intermediates. Estimates of ΔG_L (top left) and ΔG_{RL} (bottom left) over time for the ejm_31 \rightarrow ejm_45 transformation converge on the ~ 100 ns timescale, with occasional deviations that occur due to slow conformational transitions of the ligand. The WL increment in EE simulations for L (top right) and RL (bottom right) scales more quickly with fewer numbers of intermediates (shown here for a scaling factor of 0.8) .

Accurate predictions of RBFs for Tyk2 inhibitors

To test the convergence and performance of our EE protocol, we calculated RBF estimates from 400 ns of ligand-only trajectories, and 100 ns of receptor-ligand trajectories, for all 24 alchemical transformations of Tyk2 ligands. Informed by the above results, we used a WL scaling factor of 0.5 in all simulations. The predictions agree well with the experimental values, achieving an RMSE of 0.95 ± 0.17 kcal mol $^{-1}$ and MUE of 0.72 ± 0.13 kcal mol $^{-1}$ (RMSE and MUE uncertainties from 1000 bootstrapped samples of the set of ligands),

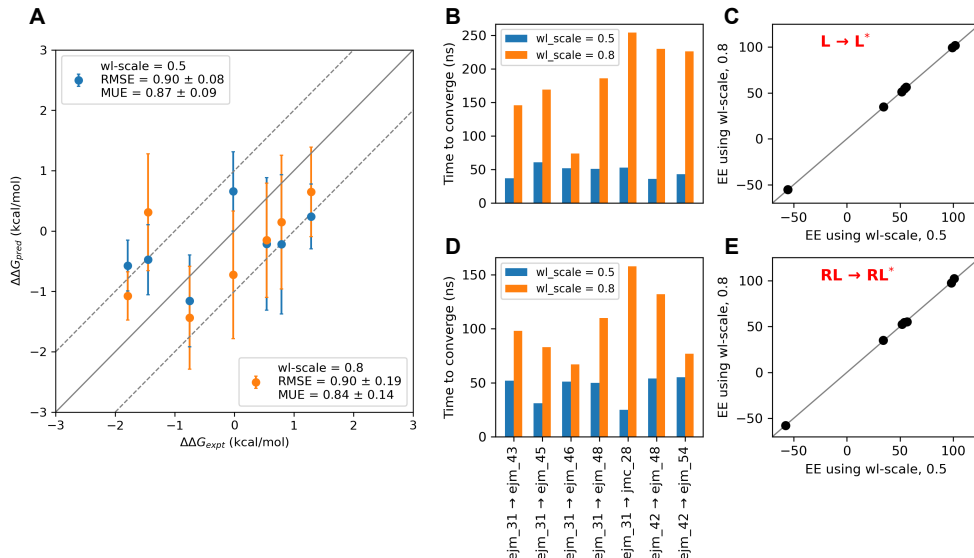


Figure 4: Comparisons of EE estimates made using a WL scaling factor of 0.5 vs. 0.8. Relative binding free estimates $\Delta\Delta G = \Delta G_{RL} - \Delta G_L$ were performed for seven alchemical transformations (the first seven in Table 1), using 400 ns of ligand-only trajectory data, and 150 ns of receptor ligand trajectory data. (A) Comparison of experiment and predicted $\Delta\Delta G$ values for WL scaling factors of 0.5 vs. 0.8. (B) Convergence times and (C) comparison of ΔG_L estimates for WL scaling factors of 0.5 vs. 0.8. (D) Convergence times and (E) comparison of ΔG_{RL} estimates for WL scaling factors of 0.5 vs. 0.8.

respectively.

The convergence profiles of the simulations, however, suggest that shorter trajectories might suffice for making accurate predictions. We calculated “normalized” convergence profiles of ΔG_L and ΔG_{RL} estimates as a function of simulation time T , as

$$\frac{1}{\Delta \tilde{f}_{1N}(T)} \left(\frac{1}{T} \int_0^T \Delta \tilde{f}_{1N}(t) dt \right) \quad (3)$$

The normalized profiles all converge to unity, enabling objective comparison. The profiles appear to equilibrate by around 150 ns (Figure 5). On average, ligand-only systems reach WL increments of $0.01 kT$ in about 100 ns (Figure 5B), while receptor-ligand systems reach this value on average in about 40 ns (Figure 5D).

To assess how the length of EE trajectory affects the accuracy of the predictions, we calculated RBE estimates for ligand-only trajectory lengths of 100, 150, 200, 250, 300, 350

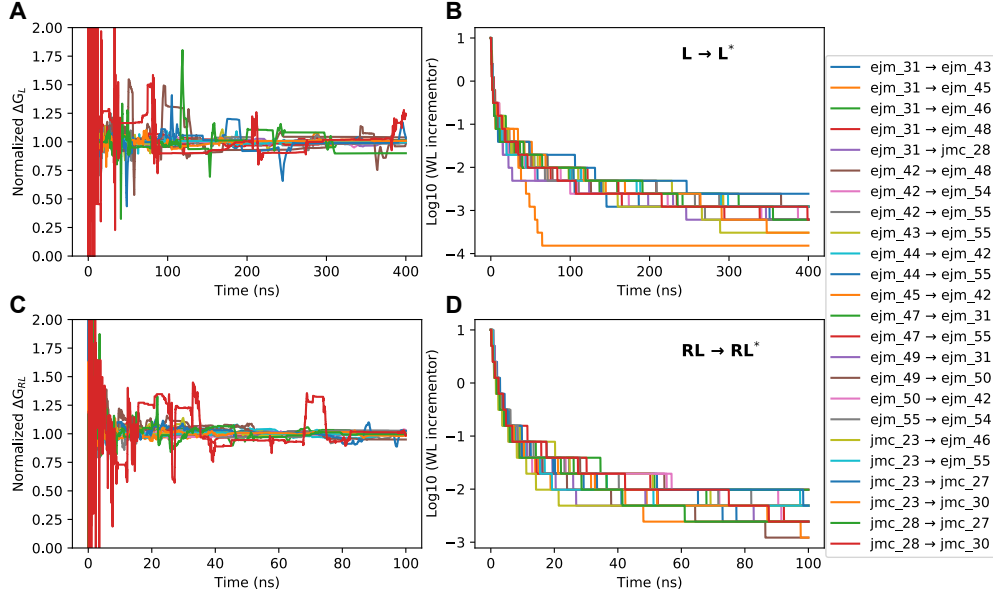


Figure 5: Traces of normalized free energy estimates show similar convergence profiles for all alchemical transformations. Free energy estimates over time, normalized by their trajectory-averaged values after convergence (when the WL increment decreases below $0.01 kT$) result in a series of profiles that all converge to unity, for the purposes of comparison. Normalized ΔG_L (A) and ΔG_{RL} (C) estimates over time, shown similar convergence behavior. The \log_{10} of the WL increment is shown over time for ΔG_L (B) and ΔG_{RL} (D). Different colors represent different systems.

and 400 ns, while using 100 ns of receptor-ligand trajectory data. We find that using 100 ns of trajectory data increases the RMSE to $0.98 \text{ kcal mol}^{-1}$ and MUE to about $0.77 \text{ kcal mol}^{-1}$, but beyond this, estimates approach those using all 400 ns of the trajectory data (Figure 6, blue lines).

When making free energy estimates from simulation trajectory data, it is imperative to properly consider the extent of time correlation in the trajectory to make accurate estimates of uncertainty.^{49,50} Since EE methods make inherently history-dependent estimates, subsampling trajectory data to remove time correlation may be an important consideration.

To determine a statistically optimal subsampling interval, we computed the autocorrelation time τ_c of the estimate $\Delta \tilde{f}_{1N}(t)$ as $\tau_c = \int_{T_0}^T g(\tau) d\tau$ where

$$g(\tau) = \frac{\frac{1}{T-T_0-\tau} \int_{T_0}^{T-\tau} (\Delta \tilde{f}_{1N}(t+\tau) - \langle \Delta \tilde{f}_{1N} \rangle) (\Delta \tilde{f}_{1N}(t) - \langle \Delta \tilde{f}_{1N} \rangle) dt}{\frac{1}{T-T_0} \int_{T_0}^T (\Delta \tilde{f}_{1N}(t) - \langle \Delta \tilde{f}_{1N} \rangle)^2 dt}. \quad (4)$$

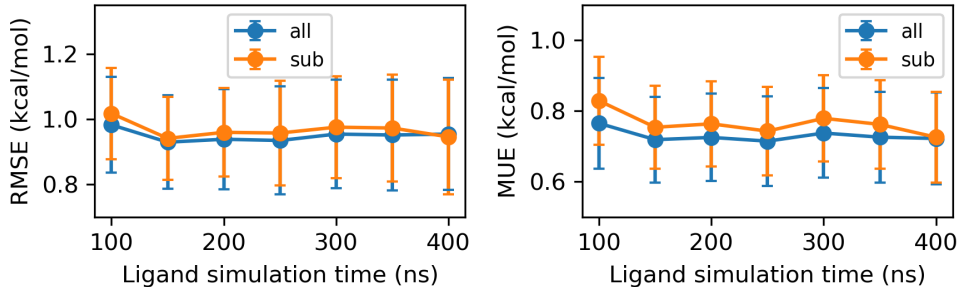


Figure 6: Accuracy of predicted RBFs for Tyk2 inhibitors versus ligand-only (L) simulation time. Shown are the RMSE (left) and MUE (right) using all trajectory data after convergence (blue line) and trajectory data subsampled to remove time correlation (orange line). In all calculations, 100 ns of receptor-ligand trajectory data is used.

In this formula, T is the trajectory length, T_0 is the time at which data is first recorded, and $\langle \tilde{f}_N \rangle$ is the time-averaged value of all recorded \tilde{f}_N values. The trajectory data is then subsampled in intervals of $2n(\tau_c) + 1$ steps, where $n(\tau_c)$ is the number of samples in a trajectory of length τ_c . For our protocol, $n(\tau_c) = \tau_c / (1 \text{ ps})$, which results in typical subsampling intervals of 10-20k steps (Table S1).

As expected, we find that while subsampling increases the values of computed uncertainties, making these uncertainties more accurate because time-correlated data is removed, it does not statistically impact the overall accuracy of estimated RBFs (Figure 7). Using 150 ns of ligand-only trajectory data and 100 ns of receptor-ligand trajectory data, we find that across all 24 alchemical transformations of Tyk2 inhibitors, using unsubsampled data results in an RMSE of $0.93 \pm 0.14 \text{ kcal mol}^{-1}$ and MUE of $0.72 \pm 0.12 \text{ kcal mol}^{-1}$, while subsampled data results in an RMSE of $0.94 \pm 0.13 \text{ kcal mol}^{-1}$ and MUE of $0.75 \pm 0.12 \text{ kcal mol}^{-1}$.

We also checked to see whether there was any significant conformational change in the receptor during RL simulations. To assess this, we computed the RMSD of the protein backbone coordinates with those of the first frame, over time. We find that all RMSD values are less than 0.3 nm, suggesting the protein receptor is relatively stable throughout the simulation (Figure S8).

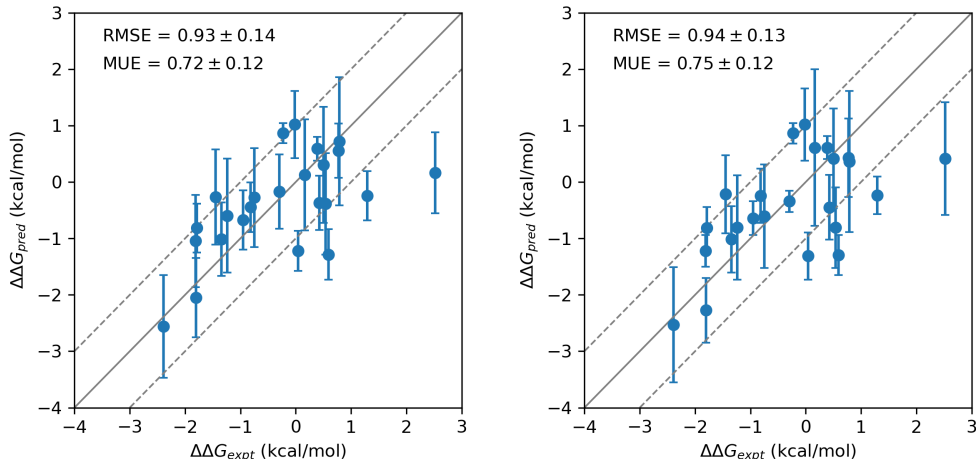


Figure 7: Comparison of predicted and experimental RBFEs for Tyk2 ligands, using all trajectory data taken after convergence (left), and trajectory data subsampled to remove time-correlation (right). In all calculations, 150 ns of ligand-only trajectory data and 100 ns of receptor-ligand data are used. The uncertainties of RMSE and MUE were calculated from a bootstrap analysis of 1000 trials over the set of 24 transformations.

We also assessed whether backbone restraints had any effect on the convergence time or the accuracy of the RBFE predictions. Therefore, we repeated all 24 $\text{RL} \rightarrow \text{RL}^*$ transformations using 100 ns trajectories, in the presence of backbone restraints with a force constant of $1000 \text{ kJ mol}^{-1}\text{nm}^{-2}$ (Figure S7). While the calculated free energies for restrained vs. unrestrained simulations are statistically indistinguishable, we find that backbone restraints decrease the convergence time, from an average convergence time of 32.8 ns (unrestrained), to an average convergence time of 20.2 ns (restrained).

Averaging multiple independent EE estimates is more accurate than a single estimate

There are several reasons why independent EE simulations will give varying estimates of free energies. The first is uncertainty due to finite sampling as occurs for all sampling algorithms. The second is that the Wang-Landau flat-histogram algorithm, as implemented in GROMACS with geometric scaling of the WL increment, can result in saturation of the error, where the WL increment may become vanishingly small but the biases nevertheless

converged to the incorrect value in the limit of infinite sampling.^{51,52} While there are ways to avoid saturation of the error, such as using “1/ t ”-scaling of the WL increment,⁵¹ this algorithm is not implemented in GROMACS, and we do not pursue the issue here.

To examine the reproducibility of RBFE estimates from EE, we performed five independent EE simulations for each of the 24 alchemical transformations of Tyk2 inhibitors. We found that the standard deviation of the estimates varies between 0.14 and 0.83 kcal mol⁻¹, suggesting that EE estimates for at least these systems are reproducible (Figure 8).

An interesting question is whether the uncertainty estimates σ_i from each single EE simulation (estimated from the variation in $\Delta\tilde{f}_{1N}$) are able to capture the run-to-run variation in $\Delta\Delta G$ estimates observed for multiple calculations. If so, such uncertainties could be trusted to provide estimates of the overall uncertainty in EE predictions. To see if this was the case, we compared the uncertainties in $\Delta\Delta G$ estimated using error propagation from single replicas ($\sigma_{PE} = \sqrt{\frac{1}{N} \sum_i \sigma_i^2}$) to the uncertainties estimated from the standard error of the mean across the five replicas, computed as the corrected sample standard deviation, $\sigma_{SEM} = \sqrt{\frac{1}{N-1} \sum_{i=1}^N (\Delta\Delta G_i - \overline{\Delta\Delta G})^2}$, where $N = 5$, and $\overline{\Delta\Delta G}$ is the mean of $\Delta\Delta G$ estimates across all replicas (Figure S10). We find that, across the 24 transformations, the computed uncertainties σ_{PE} tend to be slightly larger than those estimated from the standard error of the mean, σ_{SEM} , by about 0.22 kcal/mol (Figure S11). This suggests that the uncertainty estimates from single EE replicas do a reasonable if not perfect job of estimating the overall uncertainty in $\Delta\Delta G$ estimates, and can be prudently used for this purpose.

If we make the assumption that the force field is sufficiently accurate, then improved sampling will improve accuracy. We find in this case that prediction accuracy is improved when RBFE estimates are made by averaging the results of multiple replicates. Averaging five replicas achieves an RMSE of 0.83 ± 0.14 kcal mol⁻¹ and a MUE of 0.66 ± 0.11 kcal mol⁻¹; these values are smaller than any of the single-replicate EE estimates. (Figure 9)

To assess the statistical significance of this change in accuracy, we computed a non-parametric null distribution for the RMSE and MUE statistics in each of the above cases,

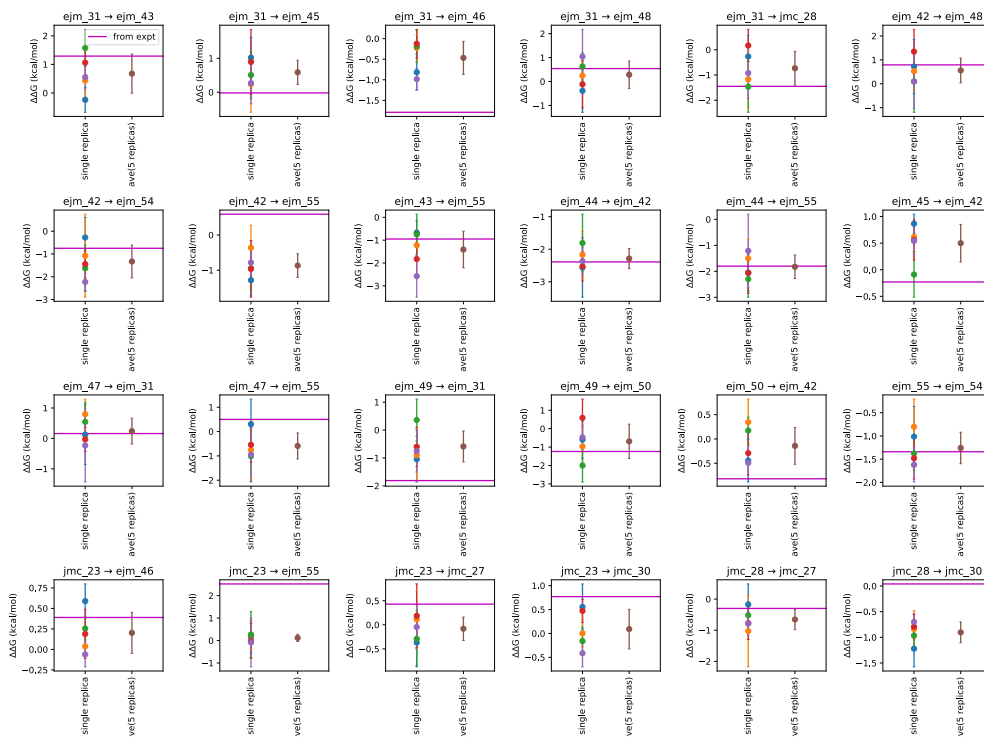


Figure 8: Comparisons of RBEFs calculated from five independent replicates, for all 24 alchemical transformations of Tyk2 inhibitors. Shown are five independent estimates and their uncertainties (blue, orange, green, red, purple circles), and the average of five replicas (brown circles). The magenta horizontal lines represent the experimental values.

using 1000 permutations of the experimental labels. In all cases, we find p -values less or equal to 0.006 (Figure S9), which suggests that these improved accuracy results are statistically significant.

To further explore the apparent accuracy gains from averaging multiple replicas, we computed predicted RBEFs for a test set of 16 Tyk2 inhibitor alchemical transformations (those with ligand-only simulations that converged within 100 ns), using n replicas of length $(400 \text{ ns})/n$ for L simulations and $(160 \text{ ns})/n$ for RL simulation, for $n = 1, 2, 3$ and 4 (Figure 10). The results of these calculations suggests that, given a fixed amount of available simulation time, lower RMSE and MUE can be achieved by averaging the results of n replicas, versus performing one simulation n times as long, in the case when trajectories are long enough to reach convergence.

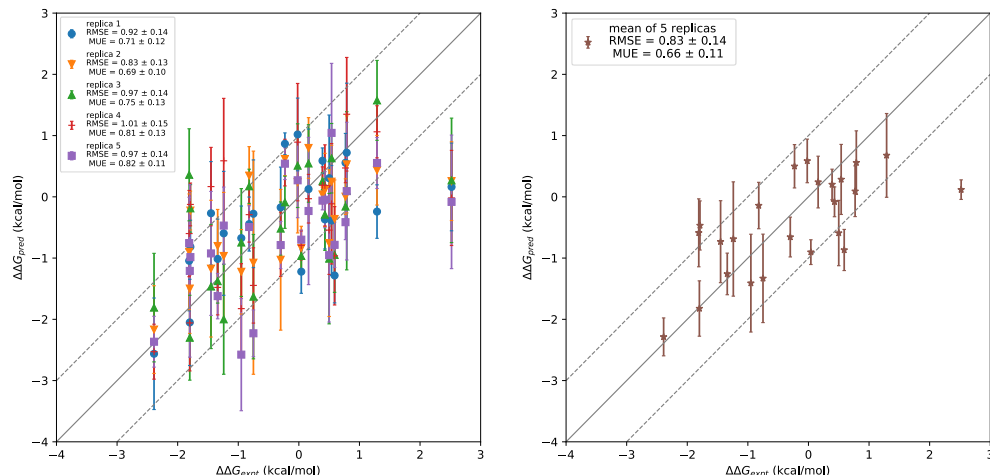


Figure 9: Comparison of predicted versus experimental RBFES for all twenty-four transformations, using five independent single-replica EE estimates (blue, orange, green, red, purple colored markers in left panel), and the mean of five replicas (brown stars in right panel). Averaging multiple replicas yields RBFES estimates within $1.0 \text{ kcal mol}^{-1}$ of the experimental value for most transformations. The uncertainties of calculated RMSE and MUE were estimated from a bootstrap analysis of 1000 trials over the set of 24 transformations.

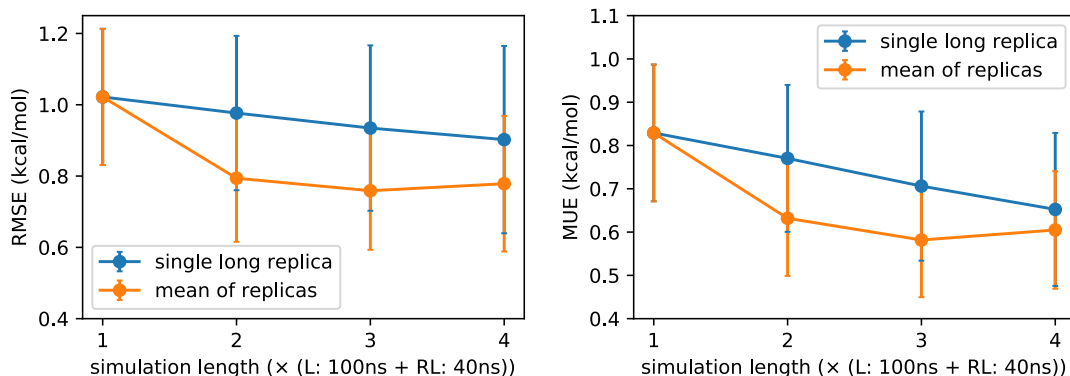


Figure 10: Accuracy of predicted RBFES for 16 Tyk2 inhibitors versus simulation length. Shown are the RMSE (left) and MUE (right) calculated using single long L and RL trajectories (blue) versus multiple replicas (orange). In each case, the total simulation time of the single long trajectories is equal to the sum of the lengths of multiple replicas: 400 ns of L trajectory data, and 160 ns of RL trajectory data. The uncertainties of RMSE and MUE were calculated from an analysis of 1000 bootstrap trials over the set of 24 transformations.

Discussion

In this work, we have shown that expanded ensemble methods, coupled with the latest Open Force Field Initiative force field, can accurately predict relative binding free energy

predictions for Tyk2 inhibitors. The main computational expense of these calculations is the simulation time required for convergence of the relative binding free energies, which depends on the number of alchemical intermediates used. Compared to a typical FEP+ protocol⁵ using 20 alchemical intermediates and 5 ns of sampling per intermediate (100 ns of total simulation time each for L and RL), the EE approach used here is only slightly more expensive, but can be run in fully independent runs. In many situations, we expect the advantages of performing such simulations without parallelizations across replicas may outweigh the disadvantages of computational cost.

Most of the results we have presented in this work use a large number of alchemical intermediate (109), despite the fact that we have demonstrated that fewer alchemical intermediates (e.g. 21) leads to faster convergence. Why use so many intermediates? One reason is that it is very hard to predict the optimal schedule of λ_i values without performing preliminary simulations. In practice, using more λ values is simple and effective. A disadvantage to this strategy is the longer simulation time it takes to achieve flat histograms. As we show above, however, more aggressive WL increment scaling (using a scaling factor of 0.5 rather than 0.8) can help accelerate convergence in this case.

Another reason to use a large number of alchemical intermediates is to avoid the risk of sampling bottlenecks. When the MC acceptance probability between two intermediates becomes low enough, we have observed that a kind of hysteresis can develop, where the more time spent in a given intermediate, the less probable it is to make outgoing transitions. The net effect is that EE sampling appears to “get stuck” intermittently at particular values of λ_i , for increasingly longer periods of time. In future work we are studying this phenomenon and its relation to the saturation of the error that is known to occur with the current geometric scaling of the WL increment.^{51,52} It should be possible to determine conditions under which such hysteresis will occur, so steps can be taken to avoid it.

There are several future directions to explore that might help improve sampling and convergence beyond the EE approach we have pursued here. Algorithms such as the accel-

erated weight histogram (AWH) method¹⁷ and self-adjusted mixture sampling (SAMS),¹⁵ which modify biases for multiple intermediates simultaneously using more sophisticated estimators, may lead to better convergence. Versions of these algorithms have recently been implemented in GROMACS and can be readily applied. More frequent swapping attempts (every 50 steps, perhaps) might also lead to faster convergence. Moreover, it is clear that “hidden barriers”, which arise due to rare conformational transitions (see Figures 3 and S6), are responsible for slow convergence. Future work should focus on automatically identifying the conformational degrees of freedom corresponding to these slow motions, and employing novel alchemical intermediates to help overcome these barriers.

A key lesson from this study is that averaging EE free energy estimates over multiple short simulations gives more accurate predictions than a single long simulations, presumably due to better sampling. This finding suggests strong benefits for using EE to perform massively parallel virtual screening efforts on cloud computing and other distributed platforms on which tight-coupling algorithms like Hamiltonian replica exchange are unfeasible. EE approaches are feasible on such platforms, with the additional benefit that better accuracy may be achieved by performing more simulations with shorter trajectory lengths.

Conclusion

In this work we use an expanded ensemble (EE) method, implemented in GROMACS,⁴⁶ along with the current Open Force Field potential,⁴⁵ to accurately predict the relative binding free energies (RBFES) of twenty-four Tyk2 inhibitors (RMSE of 0.94 ± 0.13 kcal mol⁻¹ and MUE of 0.75 ± 0.12 kcal mol⁻¹). We found that EE convergence times can be accelerated by decreasing the number of alchemical intermediates (provided sufficient MC acceptance can be maintained) and by using a more aggressive Wang-Landau scaling factor of 0.5. We also find a statistically significant benefit in estimating RBFES as the average over multiple independent EE replicates. These results suggest that while EE methods may be currently

underutilized for RBFE estimation, they are poised to play a bigger role in virtual screening especially on large-scale cloud computing platforms.

Code and data availability

Input files, mdp files and examples of ligand and receptor-ligand EE simulations can be found on GitHub at https://github.com/Sizhang92190/RBFE_EE_TYK2

Acknowledgement

We thank Temple’s High Performance Computing team (especially Richard Berger) for their support with Owlsnest. S.Z. and V.A.V. were supported was supported by NIH 1R01GM123296. This research includes calculations carried out on Temple University’s HPC resources and thus was supported in part by the National Science Foundation through major research instrumentation grant number 1625061 and by the US Army Research Laboratory under contract number W911NF-16-2-0189. Computing resources were supported in part through NIH 1S10OD020095-01, and XSEDE allocation MCB200214 (TACC Stampede). MRS was supported by NSF OAC-1835720.

References

- (1) Chodera, J. D.; Mobley, D. L.; Shirts, M. R.; Dixon, R. W.; Branson, K.; Pande, V. S. Alchemical free energy methods for drug discovery: progress and challenges. *Current opinion in structural biology* **2011**, *21*, 150–160.
- (2) Aldeghi, M.; Heifetz, A.; Bodkin, M. J.; Knapp, S.; Biggin, P. C. Accurate calculation of the absolute free energy of binding for drug molecules. *Chemical science* **2016**, *7*, 207–218.

- (3) Song, L. F.; Merz Jr, K. M. Evolution of Alchemical Free Energy Methods in Drug Discovery. *Journal of Chemical Information and Modeling* **2020**, *60*, 5308–5318.
- (4) Mey, A. S.; Allen, B. K.; Macdonald, H. E. B.; Chodera, J. D.; Hahn, D. F.; Kuhn, M.; Michel, J.; Mobley, D. L.; Naden, L. N.; Prasad, S., et al. Best Practices for Alchemical Free Energy Calculations [Article v1. 0]. *Living Journal of Computational Molecular Science* **2020**, *2*, 18378.
- (5) Wang, L.; Wu, Y.; Deng, Y.; Kim, B.; Pierce, L.; Krilov, G.; Lupyan, D.; Robinson, S.; Dahlgren, M. K.; 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.
- (6) Gapsys, V.; Pérez-Benito, L.; Aldeghi, M.; Seeliger, D.; Van Vlijmen, H.; Tresadern, G.; de Groot, B. L. Large scale relative protein ligand binding affinities using non-equilibrium alchemy. *Chemical Science* **2020**, *11*, 1140–1152.
- (7) Cournia, Z.; Allen, B. K.; Beuming, T.; Pearlman, D. A.; Radak, B. K.; Sherman, W. Rigorous free energy simulations in virtual screening. *Journal of Chemical Information and Modeling* **2020**, *60*, 4153–4169.
- (8) Loeffler, H.; Michel, J.; Woods, C. FESetup: Automating Setup for Alchemical Free Energy Simulations. *Journal of Chemical Information and Modeling* **2015**, *55*, 2485–2490.
- (9) Abel, R.; Wang, L.; Harder, E. D.; Berne, B.; Friesner, R. A. Advancing drug discovery through enhanced free energy calculations. *Accounts of chemical research* **2017**, *50*, 1625–1632.
- (10) Gapsys, V.; de Groot, B. L. pmx Webserver: a user friendly interface for alchemistry. 2017.

- (11) Eastman, P.; Swails, J.; Chodera, J. D.; McGibbon, R. T.; Zhao, Y.; Beauchamp, K. A.; Wang, L.-P.; Simmonett, A. C.; Harrigan, M. P.; Stern, C. D., et al. OpenMM 7: Rapid development of high performance algorithms for molecular dynamics. *PLoS computational biology* **2017**, *13*, e1005659.
- (12) He, X.; Liu, S.; Lee, T.-S.; Ji, B.; Man, V. H.; York, D. M.; Wang, J. Fast, Accurate, and Reliable Protocols for Routine Calculations of Protein–Ligand Binding Affinities in Drug Design Projects Using AMBER GPU-TI with ff14SB/GAFF. *ACS omega* **2020**, *5*, 4611–4619.
- (13) Lee, T.-S.; Allen, B. K.; Giese, T. J.; Guo, Z.; Li, P.; Lin, C.; McGee Jr, T. D.; Pearlman, D. A.; Radak, B. K.; Tao, Y., et al. Alchemical binding free energy calculations in AMBER20: Advances and best practices for drug discovery. *Journal of Chemical Information and Modeling* **2020**, *60*, 5595–5623.
- (14) Lyubartsev, A.; Martsinovski, A.; Shevkunov, S.; Vorontsov-Velyaminov, P. New approach to Monte Carlo calculation of the free energy: Method of expanded ensembles. *The Journal of chemical physics* **1992**, *96*, 1776–1783.
- (15) Tan, Z. Optimally adjusted mixture sampling and locally weighted histogram analysis. *Journal of Computational and Graphical Statistics* **2017**, *26*, 54–65.
- (16) Ross, G. A.; Rustenburg, A. S.; Grinaway, P. B.; Fass, J.; Chodera, J. D. Biomolecular simulations under realistic macroscopic salt conditions. *The Journal of Physical Chemistry B* **2018**, *122*, 5466–5486.
- (17) Lindahl, V.; Lidmar, J.; Hess, B. Accelerated weight histogram method for exploring free energy landscapes. *The Journal of chemical physics* **2014**, *141*, 044110.
- (18) Fukunishi, H.; Watanabe, O.; Takada, S. On the Hamiltonian replica exchange method for efficient sampling of biomolecular systems: Application to protein structure prediction. *The Journal of chemical physics* **2002**, *116*, 9058–9067.

- (19) Jiang, W.; Roux, B. Free energy perturbation Hamiltonian replica-exchange molecular dynamics (FEP/H-REMD) for absolute ligand binding free energy calculations. *Journal of chemical theory and computation* **2010**, *6*, 2559–2565.
- (20) Shirts, M.; Pande, V. S. Screen savers of the world unite! *Science* **2000**, *290*, 1903–1904.
- (21) Zimmerman, M. I. et al. SARS-CoV-2 Simulations Go Exascale to Capture Spike Opening and Reveal Cryptic Pockets Across the Proteome. *bioRxiv* **2020**,
- (22) Balasubramanian, V.; Jensen, T.; Turilli, M.; Kasson, P.; Shirts, M.; Jha, S. Adaptive ensemble biomolecular applications at scale. *SN Computer Science* **2020**, *1*, 1–15.
- (23) Escobedo, F. A.; Martinez-Veracoechea, F. J. Optimized expanded ensembles for simulations involving molecular insertions and deletions. I. Closed systems. *The Journal of chemical physics* **2007**, *127*, 174103.
- (24) Escobedo, F. A.; Martinez-Veracoechea, F. J. Optimization of expanded ensemble methods. *The Journal of chemical physics* **2008**, *129*, 154107.
- (25) Martinez-Veracoechea, F. J.; Escobedo, F. A. Variance minimization of free energy estimates from optimized expanded ensembles. *The Journal of Physical Chemistry B* **2008**, *112*, 8120–8128.
- (26) Athènes, M.; Terrier, P. Estimating thermodynamic expectations and free energies in expanded ensemble simulations: Systematic variance reduction through conditioning. *The Journal of chemical physics* **2017**, *146*, 194101.
- (27) Monroe, J. I.; Shirts, M. R. Converging free energies of binding in cucurbit [7] uril and octa-acid host–guest systems from SAMPL4 using expanded ensemble simulations. *Journal of computer-aided molecular design* **2014**, *28*, 401–415.
- (28) Rizzi, A.; Jensen, T.; Slochower, D. R.; Aldeghi, M.; Gapsys, V.; Ntekoumes, D.; Bosisio, S.; Papadourakis, M.; Henriksen, N. M.; De Groot, B. L., et al. The SAMPL6

- SAMPLing challenge: assessing the reliability and efficiency of binding free energy calculations. *Journal of computer-aided molecular design* **2020**, *34*, 601–633.
- (29) Liang, J.; Tsui, V.; Van Abbema, A.; Bao, L.; Barrett, K.; Beresini, M.; Berezhkovskiy, L.; Blair, W. S.; Chang, C.; Driscoll, J., et al. Lead identification of novel and selective TYK2 inhibitors. *European journal of medicinal chemistry* **2013**, *67*, 175–187.
- (30) Liang, J.; van Abbema, A.; Balazs, M.; Barrett, K.; Berezhkovsky, L.; Blair, W.; Chang, C.; Delarosa, D.; DeVoss, J.; Driscoll, J., et al. Lead optimization of a 4-aminopyridine benzamide scaffold to identify potent, selective, and orally bioavailable TYK2 inhibitors. *Journal of medicinal chemistry* **2013**, *56*, 4521–4536.
- (31) Wang, F.; Landau, D. P. Efficient, multiple-range random walk algorithm to calculate the density of states. *Physical review letters* **2001**, *86*, 2050.
- (32) Chodera, J. D.; Shirts, M. R. Replica exchange and expanded ensemble simulations as Gibbs sampling: Simple improvements for enhanced mixing. *The Journal of chemical physics* **2011**, *135*, 194110.
- (33) Zwanzig, R. W. Erratum: High-Temperature Equation of State by a Perturbation Method. I. Nonpolar Gases. *The Journal of Chemical Physics* **1954**, *22*, 2099–2099.
- (34) Li, H.; Fajer, M.; Yang, W. Simulated scaling method for localized enhanced sampling and simultaneous “alchemical” free energy simulations: A general method for molecular mechanical, quantum mechanical, and quantum mechanical/molecular mechanical simulations. *The Journal of chemical physics* **2007**, *126*, 01B606.
- (35) Kuhn, M.; Firth-Clark, S.; Tosco, P.; Mey, A. S.; Mackey, M.; Michel, J. Assessment of binding affinity via alchemical free-energy calculations. *Journal of Chemical Information and Modeling* **2020**, *60*, 3120–3130.

- (36) Song, L. F.; Lee, T.-S.; Zhu, C.; York, D. M.; Merz Jr, K. M. Using AMBER18 for relative free energy calculations. *Journal of chemical information and modeling* **2019**, *59*, 3128–3135.
- (37) Sastry, G. M.; Adzhigirey, M.; Day, T.; Annabhimoju, R.; Sherman, W. Protein and ligand preparation: parameters, protocols, and influence on virtual screening enrichments. *Journal of computer-aided molecular design* **2013**, *27*, 221–234.
- (38) Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C. Comparison of multiple Amber force fields and development of improved protein backbone parameters. *Proteins: Structure, Function, and Bioinformatics* **2006**, *65*, 712–725.
- (39) Best, R. B.; Hummer, G. Optimized molecular dynamics force fields applied to the helix-coil transition of polypeptides. *The journal of physical chemistry B* **2009**, *113*, 9004–9015.
- (40) Lindorff-Larsen, K.; Piana, S.; Palmo, K.; Maragakis, P.; Klepeis, J. L.; Dror, R. O.; Shaw, D. E. Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins: Structure, Function, and Bioinformatics* **2010**, *78*, 1950–1958.
- (41) Gapsys, V.; Michielssens, S.; Seeliger, D.; de Groot, B. L. pmx: Automated protein structure and topology generation for alchemical perturbations. 2015.
- (42) Hahn, D. F. Pmx Workflow. 2020.
- (43) Wagner, J. et al. openforcefield/openff-toolkit: 0.9.1 Minor feature and bugfix release. 2021; <https://doi.org/10.5281/zenodo.4592294>.
- (44) Seeliger, D.; De Groot, B. L. Protein thermostability calculations using alchemical free energy simulations. *Biophysical journal* **2010**, *98*, 2309–2316.
- (45) Qiu, Y.; Smith, D.; Boothroyd, S.; Jang, H.; Wagner, J.; Bannan, C. C.; Gokey, T.;

- Lim, V. T.; Stern, C.; Rizzi, A., et al. Development and Benchmarking of Open Force Field v1. 0.0, the Parsley Small Molecule Force Field. **2020**,
- (46) Abraham, M. J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J. C.; Hess, B.; Lindahl, E. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **2015**, *1*, 19–25.
- (47) Fort, G.; Jourdain, B.; Kuhn, E.; Lelièvre, T.; Stoltz, G. Convergence of the Wang-Landau algorithm. *Mathematics of Computation* **2015**, *84*, 2297–2327.
- (48) Shenfeld, D. K.; Xu, H.; Eastwood, M. P.; Dror, R. O.; Shaw, D. E. Minimizing thermodynamic length to select intermediate states for free-energy calculations and replica-exchange simulations. *Physical Review E* **2009**, *80*, 046705.
- (49) Chodera, J. D.; Swope, W. C.; Pitera, J. W.; Seok, C.; Dill, K. A. Use of the weighted histogram analysis method for the analysis of simulated and parallel tempering simulations. *Journal of Chemical Theory and Computation* **2007**, *3*, 26–41.
- (50) Grossfield, A.; Patrone, P. N.; Roe, D. R.; Schultz, A. J.; Siderius, D. W.; Zuckerman, D. M. Best practices for quantification of uncertainty and sampling quality in molecular simulations [Article v1. 0]. *Living journal of computational molecular science* **2018**, *1*.
- (51) Belardinelli, R.; Pereyra, V. Wang-Landau algorithm: A theoretical analysis of the saturation of the error. *The Journal of chemical physics* **2007**, *127*, 184105.
- (52) Belardinelli, R. E.; Manzi, S.; Pereyra, V. D. Analysis of the convergence of the $1/t$ and Wang-Landau algorithms in the calculation of multidimensional integrals. *Phys. Rev. E* **2008**, *78*, 067701.