

TIES 20: Relative Binding Free Energy with a Flexible Superimposition Algorithm and Partial Ring Morphing

Mateusz K. Bieniek, Agastya P. Bhati, Shunzhou Wan, and Peter V. Coveney*

Cite This: *J. Chem. Theory Comput.* 2021, 17, 1250–1265

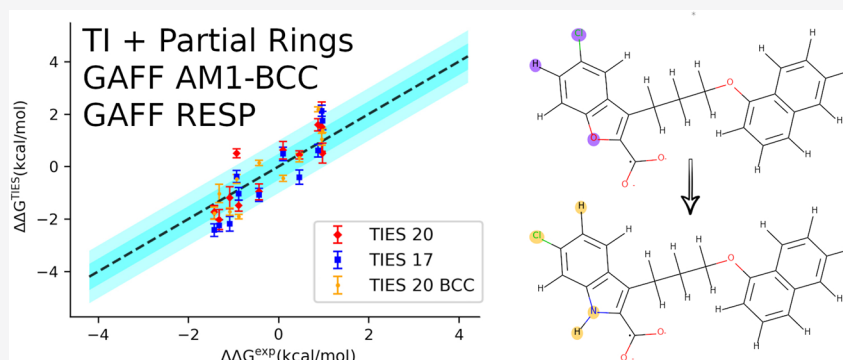
Read Online

ACCESS |

Metrics & More

Article Recommendations

Supporting Information



ABSTRACT: The TIES (Thermodynamic Integration with Enhanced Sampling) protocol is a formally exact alchemical approach in computational chemistry to the calculation of relative binding free energies. The validity of TIES relies on the correctness of matching atoms across compared pairs of ligands, laying the foundation for the transformation along an alchemical pathway. We implement a flexible topology superimposition algorithm which uses an exhaustive joint-traversal for computing the largest common component(s). The algorithm is employed to enable matching and morphing of partial rings in the TIES protocol along with a validation study using 55 transformations and five different proteins from our previous work. We find that TIES 20 with the RESP charge system, using the new superimposition algorithm, reproduces the previous results with mean unsigned error of 0.75 kcal/mol with respect to the experimental data. Enabling the morphing of partial rings decreases the size of the alchemical region in the dual-topology transformations resulting in a significant improvement in the prediction precision. We find that increasing the ensemble size from 5 to 20 replicas per λ window only has a minimal impact on the accuracy. However, the non-normal nature of the relative free energy distributions underscores the importance of ensemble simulation. We further compare the results with the AM1-BCC charge system and show that it improves agreement with the experimental data by slightly over 10%. This improvement is partly due to AM1-BCC affecting only the charges of the atoms local to the mutation, which translates to even fewer morphed atoms, consequently reducing issues with sampling and therefore ensemble averaging. TIES 20, in conjunction with the enablement of ring morphing, reduces the size of the alchemical region and significantly improves the precision of the predicted free energies.

1. INTRODUCTION

While the age of exascale computing is emerging, the field of drug discovery is increasingly using the maximum computational means available to complement its life-cycle.¹ This enormous computational power can aid in identifying new potential therapeutics, a process which involves evaluating a large number of compounds that not only are often sourced from databases of (approved) drugs^{2,3} and chemical libraries⁴ but also increasingly are generated virtually from the vast chemical space.⁵

Whereas physics based free energy computations have been historically prohibitively expensive, they are now becoming increasingly viable, bringing their application to not only the different stages of drug development, particularly hit to lead and lead optimization stages, but also increasingly virtual high throughput screening. These calculations can offset the large

costs associated with bringing novel drugs to market, while their combination of scalability and performance is particularly relevant in urgent situations such as that brought about by the COVID-19 pandemic. In addition to the computational power, there is a continuous improvement due to advances in the underlying force fields.^{6–9} Furthermore, promising initiatives such as the Open Force Field Initiative¹⁰ and blind experiments such as Drug Design Data Resource (D3R)

Received: November 10, 2020

Published: January 25, 2021



grand challenges are having a noticeable impact on progress in the field.¹¹

Relative binding free energy (RBF) calculations are one of the increasingly utilized physics based methods. In these calculations, one molecule is morphed into another in nonphysical, or “alchemical”, space. RBFs contain a plethora of postprocessing methods including thermodynamic integration,^{12,13} free energy perturbation using the Zwanzig formula,¹⁴ the Bennett acceptance ratio (BAR),¹⁵ or its variation called multistate BAR (MBAR).^{16,17} Further discussion is provided by Williams-Noonan et al. in a review of applied studies.¹⁸ It is worth pointing out here that, unlike FEP, TI has no formal constraint of being a perturbative method. Although the practice has been to only apply TI for studying relatively small transformations, in principle, it should be able to handle larger structural changes than are possible with FEP.

Notable software toolsets facilitating the calculation of RBF include, among others, the open source GROMACS-based pmx with thermodynamic integration, BAR or Jarzynski estimator,⁹ the proprietary software FEP+ which is based on replica exchange with a solute tempering approach,^{7,8} and the system builder FESetup¹⁹ that supports several molecular dynamics engines. RBFs are particularly useful in the lead optimization stage where they are used to increase affinity, or selectivity, helping to find the most promising variants of the leading compounds. Their usefulness was highlighted when FEP+ predictions outperformed human expertise in ranking lead compound variants.⁸

Here, we focus on our protocol named TIES (Thermodynamic Integration with Enhanced Sampling) for the calculation of RBFs.²⁰ TIES uses a modified dual-topology approach to alchemical transformations and has been shown previously to perform well in the ranking of compounds²¹ and the prediction of relative binding free energies²⁰ with or without replica-exchange solute tempering (REST2).^{22–24} TIES uses an ensemble of MD simulations at every λ window in order to control uncertainty in molecular dynamics simulations. The different starting velocities in each replica enhance sampling and therefore improve reproducibility. This is in contrast to single trajectory approaches, which are frequently not representational, leading to false positives.^{25,26} Such incorrect conclusions further undermine the field of molecular dynamics and its ability to produce reliable predictions. The combination of sufficient sampling and the quantification of uncertainty for the observables is therefore paramount and should be employed in all MD-based methods.²⁷

It is worth mentioning that the recent advances in machine learning are also bearing fruit.^{28–30} In a notable case, a molecular-mechanics/machine-learning approach was used to introduce a correction as a postprocessing step. This correction employed a nonequilibrium simulation of the ligands and neural-network potentials that reproduce quantum mechanics energetics at a smaller computational cost. The approach significantly improved the accuracy of the alchemical free energy calculations.³¹ In another study, machine learning was used to correct the free energy perturbation calculations. The accuracy of the predictions improved when applied on the SAMPL4 data set retrospectively.³² However, not all machine-learning approaches involve physics based free energy calculations.²⁸ A combination of deep learning and algebraic topology scored highest in binding affinity prediction in the D3R Grande Challenge 3 competition.³³

The purpose of the present work is to report on our development of a new version of TIES which we name TIES 20 because of the year of release. It should be noted that the original TIES protocol was released in 2017 and will be referred to as TIES 17 hereafter. TIES 17 relies on an external package named RDKit (<http://www.rdkit.org>) for the selection of alchemical region. TIES 20 gets rid of this external dependency and uses an in-built module based on a flexible and exhaustive joint-traversal superimposition algorithm that enables greater flexibility in the selection of the alchemical region. This flexibility is used to morph partial rings, thereby further reducing the size of the alchemical region. The relative binding free energy is computed for 55 ligand transformations across five proteins to replicate the previous study using TIES 17²⁰ that invoked the RESP charge system. The results are employed as a basis for the next step where we investigate how an increase in sampling affects the relative binding free energies. Furthermore, a corresponding full set of TIES calculations is carried out with the AM1-BCC charge system, and the results are compared along with the others.

2. THEORY

Thermodynamic integration (TI) is one of the most common methods for calculating relative binding free energies, and its complete formalism is available elsewhere.^{12,34} TI relies on a control variable λ which defines how intermolecular and intramolecular interaction potentials change from one state to another. Here the two states refer to the initial state of a ligand L1 and the final state of a ligand L2. The transition from L1 to L2 is described with $0 \leq \lambda \leq 1$. The Gibbs free energy, G , is a state function which can be written in terms of the λ control variable.³⁴ The free energy change corresponding to this transition is given by the following equation:

$$\Delta G = \int_0^1 \frac{\partial G(\lambda)}{\partial \lambda} d\lambda \quad (1)$$

Using statistical thermodynamics, it can be shown that

$$\frac{\partial G(\lambda)}{\partial \lambda} = \left\langle \frac{\partial V(\lambda, x)}{\partial \lambda} \right\rangle_{\lambda}$$

whence
$$\Delta G = \int_0^1 \left\langle \frac{\partial V(\lambda, x)}{\partial \lambda} \right\rangle_{\lambda} d\lambda \quad (2)$$

where V refers to the potential energy of the system with the atom positions x . It should be noted that the above equation is only strictly valid in the thermodynamic limit, when both left-hand-side and right-hand-side terms are unique numbers with no fluctuations. However, when working with finite systems and sampling only a fraction of its conformational space, these quantities are stochastic variables.^{25,35} This implies that free energy as well as its derivative will have a range of values with an associated probability distribution. Therefore, we must aim to calculate the expectation value of these quantities using ensemble methods. The foremost requirement for this is performing the calculation multiple times to generate a distribution of ΔG values for which mean and variance can be determined. The ensemble of phase space trajectories required in the above equation can be computed using molecular dynamics simulations. The traditional practice has been to perform a single MD simulation at each λ value to get the desired ΔG . However, given that the terms involved in the above equations are stochastic quantities, the ΔG values

recalculated in this manner vary significantly on repeating the simulation with a different set of starting conditions.^{20,25,35} This stems from the extreme sensitivity of MD simulations to their initial conditions. To account for such randomness in the observed energy derivatives, it is essential to perform an ensemble simulation, that is multiple replica simulations with different starting conditions, to generate the desired ensemble. In addition, one must take into account the shape of the resultant distribution of potential energy derivatives as well as the calculated free energy changes obtained. TIES implements ensemble simulations precisely for these reasons.

Using the free energy cycle we can calculate the relative binding free energy

$$\Delta\Delta G = \Delta G_{L2}^{\text{binding}} - \Delta G_{L1}^{\text{binding}} = \Delta G_{\text{alch}}^{\text{bound}} - \Delta G_{\text{alch}}^{\text{aq}} \quad (3)$$

where $\Delta G_{L1/L2}^{\text{binding}}$ represents the binding free energies of L1 and L2, $\Delta G_{\text{alch}}^{\text{aq}}$ is the difference in solvation free energy of the ligand transformation in an aqueous environment, and $\Delta G_{\text{alch}}^{\text{bound}}$ corresponds to the free energy change of the ligand transformation bound to the complex.

Ensemble Sampling. Given the chaotic nature of molecular dynamics, individual (one-off) trajectories are not reproducible and, moreover, become increasingly inaccurate the longer the duration of a simulation.³⁵ However, we expect the statistical properties of such simulations to be robust, and therefore we use ensembles as the basis for reproducibility and hence reliability of such studies.^{25,27,36} In order to meet these requirements, while also allowing for versatile execution of the protocol, TIES employs an ensemble approach. For each λ window along the alchemical pathway, a number of replicas are used, and the integration in eq 2 performed using the principles of stochastic calculus. The ensemble approach allows for quantification of the error in the computed $\Delta\Delta G$ by considering the variance in the underlying potential energy derivatives across the replicas.²⁰

Other authors have acknowledged the necessity of repeating the alchemical free energy calculations in order to estimate the associated uncertainty.^{9,37,38} However, most published articles still overestimate confidence in results based on non-representative undersampled simulation sets.²⁶ By contrast, the ensemble approach provides a straightforward means for discussing convergence and uncertainty in predicted values.³⁶ Although the ensemble-based approach leads to better agreement with experimental data, primarily through increasing precision, it will not necessarily help with the accuracy of results.

It is important to mention that currently feasible sampling in RBE calculations is carried out within the constrained context of the ideal docked pose. Using long simulations poses different challenges. Increasingly complex systems require significantly more computational resources and more rigorous equilibration analysis, which take us into the realm of Markov State Modeling.³⁶ Sampling and equilibration will persist as a challenge particularly if one considers sufficiently complex systems, allostery, binding pathway, disordered regions, and so on. For the relatively short duration simulations as used here, the possibility and ensuing consequences of different conformations being sampled when using enhanced sampling such as replica-exchange without a clear definition of relative weights have been pointed out by us.²⁴

3. METHODS

We begin by reiterating that TIES is a method that is used to calculate binding free energies corresponding to an alchemical transformation based on ensemble simulations. On the other hand, TIES 17 and TIES 20 refer to the two versions of a software program that are used to prepare the structure and parameters for a hybrid ligand to be used as input to a TIES calculation. Therefore, the key difference between TIES 17 and TIES 20 lies in the way the input for TIES is prepared but not in the implementation of our method or the subsequent analysis of the simulation output.

Superimposition. The principal difference between TIES 17 and TIES 20 resides in the way in which the two ligands corresponding to the two end-points of the alchemical transformation are superimposed and their maximum common substructure (MCS) is identified. TIES 17 uses RDKit for this purpose and relies on a set of heuristics whereby the molecules are converted first into SMARTS format and then the largest substructure found is extended using a network search. In TIES 20, an exhaustive recursive joint-traversal superimposition algorithm has been implemented for superimposing the ligands to map which atoms remain the same and which mutate, taking into consideration the assigned force field parameters including the atom types and charges. The algorithm carries out an MCS search for two ligands using recursion. During this stage, the electronic charges on the atoms are ignored by both TIES 17 and TIES 20. For every starting combination in the ligands L1 and L2, a new search is applied, and the next atom pair is added to the MCS if three conditions are true: 1) the next compared pair of atoms is the same chemical elements, and in the case of ambiguity, a preference for the assigned force field atom types is used; 2) either the new atom-pair completes an aromatic ring in both L1 and L2 or neither atom does, such that at no point can any MCS create a different number of rings in each of the original molecules; 3) traversal of a mutual ring in L1 and L2 cannot continue if L1 proceeds to a fused ring but L2 does not (and *vice versa*). On the return stage of the recursion, the largest MCS is selected. If two or more MCS are of the same length, the program checks if they are mirror images of each other. Then, the MCS with the smallest RMSD value after superimposition is selected. Overall, TIES 20 provides greater flexibility in the selection of MCS by allowing partial rings that are not permitted by TIES 17.

Once the MCS has been selected, all the following steps as described below are common between TIES 17 and TIES 20. After generating the MCS, the alchemical region is defined to initially include the nonmatched atoms. Then, the algorithm continues to ensure that the two ligands are chemically identical using the following rules. First, the paired atoms are checked to use the same force field atom types. Then, any atom pair with more than 0.1e charge difference is added to the alchemical region. Next, the net charge of the MCS atoms that belong to L1 must be less than 0.1e different than the net charge of MCS atoms in L2. If this condition is not true, the paired atoms with the largest charge difference are repeatedly added to the alchemical region until this condition is met. If there are multiple disconnected common substructures, the smaller ones are removed. Finally, the charges between the paired atoms are averaged. This is because each matched pair in the MCS is represented with a unique atom in the dual-topology approach as implemented in TIES. Therefore, unique

charges are needed for each of the matched pairs. This averaging can introduce a small charge imbalance which is redistributed uniformly over the alchemical region. Note that, in the case of RESP charges in TIES 17, the redistribution is handled via reparametrization after constraining the charges on the atoms belonging to the MCS to their new (average) values. In TIES 20, the charge redistribution in both cases is handled manually, and the charges are redistributed evenly across the alchemical region. The software has been implemented in python using modular object oriented principles with unit testing and continuous integration for quality assurance.

Simulations. Simulations were carried out with NAMD 2.12,³⁹ and models were assembled using tleap in AmberTools.⁴⁰ The truncated octahedron periodic box was created with a minimum of 14 Å away from the protein/ligand atom. The ligand and the protein were modeled using GAFF⁴¹ and the FF99-SBILD force field,⁴² respectively. Water was modeled with TIP3P.⁴³ Ions were added to neutralize the system.

Each simulation was first minimized in 5k steps using the NAMD conjugate gradient and line search algorithm. The system's temperature was set to 300 K, and restraining of the heavy atoms in the ligand and protein was applied in 4-step equilibration, with each stage relaxing the restraints. The restraints used the harmonic potential $k(x-x_0)^2$ with k equal to 4, 3, 2, and 1 kcal/mol/Å².

Both the equilibration and production time step was 2 fs. Nonbonded 1–4 interactions were scaled to 0.8(3). The nonbonded cutoff distance was set to 12 Å with the switch distance set to 10 Å. The long-range electrostatics was treated with the particle-mesh Ewald (PME) method. The NPT ensemble was controlled with a Langevin piston and Langevin dynamics during equilibration. In the production stage, a Berendsen barostat was used in order to match the previous configuration in the reproduced study,²⁰ whereas the temperature was controlled with Langevin dynamics.

TIES. Thermodynamic integration with enhanced sampling (TIES) is an ensemble-based approach which is explained in detail elsewhere.²⁰ Scaling of nonbonded interactions within the alchemical region (intramolecular annihilation) was turned off in favor of decoupling. The van der Waals (vdW) interactions were represented with the soft core variant of the Lennard-Jones potential. The parameter $0 \leq \lambda \leq 1$ was used to scale the potential energy terms to control (de)-coupling. For each transformation, 13 λ windows were used with fixed values 0.00, 0.05, 0.1, 0.2, ..., 0.9, 0.95, 1.00. The vdW potential energy was scaled by λ in the appearing atoms and by $1 - \lambda$ in the disappearing atoms. The electrostatic potential energy was scaled by $\frac{0.55 - \lambda}{0.55}$ for $0 \leq \lambda \leq 0.55$ for the disappearing atoms, after which they stay decoupled, and by $\frac{\lambda - 0.45}{0.55}$ for $0.45 \leq \lambda \leq 1$ for appearing atoms, before which they are considered uncharged. Consequently, coupling of the electrostatic interactions begins halfway through the scaling of the vdW interactions ($\lambda = 0.45$) in the appearing atom partition, whereas decoupling of the electrostatic interactions is finished before the vdW interactions are decoupled in the disappearing atom partition, thus avoiding the “end-point catastrophe”.

Every λ window used an ensemble of 5 replicas, which was previously found to ensure that the computed value is almost always within ± 0.8 kcal/mol of the other predictions.²⁰ Each replica was simulated for 6 ns. The first 2 ns were subsequently discarded as part of the equilibration period, and the remaining

4 ns were used in calculating the relative binding energy. In the case of the “enhanced” data set (referred to later as TIES 20E), the number of replicas per λ window was increased to 20.

A modified dual-topology thermodynamic integration protocol was used in this study.²⁰ In this approach, the MCS region is represented with one set of atoms and is therefore shared between the two ligands, whereas the alchemical region has two separate representations: one for ligand L1 and one for ligand L2. The alchemical data was output every 2 ps. The $\Delta\Delta G$ and error calculations are based on the spread of the mean values from each replica, as previously explained in detail.²⁰ The ensemble average of the potential energy derivative with respect to λ was integrated using the trapezoidal rule.

For the RESP charge system, the previously computed charges were invoked.²⁰ The AM1-BCC charges were computed using *antechamber* from the AmberTools package.^{40,44} The data set contained 55 transformations across five proteins.²⁰

The simulations were carried out on three supercomputers: Scafell Pike (hartree.stfc.ac.uk), SuperMUC-NG (<http://doku.lrz.de/display/PUBLIC/SuperMUC-NG>), and Frontera (frontera-portal.tacc.utexas.edu). TIES typically requires performing an aggregate of 390 ns of MD simulations when using 13 λ windows and 5 replicas per λ window (65 replicas of 6 ns each). However, since each replica is an independent simulation, they can all be run in parallel, and thus the entire TIES calculation only requires as much wall-clock time as required to run a single MD simulation of the same duration. To get the relative binding affinity for a pair of ligands, the TIES protocol needs to be run twice—in a protein as well as in an aqueous environment. Both of these can also be run in parallel, and hence the wall-clock time requirement is unaffected. The solvated systems in a protein environment varied in the number of atoms; approximately 40k atoms for MCL1, 50k for thrombin and TYK2, and 70k for CDK2 and PTP1B. For these systems, we are able to complete the entire set of calculations in 5 to 8 h. For example, all simulations for a system of size 65k atoms were completed in about 7 h on SuperMUC-NG using 2 nodes (96 CPUs). In the aqueous environment, the system size is only about 6k atoms, and the wall-clock time required is 2 to 3 h. Note that there is a recently released GPU acceleration for alchemical transformations in the NAMD package.⁴⁵ Similar features were previously made available in the AMBER package,⁴⁶ while we are currently approaching the completion of the first OpenMM-based version of TIES.

It is important to note that CDK2 was simulated without the presence of cyclin in line with the previous study.²⁰

Analysis. The python packages MDAnalysis,⁴⁷ NumPy,⁴⁸ and SciPy⁴⁹ were used for analysis, whereas Matplotlib was invoked for plotting.⁵⁰ Bootstrapping was carried out with replacement with 5000 samples for any estimator.

The TIES error (σ_{TIES}) is calculated by adding the uncertainty at each λ window in quadrature. The uncertainty at each λ window is given by the variance of the bootstrapped distribution of sample means of the average potential energy derivatives from each replica as shown below

$$\sigma^2(\lambda) = \text{Var}(\text{BS}(U'(\lambda, r_1), U'(\lambda, r_2), \dots, U'(\lambda, r_n))) \quad (4)$$

where BS denotes the bootstrapping function, Var denotes the variance function, and U' denotes the potential energy derivative. The r_i ($i = 1, 2, \dots, n$) refers to the replica numbers.

Table 1. Validation of the TIES 20 Implementation along with the Previous Results (TIES 17) with Respect to the Experimental $\Delta\Delta G^a$

protein	property	TIES 17	TIES 20	TIES 20E	TIES 20 BCC
TYK2	MUE	0.44 (0.14)	0.48 (0.14)	0.43 (0.14)	0.66 (0.19)
	MSE	-0.03 (0.24)	-0.00 (0.24)	0.00 (0.23)	-0.02 (0.31)
	RMSE	0.56 (0.17)	0.58 (0.16)	0.53 (0.16)	0.86 (0.22)
	Pearson's ρ	0.94	0.93	0.94	0.84
	slope	0.83	0.86	0.84	0.66
	intercept	0.12	0.07	0.08	0.20
MCL1	MUE	1.20 (0.23)	1.08 (0.23)	1.09 (0.22)	0.81 (0.17)
	MSE	0.38 (0.38)	0.09 (0.38)	0.18 (0.37)	0.22 (0.28)
	RMSE	1.41 (0.26)	1.34 (0.27)	1.32 (0.26)	0.98 (0.18)
	Pearson's ρ	0.80	0.68	0.70	0.83
	slope	1.18	0.79	0.83	0.93
	intercept	-0.32	-0.16	-0.23	-0.24
thrombin	MUE	0.71 (0.17)	0.54 (0.17)	0.53 (0.17)	0.52 (0.15)
	MSE	0.19 (0.29)	-0.04 (0.28)	-0.12 (0.27)	-0.04 (0.26)
	RMSE	0.78 (0.18)	0.65 (0.20)	0.67 (0.19)	0.63 (0.17)
	Pearson's ρ	0.91	0.87	0.86	0.90
	slope	1.50	1.20	1.20	1.29
	intercept	-0.07	0.09	0.17	0.11
PTP1B	MUE	0.37 (0.15)	0.68 (0.22)	0.69 (0.22)	0.66 (0.21)
	MSE	0.11 (0.25)	-0.24 (0.36)	-0.25 (0.35)	-0.54 (0.30)
	RMSE	0.48 (0.17)	0.83 (0.26)	0.84 (0.26)	0.83 (0.26)
	Pearson's ρ	0.84	0.42	0.41	0.66
	slope	0.86	0.30	0.31	0.53
	intercept	-0.17	-0.08	-0.07	0.33
CDK2	MUE	0.74 (0.21)	0.86 (0.27)	0.85 (0.26)	0.61 (0.24)
	MSE	0.17 (0.38)	0.14 (0.46)	0.21 (0.44)	0.02 (0.38)
	RMSE	0.81 (0.21)	1.06 (0.29)	1.02 (0.27)	0.83 (0.30)
	Pearson's ρ	0.87	0.75	0.76	0.86
	slope	1.28	1.15	1.12	1.28
	intercept	-0.10	-0.10	-0.18	0.05

^aIn TIES 20E, sampling is enhanced by increasing the number of replicas per λ window to 20. All TIES use RESP charges except for TIES 20 BCC which uses the AM1-BCC charges. Due to the small number of cases in PTP1B, one outlier was removed as it was skewing the statistics. The original data can be found in Table S1. The differences (MUE, MSE, RMSE) between calculated and experimental binding free energies are calculated by resampling the data from simulations and experiments (see the Methods section). The uncertainties of these metrics (given in parentheses) are the standard deviations of the resampling distributions. All energies are in kcal/mol.

Then, we add the uncertainties in quadrature, while scaling each to the size of the λ window

$$\sigma_{\text{bound/aq}}^2 = \sum_{\lambda} \sigma^2(\lambda) \Delta\lambda^2 \quad (5)$$

where the subscript refers to the uncertainty in either the bound or unbound environment. Finally, we add the uncertainty in a water and in a protein environment

$$\sigma_{\text{TIES}}^2 = \sigma_{\text{bound}}^2 + \sigma_{\text{aq}}^2 \quad (6)$$

As described earlier, the terms in eq 2 are stochastic quantities owing to the sensitivity of MD simulations to their initial conditions. Therefore, simply performing a single MD simulation in order to estimate the potential energy derivatives, a practice which is widespread, is not reliable as it does not account for the randomness associated with the observed quantities. In fact, the free energy change calculated using the single replica approach is not reproducible as the results vary substantially on repeating the calculation with different initial conditions.^{20,23} We have reported previously that the distribution of free energy differences obtained from TIES calculations is Gaussian,²⁰ which is a reasonable first-approximation. Here, we have analyzed these distributions

more carefully and find that they deviate from Gaussian behavior. This is true for both the potential energy derivatives at any random λ value as well as for the final relative binding affinities. A set of $\Delta\Delta G$ values was generated, individual values within which were determined using the traditional single replica approach, and kurtosis and skewness of the resultant distribution were calculated (see Figure 7). The number of points in the distribution is equal to the number of replicas per λ window as each MD simulation contributes only one data point to the sample. Due to the small number of the transformations reported here, this data was further complemented with data from our previous TIES studies²⁰ as well as some ongoing work. These include 7 transformations with CDK2 (20 replicas per λ) from Bhati et al.,²⁰ 10 transformations with MCL1 (40 replicas per λ), and 7 transformations in retinoic acid receptor alpha (RAR- α) (20 replicas per λ), both of which come from our ongoing work.³⁶

In order to calculate errors in the statistical measures in Tables 1 and S1 as displayed in parentheses, the resampling of experimental data was performed by randomly generating values in proportion to the probability density function of a normal distribution. The reported experimental average and the standard deviations (if available) are used to generate the

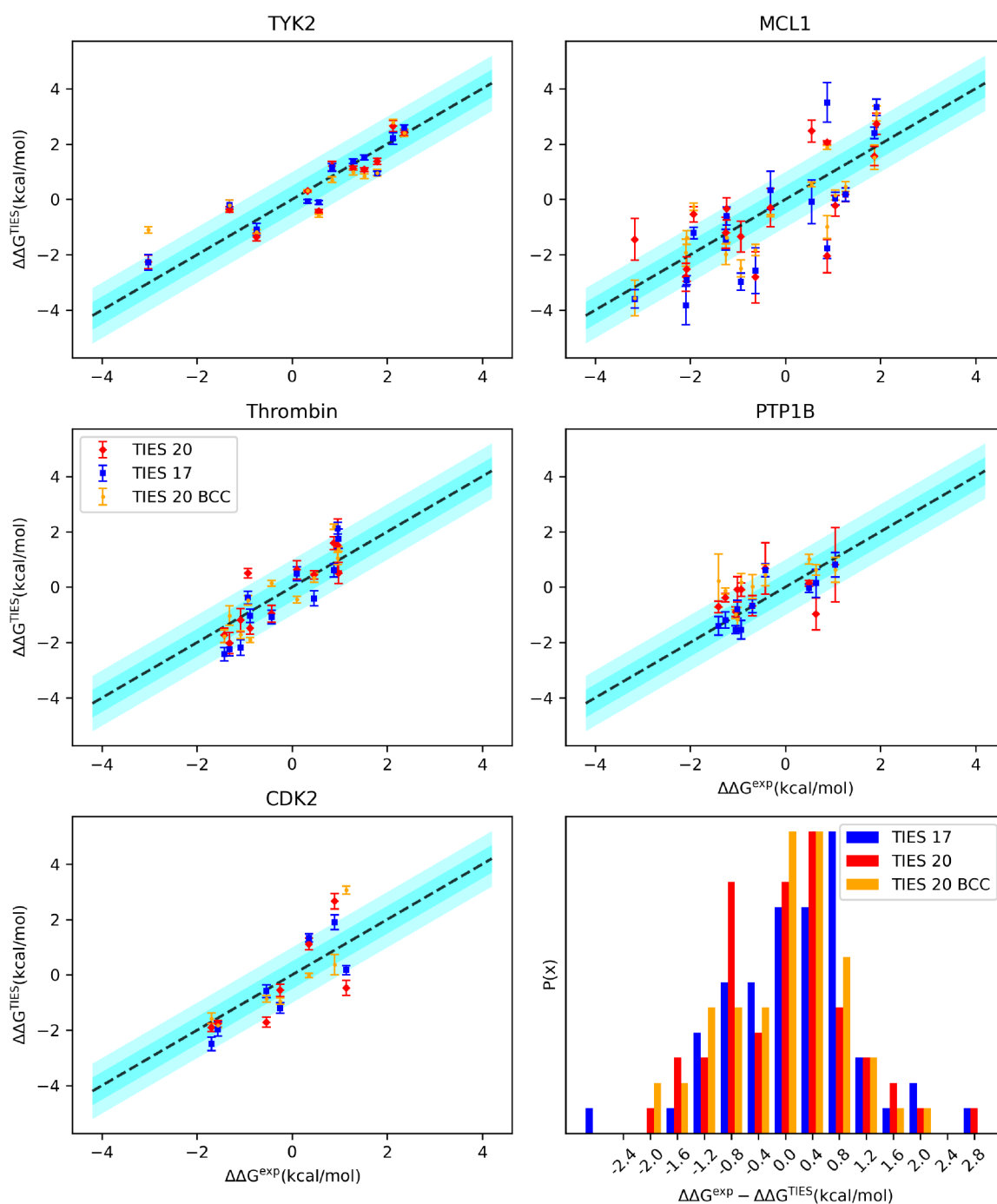


Figure 1. Computed $\Delta\Delta G$ with respect to the experimental results representing TIES 17, TIES 20 with RESP, and TIES 20 with BCC charges. The deeper cyan color represents 1 kcal/mol range, whereas the lighter cyan represents 2 kcal/mol range. The distributions of the $\Delta\Delta G$ s with respect to the experimental results are further summarized in the bottom-right figure.

distribution. Many experimental data are reported as single numbers with no standard deviations. In this case, a value of 0.55 kcal/mol is used for the standard deviation, which is an estimation of the experimental error for relative binding free energies between two compounds.⁸ The resampling is carried out by randomly selecting a subset of the data 10000 times in order to calculate the error $SE[X^{\text{exp}} \sim \mathcal{N}(\mu, \sigma^2) - X^{\text{TIES}} \sim \mathcal{N}(\mu, \sigma^2)]$.

Although a normal distribution is frequently assumed in the experimental values, there is no evidence to support it. A

recent investigation of experimental binding free energies has revealed that the distributions from a large number of independent measurements do exhibit non-normal properties (Ian Wall and Alan Graves, private communication, 2020).

4. RESULTS

We compare the results from TIES 20 with those from TIES 17. As mentioned earlier, the overall protocol remains the same except that TIES 20 allows finer control over the alchemical region by enabling the morphing of partial rings whose effect

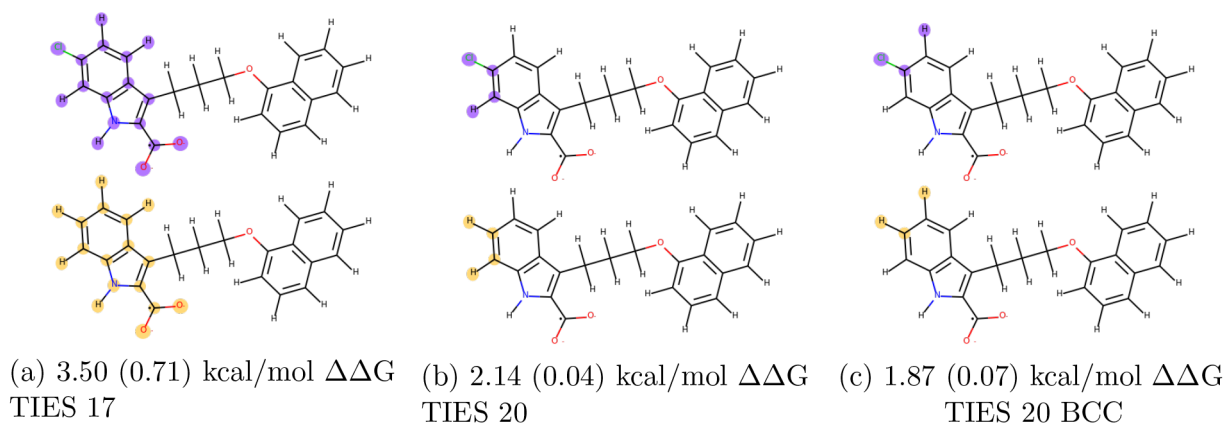


Figure 2. An example transformation MCL1 L18 to L39 showing how allowing partial rings can impact the alchemical region. The experimental $\Delta\Delta G$ is 0.88 kcal/mol. Across a), b), and c), the same TIES protocol is utilized. In a) and b), the charges are the same, and c) uses the AM1-BCC charges. L18 is depicted in the upper portion of each figure, where the purple-highlighted atoms disappear as λ increases. In the lower part, L39 is shown, where the yellow-highlighted atoms appear as λ increases.

we investigate in this work. Specifically, the partial rings were matched together only if they were of the same atom types, as determined by GAFF.

Altogether, 55 TIES transformations across five different proteins were computed to reproduce the previous study.²⁰ The results are summarized in Table 1. We first compare TIES 17 to TIES 20. It is noted that in the previous study the tolerance limit of 0.1e for the difference in net charges of MCS atoms between the two ligands was relaxed slightly in some cases. It was explained that this approach was employed in order to decrease the size of the alchemical region when this criterion led to the deletion of a link atom dividing the MCS into two or more disjoint fragments. The more notable cases include MCL1 and PTP1B (see Bhati et al.,²⁰ SI). Although such small issues were mostly avoided here, in several cases the net charge limit had to be adjusted in line with the previous study, without which the size of the alchemical region represented most of the molecule. None of these changes were necessary when invoking the AM1-BCC charge system.

In terms of mean unsigned error (MUE) and root-mean-square error (RMSE), the TIES 20 results are either similar (TYK2) and show moderate improvements (MCL1 and thrombin) or moderate regression (PTP1B and CDK2). Accordingly, the global MUE value across all transformations is 0.74 (0.09) kcal/mol for TIES 17 and 0.75 (0.10) kcal/mol for TIES 20, showing that both protocols perform, on average, equally well.

Increasing the number of replicas to 20 per λ window (TIES 20E) shows negligible impact in comparison to TIES 20 across the five proteins. However, an improved performance was seen in the case of the PTP1B outlier, where the poor initial estimate is improved by the new simulations (see below and Table S1). In that single case, due to the sampling outlier, the agreement with experiment improves significantly over TIES 20.

The TIES 20 BCC variant compares favorably, on average, to TIES 17/20 where the RESP charges were used. The MUE and RMSE metrics were moderately worse in TYK2 but showed an improvement in MCL1, thrombin, and CDK2. In the case of PTP1B, the system performed moderately worse than TIES 17 while being similar to TIES 20. The global MUE value for TIES 20 BCC was 0.67 (0.09) kcal/mol showing a slight overall improvement over TIES 17/20.

The Pearson's ρ correlation is slightly smaller in TIES 20 than in TIES 17 except for PTP1B where the difference is larger. This decrease is attributed to the problem in sampling in PTP1B stemming from mutating the carboxylates, which were described and manually corrected in the previous work.²⁰ TIES 20 BCC similarly shows a lower correlation in PTP1B; however, it is largely the same as TIES 17 in other proteins.

The computed relative binding free energies (RBFES) are displayed in Figure 1. The TYK2 transformations show very good agreement with the experimental data and are frequently located in the 1 kcal/mol range, with several cases in the 2 kcal/mol area and just two cases outside of it. This trend deteriorates in thrombin, CDK2, and PTP1B, while the number of transformations outside of the 2 kcal/mol area is largest in MCL1.

The computed $\Delta\Delta G$ with respect to the experiment is further summarized in the form of distributions in the bottom-right corner of Figure 1. The predictions are largely centered around the experimental values. A slight asymmetry appears in the distribution of TIES 17 and TIES 20 BCC, showing that the $\Delta\Delta G$ tends to be too favorable, on average, in these two data sets. However, the mean signed error (MSE) depends strongly on the protein, as was shown in Table 1. In the four proteins excluding TYK2, TIES 17, on average, overestimates the change in the binding affinity. Whereas TIES 20 and TIES 20 BCC show fewer asymmetries, both underestimate the change in the binding affinity in PTP1B protein, with TIES 20 BCC being particularly noticeable. Although it is noted that MUE is affected by the randomness of the relatively small number of transformations in each protein, it is interesting to consider the relatively large differences across the three data sets, particularly TIES 17/20. However, the MUE and MSE errors at the protein level might mean that a larger sample is needed to understand the differences between the methods.

In most of the TYK2, CDK2, and thrombin transformations, the corresponding RBFES values are also close to each other, with a few exceptions in TIES BCC 20, which produce quite different results (Figure 1). The MCL1 transformations show larger error values and spread in TIES 17/20. Interestingly, in MCL1, the errors are smaller when BCC charges are used. In the last protein, PTP1B, the errors are larger in several cases in TIES 20. However, one TIES 20 BCC transformation with a

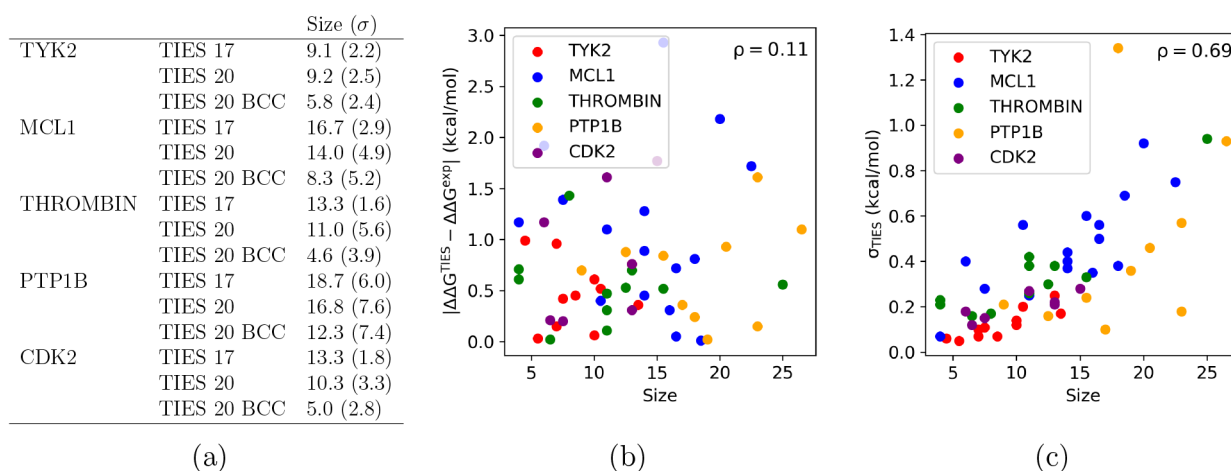


Figure 3. Size of the alchemical regions for each protein is shown in the table (a). Also shown are the correlations between the size of the alchemical region with (b) the $\Delta\Delta G$ prediction with respect to the experimental value and (c) the TIES σ error. The Pearson's ρ correlation is displayed in the upper-right corner in b) and c). The value ρ for AM1-BCC charges that corresponds to c) is 0.42. The size of the alchemical region includes hydrogen atoms and is defined as the average of the appearing and disappearing atoms.

large error is present, showing that the BCC charge system is not immune to large errors.

PTP1B Outlier. One of the transformations in the TIES 20 set, PTP1B transformation L6-L14, is an outlier with a large $\Delta\Delta G$ error. The transformation has a large alchemical region created via a combination of two superimposition rules. The first rule states that the net charge difference between the appearing and disappearing partitions should be reduced to less than 0.1e, which leads to a mutation of a pair of atoms at a junction, separating the compound into two components. The second rule states that there should not be any disjoint components, which leads to a removal of the smaller component. The error is thought to be due to the large region being mutated, because it contains the two appearing and disappearing carboxylate groups, leading to unrepresentative behavior. Specifically, the appearing and disappearing partitions sample different regions. In this case, an increase in the number of replicas helps to improve the initial results. The problem was encountered and described in more detail in the previous study (Bhati et al.,²⁰ SI).

4.1. Superimposition. The superimposition algorithm enables automated use of partial rings which can lead to a significant decrease in the size of the alchemical region. A smaller alchemical region helps avoid sampling issues which has the potential to improve the precision of the TIES ensemble calculations. A prominent example of this is shown in Figure 2 with the MCL1 transformation L18-L39 where a chlorine atom replaces a hydrogen on the dual-ring. In this case, the number of mutated atoms has decreased from 17 atoms including the carboxylate group in the original version of TIES (Figure 2a) to just 4 and 3 atoms in the case of TIES 20 and TIES 20 BCC, respectively (Figures 2b and 2c).

In this transformation in Figure 2, the computed $\Delta\Delta G$ values are significantly closer to the experimental value of 0.88 kcal/mol with TIES 20 predicting 2.14 (0.04) kcal/mol and TIES 20 BCC predicting 1.87 (0.07), an improvement over TIES 17 which predicted 3.50 (0.71) kcal/mol. However, the arguably more important improvement is in the prediction precision as captured with the TIES errors (σ_{TIES}), which drop to less than 0.1 kcal/mol in both cases from 0.71 kcal/mol in TIES 17. This trend is present across the data set with the

average $\sigma = 0.20$ kcal/mol for both variants of TIES 20, a decrease from $\sigma = 0.31$ kcal/mol in TIES 17.

The selection of the atoms for the transformation along the alchemical pathway depends on several properties of the ligands with atomic partial charges being one of them. In some cases, similarly to MCL1 L18-L39 transformation, significant differences were found in the size of the alchemical region being mutated. The number of mutated atoms is reduced substantially due to enabling of mutations in the ring.

The number of mutating atoms in the alchemical region was compiled in order to study its relationship to the $\Delta\Delta G$ prediction and its sampling error (Figure 3). A comparison of the new version of TIES with the original version shows that, on average, fewer atoms are mutated in the case of all proteins but TYK2. However, the difference is small, highlighting how the RESP charges affect the superimposition. In stark contrast, when BCC charges are used, the size of the alchemical region is often halved (Figure 3a). We attribute this to partial atom charges being modified only near the mutation site in the BCC charge system.

We tested whether the size of the alchemical region is related to the accuracy of the computed $\Delta\Delta G$. We found no correlation between the two with the Pearson's ρ equal to 0.11 (see Figure 3b). However, we found that the size of the alchemical region is well correlated with the calculated TIES error (σ_{TIES}) as shown in Figure 3c. This trend is present for all proteins, even for the TYK2 protein, which, on average, has the smallest alchemical region. In other words, an increase in the size of the alchemical region leads to larger variance in the computed $\Delta\Delta G$.

However, the increased error due to the larger alchemical size does not translate to a correlation with the $\Delta\Delta G$ prediction accuracy. This is probably because the ensemble averaging employed for each transformation uses 5 replicas per λ window, compensating for the possible increase in the sampling needs.

4.2. Number of Replicas. We investigated the impact of increased sampling on the computed $\Delta\Delta G$. For this, the data set with 20 replicas per λ window was used. We measure how much the $\Delta\Delta G$ values are affected when one extra replica is added. The results are displayed in Figure 4.

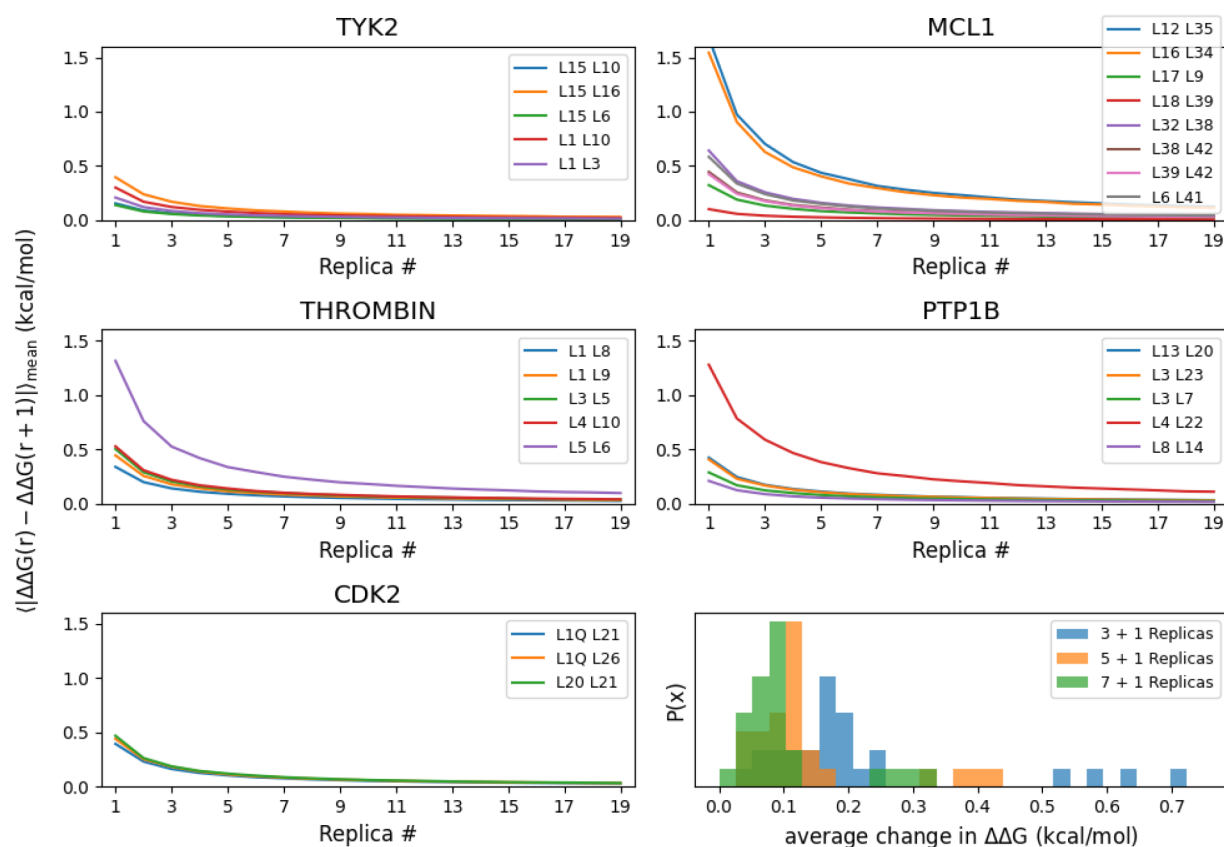


Figure 4. Average impact of an extra replica on the computed $\Delta\Delta G$. Units are in kcal/mol. The subset of the data with 20 replicas per λ window was used. For the specific cases of 3, 5, and 7 replicas, a distribution is shown in the bottom-right panel.

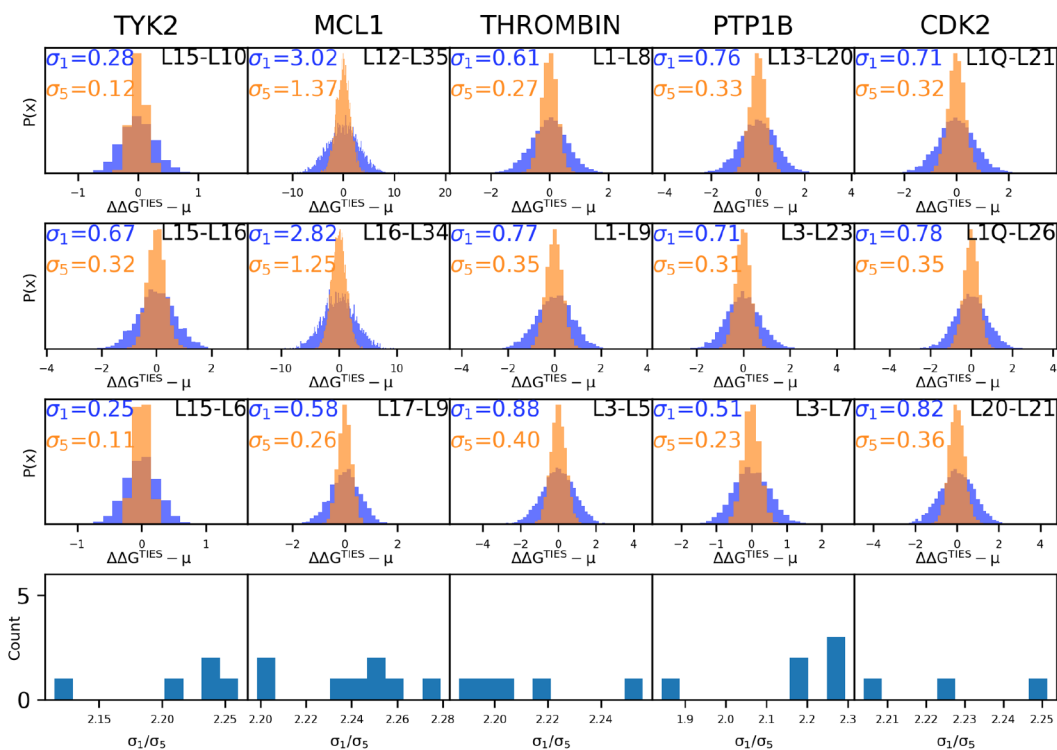


Figure 5. Distribution of the $\Delta\Delta G$ values with TIES using 1 (blue) and 5 (orange) replicas generated by drawing randomly replicas from the data set of 20 replica per λ window. The “reference” mean μ was computed using the full data set. For each protein, only three example cases are shown in a column. The lowermost plot summarizes all transformations for any given protein. All energies are in kcal/mol.

The change in $\Delta\Delta G$ was calculated using bootstrapping. Five thousand samples were generated for each replica number. Then, one additional replica was added to each sample, and the change in $\Delta\Delta G$ was recorded. All 20 replicas were used to represent the sampling space and improve the validity of the results.

A single replica is shown to be unreliable in all cases. Adding the second replica changes $\Delta\Delta G$, on average, by around 0.4 kcal/mol (Figure 4). There is only one case where a single replica is sufficient: MCL1 L18-L39. In this particular transformation, a hydrogen was mutated into chlorine, and the size of the alchemical region is minimal. In this case, adding more replicas has negligible impact on the final value. Overall, however, the changes in the computed values are substantial if only one or two replicas are used. In the case of the most challenging targets, such as MCL1 L16-L34 or PTP1B L4-L22, they are changed, on average, by 1 kcal/mol when adding the third replica.

In Figure 4, bottom right, the data is summarized for three cases (3, 5, and 7 replicas). Having 3 replicas shows that we can expect the addition of a single replica to sway our results often by 0.2 kcal/mol, although in the worst cases it could be more than 0.7 kcal/mol. However, when using 5 replicas, adding the sixth replica significantly reduces the gains, changing the value, on average, by 0.12 kcal/mol. Nevertheless, in the worst case, the value can change by more than 0.4 kcal/mol. The improvement continues with 7 replicas, where adding one more replica changes the final value by less than 0.1 kcal/mol, on average.

The choice of 5 replicas is a good trade-off between the computational resources and our confidence in the final $\Delta\Delta G$ value. For example, the TYK2 and CDK2 transformations are negligibly affected after 5 replicas. At this point, adding any more replicas changes the final value by slightly over 0.1 kcal/mol, on average. However, the analysis here shows the importance of looking at each specific case separately to decide whether more sampling is justified. This is particularly clear for several cases where the range of $\Delta\Delta G$ error is very large, such as L5-L6 in thrombin or L4-L22 in PTP1B. For example, the average change with 7 replicas was around 0.3 kcal/mol for these cases. Conversely, in many cases even 3 replicas might be sufficient, which can save computational resources.⁵¹

4.2.1. Ensemble of 5 vs 1 Replicas. The impact of the ensemble size on the shape of the $\Delta\Delta G$ distribution is further compared for 1 and 5 replicas per λ window. The data set with enhanced sampling counting 20 replicas per λ window is used, with the reference value of the $\Delta\Delta G$ computed using the full data set (TIES 20E). This comparison helps estimate how likely one is to reproduce the same results using 1 or 5 replicas.

The top three rows in Figure 5 show the distributions of $\Delta\Delta G$ values drawn using 1 or 5 replicas per λ window for each protein. The trend summarized in the bottom row was computed using all the transformations in the enhanced data set.

In the best cases, namely the TYK2 L15-L10 and L15-L6 transformations, thermodynamic integration carried out with a single replica has a 68% chance of its $\Delta\Delta G$ being within 0.28 kcal/mol of the reference value. However, using 5 replicas per λ window reduces the range to around 0.12 kcal/mol. For CDK2, the $\Delta\Delta G$ will be within 0.71 or 0.32 kcal/mol for 1 and 5 replica TIES, respectively. These two cases represent the same trend across all the transformations. In the most challenging case, MCL L12-L35, there is a 68% chance that

the $\Delta\Delta G$ will be within 3.02 and 1.37 kcal/mol for 1 and 5 replica variant. In other words, using 5 replicas ensemble reduces the dispersion of calculated $\Delta\Delta G$ by a factor of 2.2 which has a significant impact on the reproducibility of the computed free energies.

The summary of average improvement in the dispersion is presented in the bottom row of Figure 5, for which all transformations were used. There appears to be a 2.2 factor improvement, on average, when using 5 replica ensemble as opposed to 1. In other words, the likelihood of reproducing the $\Delta\Delta G$ within one standard deviation is over twice larger. Conversely, using only 1 replica per λ window means that the values will be more than twice as often outside of one standard deviation of the “reference” mean (which was computed with 20 replicas per λ window).

4.3. ΔG Variability. We examined the relationship between the ΔG variance in a water environment and in a protein environment. It is widely accepted that the presence of protein is a source of variation in the ΔG value. We similarly use the data set with an increased number of replicas per λ window. The value ΔG was bootstrapped by randomly selecting a single replica to obtain the ΔG distribution. The distribution was generated separately for the transformation in bulk water and in a protein environment.

The relationship between the standard deviation of ΔG in water and protein is shown in Figure 6. The Pearson's ρ

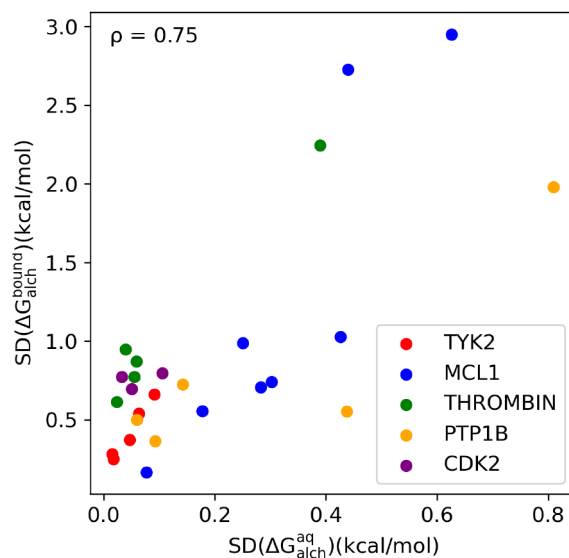


Figure 6. Relationship between the ΔG variation (standard deviation) of the transformation in water and in protein for TIES 20. The ΔG values were bootstrapped. Pearson's ρ correlation is shown in the upper-left corner.

correlation is equivalent to 0.75 showing a good correlation between the two. One possible rationale for this is that any sampling challenge posed by the ligand in water will likely be present in a protein environment as well. It might further be suggested that, to some extent, the binding pocket of the protein does not prevent the ligand from sampling the different states, at least in the case of the five proteins employed in this study. A deeper binding pocket likely constrains the ligand leading to different behavior. Therefore, such behavior is highly dependent on the structure of the binding pocket as well as the stability of the complex. Furthermore, the same size of

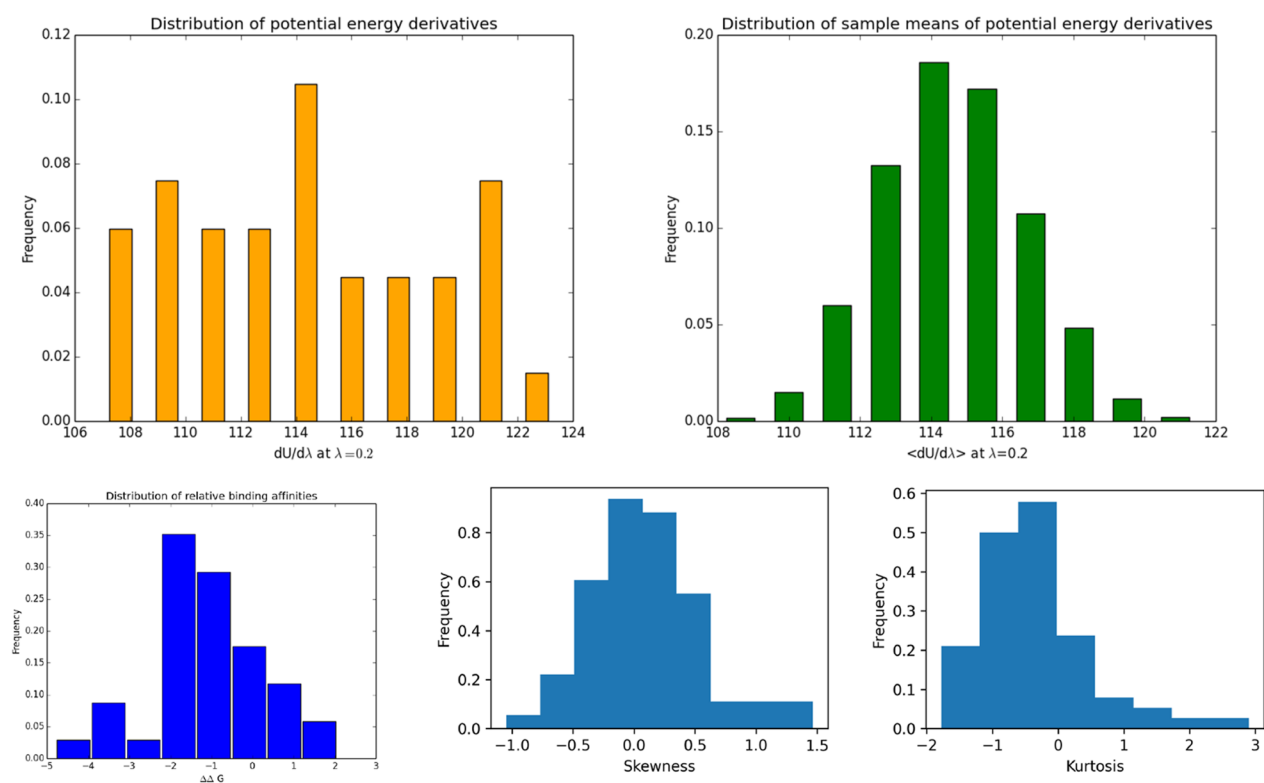


Figure 7. Non-Gaussian properties of potential energy derivatives and $\Delta\Delta G$ distributions from ensemble-based TIES simulations. The top panel displays the distributions of potential energy derivatives at $\lambda = 0.2$ for an MCL1 complex with 40 replicas per λ value; actual values in the left panel are compared to means of bootstrap resamples of size 5 in the right panel. The bottom left panel displays the distribution of $\Delta\Delta G$ values for the same MCL1 complex, whereas the other two panels show distributions of skewness and kurtosis coefficients of similar $\Delta\Delta G$ distributions for all transformations studied. All energies and energy derivatives are in kcal/mol.

the alchemical region used in both bound and unbound states can further contribute to the correlation here.

Finally, this correlation can be used in adaptive sampling to adjust the size of the TIES ensemble. The TIES error from the transformation in an aqueous environment can help estimate the initial size of the TIES ensemble needed to sufficiently sample the transformation in a protein environment. Sampling in a protein environment is about an order of magnitude more expensive computationally as compared to sampling in an aqueous environment, and hence predicting an appropriate number of replicas for the former may save us substantial resources. Similarly, alchemically transformed ligands often share a common scaffold, information that would likely help estimate the optimal number of replicas needed to achieve sufficient sampling.

4.4. Non-Normal Nature of the Distributions. The properties of the underlying data are further investigated by analyzing the distributions of $\partial V/\partial\lambda$ as well as $\Delta\Delta G$ across the transformations. Figure 7 displays the various distributions. The top panel shows how much $\partial V/\partial\lambda$ varies on simply repeating the MD simulation. In the left panel, we have shown the distribution of 40 potential energy derivatives at $\lambda = 0.2$ (a random choice) for one of the MCL1 complexes (L26-L64) for which we performed 40 replicas for each λ value (taking the average $\partial V/\partial\lambda$ for each replica). It is very clear that this distribution is nowhere near Gaussian. In the right panel, we show the distribution of sample means for the same data obtained by resampling 5 data points using bootstrapping with replacement. The resultant distribution is Gaussian which is a direct consequence of the Central Limit Theorem. The

leftmost figure in the bottom panel displays the distribution of 40 $\Delta\Delta G$ values for the same MCL1 system obtained using the traditional single replica approach. Again, this distribution is a heavy tailed one. Although the three figures discussed so far only capture the behavior of a single ligand transformation, they represent the typical behavior we would observe for any TI calculation. This is further confirmed by the remaining two figures in the bottom panel that display the distributions of Fisher-Pearson skewness and Fisher kurtosis coefficients of similar $\Delta\Delta G$ distributions for all ligand transformations included in this analysis. This includes the additional data sets from our previous and ongoing TIES studies (see Methods for more details). Such behavior is not just confined to TI calculations. It is true for any MD based method including other free energy methods like MMPBSA (refer to Figure 1 of Wan et al.^{25,36}). Furthermore, several figures provided by Knapp et al.²⁶ display skewed (i.e., non-normal) distributions of geometrical quantities emanating from large ensembles of protein MD trajectory data.

The overall trend in the $\Delta\Delta G$ distributions shows a considerable skewness and kurtosis present across the transformations (Figure 7). The skewness ranges in both value and direction, although the mean value is equal to 0.10. Similarly, kurtosis is also found to vary significantly and also appears to be non-normally distributed. The mean kurtosis is equal to -0.34 .

The relatively large skewness indicates that the $\Delta\Delta G$ distributions are not normal: they have long tails at both sides. This means that the probability of a single calculation being outside of the mean value by $\pm\sigma$ or $\pm 2\sigma$ is larger than in

the case of a normal distribution. The larger number of such expected “outliers” implies that, when comparing theoretical binding affinities with those from experiments, we should not expect an ideal linear correlation plot of slope 1 and intercept 0. Statistical tools like bootstrapping and linear regression do not have a prerequisite that the data be normally distributed and hence are also applicable to non-normal distributions. However, their quantitative reliability is questionable in such cases in the absence of sufficient quantities of data and manifestly represent an egregious problem when using only a single MD simulation. Certain estimators like median-of-means that accommodate heavy tails and handle outliers robustly may be useful to handle non-normal distributions.

As an example of how erroneous predictions from a single replica approach can be, the distribution of $\Delta\Delta G$ values for a representative ligand transformation in Figure 7 shows that the results may vary substantially (by about 7 kcal/mol in this case) if the widespread approach of running a single MD simulation per λ window is used. This illustration yet again demonstrates that one-off simulations using molecular dynamics are not reliable.

As is evident from the top right panel of Figure 7, when using simulations comprised of random multiple replicas in combination with bootstrapping, no kurtosis or skewness was found for any of the transformations. Thanks to the central limit theorem, the distribution from the bootstrapping approach then approximates a normal distribution even though the underlying data distribution is skewed. This demonstrates yet another consequence of ensemble simulations as the number of replicas increases: the bootstrapped sampling distribution usually converges to a normal distribution regardless of the distribution of predictions from the replicas.

5. DISCUSSION

The newly implemented TIES 20 closely reproduces the previous results found in TIES 17²⁰ (Table 1). However, the inclusion of partial rings in the matched areas in TIES 20 leads to fewer manual interventions. For example, no corrections were needed in the transformation in the MCL1 protein, and only one outlier transformation was encountered in the PTP1B protein, which was also previously described.²⁰ In this transformation L6-L14 in the PTP1B protein, during the superimposition, one atom was added to the alchemical region to account for the charges, which separates the morphing molecule into two parts. In a cascade, the alchemical region is enlarged by adding the smaller disconnected component. In effect, a large part of the molecule is found in the alchemical region, including the two charged carboxylate groups, leading to large errors in the TIES 20 ensemble.

Overall, however, the inclusion of partial rings helps prevent some of these cases by decreasing the size of the alchemical region. On average, the alchemical region shrank by 3 atoms in comparison to TIES 17 (Figure 3a). Furthermore, the AM1-BCC charge variant of TIES 20 decreased the size of the alchemical region to half of the size found in TIES 17/20 that used the RESP charge system (Figure 3a). The importance of the alchemical region size was confirmed by highlighting its relationship with the TIES error, which captures the variance in the ensemble (Figure 3c). The number of mutated atoms is strongly correlated with the TIES error. Consequently, the average TIES error has almost halved in comparison to TIES 17. These findings highlight further potential for improvement in the case of RESP charges. For example, by avoiding the

cascading effect which enlarges the alchemical region, it is possible to reduce the variation in the ensemble, which translates to the same results with less computational resources. One promising approach to accomplish this is the recently released NAMD implementation of the hybrid single-dual topology approach.⁵² In this method, the structurally similar region across the compared ligands can utilize the single-topology transformation, in which the matched atoms are morphed directly into each other. In effect, just like in the case of our modified dual topology approach in TIES, the two structures do not diverge in the configurational space which can happen when the dual topology region is used. However, this implementation requires that the dual topology region is connected to the single topology region via a single bond, reducing its applicability.⁵²

Interestingly, despite this clear correlation between the size of the alchemical region and the TIES error, we found no overall impact on the accuracy of the predicted $\Delta\Delta G$ values (Figure 3b). This we attribute to the ensemble approach, where the improved sampling compensates for the larger TIES errors. Nevertheless, the improvement in precision means that better estimates can be achieved with less computational resources.

The impact of further sampling, that is increasing the size of the TIES ensemble, varies and depends on the transformation (Figure 4). Two sets of transformations, TYK2 and CDK2, show minuscule changes even after 3 replicas, with 5 replicas per λ window guaranteeing good results. It is interesting to note that these two sets of transformations also happen to have the smallest alchemical regions, on average. In the other three proteins, 5 replicas per λ window suffice in the majority of the cases. The more challenging cases require larger ensembles. However, once again, this is likely due to the larger alchemical regions which create a problem with sampling, considering that the most challenging transformations in Figure 4, namely MCL1 L12-L35 and L16-L34, PTP1B L4-L22, and thrombin L5-L6, also have large alchemical regions.

A clear picture emerges showing the consequences of having a large alchemical size in a transformation. However, the source of the TIES error, or the ensemble variation, is not solely due to the alchemical region. TIES with AM1-BCC charges displays a moderate correlation between the error and the size of the alchemical region (Pearson's $\rho = 0.42$). This is because in many of these transformations the alchemical region is small. In these cases, the variation in the ensemble is likely due to other factors such as the stereochemistry in the protein environment and multiple free energy minima.

It is worth mentioning that a single-topology alchemical transformation of rings could induce bond strain with an asymmetric impact on the free energy in a water environment and in a protein environment.⁵³ However, the magnitude of the asymmetry depends on the type of the ring transformation⁵³ and whether it involves scaffold hopping. The FEP+ package addresses this issue by implementing soft bonds to avoid the divergence in the harmonic potential.⁵⁴ In our work, however, the dual topology approach avoids the bond strains by representing the mutating region separately for the initial and final molecule.

There is a strong positive correlation between the variance in ΔG in a protein environment and a water environment. This relation creates the possibility of calculating the ΔG in a water environment and using the results to estimate the size of the TIES ensemble for the transformation in the protein

environment. While it should be noted that these values would also be affected by the size of the alchemical region in the dual topology alchemical transformation, in the case of TYK2 which mutates smaller regions, this correlation is still present. This relationship might be conditional upon the size of the alchemical region being the same in any given transformation, which could be a causal factor in this correlation. This could mean that it is the size of the alchemical region itself that might help estimate the initial number of replicas required to achieve sufficient sampling.

These findings highlight the somewhat unique nature of each alchemical transformation. One example of middleware that facilitates a flexible automated workflow is RADICAL-Cybertools,⁵⁵ which can be used to tailor the protocol to meet the needs of each transformation. This can be done by adjusting the number of replicas per λ window and other parameters online as needed with the ultimate goal being to decrease uncertainty in the final binding affinity.¹³ Indeed, this approach has previously been shown to yield promising results.⁵¹ An adaptive quadrature was used to guide the selection of the λ windows in the TIES protocol while also allowing for the early termination of each replica if convergence was observed. It was shown that a substantial increase in accuracy at lower computational cost is possible with a sample of 5 transformations.⁵¹

Furthermore, it is important to consider the errors in the experimental data. For isothermal titration calorimetry (ITC), the errors are claimed to be within 0.3–0.7 kcal/mol.^{8,56,57} In another study, analysis was performed on a database from AstraZeneca that spanned 9 years of accumulated assay data. While the assay errors are often relatively small,⁵⁸ there are expected to be systematic differences across laboratories and methods. Furthermore, the error is dependent on the number of measurements taken in an assay.⁵⁸ The experimental binding affinities for the proteins thrombin⁵⁹ and CDK2⁶⁰ were based on the IC₅₀ assays, while for PTP1B,⁶¹ MCL1,⁶² and TYK2^{63,64} they were based on the inhibitory constant K_i . It is noted that a different number of measurement sets were taken across the proteins. For TYK2 and thrombin, at least three were taken, and for MCL1, three were taken; whereas for PTP1B only two or three assay sets were used to determine the average. The smaller number of measurements for PTP1B and MCL1 coincide with a decrease in computational prediction accuracy here and in our previous studies.^{20,22} Furthermore, accuracy among the less promising inhibitors might be less thoroughly pursued, creating a bias. Whereas computational methods have plenty of potential for improvement, consistent higher quality experimental data will allow for more meaningful comparisons.

6. CONCLUSIONS

The newly implemented python software TIES 20 employs the latest coding standards including modular object oriented design, unit testing, and continuous integration. The molecule-superimposition functionality from the RDKit package used in TIES 17 was replaced with a flexible and exhaustive recursive joint-traversal superimposition algorithm. This transition permitted finer control over matching atoms across compared ligands, such as the aforementioned enabling of partial ring morphing.

The matching of partial rings reduced the size of the alchemical regions, an effect that was particularly pronounced in TIES 20 with the BCC charges because of the local-to-the-

mutation changes in charges. We documented a clear relationship between the size of the alchemical region and the variation in the ensemble as described by the TIES error. Consequently, the smaller alchemical region manifests a global decrease in the TIES 20 and TIES 20 BCC errors to 0.20 kcal/mol, from $\sigma = 0.31$ kcal/mol in TIES 17. This decrease in the σ value means that TIES 20 achieves higher precision for the same amount of computational resources. This improvement further reduces the occurrence of outliers created when sampling transformations involving large alchemical regions.

We have shown that TIES 20 with the RESP charges and enabled partial rings morphing performs equally well in terms of accuracy to TIES 17, despite the variations from protein to protein, while TIES 20 with the AM1-BCC charges improves the performance by slightly over 10%, on average. It should be noted that whether RESP or BCC charges produce more accurate prediction is protein dependent. Furthermore, the improvement is in BCC accuracy.

An increase in sampling where the number of replicas per λ was increased to 20 had no overall effect on the TIES accuracy (TIES values computed relative binding free energies with respect to experiment). This likely means that the TIES ensemble counting 5 replicas per λ window compensates for the sampling issues created by the larger alchemical regions. However, the non-normal nature of the relative free energy distribution has important consequences.

We also showed that different alchemical transformations vary in the number of replicas per λ window they require and discussed the possibility of adjusting the ensemble size on-the-fly for each transformation in order to maximize gains in precision per computational costs. However, our simpler approach with an ensemble size of 5 replicas per λ window is shown to be a good trade-off between computational cost and precision. Increasing the number of replicas per λ window up to 20 has little impact in most cases, with similar results shown in TIES 17.²⁰ There was only one exception to this, PTP1B L6-L14, where the initially poor estimate was significantly improved with more replicas. However, for a large majority of the cases, a very good performance was achieved with TIES 20 using the ensemble approach counting 5 replicas per λ window.

Furthermore, we showed how employing TIES with an ensemble of 5 replicas improves the reproducibility and reliability of the results. The standard deviation of the distribution of $\Delta\Delta G$ values obtained with 5 replicas was, on average, 2.2 ($\approx\sqrt{5}$) smaller than when 1 replica per λ is used. For example, the transformation L3-L23 in PTP1B has a 95% chance of reproducing the $\Delta\Delta G$ within 0.62 kcal/mol using TIES with 5 replicas. However, using only one replica per λ window increases that to 1.42 kcal/mol. In addition, the results vary by as much as 7 kcal/mol for one representative case using single replica approach. These findings further highlight the dispersion in the underlying data and the necessity of both, sufficient sampling and discussion of uncertainty in the results.

Finally, we have discussed further potential improvements that TIES 20 could incorporate. With the new underlying flexible maximum common substructure implementation it is possible to evaluate and support new features such as the hybrid single-dual topology alchemical transformations, as well as other schemes that further decrease the size of the alchemical region.

■ ASSOCIATED CONTENT

SI Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/acs.jctc.0c01179>.

Original version of Table 1 where descriptors were calculated using data set without excluding PTP1B outlier (PDF)

■ AUTHOR INFORMATION

Corresponding Author

Peter V. Coveney – Centre for Computational Science, Department of Chemistry, University College London, London WC1H 0AJ, United Kingdom; orcid.org/0000-0002-8787-7256; Email: p.v.coveney@ucl.ac.uk

Authors

Mateusz K. Bieniek – Centre for Computational Science, Department of Chemistry, University College London, London WC1H 0AJ, United Kingdom

Agastya P. Bhati – Centre for Computational Science, Department of Chemistry, University College London, London WC1H 0AJ, United Kingdom

Shunzhou Wan – Centre for Computational Science, Department of Chemistry, University College London, London WC1H 0AJ, United Kingdom; orcid.org/0000-0001-7192-1999

Complete contact information is available at: <https://pubs.acs.org/doi/10.1021/acs.jctc.0c01179>

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

The authors would like to acknowledge funding support from (i) the European Union's Horizon 2020 Research and Innovation Programme under grant agreement 823712 (CompBioMed2, combiomed.eu) and (ii) NSF Award (https://www.nsf.gov/awardsearch/showAward?AWD_ID=1713749, Award No. NSF 1713749), (iii) the support of EPSRC via the 2020 Science programme (<http://www.2020science.net/>, EP/I017909/1), the Qatar National Research Fund (7-1083-1-191), the MRC Medical Bioinformatics project (MR/L016311/1), the EU H2020 projects ComPat (<https://www.compatproject.eu/>, Grant No. 671564), and funding from the UCL Provost. We acknowledge the Gauss Centre for Supercomputing for providing computing time on the GCS supercomputer SuperMUC-NG (<https://doku.lrz.de/display/PUBLIC/SuperMUC-NG>) at Leibniz Supercomputing Centre under project COVID-19-SNG1 and the very able assistance of its scientific support staff. Furthermore, the authors made use of the supercomputers Scafell Pike (<http://hartree.stfc.ac.uk>) at the Hartree Centre (<https://www.hartree.stfc.ac.uk/>) and Frontera (<https://www.tacc.utexas.edu/systems/frontera>), the latter being accessed via the NSF award. Thanks to Paul Smith (KCL) and David W. Wright for their advice and critical discussions and to Lisa Grant for editorial help. Simulations were carried out with NAMD, which was developed by the Theoretical and Computational Biophysics Group in the Beckman Institute for Advanced Science and Technology at the University of Illinois at Urbana-Champaign.

■ REFERENCES

- (1) Cournia, Z.; Allen, B.; Sherman, W. Relative Binding Free Energy Calculations in Drug Discovery: Recent Advances and Practical Considerations. *J. Chem. Inf. Model.* **2017**, *57*, 2911–2937.
- (2) Sterling, T.; Irwin, J. J. ZINC 15—Ligand Discovery for Everyone. *J. Chem. Inf. Model.* **2015**, *55*, 2324–37.
- (3) Wishart, D. S.; Feunang, Y. D.; Guo, A. C.; Lo, E. J.; Marcu, A.; Grant, J. R.; Sajed, T.; Johnson, D.; Li, C.; Sayeeda, Z.; Assempour, N.; Iynkkaran, I.; Liu, Y.; Maciejewski, A.; Gale, N.; Wilson, A.; Chin, L.; Cummings, R.; Le, D.; Pon, A.; Knox, C.; Wilson, M. DrugBank 5.0: a major update to the DrugBank database for 2018. *Nucleic Acids Res.* **2018**, *46*, D1074–D1082.
- (4) Davies, M.; Nowotka, M.; Papadatos, G.; Dedman, N.; Gaulton, A.; Atkinson, F.; Bellis, L.; Overington, J. P. ChEMBL web services: streamlining access to drug discovery data and utilities. *Nucleic Acids Res.* **2015**, *43*, W612–W620.
- (5) Gromski, P. S.; Henson, A. B.; Granda, J. M.; Cronin, L. How to explore chemical space using algorithms and automation. *Nature Reviews Chemistry* **2019**, *3*, 119–128.
- (6) Shivakumar, D.; Harder, E.; Damm, W.; Friesner, R. A.; Sherman, W. Improving the Prediction of Absolute Solvation Free Energies Using the Next Generation OPLS Force Field. *J. Chem. Theory Comput.* **2012**, *8*, 2553–2558.
- (7) Roos, K.; Wu, C.; Damm, W.; Reboul, M.; Stevenson, J. M.; Lu, C.; Dahlgren, M. K.; Mondal, S.; Chen, W.; Wang, L.; Abel, R.; Friesner, R. A.; Harder, E. D. OPLS3e: Extending Force Field Coverage for Drug-Like Small Molecules. *J. Chem. Theory Comput.* **2019**, *15*, 1863–1874.
- (8) Wang, L.; Wu, Y.; Deng, Y.; Kim, B.; Pierce, L.; Krilov, G.; Lupyan, D.; Robinson, S.; Dahlgren, M. K.; Greenwood, J.; Romero, D. L.; Masse, C.; Knight, J. L.; Steinbrecher, T.; Beuming, T.; Damm, W.; Harder, E.; Sherman, W.; Brewer, M.; Wester, R.; Murcko, M.; Frye, L.; Farid, R.; Lin, T.; Mobley, D. L.; Jorgensen, W. L.; Berne, B. J.; Friesner, R. A.; Abel, R. Accurate and Reliable Prediction of Relative Ligand Binding Potency in Prospective Drug Discovery by Way of a Modern Free-Energy Calculation Protocol and Force Field. *J. Am. Chem. Soc.* **2015**, *137*, 2695–2703.
- (9) 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.
- (10) Mobley, D. L.; Bannan, C. C.; Rizzi, A.; Bayly, C. I.; Chodera, J. D.; Lim, V. T.; Lim, N. M.; Beauchamp, K. A.; Slochow, D. R.; Shirts, M. R.; Gilson, M. K.; Eastman, P. K. Escaping Atom Types in Force Fields Using Direct Chemical Perception. *J. Chem. Theory Comput.* **2018**, *14*, 6076–6092.
- (11) Mobley, D. L.; Gilson, M. K. Predicting Binding Free Energies: Frontiers and Benchmarks. *Annu. Rev. Biophys.* **2017**, *46*, 531–558.
- (12) Straatsma, T. P.; Berendsen, H. J. C. Free energy of ionic hydration: Analysis of a thermodynamic integration technique to evaluate free energy differences by molecular dynamics simulations. *J. Chem. Phys.* **1988**, *89*, 5876–5886.
- (13) Bruckner, S.; Boresch, S. Efficiency of alchemical free energy simulations. II. Improvements for thermodynamic integration. *J. Comput. Chem.* **2011**, *32*, 1320–1333.
- (14) Zwanzig, R. W. High-Temperature Equation of State by a Perturbation Method. II. Polar Gases. *J. Chem. Phys.* **1955**, *23*, 1915–1922.
- (15) Bennett, C. H. Efficient estimation of free energy differences from Monte Carlo data. *J. Comput. Phys.* **1976**, *22*, 245–268.
- (16) Shirts, M. R.; Bair, E.; Hooker, G.; Pande, V. S. Equilibrium Free Energies from Nonequilibrium Measurements Using Maximum-Likelihood Methods. *Phys. Rev. Lett.* **2003**, *91*, 140601.
- (17) Shirts, M. R.; Chodera, J. D. Statistically optimal analysis of samples from multiple equilibrium states. *J. Chem. Phys.* **2008**, *129*, 124105.
- (18) Williams-Noonan, B. J.; Yuriev, E.; Chalmers, D. K. Free Energy Methods in Drug Design: Prospects of “Alchemical

Perturbation" in Medicinal Chemistry. *J. Med. Chem.* **2018**, *61*, 638–649.

(19) Loeffler, H. H.; Michel, J.; Woods, C. FESetup: Automating Setup for Alchemical Free Energy Simulations. *J. Chem. Inf. Model.* **2015**, *55*, 2485–2490.

(20) Bhati, A. P.; Wan, S.; Wright, D. W.; Coveney, P. V. Rapid, Accurate, Precise, and Reliable Relative Free Energy Prediction Using Ensemble Based Thermodynamic Integration. *J. Chem. Theory Comput.* **2017**, *13*, 210–222.

(21) Wan, S.; Bhati, A. P.; Zasada, S. J.; Wall, I.; Green, D.; Bamborough, P.; Coveney, P. V. Rapid and Reliable Binding Affinity Prediction of Bromodomain Inhibitors: A Computational Study. *J. Chem. Theory Comput.* **2017**, *13*, 784–795.

(22) Wan, S.; Tresadern, G.; Pérez-Benito, L.; Vlijmen, H.; Coveney, P. V. Accuracy and Precision of Alchemical Relative Free Energy Predictions with and without Replica-Exchange. *Advanced Theory and Simulations* **2020**, *3*, 1900195.

(23) Bhati, A. P.; Wan, S.; Hu, Y.; Sherborne, B.; Coveney, P. V. Uncertainty Quantification in Alchemical Free Energy Methods. *J. Chem. Theory Comput.* **2018**, *14*, 2867–2880.

(24) Bhati, A. P.; Wan, S.; Coveney, P. V. Ensemble-Based Replica Exchange Alchemical Free Energy Methods: The Effect of Protein Mutations on Inhibitor Binding. *J. Chem. Theory Comput.* **2019**, *15*, 1265–1277.

(25) Wan, S.; Sinclair, R. C.; Coveney, P. V. Uncertainty Quantification in Classical Molecular Dynamics. Philosophical Transactions of the Royal Society A: 2020. <https://arxiv.org/abs/2006.07104> (accessed 2021-01-20).

(26) Knapp, B.; Ospina, L.; Deane, C. M. Avoiding False Positive Conclusions in Molecular Simulation: The Importance of Replicas. *J. Chem. Theory Comput.* **2018**, *14*, 6127–6138.

(27) 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* **2019**, *1*, 5067.

(28) Wang, D. D.; Zhu, M.; Yan, H. Computationally predicting binding affinity in protein–ligand complexes: free energy-based simulations and machine learning-based scoring functions. *Briefings Bioinf.* **2020**, bbaa107.

(29) Saadi, A. A.; Alfe, D.; Babuji, Y.; Bhati, A.; Blaiszik, B.; Brettin, T.; Chard, K.; Chard, R.; Coveney, P.; Trifan, A.; Brace, A.; Clyde, A.; Foster, I.; Gibbs, T.; Jha, S.; Keipert, K.; Kurth, T.; Kranzlmüller, D.; Lee, H.; Li, Z.; Ma, H.; Merzky, A.; Mathias, G.; Partin, A.; Yin, J.; Ramanathan, A.; Shah, A.; Stern, A.; Stevens, R.; Tan, L.; Titov, M.; Tsaris, A.; Turilli, M.; Dam, H. V.; Wan, S.; Wifling, D. IMPECCABLE: Integrated Modeling PipelinE for COVID Cure by Assessing Better LEads. 2020. <https://arxiv.org/abs/2010.06574> (accessed 2021-01-20).

(30) Lee, H.; Merzky, A.; Tan, L.; Titov, M.; Turilli, M.; Alfe, D.; Bhati, A.; Brace, A.; Clyde, A.; Coveney, P.; Ma, H.; Ramanathan, A.; Stevens, R.; Trifan, A.; Dam, H. V.; Wan, S.; Wilkinson, S.; Jha, S. Scalable HPC and AI Infrastructure for COVID-19 Therapeutics. 2020. <https://arxiv.org/abs/2010.10517> (accessed 2021-01-20).

(31) Rufa, D. A.; Bruce Macdonald, H. E.; Fass, J.; Wieder, M.; Grinaway, P. B.; Roitberg, A. E.; Isayev, O.; Chodera, J. D. Towards chemical accuracy for alchemical free energy calculations with hybrid physics-based machine learning/molecular mechanics potentials. *bioRxiv* 2020, 2020.07.29.227959. <https://www.biorxiv.org/content/10.1101/2020.07.29.227959v1> (accessed 2021-01-20).

(32) Scheen, J.; Wu, W.; Mey, A. S. J. S.; Tosco, P.; Mackey, M.; Michel, J. Hybrid Alchemical Free Energy/Machine-Learning Methodology for the Computation of Hydration Free Energies. *J. Chem. Inf. Model.* **2020**, *60*, 5331.

(33) Nguyen, D. D.; Cang, Z.; Wu, K.; Wang, M.; Cao, Y.; Wei, G.-W. Mathematical deep learning for pose and binding affinity prediction and ranking in D3R Grand Challenges. *J. Comput.-Aided Mol. Des.* **2019**, *33*, 71–82.

(34) Straatsma, T. P.; McCammon, J. A. Multiconfiguration thermodynamic integration. *J. Chem. Phys.* **1991**, *95*, 1175–1188.

(35) Coveney, P. V.; Wan, S. On the calculation of equilibrium thermodynamic properties from molecular dynamics. *Phys. Chem. Chem. Phys.* **2016**, *18*, 30236–30240.

(36) Wan, S.; Bhati, A. P.; Zasada, S. J.; Coveney, P. V. Rapid, accurate, precise and reproducible ligand–protein binding free energy prediction. *Interface Focus* **2020**, *10*, 20200007.

(37) Zagrovic, B.; van Gunsteren, W. F. Computational Analysis of the Mechanism and Thermodynamics of Inhibition of Phosphodiesterase 5A by Synthetic Ligands. *J. Chem. Theory Comput.* **2007**, *3*, 301–311.

(38) Lawrenz, M.; Baron, R.; McCammon, J. A. Independent-Trajectories Thermodynamic-Integration Free-Energy Changes for Biomolecular Systems: Determinants of H5N1 Avian Influenza Virus Neuraminidase Inhibition by Peramivir. *J. Chem. Theory Comput.* **2009**, *5*, 1106–1116.

(39) Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kalé, L.; Schulten, K. Scalable molecular dynamics with NAMD. *J. Comput. Chem.* **2005**, *26*, 1781–1802.

(40) Case, D. A.; Belfon, K.; Ben-Shalom, I. Y.; Brozell, S. R.; Cerutti, D. S.; Chaeatham, T. E., III; Cruzeiro, V. W. D.; Darden, T. A.; Duke, R. E.; Giambasu, G.; Gilson, M. K.; Gohlke, H.; Goetz, A. W.; Harris, R.; Izadi, S.; Izmailov, S. A.; Kasavajhala, K.; Kovalenko, A.; Krasny, R.; Kurtzman, T.; Lee, T. S.; LeGrand, S.; Li, P.; Lin, C.; Liu, J.; Luchko, T.; Luo, R.; Man, V.; Merz, K. M.; Miao, Y.; Mikhailovskii, O.; Monard, G.; Nguyen, H.; Onufriev, A.; Pan, F.; Pantano, S.; Qi, R.; Roe, D. R.; Roitberg, A.; Sagui, C.; Schott-Verdugo, S.; Shen, J.; Simmerling, C. L.; Skrynnikov, N. R.; Smith, J.; Swails, J.; Walker, R. C.; Wang, J.; Wilson, L.; Wolf, R. M.; Wu, X.; Xiong, Y.; Xue, Y.; York, D. M.; Kollman, P. A. AMBER; 2020. <http://ambermd.org> (accessed 2021-01-20).

(41) Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. Development and testing of a general amber force field. *J. Comput. Chem.* **2004**, *25*, 1157–1174.

(42) Lindorff-Larsen, K.; Piana, S.; Palmo, K.; Maragakis, P.; Klepeis, J. L.; Dror, R. O.; Shaw, D. E. Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins: Struct., Funct., Bioinf.* **2010**, *78*, 1950–8.

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

(44) Jakalian, A.; Jack, D. B.; Bayly, C. I. Fast, efficient generation of high-quality atomic charges. AM1-BCC model: II. Parameterization and validation. *J. Comput. Chem.* **2002**, *23*, 1623–1641.

(45) Chen, H.; Maia, J.; 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.* **2020**, *60*, 5301.

(46) Lee, T.-S.; Hu, Y.; Sherborne, B.; Guo, Z.; York, D. M. Toward Fast and Accurate Binding Affinity Prediction with pmemdGTI: An Efficient Implementation of GPU-Accelerated Thermodynamic Integration. *J. Chem. Theory Comput.* **2017**, *13*, 3077–3084.

(47) Gowers, R.; Linke, M.; Barnoud, J.; Reddy, T.; Melo, M.; Seyler, S.; Domański, J.; Dotson, D.; Buchoux, S.; Kenney, I.; Beckstein, O. MDAnalysis: A Python Package for the Rapid Analysis of Molecular Dynamics Simulations. *Proceedings of the 15th Python in Science Conference* **2016**, 98–105.

(48) van der Walt, S.; Colbert, S. C.; Varoquaux, G. The NumPy Array: A Structure for Efficient Numerical Computation. *Comput. Sci. Eng.* **2011**, *13*, 22–30.

(49) Virtanen, P.; Gommers, R.; Oliphant, T. E.; Haberland, M.; Reddy, T.; Cournapeau, D.; Burovski, E.; Peterson, P.; Weckesser, W.; Bright, J.; van der Walt, S. J.; Brett, M.; Wilson, J.; Jarrod Millman, K.; Mayorov, N.; Nelson, A. R. J.; Jones, E.; Kern, R.; Larson, E.; Carey, C.; Polat, İ.; Feng, Y.; Moore, E. W.; Vand erPlas, J.; Laxalde, D.; Perktold, J.; Cimrman, R.; Henriksen, I.; Quintero, E. A.; Harris, C. R.; Archibald, A. M.; Ribeiro, A. H.; Pedregosa, F.; van Mulbregt, P.;

Contributors, S. SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nat. Methods* **2020**, *17*, 261–272.

(50) Hunter, J. D. Matplotlib: A 2D Graphics Environment. *Comput. Sci. Eng.* **2007**, *9*, 90–95.

(51) Dakka, J.; Farkas-Pall, K.; Turilli, M.; Wright, D. W.; Coveney, P. V.; Jha, S. Concurrent and Adaptive Extreme Scale Binding Free Energy Calculations. *IEEE 14th International Conference on e-Science (e-Science)*; 2018; pp 189–200, DOI: 10.1109/eScience.2018.00034.

(52) Jiang, W.; Chipot, C.; Roux, B. Computing Relative Binding Affinity of Ligands to Receptor: An Effective Hybrid Single-Dual-Topology Free-Energy Perturbation Approach in NAMD. *J. Chem. Inf. Model.* **2019**, *59*, 3794–3802.

(53) Liu, S.; Wang, L.; Mobley, D. L. Is Ring Breaking Feasible in Relative Binding Free Energy Calculations? *J. Chem. Inf. Model.* **2015**, *55*, 727–735.

(54) Wang, L.; Deng, Y.; Wu, Y.; Kim, B.; LeBard, D. N.; Wandschneider, D.; Beachy, M.; Friesner, R. A.; Abel, R. Accurate Modeling of Scaffold Hopping Transformations in Drug Discovery. *J. Chem. Theory Comput.* **2017**, *13*, 42–54.

(55) Turilli, M.; Balasubramanian, V.; Merzky, A.; Paraskevagos, I.; Jha, S. Middleware Building Blocks for Workflow Systems. *Comput. Sci. Eng.* **2019**, *21*, 62–75.

(56) Brown, S. P.; Muchmore, S. W.; Hajduk, P. J. Healthy skepticism: assessing realistic model performance. *Drug Discovery Today* **2009**, *14*, 420–427.

(57) Kramer, C.; Kalliokoski, T.; Gedeck, P.; Vulpetti, A. The Experimental Uncertainty of Heterogeneous Public Ki Data. *J. Med. Chem.* **2012**, *55*, 5165–5173.

(58) Kramer, C.; Dahl, G.; Tyrchan, C.; Ulander, J. A comprehensive company database analysis of biological assay variability. *Drug Discovery Today* **2016**, *21*, 1213–1221.

(59) Baum, B.; Mohamed, M.; Zayed, M.; Gerlach, C.; Heine, A.; Hangauer, D.; Klebe, G. More than a Simple Lipophilic Contact: A Detailed Thermodynamic Analysis of Nonbasic Residues in the S1 Pocket of Thrombin. *J. Mol. Biol.* **2009**, *390*, 56–69.

(60) Hardcastle, I. R.; Arris, C. E.; Bentley, J.; Thomas Boyle, F.; Chen, Y.; Curtin, N. J.; Endicott, J. A.; Gibson, A. E.; Golding, B. T.; Griffin, R.; Jewsbury, P.; Menyerol, J.; Mesguiche, V.; Newell, D.; Noble, M. E. M.; Pratt, D. J.; Wang, L.-Z.; Whitfield, H. J. N2-Substituted O6-Cyclohexylmethylguanine Derivatives: Potent Inhibitors of Cyclin-Dependent Kinases 1 and 2. *J. Med. Chem.* **2004**, *47*, 3710–3722.

(61) Wilson, D. P.; Wan, Z.-K.; Xu, W.-X.; Kirincich, S. J.; Follows, B. C.; Joseph-McCarthy, D.; Foreman, K.; Moretto, A.; Wu, J.; Zhu, M.; Binnun, E.; Zhang, Y.-L.; Tam, M.; Erbe, D. V.; Tobin, J.; Xu, X.; Leung, L.; Shilling, A.; Tam, S. Y.; Mansour, T. S.; Lee, J. Structure-Based Optimization of Protein Tyrosine Phosphatase 1B Inhibitors: From the Active Site to the Second Phosphotyrosine Binding Site. *J. Med. Chem.* **2007**, *50*, 4681–4698.

(62) Friberg, A.; Vigil, D.; Zhao, B.; Daniels, R. N.; Burke, J. P.; Garcia-Barrantes, P. M.; Camper, D.; Chauder, B. A.; Lee, T.; Olejniczak, E. T.; Fesik, S. W. Discovery of Potent Myeloid Cell Leukemia 1 (Mcl-1) Inhibitors Using Fragment-Based Methods and Structure-Based Design. *J. Med. Chem.* **2013**, *56*, 15–30.

(63) Liang, J.; Tsui, V.; Van Abbema, A.; Bao, L.; Barrett, K.; Beresini, M.; Berezhkovskiy, L.; Blair, W. S.; Chang, C.; Driscoll, J.; Eigenbrot, C.; Ghilardi, N.; Gibbons, P.; Halladay, J.; Johnson, A.; Kohli, P. B.; Lai, Y.; Liimatta, M.; Mantik, P.; Menghrajani, K.; Murray, J.; Sambrone, A.; Xiao, Y.; Shia, S.; Shin, Y.; Smith, J.; Sohn, S.; Stanley, M.; Ultsch, M.; Zhang, B.; Wu, L. C.; Magnuson, S. Lead identification of novel and selective TYK2 inhibitors. *Eur. J. Med. Chem.* **2013**, *67*, 175–187.

(64) Liang, J.; van Abbema, A.; Balazs, M.; Barrett, K.; Berezhkovskiy, L.; Blair, W.; Chang, C.; Delarosa, D.; DeVoss, J.; Driscoll, J.; Eigenbrot, C.; Ghilardi, N.; Gibbons, P.; Halladay, J.; Johnson, A.; Kohli, P. B.; Lai, Y.; Liu, Y.; Lyssikatos, J.; Mantik, P.; Menghrajani, K.; Murray, J.; Peng, I.; Sambrone, A.; Shia, S.; Shin, Y.; Smith, J.; Sohn, S.; Tsui, V.; Ultsch, M.; Wu, L. C.; Xiao, Y.; Yang, W.; Young, J.; Zhang, B.; Zhu, B.-y.; Magnuson, S. Lead Optimization of a

4-Aminopyridine Benzamide Scaffold To Identify Potent, Selective, and Orally Bioavailable TYK2 Inhibitors. *J. Med. Chem.* **2013**, *56*, 4521–4536.