Assessment of Binding Affinity via Alchemical Free Energy Calculations

OPEN ACCESS

Assessment of Binding Affinity via Alchemical Free Energy Calculations

Maximilian Kuhn,[a,b] Stuart Firth-Clark,[a] Paolo Tosco,[a] Antonia S.J.S. Mey,[b] Mark Mackey,[a,] * Julien Michel[b]

[a] Cresset, New Cambridge House, Bassingbourn Road, Litlington SG8 0O5, Cambridgeshire, United Kingdom

[b] EaStCHEM School of Chemistry, University of Edinburgh, David Brewster Road, Edinburgh EH9 3FJ, United Kingdom

ABSTRACT

Free energy calculations have seen increased usage in structure-based drug design. Despite the rising interest, automation of the complex calculations and subsequent analysis of their results are still hampered by the restricted choice of available tools. In this work, an application for automated setup and processing of free energy calculations is presented. Several sanity checks for assessing the reliability of the calculations were implemented, constituting a distinct advantage over existing open-source tools. The underlying workflow is built on top of the software Sire, SOMD, BioSimSpace and OpenMM and uses the AMBER14SB and GAFF2.1

force fields. It was validated on two datasets originally composed by Schrödinger, consisting of 14 protein structures and 220 ligands. Predicted binding affinities were in good agreement with experimental values. For the larger dataset the average correlation coefficient R_p was 0.70 ± 0.05 and average Kendall's τ was 0.53 ± 0.05 which is broadly comparable to or better than previously reported results using other methods.

1. INTRODUCTION

Reliable prediction of ligand binding affinities is highly desirable in the early stages of drug discovery projects. During the last few years alchemical free energy calculations have emerged as a prominent tool for this task.[1–3] Such calculations have been applied to predict ligand binding affinities of large datasets, yielding on average predictions accurate to within 0.8 kcal/mol, thus adding significant value to drug discovery projects.[4–7] However, lack of automation is still a major obstacle to a more routine application of these methods, despite several tools to address this being in development.[8–11] Furthermore, carrying out alchemical free energy calculations requires expert knowledge, since assessing the quality and validity of these simulations can be non-trivial.[12,13] Different approaches in free energy prediction methods have been assessed through community challenges such as D3R[14] or SAMPL[15] using various datasets.

Alchemical free energy calculations for computing the binding energy of a compound A relative to a reference compound B generally involve a thermodynamic cycle in which A is transformed into B in the protein environment and in simple solvent (Figure 1). The corresponding states of A and B need to have significant overlap in their potential energy distributions, as otherwise the configurations sampled for A may be of low probability for B, which leads to slow convergence of the free energy change and computation of the result may not be possible.[16] This limitation is typically overcome by introducing a coupling parameter λ

which takes on values between 0 and 1 to interpolate the potential energy functions of the end-states. The transformation from A to B is then achieved through a varying number of artificial ("alchemical") states characterized by intermediate λ values (the so-called λ-windows) providing an improved overlap of potential energy distributions between neighboring states.

There are multiple different approaches available for computing binding free energies from the potential function distributions. These include Free Energy Perturbation (FEP)[17], Thermodynamic Integration (TI)[18,19] and the Multistate Bennett Acceptance Ratio (MBAR) estimator.[20,21] For a full review of the advantages and disadvantages of each method see [12,13,22].
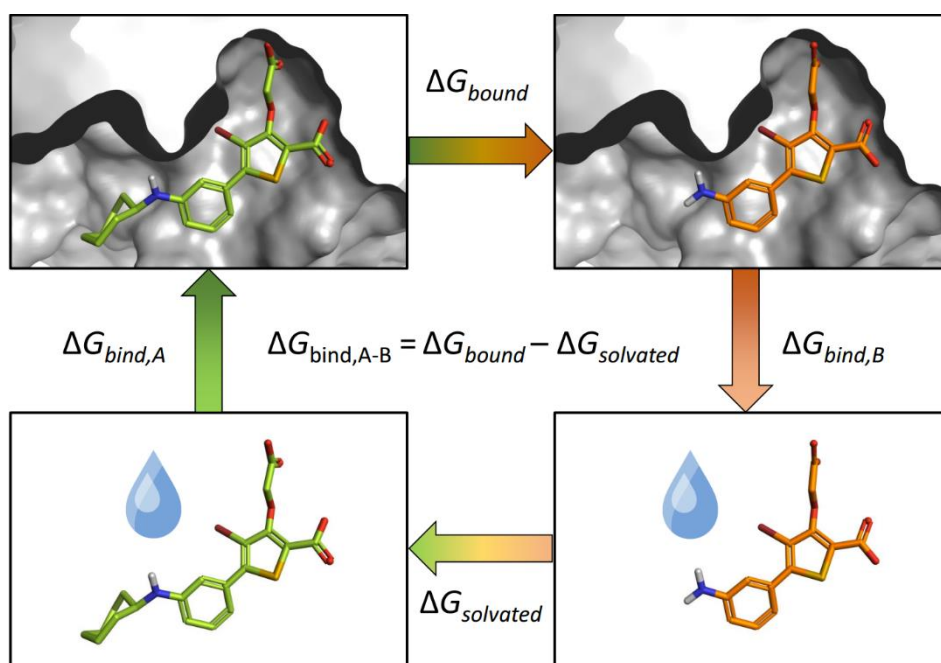

Figure 1. Schematic representation of free energy calculations. The binding free energy of a compound is calculated according to the thermodynamic cycle which consists of four different states: the protein-ligand complexes of molecules A and B, as well as A and B in a water box. A is transformed into B through a number of alchemical intermediates. As the molecular dynamics

simulations are carried out in a isothermal-isobaric ensemble, Gibb's free energy ΔG is used instead of Helmholtz' free energy ΔF.

There are several commercial and open-source options available for carrying out free energy perturbation (FEP) calculations, the most well-known being the package provided by Schrödinger (referred to as FEP+). FEP+ was originally validated on a large dataset of eight proteins and 199 ligands using the OPLS2.1 forcefield and replica exchange with solute tempering (REST).[1] More recently, two other approaches utilizing AMBER[23] and GROMACS[24] have also been used to study this dataset. Comparative analysis of the performance of different FEP implementations is important to identify areas where further methodological efforts can be concentrated to improve FEP methodologies. Recent efforts from the field have highlighted that the results of an FEP calculation depend on several parameters other than the potential energy function, for instance input preparation, simulation conditions and choice of the free energy analysis method.[25,26]

Most recently, relative binding free energy (RBFE) calculations were implemented in Cresset's structure-based drug design suite, Flare.[27,28] This approach is based on the Sire[29] framework, and leverages several open source tools, namely Biosimspace,[11] SOMD,[30] OpenMM,[31] AmberTools[32] and a customized version of LOMAP.[8] Automated setup, deployment on remote computing instances, and analysis of free energy calculations can be accomplished through a graphical user interface (GUI) and a Python application programming interface (API). Automation and accessibility, though highly desirable, critically depend on a robust sanity check and error analysis of simulation results. Common sanity indicators for alchemical calculations are monitored to detect potential issues; among these indicators are hysteresis of forward and

backward transformations, cycle closure errors, and poor overlap between neighboring λ windows.[12,33,34]

The SOMD package available within the Sire framework has been successfully applied to relative and absolute binding free energy studies of a range of fragments, drug-like small molecules, carbohydrates, host-guest systems.[5–7,30,35–38]

However, a systematic comparison on standard datasets has not been published to date. Here we report a large-scale validation of Flare's FEP implementation on the Wang et al.[1] dataset, as well as on a smaller dataset of more challenging scaffold hopping modifications.[39]

2. METHODS

Figure 2. Screenshot from a finished FEP project in Flare. The perturbation network and the calculated relative free energies can be found in the central panel. Arrows are color-coded according to the hysteresis of the perturbations in both directions. The other four panels contain the overlap matrices (upper left), a plot comparing computed and experimental free energies

(lower left), a 3D window containing screenshots from the molecular dynamics simulations (upper right) and a color-coded table listing the cycle closure errors (lower right).

2.1 Topology and perturbation networks. Free energy calculations may be carried out using a single or a dual topology approach.[13,40] The software packages used in this work use single topology, i.e. the parameters of the perturbed atoms from the starting molecule are changed to represent those in the target molecule. Additional atoms are treated as dummies, i.e. their charge and Lennard Jones potential are set to zero in the corresponding states. Perturbations in which the size of a ring is altered are best performed via an intermediate ringless state (e.g. -cyclopentyl to -cyclobutyl may be attempted via a common -H intermediate). If the ring systems differ substantially in their nature and substitution pattern, adding further intermediates can be beneficial. Hence, it was not considered optimal to use the perturbation networks provided in the Wang et al.[1] study and new networks were created using a customized version of LOMAP[8], in which candidate perturbations which are hard to perform in a single-topology approach such as ring size changes are assigned a zero score, and hence are not allowed in the final perturbation network.

2.2 System preparation. Protein and ligand structures were obtained from the supporting information (SI) of the publications by Wang et al. While protonation states of the ligands were maintained as in the original publication, proteins were re-prepared in Flare[27,28] and protonation states assigned using the TSAR algorithm.[41] For thrombin, only the A chain was used and the missing loop was built from entry 5LPD in the Protein Data Bank[42] using Modeller 9.20.[43] Furthermore, the tartaric acid molecule present in the BACE structure was discarded. The

AMBER 14FFSB force field[44] was used to describe the proteins and GAFF2.1[45] for the ligand, with AM1-BCC charges.[46] Missing force field parameters for the ligands were generated using the parmchk module.[47] The systems were solvated using the TIP3P water model[48] and a cuboid solvation box with a solvation buffer size of 10 Å. Systems were neutralized by adding Na^+ or Cl^- ions as required. Bonds involving hydrogen atoms were constrained using the SHAKE[49] algorithm and a time step of 2 fs was used throughout all simulations. Long-range electrostatic interactions were treated using the reaction-field method.[50] Non-bonded interactions were treated using a 10 Å cut-off. Systems were equilibrated using OpenMM 7.3.0[31] at $\lambda=0$ following a four-step protocol. Details of the equilibration protocol can be found in the SI (Table S2).

2.3 FEP calculations. In contrast to other work reported on these datasets,[1,23,24,39] each perturbation was carried out in two independent simulations, one in each direction. Though this increases the overall computational costs, it provides an opportunity to assess the reliability of the calculations as the hysteresis should ideally be zero. The perturbation networks were drafted using an extended version of LOMAP[8] as implemented in Flare and were then optimized through the user interface by manually adding or removing links from the perturbation map. For several datasets, a few additional compounds were necessary in order to connect all the prepared ligands into one perturbation map. The relative binding free energy change for each perturbation was calculated using a varying number of linearly spaced λ windows (5, 9, 14, 19 or 24), each window being subjected to a 4-ns MD simulation. Larger numbers of λ windows were used for more challenging perturbations, such as those involving ring additions, or non-polar-to-polar group mutations. Electrostatic and van der Waals interactions were described using soft-core potentials (cf. the SI of Loeffler et al.[26] for details). The input files describing the perturbation

were generated using BioSimSpace.[11] Prior to starting the simulations, an energy minimization was carried out running 1,000 steps of the steepest descent algorithm. Simulations were performed using SOMD.[29,30] The first 400 ps of the trajectories of each λ window were discarded prior to free energy estimation to allow for equilibration. Free energies were calculated using MBAR.[20] The final change in the free energy of binding for each ligand pair was calculated by averaging results obtained for the perturbation carried out in both directions; ΔG values were computed from the network according to the method of Xu et al.[51]

2.4 Analysis of the results. This work focuses on the overall correlation for each of the subsets and the mean unsigned error of each compound (MUE_c). Previous publications[1,23,24] primarily aimed at the comparison of the mean unsigned error of each perturbation (MUE_p), i.e. the mean prediction error for each ligand pair. However, strictly speaking comparing MUE_p between different approaches requires the underlying perturbation networks to be identical, which is not the case for the approach presented here (cf. above). To address this situation, the unsigned error $UE_\text{p,A-B}$ for transformations where compounds A and B are connected via an additional intermediate molecule I with unknown experimental affinity was estimated as:

$$UE_\text{p,A-B} = |\Delta\Delta G_\text{calc,A-I} + \Delta\Delta G_\text{calc,I-B} - \Delta\Delta G_\text{exp,A-B}| \qquad (1)$$

Here, $\Delta\Delta G_\text{calc,A-I}$ and $\Delta\Delta G_\text{calc,I-B}$ are the calculated free energies between A and I and I and B, respectively, while $\Delta\Delta G_\text{exp,A-B}$ is the difference in experimentally measured binding affinity between A and B.

As the work of Gapsys et al. did not contain computed ΔG values, these values were obtained from the reported $\Delta\Delta G$ values using Flare's internal protocol (data can be found in the SI).

2.5 λ Schedule. A critical factor for the success of free energy calculations is to ensure sufficient overlap of the phase space probability distributions integrals between all neighboring states.[33,34] Typically, this is guaranteed only when the two states are very similar to each other, which is usually not the case for molecules characterized by significant structural differences. Hence, a sufficiently high number of alchemical intermediate states must be chosen to connect the end states and ensure that the calculated free energy converges in a reasonable amount of time. Whether a perturbation obeys to these criteria can be checked through a so-called phase space overlap matrix.[12] The overlap matrix $\mathcal{O} \in \mathbb{R}^{n \times n}$ (where n is the number of λ windows) contains the probability that a sample from any state can be found in any other state; hence, the matrix is symmetric (with respect to its main diagonal) and all rows and columns sum to 1. The first offset diagonals of the matrix contain the information about neighboring states. It has been reported that a value as low as 0.03 in these diagonals may be sufficient to obtain reliable results.[12] In this study, the overlap matrices were used to guide the selection of the number of λ windows (cf. Figure 3).

.824	.168	.008						
.168	.617	.199	.015					
.008	.199	.549	.222	.022				
	.015	.222	.510	.235	.018			
		.022	.235	.519	.210	.015		
			.018	.210	.532	.224	.016	
				.015	.224	.532	.216	.012
					.016	.216	.550	.217
						.012	.217	.770

.994	.005						
.005	.899	.095					
	.095	0.870	.135				
		.135	.728	.136	.001		
			.136	.718	.144	.001	
			.001	.144	.712	.142	.001
				.001	.142	.732	.125
					.001	.125	.805	.070
							.070	.930

Figure 3. Two example overlap matrices. Good overlap (green cells) can be observed for all neighboring windows in the matrix on the left whereas there is poor overlap (red cells) between the first and the second window of the second perturbation in the matrix on the right. Overlap in the empty cells is smaller than 0.001.

In the previously published benchmark papers discussed above[1,23] the 12 λ-windows were spaced non-linearly (0.00922, 0.04794, 0.11505, 0.20634, 0.31608, 0.43738, 0.56262, 0.68392, 0.79366, 0.88495, 0.95206, 0.99078). In contrast, results in this paper were obtained using linear spacing (e.g. 0.000, 0.125, 0.250, 0.375, 0.500, 0.626, 0.750, 0.875, 1.000 for 9 windows). Neither approach is perfect, and as the matrix can only be calculated once the calculations of all windows have finished, a significant amount of time can be spent on simulations which are then deemed unreliable due to poor phase space overlap. We have attempted to derive heuristics to determine how many λ windows should be used for certain types of perturbations to ensure sufficient overlap using the approach presented here. However, as the windows are spaced linearly, adding more windows also affects already well-sampled areas of the matrix, which increases computation time while only marginally improving the overall quality of the free energy estimate. We are currently investigating other approaches, such as the Conveyor Belt Thermodynamic Integration scheme,[52] adaptive sampling[53] or non-linear λ functions derived from heuristics.

2.6 Protonation state of Cys215 in PTP1B. In a recently published paper[24] the authors argue that the ionization state of Cys215 in the binding pocket of PTP1B, which was set to deprotonated in the previous work by Wang et al.,[1] may be incorrect; in fact, there is only a short distance between the sulfur atom of Cys215 and the closest ligand carboxyl oxygen atom (3.5-

3.9 Å). Furthermore, two different predictors suggest that the pK_a of cysteine in the complexes ranges from 9.8 to 10.5. The authors report a better agreement with experiment for the free energy calculations using their consensus approach when Cys215 is modelled as protonated.[24] In this study, Cys215 was treated as being protonated.

3. RESULTS AND DISCUSSION

3.1 Overall comparison of the FEP+ dataset. The summary of the results of the eight datasets from Wang et al.[1] is reported in Table 1. Details can be found in the SI (Table S4; results may be recalculated using the supplied Jupyter Notebook and input files). The data for Pearson's correlation coefficient R^2 for the experimental and calculated ΔG values, the mean unsigned errors per ligand (MUE$_c$) and perturbation (MUE$_p$) as well as Kendall's τ are also shown in Figure 4. The number of perturbations used for each dataset in this study are roughly similar to those in Wang et al., though for PTP1B only 35 instead of 49 perturbations were performed.

In terms of R, results obtained with Flare showed higher correlation ($\Delta R > 0.10$) than those reported by Song et al.[23] for five of the eight datasets, and comparable for the three remaining datasets (MCL1,[54] PTP1B[55] and BACE[56]). Interestingly, Flare's MUE$_c$ is lower, particularly for the BACE dataset (0.76 ± 0.10 vs. 1.03 ± 0.14 kcal/mol; see Table S4)). As both approaches use the same force field (AMBER 14FFSB), this highlights that accuracy of an FEP implementation depends on several other aspects beyond the choice of a potential energy function.

Our results are overall more predictive than those reported by Wang et al.[1] for thrombin[57] and CDK2[58] ($\Delta R > 0.10$), comparable for MCL1 and p38,[59] and less predictive for TYK2,[60,61] PTP1B and JNK1[62] ($\Delta R < -0.10$). The most striking difference was observed for the BACE dataset, for

which the overall correlation R² was poor (0.19, 95%-CI: [0.08-0.32] vs 0.61, 95%-CI: [0.50-0.71]) but the MUE_c and MUE_p are only slightly worse (0.76 ± 0.10 vs 0.67 ± 0.09 and 0.95 ± 0.10 vs 0.85 ± 0.08 kcal/mol).

Similar performance can be observed when comparing results with the study of Gapsys et al.,[24] who used a consensus approach of GAFF[45] and CGenFF[63] as implemented in pmx.[64] Results obtained with Flare showed higher correlation ($\Delta R > 0.10$) for thrombin and PTP1B, similar for CDK2, TYK2 and MCL1, and poorer correlation for the remaining three other datasets ($\Delta R < -0.10$).

Table 1. Summary for the average correlation coefficient, τ, MUE_c and MUE_p for all datasets.

Parameter	this study	Wang et al.	Song et al.	Gapsys et al., GAFF	Gapsys et al., CGenFF	Gapsys et al., Consensus
R_{avg}	0.70 ± 0.05	0.74 ± 0.05	0.56 ± 0.05	0.65 ± 0.07	0.54 ± 0.09	0.60 ± 0.09
τ_{avg}	0.53 ± 0.05	0.57 ± 0.06	0.41 ± 0.05	0.46 ± 0.06	0.42 ± 0.06	0.47 ± 0.08
$MUE_{c,avg}$	0.83 ± 0.05	0.76 ± 0.04	1.01 ± 0.06	0.74 ± 0.04	0.84 ± 0.05	0.72 ± 0.04
$MUE_{p,avg}$	1.02 ± 0.05	0.90 ± 0.04	1.17 ± 0.05	0.91 ± 0.04	1.09 ± 0.05	0.85 ± 0.05

R_{avg} is the average correlation coefficient and τ_{avg} the average Kendall τ across all eight datasets. $MUE_{c,avg}$ and $MUE_{p,avg}$ give the average MUE per compound. Uncertainties denote the standard error of the mean.

a)

b)

c)

d)

this study	Gapsys et al., GAFF
Wang et al., FEP+	Gapsys et al., CGenFF
Song et al., AMBER-TI	Gapsys et al., GAFF+CGenFF
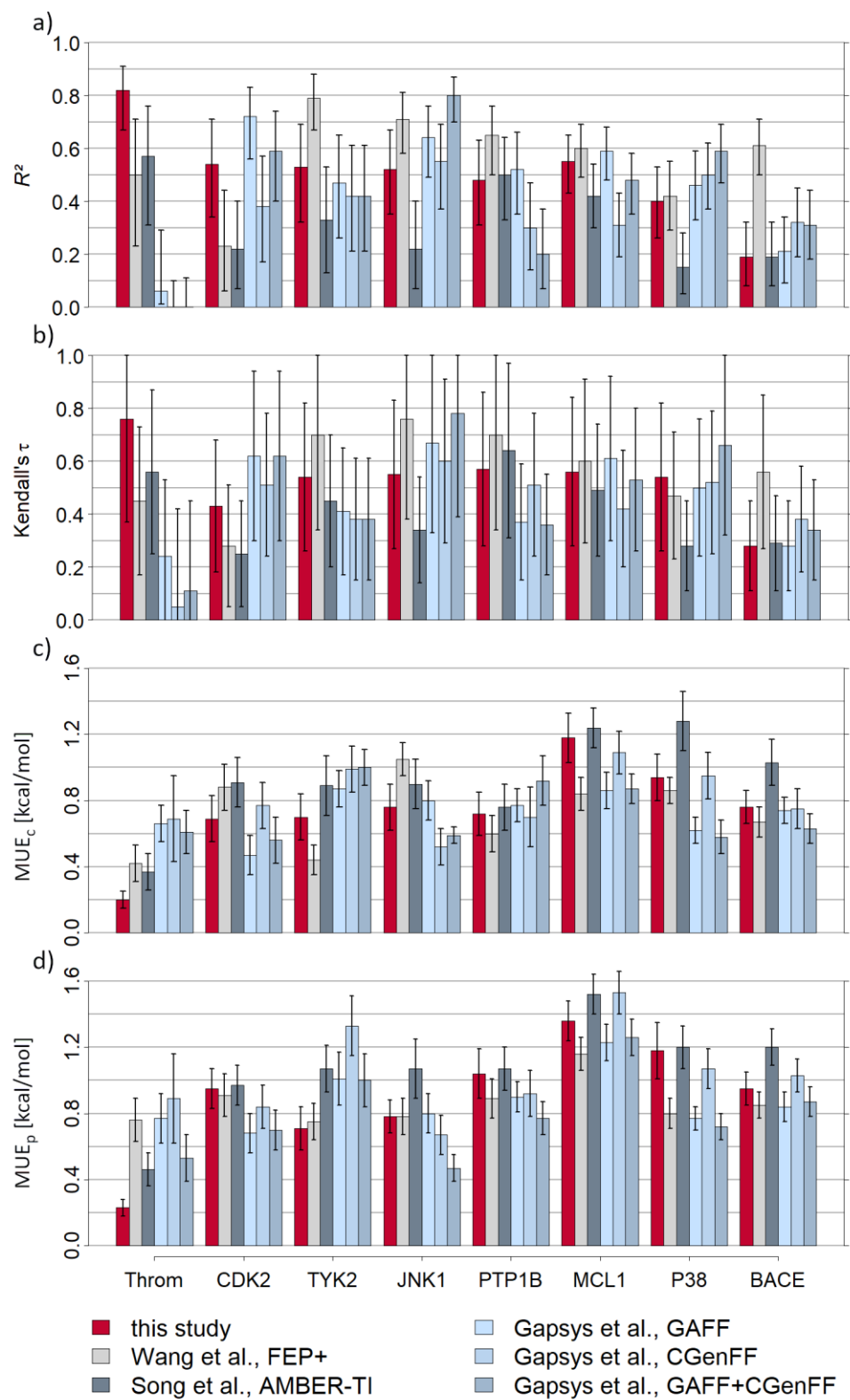
Figure 4. Charts reporting results for each of the eight datasets. a) Pearson's correlation coefficient R^2. b) Kendall's τ. c) Mean Unsigned Error per compound (MUE_c) and d) Mean Unsigned error per perturbation (MUE_p) for each of the eight datasets in kcal/mol. The error bars in a) indicate the 95% confidence interval calculated using the Fisher transformation.[65,66] For b)-d), they denote the standard deviation obtained by bootstrapping using 10,000 iterations.

To allow for easier comparison across all datasets, the average correlation coefficient R_{avg}, Kendall's τ_{avg} and the mean unsigned errors per compound ($MUE_{c,avg}$) and per perturbation ($MUE_{p,avg}$) are reported in Table 1. Overall, the results obtained are slightly worse than those reported by Wang et al.,[1] but better than those reported by Song et al.[23] Compared to Gapsys et al.,[24] results are better than those obtained using CGenFF, and similar to those obtained using GAFF or GAFF+CGenFF. There are several differences among the different free energy approaches which may help to explain these findings. An important factor is the underlying force fields applied to the proteins and ligands. While the same protein force field was used in this work as in the work published by Song et al., the ligands were parameterized using different versions of GAFF (v2.1 vs. v1.8), and different water models were used (TIP3P vs SPC/E). The present protocol used a reaction field treatment of long-range electrostatic interactions, in contrast to a Particle Mesh Ewald treatment in the other studies. However, the results appear overall more accurate than those reported by Song et al., thus there is no evidence that the use of this computationally more efficient protocol degraded accuracy of the calculations on this dataset. As there are other important differences in the perturbation network, sampling time and in the way results are analyzed (MBAR vs. TI), it is difficult to compare the results directly. Even though Gapsys et al. report results closer to those presented here using GAFF v2.1, their

14

non-equilibrium approach cannot be directly compared to the method used in our work. The perturbation network used in this work differs from the one used in the other studies reported,[1,23,24] since the networks used by Wang et al.[1] are not optimal for use with SOMD (cf. section 2.1). For the BACE dataset several intermediate structures had to be introduced to connect all subsets reliably. One subset contains a cyclopropyl group, which is replaced by a phenyl ring in another subset. In order to connect these structural motifs, two intermediate structures (each containing a hydrogen atom at the corresponding position) were introduced. This makes achieving precise results challenging as the relative binding affinity difference between the compounds in both subsets is mainly governed by only a few relatively challenging perturbations.

The sampling time per perturbation is also of interest as computational resources in drug design projects are limited. The approach by Gapsys et al. used 60 (using GAFF) or 120 ns (when using GAFF+CGenFF) for a single perturbation with a 2 fs time step, whereas Song et al. used 74 ns[23,24] and a 1 fs timestep. The method of Wang et al. used 60 ns with a 4 fs timestep, but multiple replicas were run at different temperatures. The perturbation networks used in these studies were identical and contained a total of 330 perturbations. The approach described in this work on average used 90 ns and a 2 fs timestep, with 315 perturbations in total. However, each ligand pair was calculated twice (once in each direction) to allow for an additional reliability estimate. Using only one of the two possible directions for each perturbation was found to only marginally affect the correlation coefficient and MUE, meaning that the average calculation time can be halved to 45 ns without having a major impact on the quality of the results (cf. Table S3 in the SI).

There are several other factors such as ion concentration, distance cut-offs, water models, etc. which have an impact on the predicted binding affinities. However, the impact of these parameters on the free energies cannot be assessed easily. A table summarizing the methodological details of different methodologies can be found in the SI (Table S1).

3.2 Null model. Free energy calculations are an advanced and computationally expensive method. Hence, it is worthwhile to consider the benefit of the calculations compared to simpler models. For example, Song et al. report a null model which considers all ligands to be equally potent.[23] Though this is a "true" null model, it seems oversimplified as even simple molecular descriptors may be able to yield a weak estimate of the binding affinity. An example for such a descriptor is the molecular weight (MW). Broadly speaking, the binding affinity can be expected to increase as the MW increases, since a heavier (and thus probably larger) compound can form more interactions with the binding pocket.

Interestingly, there is indeed a weak correlation between the MW and the experimental binding affinities for most datasets (cf. Table S4 in the SI), except for BACE and TYK2. R^2 for MW in PTP1B is surprisingly high (0.55), reflecting that the dataset contains various ligands of different sizes which are incrementally grown in the additional and mainly hydrophobic parts of the binding pocket. In summary, our results are clearly superior to those obtained by this null model (average $R = 0.70 \pm 0.05$ for the free energy calculations and 0.38 ± 0.11 for the molecular weight), apart from the PTP1B dataset.

3.3 Variability of the results. Another aspect to consider when running free energy calculations is how much results of a given perturbation may vary across different runs starting

from identical input molecules. In fact, the pre-FEP MD equilibration protocols uses randomly sampled velocities from the Maxwell-Boltzmann distribution, and consequently yield different starting 3D coordinates each time they are run. Likewise, different production runs started from the same equilibrated structure will yield different trajectories because velocities are re-initialized at each λ-window. Four selected transformations (named P1-4) covering conservative as well as more challenging structural modifications were performed in quintuplicate to assess the variability of the results. For P1 and P2, only a few atoms are altered during the transformation. In P3 and P4 an aliphatic ring system is introduced (cyclopropyl and cycloheptyl, respectively) posing a more difficult challenge. The results are summarized in Figure 5 and detailed information can be found in the SI (Table S5).
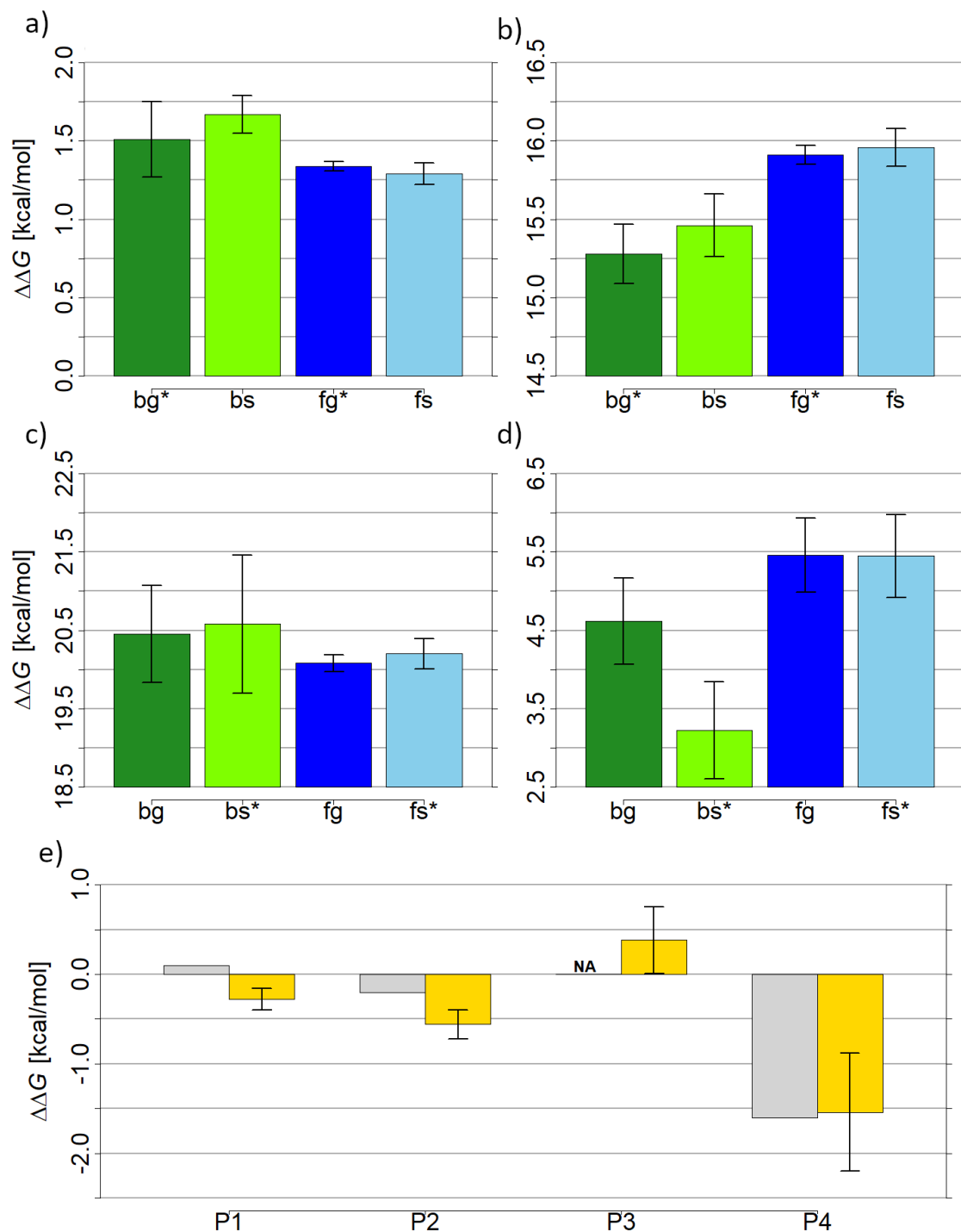
Figure 5. Calculated relative binding affinities for four perturbations from quintuplicate simulations: a) P1: -Cl ⇔ -Br (Thrombin), b) P2: -CH₂- ⇔ -O- (PTP1B), c) P3: -H ⇔ -cyclopropyl (BACE), d) P4: -H ⇔ -cycloheptyl (PTP1B). Results are shown for growing (g) and

shrinking (s) the atoms in the bound (b) and the free (f) state. The absolute values are shown to allow for easier comparison (negative values are marked with an asterisk (*)). e) Comparison of the experimental and computed binding affinity for the molecule pairs (experimental data not available for P3).

The hysteresis between corresponding "growing" and "shrinking" perturbations was found to be small (< 0.20 kcal/mol) after five repeats for the perturbations P1-3, emphasizing the reliability of the presented approach. By contrast, the calculations for the insertion of cycloheptyl in PTP1B (P4) shows a hysteresis of 1.4 kcal/mol. Nevertheless, the calculated free energy over all five replicas in both directions closely matches the experimental value. The standard deviation of the binding free energy estimate for the larger perturbations P3 and P4 is larger (ca. 0.5 kcal/mol) than for the smaller perturbations P1 and P2. Thus, it is advisable to carry out multiple repeat runs for perturbations involving growth of entire rings.

3.4 Scaffold hopping. Apart from "conventional" perturbations as contained in the datasets above, more advanced structural changes may also be of interest during the lead optimization stage of the drug design cycle. Those include macrocyclization and core replacement. In terms of free energy calculations, modification of the overall ligand formal charge is also challenging. Scaffold hopping experiments in principle are amenable to FEP simulations. As the underlying method uses the single topology approach, such perturbations can be computationally more demanding compared to dual topology approaches if the structural motif to be replaced is located in the middle of the molecule. To assess the feasibility of this type of perturbations, six series of protein-ligand complexes for which input structures have been made available[39] were investigated. For some of the datasets one or several intermediate structures had to be introduced

in order to ensure a reasonable perturbation network. These are labelled as "mcs" (maximum common substructure of all ligands) and "sub" (substructure) in the perturbation networks, respectively. The networks can be found in the SI (Figure S2). The results are shown in Figure 6 and details are specified in the SI (Table S5). The MUE_c was 0.59 ± 0.50 kcal/mol for the results obtained in this study compared to 0.28 ± 0.31 kcal/mol reported by Wang et al.[39]
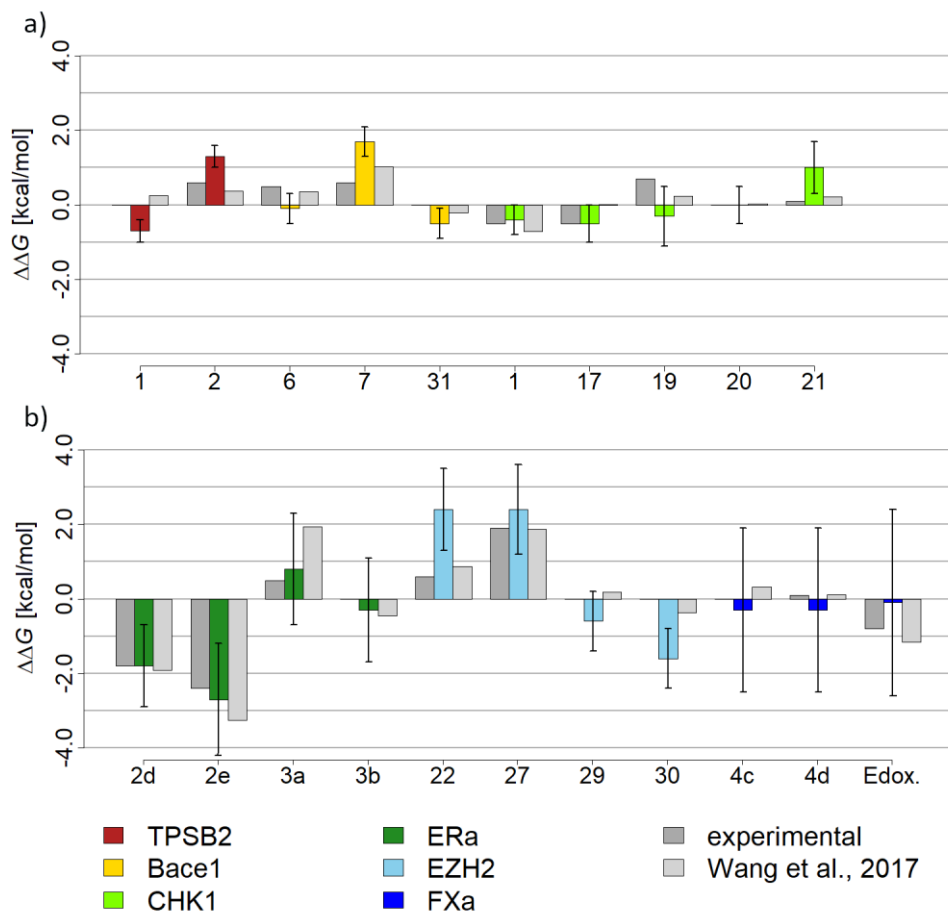

Figure 6. Experimental (dark grey) and computed (colored for this study and light grey for the report by Wang et al.[39]) relative binding affinities for the 21 ligands in the scaffold hoping dataset. Compound-wise errors were not specified in the work of Wang et al. Note that the experimental binding affinity of **3a** (ERa) may be lower than shown.[67]

In the following the checkpoint kinase 1 (CHK1)[68] dataset is described in more detail (cf. Figure 7). Both scaffold hopping and conventional R-group modifications can be found in this dataset consisting of 5 ligands. The computed free energies of compounds **1**, **17** and **20** are in excellent agreement with experiment. Though increasing the ring size is correctly predicted to be unfavorable (**21**), the predicted free energy is off by 0.9 ± 0.7 kcal/mol. The least active compound in the dataset (**19**) was not ranked last by predicted potency: The predicted change in relative affinity compared to **17** is only $+0.2 \pm 0.4$ kcal/mol, whereas the experimental affinity changes by $+1.2$ kcal/mol. Interestingly, also Wang et al.[39] report a predicted change in affinity of just $+0.3 \pm 0.1$ kcal/mol for this perturbation, which indicates that neither approach reproduces well experimental data for this pair of compounds.
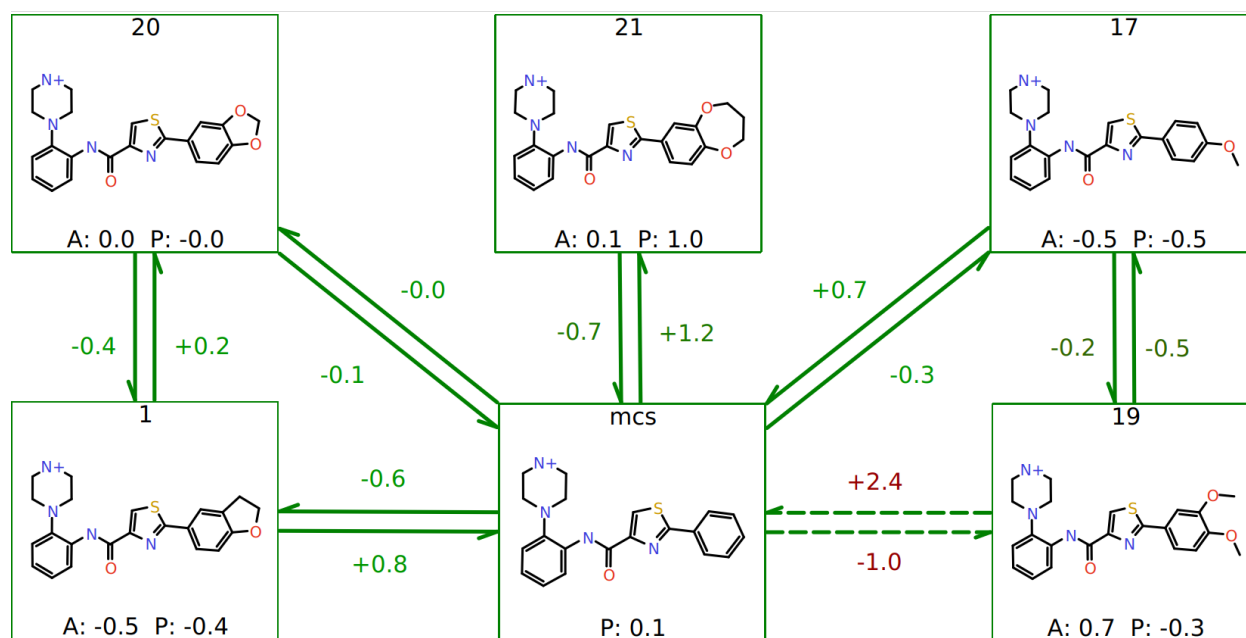
Figure 7. Depictions of the scaffold hopping perturbation network of CHK1 and the obtained results. Numbers along the arrows give the relative free energies for each perturbation. P and A give the predicted and experimental affinity relative to the reference compound, **20**. A dashed

arrow indicates that the perturbation was not considered for the final calculation of the ΔG values due to hysteresis.

In summary, the scaffold hopping experiments correctly captured the effects of the ring extension transformations (TPSB2 and BACE1), though the predicted change in affinity is higher than the experimental value. The ring opening/closing transformations in EZH2 and FXa are more challenging. This is mainly due to the large number of necessary additional intermediates which are linearly connected in the perturbation network and thus increase the overall uncertainty (cf. Figure S2). On the other hand, for a similar transformation in which the ring size was modified (ERa), results were in excellent agreement with experimental values. Lastly, results for the CHK1 dataset containing different modifications were in good agreement with measured affinities except for compound **19**. Overall, even though the calculations tend to exaggerate the impact of the modification on the free energy, they capture the trends in changes of binding affinity.

3.5 Computational cost. An important aspect of running FEP calculations is the overall calculation time of the whole dataset, especially when it is constituted by hundreds of compounds. To estimate the computational cost, we have performed benchmarks on two of the proteins contained in this study, TYK2 and CHK1, containing 288 and 252 amino acids respectively. For each system, three independent 4-ns windows were run for the protein ligand complex and the solvated ligand on a NVIDIA GTX2070. For TYK2, the runs took 84 and 17 minutes on average, corresponding to ~21 min/ns and ~4 min/ns. Similarly, times were 76 and 19 minutes for CHK1, i.e. ~19 min/ns and ~5 min/ns, respectively. These results were obtained using OpenMM's[31] mixed precision mode. For comparison, it has been reported that FEP+ takes

86 minutes to sample a 1-ns perturbation for 12 windows on 4 NVIDIA Tesla K80 for BACE1[69] complexes (containing 401 amino acids) and 34 minutes for the corresponding simulations of the solvated ligands.[70] This corresponds to ~29 min/ns for the complex and ~11 min/ns for the ligand-only simulations, respectively. The Gromacs-FEP implementation was reported to be 3 to 6 times slower than FEP+.[70] In a benchmark for AMBER-TI, a perturbation in factor Xa (total system size: 41,563 atoms) can be calculated at about 120 ns/day on a NVIDIA GTX1080, corresponding to 12 min/ns.[71]

Obviously, the data for the different FEP approaches cannot be directly compared, as the GPUs as well as the proteins used deviate. However, these figures show that the perturbations can be carried out reasonably quickly. In Flare, calculating the binding free energy for a ligand pair in TYK2 using 9 λ windows and a 4-ns calculation time can be carried out in ~15 hours on a single modern GPU. Accounting for equilibration time and with sufficient GPUs available this can be parallelized over 18 windows (free and bound legs) to yield a minimum time to answer of about 1.7 hours.

Production runs in the work presented here were 4-ns long for each λ window. However, there is evidence that shorter calculations times (usually 1 ns) may be sufficient to obtain a reasonable free energy estimate for certain classes of perturbations.[24,70] Thus, there may be scope to further decrease time to answer in a scenario where it is important to maximize throughput.

4. CONCLUSION

It has been demonstrated that the FEP method for calculating binding free energies implemented in Flare yields results which are on-par with various other published workflows. The method can also readily be applied to more challenging scaffold hopping experiments. The

software allows automatic set up of perturbation networks and implements several sanity checks of the reliability of the calculations, such as analysis of cycle closure errors and assessment of overlap matrices. Future work will be focused on enhancing the overlap between the sampled states of a perturbation by non-linearly spaced windows and by adaptive sampling. Furthermore, validation of the method will be extended considering another recently published large dataset for free energy calculations.[72]

ASSOCIATED CONTENT

Supporting Information. The following files are available free of charge.

Additional tables and figures (PDF)

ΔG values of all approaches discussed (XLSX)

Jupyter Notebook containing the calculation of R, Kendall's τ, MUE and root mean square error (RMSE) for all datasets (IPYNB)

File archive containing the experimental affinities and the $\Delta\Delta G$ values of all approaches discussed for rerunning the calculations in the Jupyter Notebook (ZIP)

AUTHOR INFORMATION

Corresponding Author

*mark@cresset-group.com

Author Contributions

The manuscript was written through contributions of all authors. All authors have given approval to the final version of the manuscript.

Notes

The authors declare the following competing financial interest(s): M. K., S. F.-C., P. T. and M. M. are Cresset employees. M. M. owns shares in Cresset. Cresset commercially distributes the Flare software used in this work.

REFERENCES

Flare, 3.0 , Cresset, Litlington, Cambridgeshire, UK; http://www.cresset-group.com/flare/;

(1)   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* (7), 2695–2703. https://doi.org/10.1021/ja512751q.

(2) Mobley, D. L.; Gilson, M. K. Predicting Binding Free Energies: Frontiers and Benchmarks. *Annu. Rev. Biophys.* **2017**, *46* (1), 531–558. https://doi.org/10.1146/annurev-biophys-070816-033654.

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

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

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

(6) Calabrò, G.; Woods, C. J.; Powlesland, F.; Mey, A. S. J. S.; Mulholland, A. J.; Michel, J. Elucidation of Nonadditive Effects in Protein–Ligand Binding Energies: Thrombin as a Case Study. *J. Phys. Chem. B* **2016**, *120* (24), 5340–5350. https://doi.org/10.1021/acs.jpcb.6b03296.

(7)   De Simone, A.; Georgiou, C.; Ioannidis, H.; Gupta, A. A.; Juárez-Jiménez, J.; Doughty-Shenton, D.; Blackburn, E. A.; Wear, M. A.; Richards, J. P.; Barlow, P. N.; Carragher, N.; Walkinshaw, M. D.; Hulme, A. N.; Michel, J. A Computationally Designed Binding Mode Flip Leads to a Novel Class of Potent Tri-Vector Cyclophilin Inhibitors. *Chem. Sci.* **2019**, *10* (2), 542–547. https://doi.org/10.1039/c8sc03831g.

(8)   Liu, S.; Wu, Y.; Lin, T.; Abel, R.; Redmann, J. P.; Summa, C. M.; Jaber, V. R.; Lim, N. M.; Mobley, D. L. Lead Optimization Mapper: Automating Free Energy Calculations for Lead Optimization. *J. Comput. Aided. Mol. Des.* **2013**, *27* (9), 755–770. https://doi.org/10.1007/s10822-013-9678-y.

(9)   Christ, C. D.; Fox, T. Accuracy Assessment and Automation of Free Energy Calculations for Drug Design. *J. Chem. Inf. Model.* **2014**, *54* (1), 108–120. https://doi.org/10.1021/ci4004199.

(10)   Fu, H.; Gumbart, J. C.; Chen, H.; Shao, X.; Cai, W.; Chipot, C. BFEE: A User-Friendly Graphical Interface Facilitating Absolute Binding Free-Energy Calculations. *J. Chem. Inf. Model.* **2018**, *58* (3), 556–560. https://doi.org/10.1021/acs.jcim.7b00695.

(11)   Hedges, L. O.; Mey, A. S. J. S.; Laughton, C. A.; Gervasio, F. L.; Mulholland, A. J.; Woods, C. J.; Michel, J. BioSimSpace : An Interoperable Python Framework for Biomolecular Simulation. **2019**, *4*, 1–10. https://doi.org/10.21105/joss.01831.

(12)   Klimovich, P. V.; Shirts, M. R.; Mobley, D. L. Guidelines for the Analysis of Free Energy Calculations. *J. Comput. Aided. Mol. Des.* **2015**, *29* (5), 397–411.

(13)   Michel, J.; Essex, J. W. Prediction of Protein-Ligand Binding Affinity by Free Energy Simulations : Assumptions, Pitfalls and Expectations. *J. Comput. Aided. Mol. Des.* **2010**, *24* (8), 639–658. https://doi.org/10.1007/s10822-010-9363-3.

(14)   Gaieb, Z.; Parks, C. D.; Chiu, M.; Yang, H.; Shao, C.; Walters, W. P.; Lambert, M. H.; Nevins, N.; Bembenek, S. D.; Ameriks, M. K.; Mirzadegan, T.; Burley, S. K.; Amaro, R. E.; Gilson, M. K. D3R Grand Challenge 3: Blind Prediction of Protein–Ligand Poses and Affinity Rankings. *J. Comput. Aided. Mol. Des.* **2019**, *33* (1), 1–18. https://doi.org/10.1007/s10822-018-0180-4.

(15)   Rizzi, A.; Murkli, S.; McNeill, J. N.; Yao, W.; Sullivan, M.; Gilson, M. K.; Chiu, M. W.; Isaacs, L.; Gibb, B. C.; Mobley, D. L.; Chodera, J. D. Overview of the SAMPL6 Host–Guest Binding Affinity Prediction Challenge. *J. Comput. Aided. Mol. Des.* **2018**, *32* (10), 937–963. https://doi.org/10.1007/s10822-018-0170-6.

(16)   Jarzynski, C. Rare Events and the Convergence of Exponentially Averaged Work Values. *Phys. Rev. E* **2006**, *73* (4), 46105. https://doi.org/10.1103/PhysRevE.73.046105.

(17)   Zwanzig, R. W. High- Temperature Equation of State by a Perturbation Method. I. Nonpolar Gases. *J. Chem. Phys.* **1954**, *22* (8), 1420–1426. https://doi.org/10.1063/1.1740409.

(18)   Kirkwood, J. G. Statistical Mechanics of Fluid Mixtures. *J. Chem. Phys.* **1935**, *3* (5), 300–313. https://doi.org/10.1063/1.1749657.

(19)   Merz, K. M.; Kollman, P. A. Free Energy Perturbation Simulations of the Inhibition of Thermolysin: Prediction of the Free Energy of Binding of a New Inhibitor. *J. Am. Chem. Soc.* **1989**, *111* (15), 5649–5658. https://doi.org/10.1021/ja00197a022.

(20)   Shirts, M. R.; Chodera, J. D. Statistically Optimal Analysis of Samples from Multiple Equilibrium States. *J. Chem. Phys.* **2008**, *129* (12). https://doi.org/10.1063/1.2978177.

(21)   Bennett, C. H. Efficient Estimation of Free Energy Differences from Monte Carlo Data. *J. Comput. Phys.* **1976**, *22* (2), 245–268. https://doi.org/https://doi.org/10.1016/0021-

9991(76)90078-4.

(22) Christ, C. D.; Mark, A. E.; Van Gunsteren, W. F. Basic Ingredients of Free Energy Calculations: A Review. *J. Comput. Chem.* **2010**, *31* (8), 1569–1582. https://doi.org/10.1002/jcc.21450.

(23) Song, L. F.; Lee, T. S.; Zhu, C.; York, D. M.; Merz, K. M. Using AMBER18 for Relative Free Energy Calculations. *J. Chem. Inf. Model.* **2019**, *59*, 3128–3135. https://doi.org/10.1021/acs.jcim.9b00105.

(24) Gapsys, V.; Pérez-Benito, L.; Aldeghi, M.; Seeliger, D.; Vlijmen, H. van; Tresadern, G.; Groot, B. L. de. Large Scale Relative Protein Ligand Binding Affinities Using Non-Equilibrium Alchemy. *Chem. Sci.* **2020**. https://doi.org/10.1021/cen-v068n020.p042.

(25) Rizzi, A.; Jensen, T.; Slochower, D. R.; Aldeghi, M.; Gapsys, V.; Ntekoumes, D.; Bosisio, S.; Papadourakis, M.; Henriksen, N. M.; de Groot, B. L.; Cournia, Z.; Dickson, A.; Michel, J.; Gilson, M. K.; Shirts, M. R.; Mobley, D. L.; Chodera, J. D. The SAMPL6 SAMPLing Challenge: Assessing the Reliability and Efficiency of Binding Free Energy Calculations. *J Comput Aided Mol Des* **2020**, https://doi.org/10.1007/s10822-020-00290-5.

(26) Loeffler, H. H.; Bosisio, S.; Duarte Ramos Matos, G.; Suh, D.; Roux, B.; Mobley, D. L.; Michel, J. Reproducibility of Free Energy Calculations Across Different Molecular Simulation Software. *J. Chem. Theory Comput.* **2018**, acs.jctc.8b00544. https://doi.org/10.1021/acs.jctc.8b00544.

(27) Cheeseright, T.; Mackey, M.; Rose, S.; Vinter, A. Molecular Field Extrema as Descriptors of Biological Activity: Definition and Validation. *J. Chem. Inf. Model.* **2006**, *46* (2), 665–676. https://doi.org/10.1021/ci050357s.

(28) Bauer, M. R.; Mackey, M. D. Electrostatic Complementarity as a Fast and Effective Tool to Optimize Binding and Selectivity of Protein-Ligand Complexes. *J. Med. Chem.* **2019**, *62* (6). https://doi.org/10.1021/acs.jmedchem.8b01925.

(29) Sire Molecular Simulation Framework. *http://siremol.org*. 2019.

(30) Mey, A. S. J. S.; Jiménez, J. J.; Michel, J. Impact of Domain Knowledge on Blinded Predictions of Binding Energies by Alchemical Free Energy Calculations. *J. Comput. Aided. Mol. Des.* **2018**, *32* (1), 199–210. https://doi.org/10.1007/s10822-017-0083-9.

(31) 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.; Wiewiora, R. P.; Brooks, B. R.; Pande, V. S. OpenMM 7: Rapid Development of High Performance Algorithms for Molecular Dynamics. *PLoS Comput. Biol.* **2017**, *13* (7), 1–17. https://doi.org/10.1371/journal.pcbi.1005659.

(32) Case, D. A.; Ben-Shalom, I. Y.; Brozell, S. R.; Cerutti, D. S.; Cheatham, III, T. E.; Cruzeiro, V. W. D.; Darden, T. A.; Duke, R. E.; Ghoreishi, D.; Gilson, M. K.; Gohlke, H.; Goetz, A. W.; Kollman, P. A. AMBER2019. University of California, San Francisco 2019.

(33) Wu, D.; Kofke, D. A. Phase-Space Overlap Measures. I. Fail-Safe Bias Detection in Free Energies Calculated by Molecular Simulation. *J. Chem. Phys.* **2005**, *123* (5), 1–10. https://doi.org/10.1063/1.1992483.

(34) Wu, D.; Kofke, D. A. Phase-Space Overlap Measures. II. Design and Implementation of Staging Methods for Free-Energy Calculations. *J. Chem. Phys.* **2005**, *123* (8), 1–10. https://doi.org/10.1063/1.2011391.

(35) Bosisio, S.; Mey, A. S. J. S.; Michel, J. Blinded Predictions of Host-Guest Standard Free

Energies of Binding in the SAMPL5 Challenge. *J. Comput. Aided. Mol. Des.* **2017**, *31* (1), 61–70. https://doi.org/10.1007/s10822-016-9933-0.

(36)   Papadourakis, M.; Bosisio, S.; Michel, J. Blinded Predictions of Standard Binding Free Energies: Lessons Learned from the SAMPL6 Challenge. *J. Comput. Aided. Mol. Des.* **2018**, *0* (0), 0. https://doi.org/10.1007/s10822-018-0154-6.

(37)   Mishra, S. K.; Calabró, G.; Loeffler, H. H.; Michel, J.; Koča, J. Evaluation of Selected Classical Force Fields for Alchemical Binding Free Energy Calculations of Protein-Carbohydrate Complexes. *J. Chem. Theory Comput.* **2015**, *11* (7), 3333–3345. https://doi.org/10.1021/acs.jctc.5b00159.

(38)   Granadino-Roldán, J. M.; Mey, A. S. J. S.; González, J. J. P. Effect of Automation on the Accuracy of Alchemical Free Energy Calculation Protocols. **2018**, 1–26.

(39)   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* (1), 42–54. https://doi.org/10.1021/acs.jctc.6b00991.

(40)   Boresch, S.; Karplus, M. The Role of Bonded Terms in Free Energy Simulations: 1. Theoretical Analysis. *J. Phys. Chem. A* **1999**, *103* (1), 103–118. https://doi.org/10.1021/jp981628n.

(41)   Stroganov, O. V; Novikov, F. N.; Zeifman, A. A.; Stroylov, V. S.; Chilov, G. G. TSAR, a New Graph–Theoretical Approach to Computational Modeling of Protein Side-Chain Flexibility: Modeling of Ionization Properties of Proteins. *Proteins Struct. Funct. Bioinforma.* **2011**, *79* (9), 2693–2710. https://doi.org/10.1002/prot.23099.

(42)   Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.;

Shindyalov, I. N.; Bourne, P. E. The Protein Data Bank. *Nucleic Acids Res.* **2000**, *28* (1), 235–242. https://doi.org/10.1093/nar/28.1.235.

(43) Šali, A.; Blundell, T. L. Comparative Protein Modelling by Satisfaction of Spatial Restraints. *Journal of Molecular Biology*. 1993, pp 779–815.

(44) Case, D. A.; Babin, V.; Berryman, J. T.; Betz, R. M.; Cai, Q.; Cerutti, D. S.; Cheatham III, T. E.; Darden, T. A.; Duke, R. E.; Gohlke, H.; Goetz, A. W.; Gusarov, S.; Homeyer, N.; Janowski, P.; Kaus, J.; Kolossváry, I.; Kovalenko, A.; Lee, T. S.; LeGrand, S.; Luchko, T.; Luo, R.; Madej, B.; Merz, K. M.; Paesani, F.; Roe, D. R.; Roitberg, A.; Sagui, C.; Salomon-Ferrer, R.; Seabra, G.; Simmerling, C. L.; Smith, W.; Swails, J.; Walker; Wang, J.; Wolf, R. M.; Wu, X.; Kollman, P. A. *Amber 14 OR - University of California, San Francisco*; 2014.

(45) Wang, J. M.; 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* (9), 1157–1174. https://doi.org/10.1002/jcc.20035.

(46) 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* (16), 1623–1641. https://doi.org/10.1002/jcc.10128.

(47) Wang, J.; Wang, W.; Kollman, P.; Case, D. A. Automatic Atom Type and Bond Type Perception in Molecular Mechanical Calculations. *J. Mol. Graph. Model.* **2006**, *25* (2), 247–260. https://doi.org/doi: 10.1016/j.jmgm.2005.12.005.

(48) 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* (2), 926. https://doi.org/10.1063/1.445869.

(49)   Ryckaert, J. P.; Ciccotti, G.; Berendsen, H. J. C. Numerical Integration of the Cartesian Equations of Motion of a System with Constraints: Molecular Dynamics of n-Alkanes. *J. Comput. Phys.* **1977**, *23* (3), 327–341. https://doi.org/10.1016/0021-9991(77)90098-5.

(50)   Tironi, I. G.; Sperb, R.; Smith, P. E.; Van Gunsteren, W. F. A Generalized Reaction Field Method for Molecular Dynamics Simulations. *J. Chem. Phys.* **1995**, *102* (13), 5451–5459. https://doi.org/10.1063/1.469273.

(51)   Xu, H. Optimal Measurement Network of Pairwise Differences. *J. Chem. Inf. Model.* **2019**, *59* (11), 4720–4728. https://doi.org/10.1021/acs.jcim.9b00528.

(52)   Hahn, D. F.; Hünenberger, P. H. Alchemical Free-Energy Calculations by Multiple-Replica -Dynamics: The Conveyor Belt Thermodynamic Integration Scheme. *J. Chem. Theory Comput.* **2019**, *15* (4), 2392–2419. https://doi.org/10.1021/acs.jctc.8b00782.

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

(54)   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* (1), 15–30. https://doi.org/10.1021/jm301448p.

(55)   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* (19), 4681–4698. https://doi.org/10.1021/jm0702478.

(56)   Cumming, J. N.; Smith, E. M.; Wang, L.; Misiaszek, J.; Durkin, J.; Pan, J.; Iserloh, U.; Wu, Y.; Zhu, Z.; Strickland, C.; Voigt, J.; Chen, X.; Kennedy, M. E.; Kuvelkar, R.; Hyde, L. A.; Cox, K.; Favreau, L.; Czarniecki, M. F.; Greenlee, W. J.; McKittrick, B. A.; Parker, E. M.; Stamford, A. W. Structure Based Design of Iminohydantoin BACE1 Inhibitors: Identification of an Orally Available, Centrally Active BACE1 Inhibitor. *Bioorg. Med. Chem. Lett.* **2012**, *22* (7), 2444–2449. https://doi.org/https://doi.org/10.1016/j.bmcl.2012.02.013.

(57)   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* (1), 56–69. https://doi.org/https://doi.org/10.1016/j.jmb.2009.04.051.

(58)   Hardcastle, I. R.; Arris, C. E.; Bentley, J.; Boyle, F. T.; Chen, Y.; Curtin, N. J.; Endicott, J. A.; Gibson, A. E.; Golding, B. T.; Griffin, R. J.; Jewsbury, P.; Menyerol, J.; Mesguiche, V.; Newell, D. R.; 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* (15), 3710–3722. https://doi.org/10.1021/jm0311442.

(59)   Goldstein, D. M.; Soth, M.; Gabriel, T.; Dewdney, N.; Kuglstatter, A.; Arzeno, H.; Chen, J.; Bingenheimer, W.; Dalrymple, S. A.; Dunn, J.; Farrell, R.; Frauchiger, S.; La Fargue, J.; Ghate, M.; Graves, B.; Hill, R. J.; Li, F.; Litman, R.; Loe, B.; McIntosh, J.; McWeeney,

D.; Papp, E.; Park, J.; Reese, H. F.; Roberts, R. T.; Rotstein, D.; San Pablo, B.; Sarma, K.; Stahl, M.; Sung, M.-L.; Suttman, R. T.; Sjogren, E. B.; Tan, Y.; Trejo, A.; Welch, M.; Weller, P.; Wong, B. R.; Zecic, H. Discovery of 6-(2,4-Difluorophenoxy)-2-[3-Hydroxy-1-(2-Hydroxyethyl)Propylamino]-8-Methyl-8H-Pyrido[2,3-d]Pyrimidin-7-One (Pamapimod) and 6-(2,4-Difluorophenoxy)-8-Methyl-2-(Tetrahydro-2H-Pyran-4-Ylamino)Pyrido[2,3-d]Pyrimidin-7(8H)-One (R1487) as Orally . *J. Med. Chem.* **2011**, *54* (7), 2255–2265. https://doi.org/10.1021/jm101423y.

(60)   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.; Bir Kohli, P.; 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. https://doi.org/10.1016/j.ejmech.2013.03.070.

(61)   Liang, J.; Van Abbema, A.; Balazs, M.; Barrett, K.; Berezhkovsky, 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* (11), 4521–4536. https://doi.org/10.1021/jm400266t.

(62)   Szczepankiewicz, B. G.; Kosogof, C.; Nelson, L. T. J.; Liu, G.; Liu, B.; Zhao, H.; Serby,

M. D.; Xin, Z.; Liu, M.; Gum, R. J.; Haasch, D. L.; Wang, S.; Clampit, J. E.; Johnson, E. F.; Lubben, T. H.; Stashko, M. A.; Olejniczak, E. T.; Sun, C.; Dorwin, S. A.; Haskins, K.; Abad-Zapatero, C.; Fry, E. H.; Hutchins, C. W.; Sham, H. L.; Rondinone, C. M.; Trevillyan, J. M. Aminopyridine-Based c-Jun N-Terminal Kinase Inhibitors with Cellular Activity and Minimal Cross-Kinase Activity. *J. Med. Chem.* **2006**, *49* (12), 3563–3580. https://doi.org/10.1021/jm060199b.

(63)   Vanommeslaeghe, K.; Hatcher, E.; Acharya, C.; Kundu, S.; Zhong, S.; Shim, J.; Darian, E.; Guvench, O.; Lopes, P.; Vorobyov, I.; MacKerell Jr., A. D. CHARMM.Pdf. *J. Comput. Chem.* **2010**, *31* (4), 671–690. https://doi.org/10.1002/jcc.21367.CHARMM.

(64)   Gapsys, V.; Michielssens, S.; Seeliger, D.; de Groot, B. L. Accurate and Rigorous Prediction of the Changes in Protein Free Energies in a Large-Scale Mutation Scan. *Angew. Chemie - Int. Ed.* **2016**, *55* (26), 7364–7368. https://doi.org/10.1002/anie.201510054.

(65)   Fisher, R. A. Frequency Distribution of the Values of the Correlation Coefficient in Samples from an Indefinitely Large Population. *Biometrika* **1915**, *10* (4), 507–521. https://doi.org/10.2307/2331838.

(66)   Nicholls, A. Confidence Limits, Error Bars and Method Comparison in Molecular Modeling. Part 1: The Calculation of Confidence Intervals. *J. Comput. Aided. Mol. Des.* **2014**, *28* (9), 887–918. https://doi.org/10.1007/s10822-014-9753-z.

(67)   Norman, B. H.; Dodge, J. A.; Richardson, T. I.; Borromeo, P. S.; Lugar, C. W.; Jones, S. A.; Chen, K.; Wang, Y.; Durst, G. L.; Barr, R. J.; Montrose-Rafizadeh, C.; Osborne, H. E.; Amos, R. M.; Guo, S.; Boodhoo, A.; Krishnan, V. Benzopyrans Are Selective Estrogen Receptor β Agonists with Novel Activity in Models of Benign Prostatic

Hyperplasia. *J. Med. Chem.* **2006**, *49* (21), 6155–6157. https://doi.org/10.1021/jm060491j.

(68)   Huang, X.; Cheng, C. C.; Fischmann, T. O.; Duca, J. S.; Yang, X.; Richards, M.; Shipps, G. W. Discovery of a Novel Series of CHK1 Kinase Inhibitors with a Distinctive Hinge Binding Mode. *ACS Med. Chem. Lett.* **2012**, *3* (2), 123–128. https://doi.org/10.1021/ml200249h.

(69)   Rombouts, F. J. R.; Tresadern, G.; Delgado, O.; Martínez-Lamenca, C.; Prokopcova, H.; Alonso, J. M.; Austin, N.; Borghys, H.; Brandt, S. Van; Gool, M. Van; García-Molina, A.; Diego, S. A. A. de; Oehlrich, D.; Surkyn, M.; Cleyn, M. De; Vos, A.; Alexander, R.; Macdonald, G.; Moechars, D.; Gijsen, H.; Trabanco, A. A. 1,4-Oxazine β‑Secretase 1 (BACE1) Inhibitors: From Hit Generation to Orally Bioavailable Brain Penetrant Leads. *J. Med. Chem.* **2015**, *58* (20), 8216–8235. https://doi.org/10.1021/acs.jmedchem.5b01101.

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

(71)   Lee, T. S.; Cerutti, D. S.; Mermelstein, D.; Lin, C.; LeGrand, S.; Giese, T. J.; Roitberg, A. E.; Case, D. A.; Walker, R. C.; York, D. M. GPU-Accelerated Molecular Dynamics and Free Energy Methods in Amber18: Performance Enhancements and New Features. *J. Chem. Inf. Model.* **2018**, acs.jcim.8b00462. https://doi.org/10.1021/acs.jcim.8b00462.

(72)   Schindler, C. E. M.; Baumann, H.; Blum, A.; Boese, D.; Buchstaller, H.; Burgdorf, L.; Cappel, D.; Chekler, E.; Dorsch, D.; Eguida, M.; Follows, B.; Grädler, U.; Gunera, J.; Johnson, T.; Jorand, C.; Karra, S.; Klein, M.; Kötzner, L.; Knehans, T.; Leiendecker, M.; Leuthner, B.; Li, L.; Mochalkin, I.; Musil, D.; Neagu, C.; Rippmann, F.; Schiemann, K.; Steinbrecher, T.; Tanzer, E.; Lopez, A. U. Large-Scale Assessment of Binding Free

Energy Calculations in Active Drug Discovery Projects. **2020**, 1–33. https://doi.org/10.26434/chemrxiv.11364884.v1.