Effect of automation on the accuracy of alchemical free energy calculation protocols

José M. Granadino-Roldán[a,], Antonia S. J. S. Mey[b], Juan J. Pérez González[c], Stefano Bosisio[b], Jaime Rubio-Martinez[d], Julien Michel,[b,*]*

a) *Departamento de Química Física, Facultad de Ciencias Experimentales, Universidad de Jaén, Campus "Las Lagunillas" s/n, 23071, Jaén, Spain*

b) EaStCHEM School of Chemistry, Joseph Black Building, The King's Buildings, Edinburgh, EH9 3FJ, UK

c) Department of Chemical Engineering, Universitat Politecnica de Catalunya, ETSEIB. Av. Diagonal, 647, 08028 Barcelona, Spain

d) *Departament de Química Física, Universitat de Barcelona (UB) and the Institut de Recerca en Quimica Teorica i Computacional (IQTCUB), Martí i Franqués 1, 08028 Barcelona, Spain*

KEYWORDS: computer-aided drug design, protein-ligand interactions, alchemical free energy calculations, ACK1

ABSTRACT

Virtual *in silico* screening in drug discovery frequently relies on a cascade of computational methods that starts with rapid calculations applied to a large number of compounds and ends with more expensive computations restricted to a subset of compounds that passed initial filters. This work focuses on protocols for alchemical free energy (AFE) scoring in the context of a Docking – MM/PBSA – AFE cascade. A dataset of 15 congeneric inhibitors of the ACK1 protein was used to evaluate the performance of AFE protocols that varied in the steps taken to prepare input files (fully automated from previously docked and scored poses, manual selection of poses, manual placement of binding site water molecules). The main finding is that use of knowledge derived from X-ray structures to model binding modes, together with the manual placement of a bridging water molecule, improves the R^2 from 0.45 ± 0.06 to 0.76 ± 0.02 and decreases the mean unsigned error from 2.11 ± 0.08 to 1.24 ± 0.04 kcal mol^{-1}. By contrast a brute force automated protocol that increased the sampling time ten-fold lead to little improvements in accuracy.

INTRODUCTION

There is continuous interest in computational methods to decrease time and costs of hit-to-lead and lead optimization efforts in preclinical drug discovery[1]. A recurring topic in computational chemistry is the use of virtual *in silico* screens to find ligands for proteins[2-3]. Typically, the goal is to filter via a cascade of computational methods a large library to focus experimental efforts on a small number of molecules. Usually inexpensive methodologies are applied first to eliminate a large number of poorly suited molecules, with more expensive calculations applied reserved to a subset of promising

ligands. For a structure-based virtual screen the main steps involve frequently library screening, docking, initial scoring, and re-scoring with diverse molecular simulation methods such as Molecular Mechanics Poisson Boltzmann (Generalized Born) Surface Area (MM/PBSA)[4], Linear Interaction Energy (LIE)[5] or Free energy Perturbation (FEP)[6] methods[7].

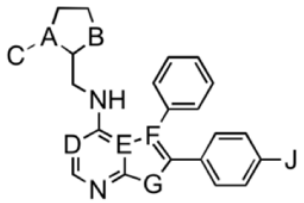
In one of our previous works[8], different re-scoring strategies based on the MM/PB(GB)SA methodologies were assessed in the context of virtual screens against protein targets. It was concluded that a compromise between quality of results and computational cost can be achieved using this method on one minimized structure from the docking procedure. The study biased the docking, as the set of ligands analysed belonged to the same scaffold and it was assumed that the core binding mode of the conserved scaffold would not deviate from that of the experimentally X-ray resolved one.

The present study explores the suitability of alchemical free energy (AFE) methods for further re-scoring of ligands processed by docking and MM/PBSA protocols. AFE methods are increasingly used for predictions of free energies of binding in blinded competitions such as such as SAMPL (Statistic Assessment of Modelling of Proteins and Ligands) and D3R grand challenges[9-15]. Some AFE protocols have even achieved predictions of binding energies with root mean square deviations (RMSD) under 1.5 kcal mol^{-1}, and Pearson Correlation coefficients (R) of around 0.7 or better[16-18]. Nevertheless, the performance varies significantly between different AFE protocols and targets and it is important to explore further the robustness of these methodologies.

Specifically, this study explores the extent to which the domain knowledge of a user may influence the accuracy of AFE calculations via careful setup, or whether issues such as binding poses selection or binding site water placement can be overcome via

brute force automation. This was investigated using a dataset of 15 congeneric inhibitors of the protein activated Cdc42-associated kinase (ACK1)[19], a potential cancer target[20-21]. The compounds span a large range of activity (K_i values ranging from more than 10 μM to 0.0003 μM), as seen in Figure 1A, and are typical of the structural modifications performed in hit-to-lead programs. The 15 ligands were first docked into the ACK1 ATP-binding site, and a set of docked poses obtained for each ligand was re-scored with a 4-step minimization protocol followed by a single-snapshot MM/PBSA re-scoring. The best scored pose is alchemically studied and the relative binding energy is compared to the experimental one. The alchemical calculations are also repeated with a 10-fold increase in sampling time. The role of a possible bridging water molecule in the binding pocket is also taken into account. Finally, thermodynamic cycle closures are analyzed as a way to detect incorrectly predicted poses without knowledge of the experimental relative binding energies.

A



	A	B	C	D	E	F	G	J	K_i (μM)	Compound
Batch 1	O	C	-	N	C	C	O	H	0.01	**2**
	O	C	-	N	C	C	S	H	> 10	**3**
	O	C	-	N	C	N	N	H	7.3	**4**
	O	C	-	N	N	C	N	H	1.8	**5**
	O	C	-	C	C	C	O	H	0.07	**6**
	O	C	-	N	C	C	NH	H	0.006	**7**
Batch 2	O	C	-	N	C	C	O	OCH_3	0.005	**8**
	O	C	-	N	C	C	O	$O(CH_2)_2NMe$	0.005	**15**
	O	C	-	N	C	C	NH	$O(CH_2)_2NMe$	0.006	**16**
	S	S	-	N	C	C	NH	$O(CH_2)_2NMe$	0.0003	**35**
	S	S	-	N	C	C	O	$O(CH_2)_2NMe$	0.0002	**36**
	S	C	-	N	C	C	O	$O(CH_2)_2NMe$	0.08	**38**
	O	O	-	N	C	C	O	$O(CH_2)_2NMe$	0.013	**39**
	C	C	OH	N	C	C	NH	$O(CH_2)_2NMe$	0.04	**44**
	C	C	OMe	N	C	C	NH	$O(CH_2)_2NMe$	0.05	**45**

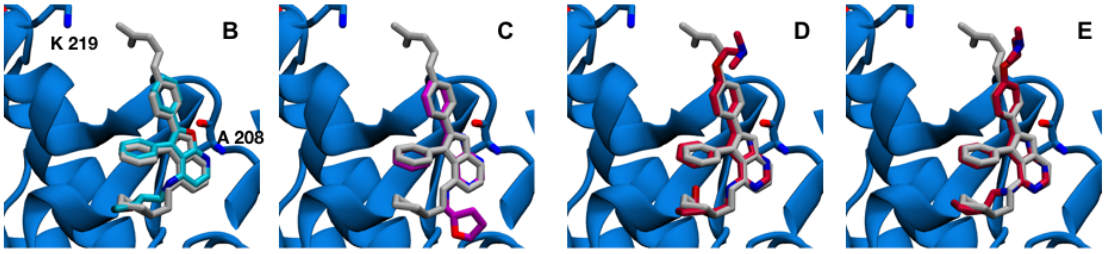


Figure 1. (A) Ligands studied in this work, along with reported K_i values[19]. (B-E) Superimposition of the X-ray diffraction derived structure of ligand **35** (grey) bound to ACK1 (PDB code 4EWH), (B) with the best predicted MM/PBSA docked poses for ligand **6** (blue), (C) with ligand **7** (purple) exhibiting a different binding mode, (D) with ligand **44** from MM/PBSA prediction using the best predicted binding mode and (E) with ligand **44** using the second best binding mode prediction. All carbon atoms of ligand **44** are colored in red. Hydrogens are omitted for clarity.

MATERIALS AND METHODS

Dataset

The dataset consists of 15 ACK1 competitive inhibitors for which inhibition constants (K_i) have been reported. The structure of only one protein-ligand complex (compound **35**) was determined by X-ray crystallography[19] (Figure 1B). This dataset was further divided into two subsets: *batch1* (6 ligands with K_i values ranging from >10 μM to 0.006 μM), and *batch 2* (9 ligands with K_i values ranging from 0.013 to 0.0003 μM).

Protein setup

The ACK1 kinase domain structure was taken from the Protein Data Bank, code 4EWH[19], using chain B of the crystal structure, which was protonated with MOE v2009.1[22]. The structure has no missing residues; Tyrosine 284 was dephosphorylated with MOE following Lougheed et al. observation that inhibitor binding is not expected to be sensitive to the phosphorylation state of this residue[23].

Docking

Docking was performed with MOE v2009.1[22]. The full docking process was done in three steps. The first one was an exhaustive conformational search of the ligands using the Systematic option of MOE together with the option Enforce chair conformations on. All other parameters were set to the standard options. A maximum of 100

conformations by compound were selected for the Placement step. In this second step the receptor was defined as those atoms within 9.0 Å from the ligand. The Rotate Bonds option was activated and the Affinity dG function employed together with the Triangle Matcher method for placement. A maximum of 30 poses for each ligand were retained. Finally, the 500 best structures were submitted to the Refinement step with the Force Field function and allowing the lateral chains of the pocket residues to move during the optimization without restriction. All other parameters were set to the standard options. The five best structures obtained for each ligand were retained for minimization and re-scoring with MM/PBSA.

MM/PBSA

A four-step minimization protocol followed by a single snapshot MM/PBSA re-scoring was performed with Amber 14[24]. Ligands were prepared with Antechamber using the GAFF force field[25] with AM1-BCC partial charges[26-27], while the ff99SB[28] force field was used for the protein. All systems were solvated in a rectangular box of TIP3P water molecules[29]. Counterions were added as necessary to neutralize the systems[30]. Energy minimization was performed under periodic boundary conditions using the particle-mesh-Ewald method for the treatment of the long-range electrostatic interactions[31]. A cut-off distance of 10 Å was chosen to compute non-bonded interactions. The four-step minimization procedure was as follows: 1) 5000 steepest descent (SD) steps applied to water molecule coordinates only; 2) 5000 SD steps applied also to protein atoms, with positional harmonic restraints (5 kcal mol^{-1} $Å^{-2}$) applied to backbone atoms only; 3) 5000 SD steps as done previously with backbone atom restraints set to 1 kcal mol^{-1} $Å^{-2}$ and 4) 5000 SD steps with no restraints.

For each of the energy minimized structures, a binding free energy was estimated following the MM/PBSA methodology as implemented in the MM/PBSA.py program[32].

No entropic contributions were taken into account, while the variables *cavity_surften* and *cavity_offset* were assigned the values of 0.00542 kcal mol^{-2} $Å^{-2}$ and -1.008, respectively, using the defaults for all remaining variables.

Alchemical free energy calculations

Relative binding free energies were calculated using a single topology molecular dynamics alchemical free energy approach[33]. Alchemical free energy calculations avoid direct computation of the free energy change associated with the reversible binding of a ligand to a protein trough an artificial morphing of a ligand *X* into another ligand *Y* by using a parameter λ which defines the change from *X* to *Y*. Thus, the relative free energy of binding ($\Delta\Delta G_{X\rightarrow Y}$) was given by equation 1 as:

$$\Delta\Delta G_{X\rightarrow Y} = \Delta G_{X\rightarrow Y}^{complex} - \Delta G_{X\rightarrow Y}^{free} \quad (1)$$

Where $\Delta G_{X\rightarrow Y}^{free}$ is the free energy change for transforming ligand *X* into ligand *Y* in solution whereas $\Delta G_{x\rightarrow y}^{complex}$ is the free energy change for the same transformation in the protein binding site.

A relative free energy perturbation network for both *batch 1* and *batch 2* was designed (Figure S1 and Figure S11). The top-scored MM/PBSA pose for each ACK1 ligand was used as input for the subsequent alchemical free energy preparation protocol using the FESetup software package[34]. The protocol used by FESetup for the automated preparation of ligands, protein and complexes was as follows:

Ligands. Atomic charges were assigned by using the Antechamber module in AmberTools 14[24], selecting the AM1-BCC method[26-27], and the GAFF2 force field[25]. Ligands were solvated with TIP3P water molecules[29], with counterions added as necessary to neutralize the system[30]. Each system was energy minimized for 100 SD cycles and equilibrated at 300 K and 1 atm pressure for 10^5 molecular dynamics (MD)

steps with a 2 fs timestep using the module Sander[24], with a positional harmonic restraint (10 kcal mol^{-1} $Å^{-2}$) applied to ligand atoms. Bonds involving hydrogen atoms were constrained.

Protein. The protein was parametrized using the Amber ff14SB force field[37].

Complexes. Each ligand was combined back with the ACK1 protein model and the complex was solvated with TIP3P water molecules[29]. Counterions were also added to neutralize the solution[30]. The system was afterwards equilibrated following the procedure already described for ligands, using now 5000 MD steps.

All alchemical free energy calculations used 11 equidistant λ windows. For each λ value MD trajectories were computed in the NPT ensemble with a pressure of 1 atm and temperature of 300 K using the software SOMD 2016.1.0[35-37]. SOMD has been used in several recent studies to model the binding energetics of enzyme inhibitors[38], carbohydrate ligands[39], and host-guest systems[13]. Each λ window was sampled for 4 ns. Pressure was regulated using a Monte Carlo barostat[40-41] with an update frequency of 25 MD steps. Temperature was kept constant using the Andersen thermostat[42], with a collision frequency of 10 ps^{-1}. A 2 fs time step was used with the leapfrog-Verlet integrator. All bonds involving hydrogens were constrained to their equilibrium distances. Non-bonded interactions were evaluated setting a cutoff distance of 12 Å. Long-range electrostatic interactions were calculated using the shifted atom-based Barker-Watts reaction field[43], with the medium dielectric constant set to 82.0. In order to avoid steric clashes at the beginning of each MD run due to modifications of the ligand parameters associated with changes in λ, each structure was energy minimized for 1000 steps prior to MD simulation.

Each simulation was repeated at least twice using different initial assignments of velocities, and both $\Delta\Delta G_{X \rightarrow Y}$ and $\Delta\Delta G_{Y \rightarrow X}$ were calculated from independent simulations.

In some cases, where poor agreement was observed between duplicates a third run was performed. Thus, relative binding energies are reported as the average of 2 or 3 runs and the reported statistical uncertainties are the standard error of the mean.

Ligand **38** was tested as a racemic mixture. Calculations were carried out for each enantiomer and the binding energies relative to this ligand were given with equation 2:

$$\Delta\Delta G_{\mathbf{38}\rightarrow X} = -kTln\left[0.5\left(exp\left(\frac{-\Delta\Delta G_{38R\rightarrow X}}{kT}\right) + exp\left(\frac{-\Delta\Delta G_{38S\rightarrow X}}{kT}\right)\right)\right] \quad (2)$$

Cycle closures were evaluated using free energy changes from both the forward (**X→Y**) and reverse (**Y→X**) perturbations. The metrics used to evaluate the datasets were the determination coefficient R^2, linear regression slope and the mean unsigned error (MUE). Experimental binding affinities were calculated from the corresponding inhibition constants[19] (K_i) using $\Delta G = RTln\left(\frac{K_i}{C^0}\right)$ with C^0 = 1 mol L^{-1}. As no uncertainties have been reported for the K_i values, an uncertainty of 0.4 kcal mol^{-1} was assumed[37,44].

Relative free energies were estimated using the multistate Bennett's acceptance ratio (MBAR) method[45], as included in the software *analyse_freenrg* from the Sire software suite. Relative free energies for complete datasets were evaluated using version 0.3.5 of the *freenrgworkflows* python module [https://github.com/michellab/freenrgworkflows]. For more details, see Mey *et al*[46]. All analysis scripts are available online at https://github.com/michellab/ACK1_Data.

Alchemical free energy Protocols

Five different alchemical free energy protocols were followed. *Protocol A* uses for each ligand the best scored pose according to MM/PBSA. This leads to a pose that differs from the X-ray crystallographic pose of **35** for several ligands (**2**, **4**, **7**, **8**, **16**, **44** and **45**). *Protocol B* assumes user intervention to select the pose that resembles the most the X-ray binding mode among the 5 MM/PBSA poses. *Protocols C* and *D* explore the effect of manually modelling a water molecule inside the ACK1 ATP-binding site (see

Figure S6). This reflects user knowledge that in other high-resolution structures of ACK1 (e.g. the 1.31 Å resolution 4HZR structure[47]) one additional binding site water molecule between the protein and ligand is apparent. *Protocol C* uses the same ligand poses as *Protocol A*, while *Protocol D* uses the same poses as *Protocol B*.

Finally, *Protocol E* is simply *Protocol A* with the per λ simulation time increased ten-fold. This was done to evaluate whether the different binding mode and ATP-binding site water rearrangements seen in *Protocols A-D* can be sampled with longer MD simulation protocols. *Protocol E* is computationally expensive and was applied to *batch 1* only (ca. 10 μs of simulation time). Figures were rendered with VMD[48], while graphs were prepared with Origin[49] and python using plotting libraries Matplotlib version 2.0.2[50] and Seaborn version 0.7.1[51].

RESULTS

Batch1

Protocol A renders (Figure 2A) modest results, with a R^2 of 0.36±0.07 and a strong underestimation of relative free energies, as shown by the slope of the regression line (0.3). Inspection of Figure 2B shows that ligands **2**, **4** and **7** are clear outliers. These ligands have a predicted docked pose which differ more from the X-ray derived binding mode (see Figures 1B and 1C). Results for *protocol B* are shown in Figures 2A, S2 and S5. This protocol gives clearly better results, although the underestimation (slope 0.4) of relative binding free energies remains high, and ligands **2**, **4** and **7** are still ranked poorly.

A

Protocol	R^2	MUE	Slope
A	0.36±0.07	1.55±0.06	0.3
B	0.67±0.03	1.29±0.03	0.4
C	0.5±0.2	1.53±0.07	0.2
D	0.84±0.03	1.29±0.03	0.5
E	0.33±0.08	1.73±0.08	0.3

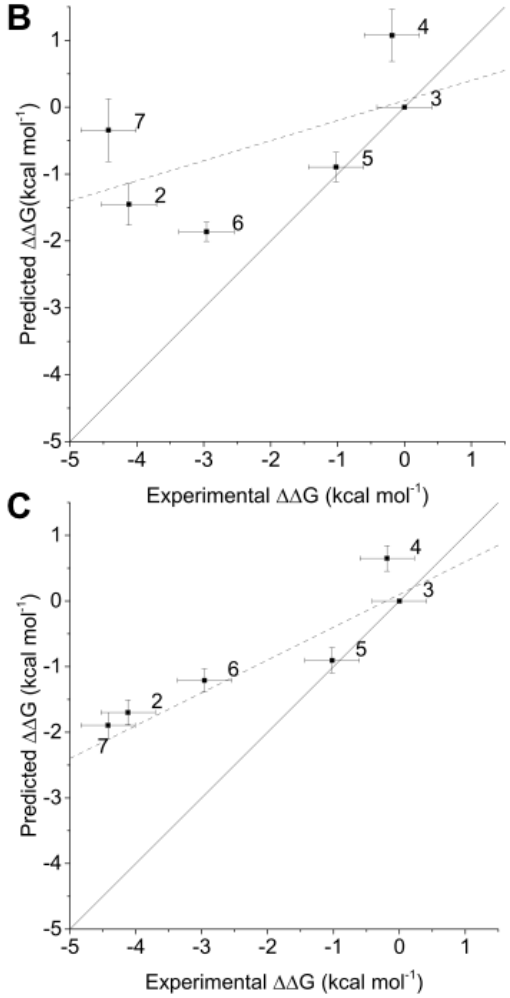


Figure 2. (A) R^2, MUE (kcal mol^{-1}) and slope metrics obtained from the comparison of experimental and predicted relative free energies of binding of *batch 1*. (B) and (C) Comparison of experimental and predicted relative free energies of binding of *batch 1* for *protocols A* and *D*, respectively. Free energies of binding are relative to ligand **3**. The linear regression line (dashed line) and a line with slope unity (solid line) are also presented.

An analysis of the relative binding energies calculated with *protocols A* and *B* (Figures S1 and S2), for ligands **2**, **4** and **7**, reveals that these ligands appear in the

perturbations that show the highest deviations between the experimental and calculated relative binding energies. Thus, for *protocol A* deviations of more than 3.0 kcal mol^{-1} are observed for **2→3**, **7→4**, **7→6**, **7→3** and **3→7**, while for *protocol B* these deviations appear for perturbations **2→3**, **4→7**, **7→4**, **7→3** and **3→7**. An analysis of the docked structures of ligands **3**, **4**, **6** and **7** suggested that a possible explanation for the inability of the protocol to reproduce the experimental relative binding affinities is due to interactions of the pyrrole NH group from ligand **7** has (see Figure 3). The NH group in the pyrrole ring could establish a hydrogen bond with THR205 (see Figure S6) if a bridging water was present. Indeed, several water molecules are present inside the ATP-binding pocket of 4HZR[47]. That possibility is explored in *protocols C* and *D*, where a water molecule has been manually placed inside the binding pocket. Results for *protocol C* are shown in Figures 2A and S7, while those for *protocol D* appear in Figures 2A, 2C and 3. *Protocol D* clearly surpass all others, with a R^2 of 0.84±0.03 and an improvement in the underestimation of relative binding energies (slope = 0.5). A comparison of the calculated relative binding energies for ligands **3** and **4** allows to conclude that using a different pose for ligand **4** does not seem to affect the results (both *protocols A* and *B* for example, give $\Delta\Delta G_{3\rightarrow4}$ = 1.3 ± 0.1 kcal mol^{-1}). Inspection of the calculated trajectories show that ligand **4** rapidly converts from its initial docked pose (*protocols A* and *C*) to one similar to that used as input for *protocols B* and *D*.

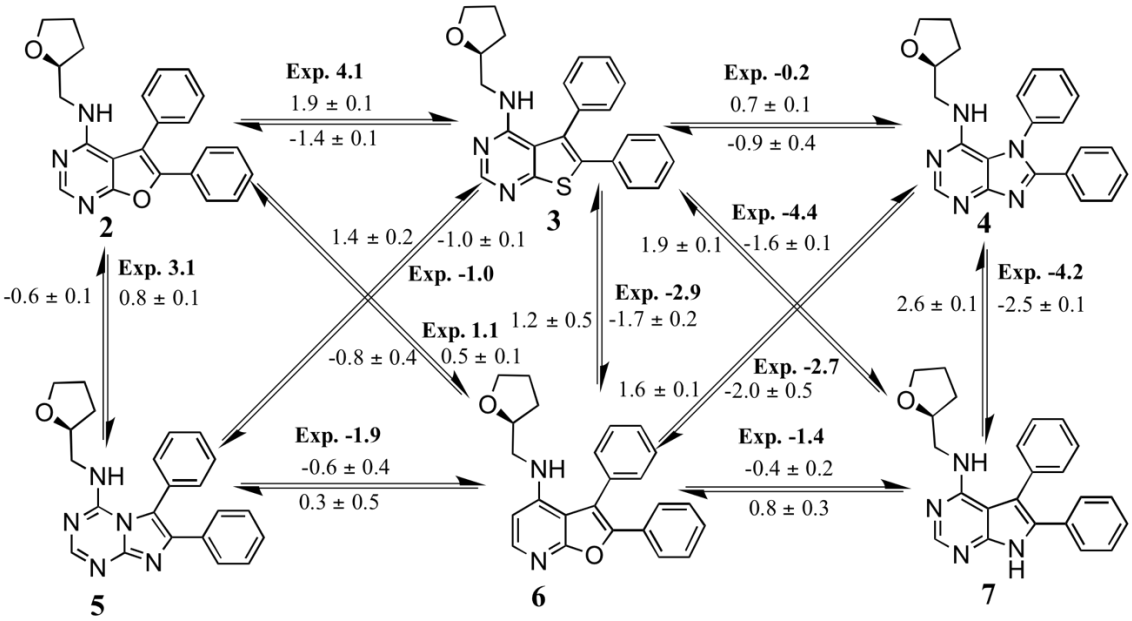


Figure 3. Calculated and experimental (in bold) relative binding affinities (in kcal mol^{-1}) for all the perturbations run in *batch 1* with *protocol D*

The possibility of resolving ambiguities in binding poses and binding site water content without user intervention was next tested by increasing the simulation sampling time to 40 ns for each λ window. The expectation was this would allow the ligand to find the correct pose and to allow water molecules diffuse in the ATP-binding site (see Figure S6). Results are shown in Figures 2A and S8. The increased simulation time does not translate into any improvement of the results. The R^2, slope and MUE values are as poor or poorer as those for *protocol A*, while the outliers remain the same. The MD trajectories show that, even with the increased simulation time, ligand **7** is not able to change its docking pose, while ligand **4** needs under 4 ns to adopt a pose that resembles the X-ray pose of **35**. Besides, a water molecule enters and remains in the ATP-binding site in 7 out of 22 MD trajectories only.

Analysis of the complete dataset

The robustness of the results obtained for *batch 1* was tested by processing *batch 2* and re-analyzing the full dataset. Ligands in *batch 2* are positively charged in the assay

conditions, whereas *batch 1* ligands are neutral. Relative free energy calculations that involve a net charge change are still technically challenging for simulations carried out with a reaction-field cutoff. Thus, the perturbations between ligands **8** and **15** were carried out assuming **15** is neutral. Results for individual perturbations in *batch 2* are shown in Figures S11 to S15.

Protocol A, as expected given the results obtained for *batch 1*, gives modest results, as can be seen in Figures 4A and B (R^2 =0.45±0.06 and slope of 0.5). The slope has improved from 0.3 to 0.5 because the relative free energies of the compounds in *batch 2* are not as under predicted as those from *batch 1* (see Table S1). Ligands **16**, **44** and **45** need further inspection. Figure S11 shows that, while the experimental $\Delta\Delta G_{44\rightarrow45}$ is -0.1 kcal mol^{-1}, the calculated $\Delta\Delta G_{44\rightarrow45}$ is 1.7 ± 0.1 kcal mol^{-1} (the reverse perturbation was calculated as $\Delta\Delta G_{45\rightarrow44}$ = -1.9 ± 0.1 kcal mol^{-1}). Similarly, while the experimental $\Delta\Delta G_{16\rightarrow45}$ is 1.2 kcal mol^{-1}, the calculated results are $\Delta\Delta G_{16\rightarrow45}$ = -1.2 ± 0.4 kcal mol^{-1} and $\Delta\Delta G_{45\rightarrow16}$ = -2.3 ± 0.1 kcal mol^{-1}.

A

Protocol	R^2	MUE	Slope
A	0.45±0.06	2.11±0.08	0.5
B	0.59±0.04	2.12±0.04	0.5
C	0.55±0.05	1.82±0.06	0.5
D	0.76±0.02	1.24±0.04	0.8

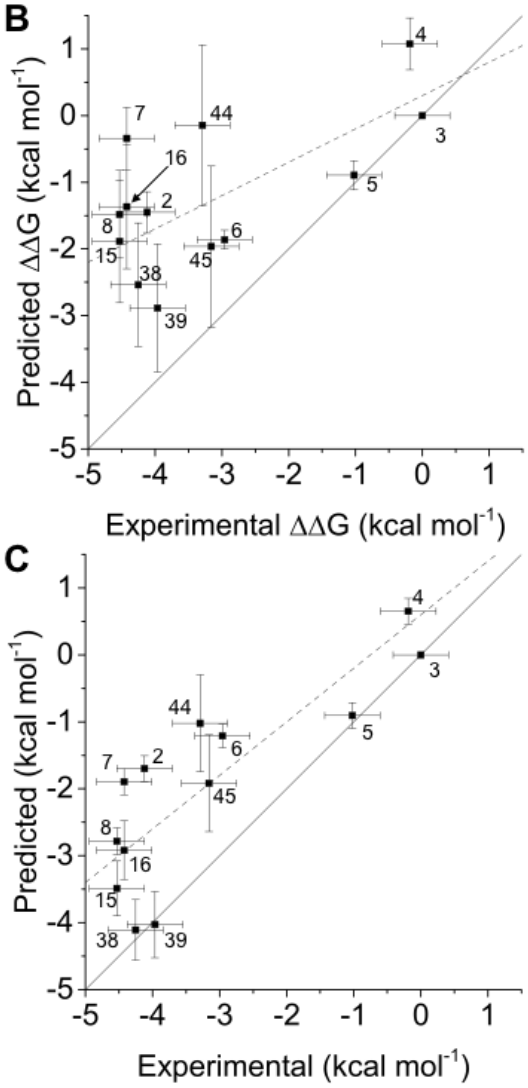


Figure 4. (A) R^2, MUE (kcal mol^{-1}) and slope metrics obtained from the comparison of experimental and predicted relative free energies of binding of the whole set. (B) and (C) Comparison of experimental and predicted relative free energies of binding of the whole set for *protocols A* and *D*, respectively. Free energies of binding are relative to ligand **3**. The linear regression line (dashed line) and a line with slope 1 (solid line) are also presented.

Interestingly, the dihedral angle defining the relative orientation of the NH group that links the pyrimidine and the cyclopentanol rings changes values rather quickly during the simulation. Figure 5A shows an example for the first repeat of the perturbation **44→45** at λ=0. For the simulations involving ligand **44** an intramolecular H-bond between its aniline NH group and its cyclopentyl hydroxyl group is established (see Figure 5B). That conformation is precisely the second-best MM/PBSA docked one (see Figure 1C), which features that intramolecular hydrogen bond. Thus, *batch 2 protocol B* includes the second-best scored MM/PBSA poses for ligands **8**, **16** and **44**. In the case of ligands **8** and **16**, this implies using a pose the pose that resembles the most the X-ray binding mode, while for ligand **44** the second-best scored MM/PBSA pose differs from the best-scored one in the aniline NH dihedral angle (see Figure 1C). The improvement, as shown in Figures 4A and S9, for *protocol B* as compared with *protocol A*, is quite modest. Results are clearly better for the **16→45** and **45→16** perturbations, with the disagreement between experimental and calculated relative binding energy decreasing from 3.5 to 0.3 kcal mol^{-1} (compare Figure S11 and S12), but ligand **44** is still an outlier. Although the experimental relative binding energy for the **44 → 45** perturbation is just 0.1 kcal mol^{-1}, ligand **45** is predicted to bind much more strongly to ACK1 (calculated $\Delta\Delta G_{44\rightarrow45}$ is -1.9 ± 0.1 and -1.4 ± 0.2 kcal mol^{-1} for *protocols A* and *B*, respectively) than **44**. This suggests possible deficiencies in the force field used for **44** in this study.

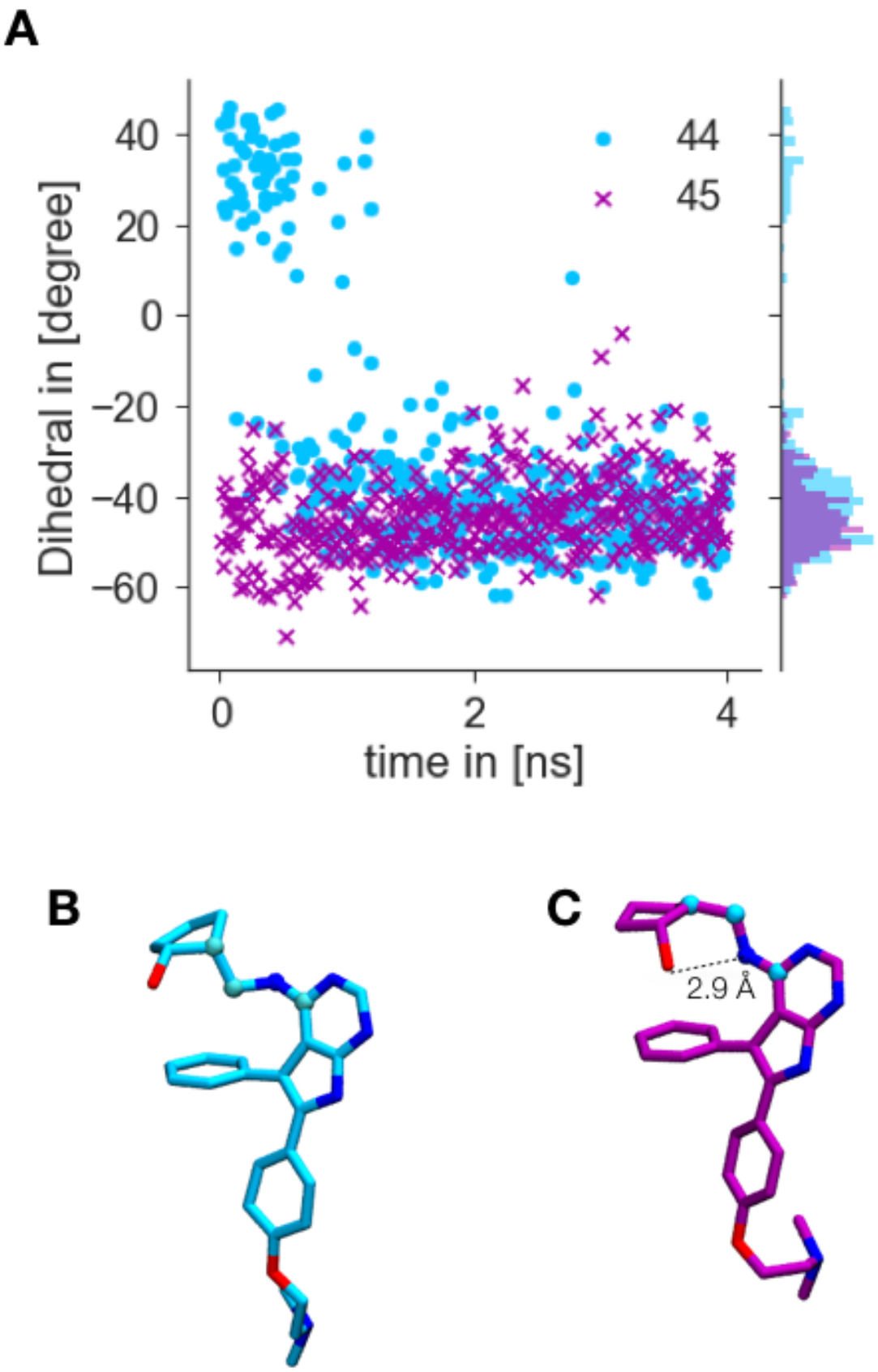

A
Dihedral in [degree]
40
20
0
−20
−40
−60
44
45
0
2
4
time in [ns]
B
C
2.9 Å

Figure 5. (A) 4 ns trajectory monitoring dihedral angle of ligand **44** (blue circles) and **45** (purple crosses) as indicated in (B) and (C) as well as probability distribution of dihedrals over the trajectory. (B) Snapshot of conformation of ligand **44** taken from a $\lambda = 0$ trajectory at t=0 ns indicating dihedral conformation monitored in (A) highlighted by the spheres. (C) Snapshot of conformation of ligand **44** taken from a $\lambda = 0$ trajectory at t=3 ns. Also representative structure of the dihedral trajectory shown in purple in (A).

Protocols C and D, follow the same trends already explained for *batch 1*, pointing to an improvement in the results when a water in the ATP-binding pocket is included (Figure 4A). An encouraging R^2 of 0.76 ± 0.02 and an improvement in the underestimation of relative binding energies (slope 0.8) is obtained, though there is still room for improvements for affinity predictions for **44** and **16**.

Thermodynamic cycle closures analysis

Hysteresis, being defined as the difference in binding energy between the forward and reverse perturbation[35, 52-53], has been proposed as useful metric to identify problematic perturbations[54-55]. Cycle closures for both *batch 1* and *batch 2* were computed to determine whether incorrectly predicted binding poses could be detected in the absence of experimental binding affinity data. Results are shown in Table 1.

Table 1. Calculated thermodynamic cycle closures. Cycle closures that exceed or equal a threshold of 0.8 kcal mol^{-1} are highlighted in bold.

		Cycle closure (kcal mol^{-1})				
	Protocol	A	B	C	D	E
Cycle	3-2-5-6	0.6 ± 0.3	0.0 ± 0.3	0.6 ± 0.6	0.2 ± 0.4	**0.8 ± 0.3**
	3-2-5	0.2 ± 0.2	-0.4 ± 0.2	0.7 ± 0.5	0.0 ± 0.1	0.5 ± 0.3
	2-6-5	**-0.8 ± 0.4**	0.0 ± 0.3	0.0 ± 0.3	-0.2 ± 0.3	-0.7 ± 0.8
	3-5-6	0.4 ± 0.2	0.4 ± 0.2	-0.1 ± 0.4	0.2 ± 0.4	0.4 ± 0.1
	3-6-2	0.2 ± 0.4	0.1 ± 0.3	-0.6 ± 0.5	0.0 ± 0.2	-0.1 ± 0.9
	3-4-7-6	**1.0 ± 0.4**	0.3 ± 0.4	**1.6 ± 0.4**	0.2 ± 0.4	0.3 ± 0.7
	3-7-6	0.2 ± 0.4	0.1 ± 0.3	**0.9 ± 0.7**	0.2 ± 0.3	-0.2 ± 0.9
	4-6-7	**0.9 ± 0.6**	0.0 ± 0.3	**-1.1 ± 0.5**	0.2 ± 0.3	-0.2 ± 0.8
	3-4-7	**0.8 ± 0.4**	0.2 ± 0.4	0.7 ± 0.7	0.0 ± 0.2	0.5 ± 0.5
	3-4-6	**1.9 ± 0.4**	0.3 ± 0.2	0.5 ± 0.5	0.2 ± 0.4	0.1 ± 0.4
	45-16-44	**-2 ± 1**	**-3 ± 1**	**-2 ± 1**	**-1.8 ± 0.9**	N/A
	38-39-35-36	0.6 ± 0.5	0.6 ± 0.5	0.2 ± 0.3	0.2 ± 0.3	N/A

As could be expected, similar conclusions can be obtained when analyzing ring cycle closures or comparing forward and reverse perturbations, although there are some cases with high deviations between experimental and calculated relative binding energies, while exhibiting almost null hysteresis for the forward and reverse perturbations (i.e. the perturbations between ligands **2** and **5** in *batch 1* and those between ligands **44** and **45** in *batch 2*).

Overall it appears that a threshold of ± 0.8 kcal mol^{-1} for cycle closure errors is useful to flag poses that need further attention even without prior knowledge of the experimental binding affinities. Thus, for *protocol A*, **3**-**4**-**7**-**6**, **3**-**4**-**6**, **3**-**4**-**7**, **4**-**6**-**7**, **2**-**6**-**5** and **45**-**16**-**44** thermodynamic cycle closures are indicative of problematic ligands. According to metric, a significant improvement when using *protocol B* (only one thermodynamic cycle closure above the threshold) is seen, while a comparison between *protocols A* (6 cycles with hysteresis above the threshold) and *C* (4 cycles) suggest a

modest improvement. Results for *batch 2* clearly indicate that ligands **44**, **45** and **16** (hysteresis of -2 ± 1 kcal mol^{-1} in their thermodynamic cycle for *protocol A*) are much more problematic than ligands **35**, **36**, **38** and **39** (hysteresis of 0.2 ± 0.3 kcal mol^{-1} for *protocol A*). The best performing *protocol D* is unable to improve the hysteresis for the **45**-**16**-**44** thermodynamic cycle.

DISCUSSION

This work has explored the viability of using alchemical free energy methods as a final filter in a cascade of computational methods for structure-based virtual screens. The two major limitations of AFE methods are the quality of the potential energy function used, and the extent to which the configurational sampling performed has captured relevant protein-ligand conformations[54-55] In principle sufficient long simulations will relax a protein-ligand complex to the ligand pose and protein conformation preferred by the force field used. However, because computing time is limited in practical scenarios, AFE simulations typically afford only a few ns per window, which can make the calculated binding affinities sensitive to the selection of the starting conformations. This work indicates that use of experimental data to bias the selection of poses and setup of binding site water content can lead to significant performance improvements. Of course, as illustrated with ligand **4**, even in cases where the MD simulations relax a previously modelled binding pose to one that closely resembles a pose inferred from X-ray data, the free energy calculations may still fail to reproduce the experimental binding affinities.

Careful analysis of literature structural data[47, 56-57] was key to identify a conserved hydration site that was not modelled in the prior docking calculations. This knowledge

was important to realize upon inspection of putative poses for ligand **7** the feasibility of a hydrogen bonding interaction via a bridging water molecule. Gratifyingly modelling of this hydration site leads to significant accuracy improvements for several perturbations. In principle, assuming an accurate potential energy function, these sampling issues could be dealt with by simply increasing the sampling time of the MD simulations. For the present dataset, we find that a one order of magnitude increase in sampling time was insufficient to bring about improvements in binding poses accuracy and binding site water content. Thus, at present it seems wise to pay attention to the starting conditions of the free energy calculations to maximize cost effectiveness. Where experimental data is lacking, a number of molecular modelling protocols have been proposed to determine location and energetics of important binding site water molecules[58-63].

ASSOCIATED CONTENT

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

Table S1 and Figures S1 to S15 (PDF)

AUTHOR INFORMATION

Corresponding Author

* José M. Granadino-Roldán. Tel.: +34953213057. E-mail jmroldan@ujaen.es

* Julien Michel. Tel.: +441316504797. E-mail julien.michel@ed.ac.uk

Author Contributions

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

ACKNOWLEDGMENT

JMGR would like to thank the Spanish Ministerio de Educación, Cultura y Deporte for funding a research stay under the Salvador de Madariaga program (PR2015-00118, 2015) at Prof. Julien Michel's lab. JM is supported by a University Research Fellowship from the Royal Society. The research leading to these results has received funding from the European Research Council under the European Union's Seventh Framework Programme (FP7/2007-2013)/ERC grant agreement No. 336289.

REFERENCES

1. PhRMA Fact Sheet "Drug Discovery and Development. Understanding theR&D process". http://www.phrma.org/graphic/four-facts-about-spending-on-prescription-medicines (accessed 30.08.2017).
2. Whitesides, G. M.; Krishnamurthy, V. M., Designing ligands to bind proteins. *Q. Rev. Biophys.* **2006,** *38* (4), 385-395.
3. Smith, A. J. T.; Zhang, X.; Leach, A. G.; Houk, K. N., Beyond Picomolar Affinities: Quantitative Aspects of Noncovalent and Covalent Binding of Drugs to Proteins. *J. Med. Chem.* **2009,** *52* (2), 225-233.
4. Kollman, P. A.; Massova, I.; Reyes, C.; Kuhn, B.; Huo, S.; Chong, L.; Lee, M.; Lee, T.; Duan, Y.; Wang, W.; Donini, O.; Cieplak, P.; Srinivasan, J.; Case, D. A.; Cheatham, T. E., Calculating Structures and Free Energies of Complex Molecules: Combining Molecular Mechanics and Continuum Models. *Acc. Chem. Res.* **2000,** *33* (12), 889-897.
5. Aqvist, J.; Marelius, J., The linear interaction energy method for predicting ligand binding free energies. *Comb. Chem. High Throughput Screen.* **2001,** *4* (8), 613-26.
6. Zwanzig, R. W., High-Temperature Equation of State by a Perturbation Method. I. Nonpolar Gases. *J. Chem. Phys.* **1954,** *22* (8), 1420-1426.
7. Michel, J.; Foloppe, N.; Essex, J. W., Rigorous Free Energy Calculations in Structure-Based Drug Design. *Mol. Inform.* **2010,** *29* (8-9), 570-8.
8. Granadino-Roldan, J. M.; Garzon, A.; Gomez-Gutierrez, P.; Pasamontes-Funez, I.; Santos Tomas, M.; Rubio-Martinez, J., A multistep docking and scoring protocol for congeneric series: Implementation on kinase DFG-out type II inhibitors. *Future Med. Chem.* **2018,** *10* (3), 297-318.
9. Yin, J.; Henriksen, N. M.; Slochower, D. R.; Shirts, M. R.; Chiu, M. W.; Mobley, D. L.; Gilson, M. K., Overview of the SAMPL5 host–guest challenge: Are we doing better? *J. Comput. Aided Mol. Des.* **2017,** *31* (1), 1-19.
10. Muddana, H. S.; Fenley, A. T.; Mobley, D. L.; Gilson, M. K., The SAMPL4 host-guest blind prediction challenge: an overview. *J. Comput. Aided Mol. Des.* **2014,** *28* (4), 305-17.
11. Deng, N.; Flynn, W. F.; Xia, J.; Vijayan, R. S.; Zhang, B.; He, P.; Mentes, A.; Gallicchio, E.; Levy, R. M., Large scale free energy calculations for blind predictions of

protein-ligand binding: the D3R Grand Challenge 2015. *J. Comput. Aided Mol. Des.* **2016,** *30* (9), 743-751.
12. Gaieb, Z.; Liu, S.; Gathiaka, S.; Chiu, M.; Yang, H.; Shao, C.; Feher, V. A.; Walters, W. P.; Kuhn, B.; Rudolph, M. G.; Burley, S. K.; Gilson, M. K.; Amaro, R. E., D3R Grand Challenge 2: blind prediction of protein-ligand poses, affinity rankings, and relative binding free energies. *J. Comput. Aided Mol. Des.* **2018,** *32* (1), 1-20.
13. Bosisio, S.; Mey, A.; 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.
14. Mey, A.; Juarez-Jimenez, J.; Hennessy, A.; Michel, J., Blinded predictions of binding modes and energies of HSP90-alpha ligands for the 2015 D3R grand challenge. *Bioorg. Med. Chem.* **2016,** *24* (20), 4890-4899.
15. Athanasiou, C.; Vasilakaki, S.; Dellis, D.; Cournia, Z., Using physics-based pose predictions and free energy perturbation calculations to predict binding poses and relative binding affinities for FXR ligands in the D3R Grand Challenge 2. *J. Comput. Aided Mol. Des.* **2018,** *32* (1), 21-44.
16. 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), 2695-703.
17. Shoichet, B. K.; Walters, W. P.; Jiang, H.; Bajorath, J., Advances in Computational Medicinal Chemistry: A Reflection on the Evolution of the Field and Perspective Going Forward. *J. Med. Chem.* **2016,** *59* (9), 4033-4.
18. Steinbrecher, T. B.; Dahlgren, M.; Cappel, D.; Lin, T.; Wang, L. L.; Krilov, G.; Abel, R.; Friesner, R.; Sherman, W., Accurate Binding Free Energy Predictions in Fragment Optimization. *J. Chem. Inf. Mod.* **2015,** *55* (11), 2411-2420.
19. Jiao, X.; Kopecky, D. J.; Liu, J.; Liu, J.; Jaen, J. C.; Cardozo, M. G.; Sharma, R.; Walker, N.; Wesche, H.; Li, S.; Farrelly, E.; Xiao, S. H.; Wang, Z.; Kayser, F., Synthesis and optimization of substituted furo[2,3-d]-pyrimidin-4-amines and 7H-pyrrolo[2,3-d]pyrimidin-4-amines as ACK1 inhibitors. *Bioorg. Med. Chem. Lett.* **2012,** *22* (19), 6212-7.
20. Mahajan, K.; Mahajan, N. P., Shepherding AKT and androgen receptor by Ack1 tyrosine kinase. *J. Cell. Physiol.* **2010,** *224* (2), 327-333.
21. Chua, B. T.; Lim, S. J.; Tham, S. C.; Poh, W. J.; Ullrich, A., Somatic mutation in the ACK1 ubiquitin association domain enhances oncogenic signaling through EGFR regulation in renal cancer derived cells. *Mol. Oncol.* **2010,** *4* (4), 323-334.
22. *Molecular Operating Environment (MOE) 2009.1. Chemical Computing Group Inc., 1010 Sherboke St. West, Suite #90, Montreal, QC, Canada, H3A 2R7, 2013.* 2013.08 ed.
23. Lougheed, J. C.; Chen, R. H.; Mak, P.; Stout, T. J., Crystal structures of the phosphorylated and unphosphorylated kinase domains of the Cdc42-associated tyrosine kinase ACK1. *J. Biol. Chem.* **2004,** *279* (42), 44039-45.
24. 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, R. C.; Wang, J.; Wolf, R. M.; Wu, X.; Kollman, P. A., *AMBER 14*. University of California, San Francisco: University of California, San Francisco, 2014.
25. 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* (9), 1157-1174.
26. Jakalian, A.; Bush, B. L.; Jack, D. B.; Bayly, C. I., Fast, efficient generation of high-quality atomic charges. AM1-BCC model: I. Method. *J. Comput. Chem.* **2000,** *21* (2), 132-146.
27. 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.
28. Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C., Comparison of multiple Amber force fields and development of improved protein backbone parameters. *Proteins: Struct. Funct. Bioinform.* **2006,** *65* (3), 712-725.
29. 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), 926-935.
30. Åqvist, J., Ion-water interaction potentials derived from free energy perturbation simulations. *The Journal of Physical Chemistry* **1990,** *94* (21), 8021-8024.
31. Darden, T.; York, D.; Pedersen, L., Particle mesh Ewald: An N log(N) method for Ewald sums in large systems. *J. Chem. Phys.* **1993,** *98* (12), 10089-10092.
32. Miller, B. R.; McGee, T. D.; Swails, J. M.; Homeyer, N.; Gohlke, H.; Roitberg, A. E., MMPBSA.py: An Efficient Program for End-State Free Energy Calculations. *Journal of Chemical Theory and Computation* **2012,** *8* (9), 3314-3321.
33. 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), 639-58.
34. Loeffler, H. H.; Michel, J.; Woods, C., FESetup: Automating Setup for Alchemical Free Energy Simulations. *J Chem Inf Model* **2015,** *55* (12), 2485-90.
35. Calabro, G.; Woods, C. J.; Powlesland, F.; Mey, A. 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), 5340-50.
36. Mey, A. S. J. S.; Juárez-Jiménez, J.; Hennessy, A.; Michel, J., Blinded predictions of binding modes and energies of HSP90-α ligands for the 2015 D3R grand challenge. *Biorg. Med. Chem.* **2016,** *24* (20), 4890-4899.
37. Brown, S. P.; Muchmore, S. W.; Hajduk, P. J., Healthy skepticism: assessing realistic model performance. *Drug Discovery Today* **2009,** *14* (7), 420-427.
38. 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.
39. 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. *Journal of Chemical Theory and Computation* **2015,** *11* (7), 3333-3345.
40. Chow, K.-H.; Ferguson, D. M., Isothermal-isobaric molecular dynamics simulations with Monte Carlo volume sampling. *Comput. Phys. Commun.* **1995,** *91* (1), 283-289.
41. Åqvist, J.; Wennerström, P.; Nervall, M.; Bjelic, S.; Brandsdal, B. O., Molecular dynamics simulations of water and biomolecules with a Monte Carlo constant pressure algorithm. *Chem. Phys. Lett.* **2004,** *384* (4), 288-294.

42. Andersen, H. C., Molecular dynamics simulations at constant pressure and/or temperature. *J. Chem. Phys.* **1980,** *72* (4), 2384-2393.
43. Barker, J. A.; Watts, R. O., Monte Carlo studies of the dielectric properties of water-like models. *Mol. Phys.* **1973,** *26* (3), 789-792.
44. Mikulskis, P.; Genheden, S.; Ryde, U., A Large-Scale Test of Free-Energy Simulation Estimates of Protein–Ligand Binding Affinities. *J. Chem. Inf. Mod.* **2014,** *54* (10), 2794-2806.
45. Shirts, M. R.; Chodera, J. D., Statistically optimal analysis of samples from multiple equilibrium states. *J. Chem. Phys.* **2008,** *129* (12), 124105.
46. Mey, A.; Jimenez, 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.
47. Gajiwala, K. S.; Maegley, K.; Ferre, R.; He, Y.-A.; Yu, X., Ack1: Activation and Regulation by Allostery. *PLoS One* **2013,** *8* (1), e53994.
48. Humphrey, W.; Dalke, A.; Schulten, K., VMD: visual molecular dynamics. *J. Mol. Graph.* **1996,** *14* (1), 33-8, 27-8.
49. *Origin (OriginLab, Northampton, MA)* OriginLab, Northampton, MA.
50. Hunter, J. D., Matplotlib: A 2D graphics environment. *Comput Sci Eng* **2007,** *9* (3), 90-95.
51. https://github.com/mwaskom/seaborn.
52. König, G.; Hudson, P. S.; Boresch, S.; Woodcock, H. L., Multiscale Free Energy Simulations: An Efficient Method for Connecting Classical MD Simulations to QM or QM/MM Free Energies Using Non-Boltzmann Bennett Reweighting Schemes. *Journal of Chemical Theory and Computation* **2014**, 140303161846003.
53. Michel, J.; Tirado-Rives, J.; Jorgensen, W. L., Prediction of the water content in protein binding sites. *J. Phys. Chem. B* **2009,** *113* (40), 13337-46.
54. 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.
55. 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* (3), 638-649.
56. Kopecky, D. J.; Hao, X.; Chen, Y.; Fu, J.; Jiao, X.; Jaen, J. C.; Cardozo, M. G.; Liu, J.; Wang, Z.; Walker, N. P. C.; Wesche, H.; Li, S.; Farrelly, E.; Xiao, S.-H.; Kayser, F., Identification and optimization of N3,N6-diaryl-1H-pyrazolo[3,4-d]pyrimidine-3,6-diamines as a novel class of ACK1 inhibitors. *Biorg. Med. Chem. Lett.* **2008,** *18* (24), 6352-6356.
57. Lougheed, J. C.; Chen, R.-H.; Mak, P.; Stout, T. J., Crystal Structures of the Phosphorylated and Unphosphorylated Kinase Domains of the Cdc42-associated Tyrosine Kinase ACK1. *J. Biol. Chem.* **2004,** *279* (42), 44039-44045.
58. Luccarelli, J.; Michel, J.; Tirado-Rives, J.; Jorgensen, W. L., Effects of Water Placement on Predictions of Binding Affinities for p38alpha MAP Kinase Inhibitors. *J. Chem. Theory Comput.* **2010,** *6* (12), 3850-3856.
59. Bodnarchuk, M. S.; Viner, R.; Michel, J.; Essex, J. W., Strategies to calculate water binding free energies in protein-ligand complexes. *J Chem Inf Model* **2014,** *54* (6), 1623-33.
60. Michel, J.; Henchman, R. H.; Gerogiokas, G.; Southey, M. W. Y.; Mazanetz, M. P.; Law, R. J., Evaluation of Host–Guest Binding Thermodynamics of Model Cavities with Grid Cell Theory. *Journal of Chemical Theory and Computation* **2014,** *10* (9), 4055-4068.

61. Sindhikara Daniel, J.; Yoshida, N.; Hirata, F., Placevent: An algorithm for prediction of explicit solvent atom distribution—Application to HIV-1 protease and F-ATP synthase. *J. Comput. Chem.* **2012,** *33* (18), 1536-1543.
62. Sridhar, A.; Ross, G. A.; Biggin, P. C., Waterdock 2.0: Water placement prediction for Holo-structures with a pymol plugin. *PLoS One* **2017,** *12* (2), e0172743.
63. Hu, B.; Lill Markus, A., WATsite: Hydration site prediction program with PyMOL interface. *J. Comput. Chem.* **2014,** *35* (16), 1255-1260.