Supporting Information: Alchemical Absolute Protein-Ligand Binding Free Energies for Drug Design

Y. Khalak[1], G. Tresadern[2], M. Aldeghi[3], H. M. Baumann[4], D. L. Mobley[4,5], B. L. de Groot[1], and V. Gapsys[1,*]

[1] Computational Biomolecular Dynamics Group, Department of Theoretical and Computational Biophysics, Max Planck Institute for Biophysical Chemistry, D-37077, Göttingen, Germany.
[2] Computational Chemistry, Janssen Research & Development, Janssen Pharmaceutica N. V., Turnhoutseweg 30, 2340, Beerse, Belgium.
[3] MIT Department of Chemical Engineering, Massachusetts Institute of Technology, Cambridge, MA 02139, USA.
[4] Department of Pharmaceutical Sciences, University of California, Irvine, CA 92697, USA.
[5] Department of Chemistry, University of California, Irvine, CA 92697, USA.
[*] vgapsys@gwdg.de

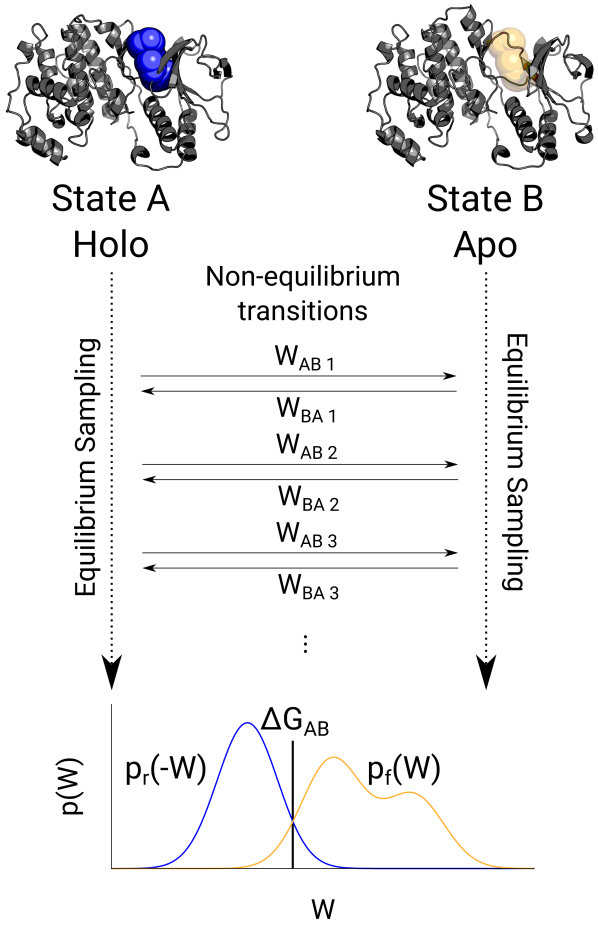

Figure S1: Scheme for non-equilibrium free energy calculations. Firstly, molecular dynamics simulations are performed to obtain equilibrium ensembles of the protein's holo and apo states. Subsequently, conformations from the equilibrium ensembles are extracted and rapid out-of-equilibrium simulations are started driving the system from holo to apo state and vice versa. Work values from these transitions are recorded and Crooks Fluctuation Theorem[5] is applied to relate the work distributions to the free energy difference between the two end-states. This protocol is particularly suited for the absolute ligand-protein binding free energy calculations. The two states, apo and holo, can be initialized independently with the crystallographically resolved structures (in case they are available) and the non-equilibrium transitions, in turn, allow connecting the two states.

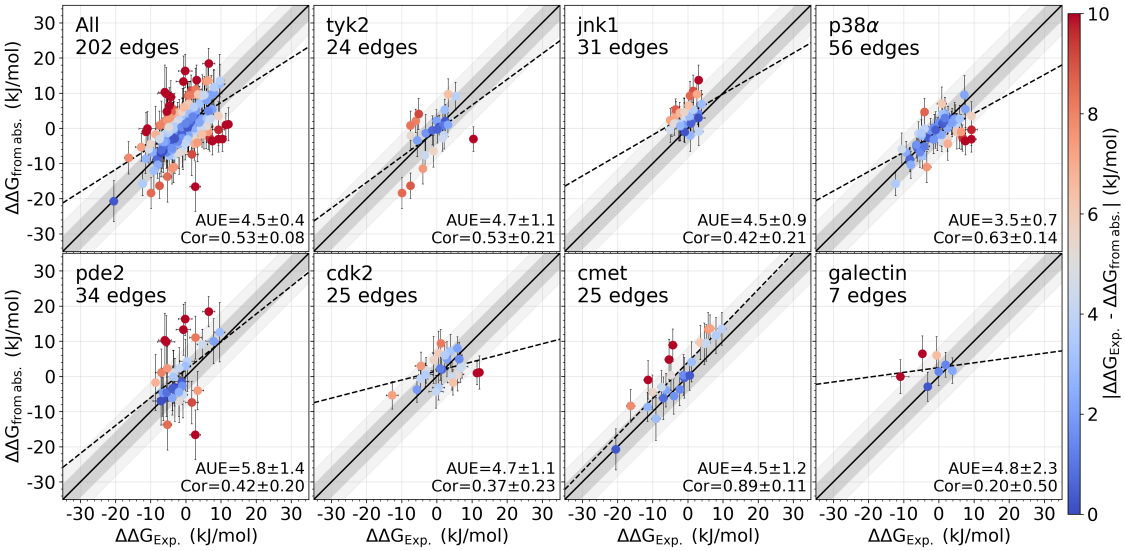

Figure S2: Accuracy of relative binding free energies recalculated from absolute ones compared to experimental measurements and broken down by system. Average unsigned error (AUE) is in kJ/mol. Dark and light shaded areas represent regions deviating from experiment by at most 1 and 2 kcal/mol.

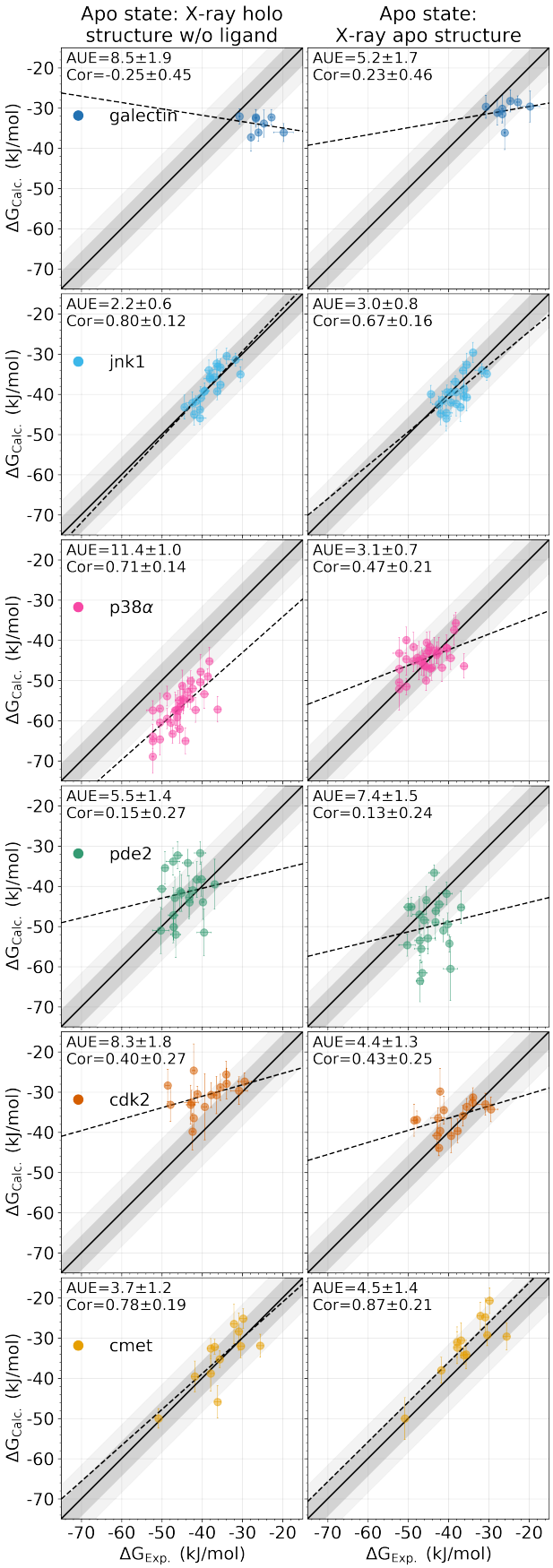

Figure S3: Calculated binding ΔG estimates plotted against the experimental values with data separated by system. Representing the protein in decoupled ligand systems with holo (left) and apo (right) structures can lead to different accuracies and systematic shifts in the predicted ΔG values. Dark and light shaded areas represent regions deviating from experiment by at most 1 and 2 kcal/mol.

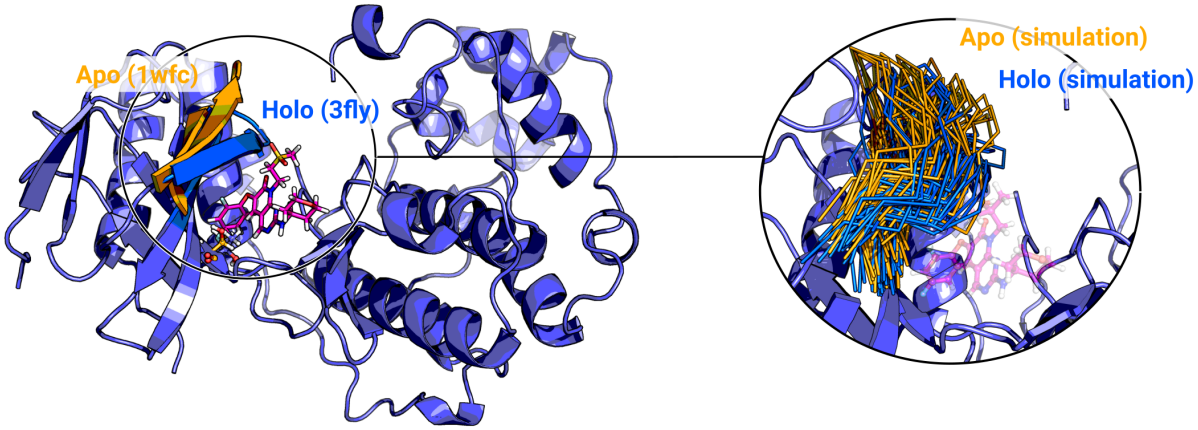

Figure S4: The experimentally resolved apo (1wfc) and holo (3fly) structures differ in a substantial loop motion close to the ligand's binding site. The inset illustrates that in the simulations without the ligand that are started either from the 1wfc or from 3fly structure, the loop motion is sufficiently sampled.

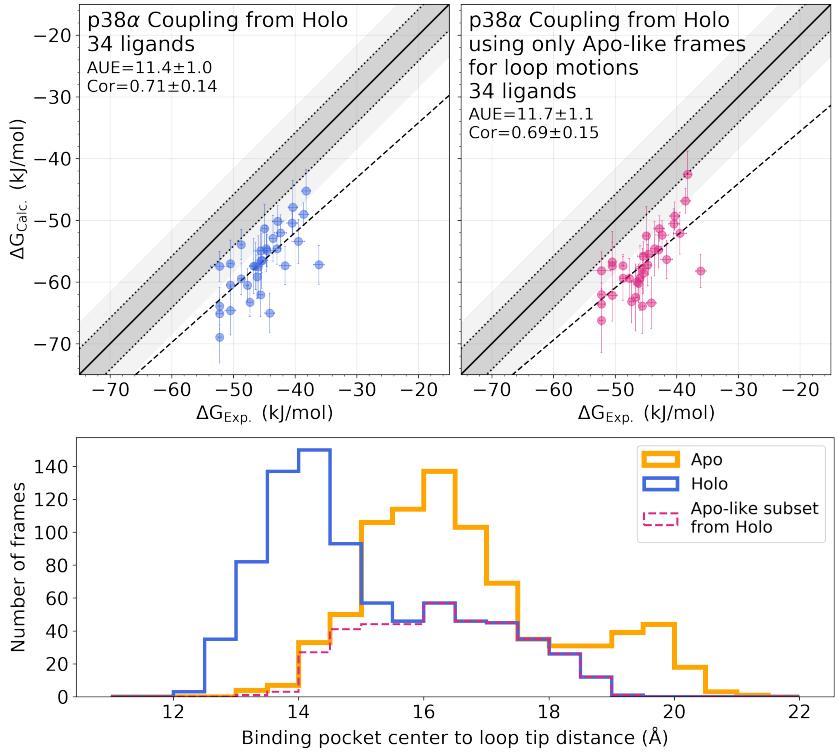

Figure S5: Calculated binding ΔG estimates for the p38α system plotted against experimental values. Left panel: the simulations in the apo state were started from the experimental holo (3fly) structure after removing the ligand. Right panel: the structural ensemble for the apo state that was used for the calculations in the left panel was filtered by retaining only those structures that are more similar to the experimentally resolved apo (1wfc) structure in terms of the loop motion at the binding site. Bottom panel: Histogram quantifying the loop motion in terms of the distance between the tip of the loop (C_α of Tyr 35) and the centroid of the binding pocket. Filtering captures only the extended states of the loop. Dark and light shaded areas represent regions deviating from experiment by at most 1 and 2 kcal/mol.

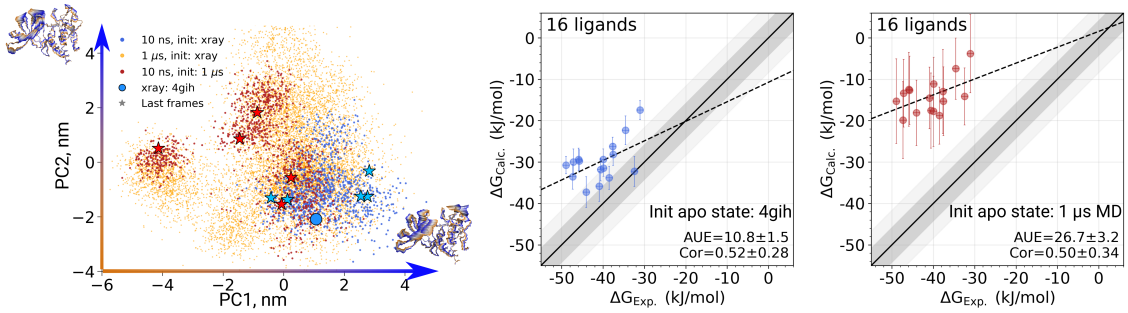

Figure S6: Sampling effects on the trajectories and calculated ligand binding ΔG for the tyk2 kinase. Left panel: projection of the simulated apo tyk2 trajectories on the two principal components with the largest eigenvalues. The short (10 ns) equilibrium ensembles initialized with the 4gih[11] structure with the ligand removed are shown in blue, 1 μs simulations in orange and 10 ns equilibrations started from the end-states of the 1 μs runs are in red. Middle and right panels: binding ΔG calculated by initializing apo state simulations with the 4gih structure and the final configurations from 1 μs simulations, respectively. Dark and light shaded areas represent regions deviating from experiment by at most 1 and 2 kcal/mol.

Long sampling of the tyk2 kinase apo state

As no crystallographic structure for the tyk2 kinase in its apo state is available, we attempted to further relax the experimentally resolved holo state 4gih[11] with the removed ligand. We performed 5 independent 1 μs unbiased molecular dynamics simulations. The generated trajectories have explored a broad range of conformations, deviating substantially from the starting structure (Fig S11, left panel). The ΔG estimates obtained using 5 structural apo ensembles equilibrated for 1 μs each (right panel) highlight how the significant increase in sampling deteriorates the prediction accuracy.

Effects of using crystallographic water

Solvation of binding pocket in the protein-ligand system may play a significant role in determining the binding affinity. In our current protocol initial solvent positions around the protein and the ligand did not include crystallographically resolved waters. To probe the influcence of crystallographic water placement on the final ΔG estimates we have repeated the procedure by including the experimentally resolved waters (Fig. S7).

Our results show that while adding crystallographic water can improve accuracy, like in the case of tyk2, often it provides no benefit, and in the cases of galectin and pde2 is even detrimental.

For example, ligand binding free energies for the pde2 system increased by on average 8 kJ/mol when crystallographic water was used. Here, both the apo and holo states were initialized from the 6ezf crystal structure, which was previously shown to better reproduce relative binding free energies than the alternatives [12] and which we showed to better represent the apo state. However, 6ezf was resolved with a much smaller ligand than those used herein. In turn, incorporating crystallographic waters into the currently simulated protein-ligand system, results in more water molecules in the binding pocket and close to the ligands (Fig. S8A), with two waters even starting out in a sterical clash with some of the molecules of the probed set. While the clash is resolved during energy minimization, the number of waters in the binding pocket remains higher than in the simulations without crystallographic water (Fig. S8B). This increases water contacts with hydrophobic areas of the ligands, as seen from the radial distribution function of water oxygens around the ligand carbon atoms (Fig. S8C). The apo simulations meanwhile are not significantly affected, resulting in reduced binding affinity.

On the other hand, using crystallographic water in the tyk2 system increased the binding affinity of most ligands, leading to better agreement with the experimental measurement. We traced this effect to an increase in water density in areas where conserved parts of the ligands hydrogen bond with water (Fig. S9A). This results in more hydrogen bonds between the ligands and solvent when using crystallographic water (Fig. S9B), leading to more stable holo states and tighter binding.

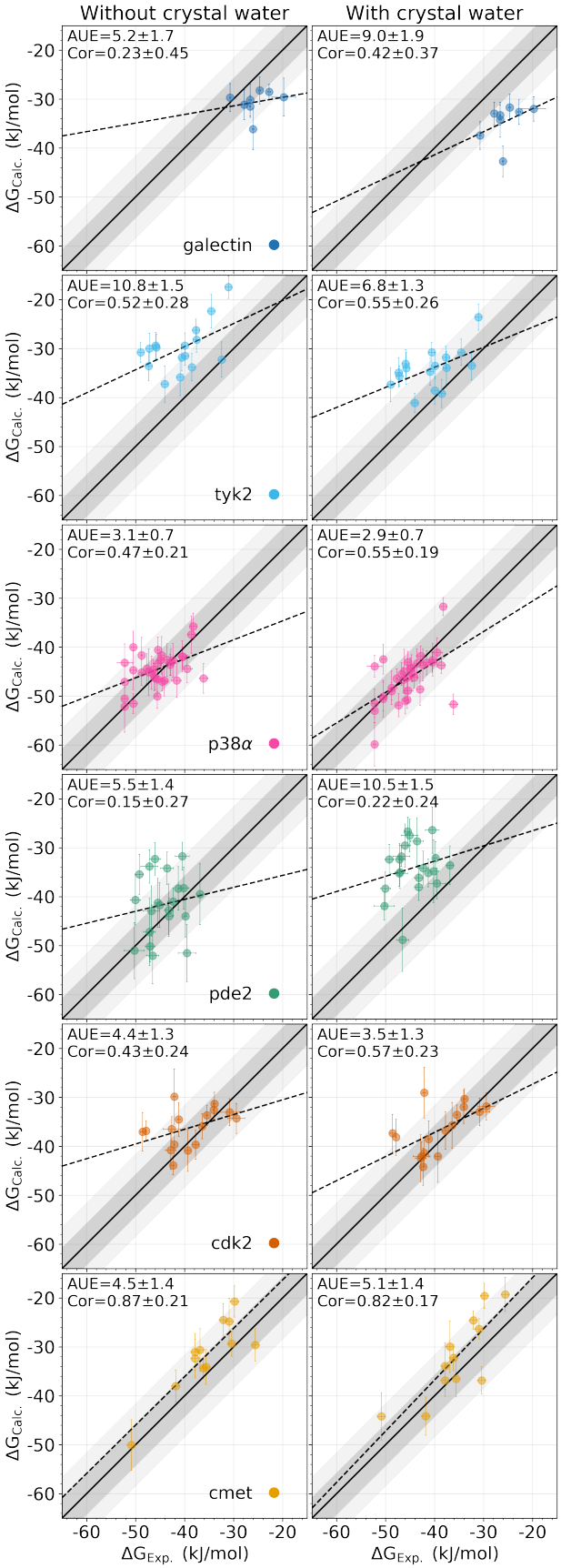

Figure S7: Effects of crystallographic water on predicted absolute binding free energies. The left column shows the simulations started from structures where no crystallographically resolved water was used to initialize the simulations. In the column on the right, crystallographic water positions were retained for both apo and holo crystal structures. Dark and light shaded areas represent regions deviating from experiment by at most 1 and 2 kcal/mol.

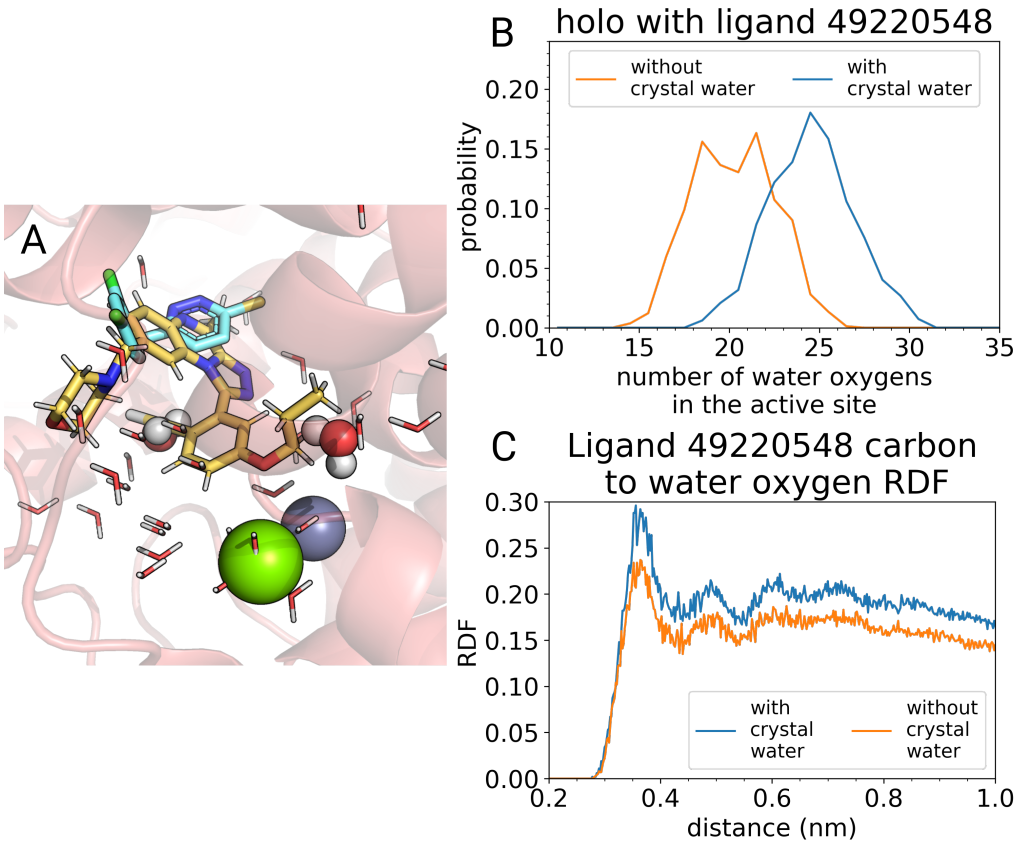

Figure S8: Effects of crystallographic water on pde2-ligand systems. **A** Ligand 49220548 is shown in yellow, while the ligand from the 6ezf crystal structure is shown in cyan. Initial positions of some crystalographic waters are close to the studied ligands. Water molecules that directly clash with the ligand are highlighted as spheres. Crystallographic ions are also shown as larger spheres. **B** Distributions of the number of water oxygens within the active site (approximated here as volume within 5 Å of the starting coordinates of ligand 49220548) show an increased number of waters in holo equilibrium simulations. This holds true for all other studied pde2 ligands as well (not shown). **C** Radial distribution functions (RDF) of water oxygens around the carbon atoms of the ligand 49220548 show an increase of the hydrophobic contacts for the simulations initialized using crystallographic waters.

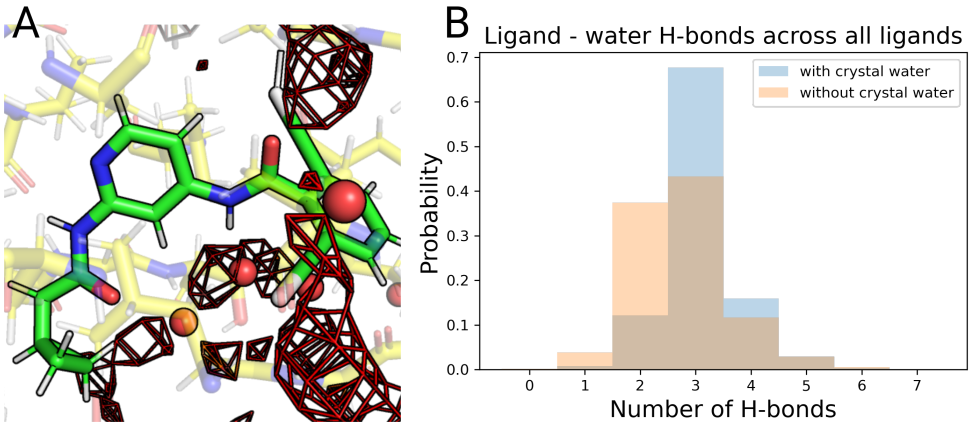

Figure S9: Effects of crystallographic water on tyk2-ligand systems. **A** Difference in water density around ligand ejm_45 between simulations initialized with crystallographic water and without. The red mesh indicates an isosurface where the water density is 0.04 arb. units higher when using crystallographic water, as calculated with GROmaps[4]. Crystallographic water oxygen positions are highlighted as red as spheres. **B** Distributions of the number of ligand-water hydrogen bonds tabulated across all tyk2 ligands.

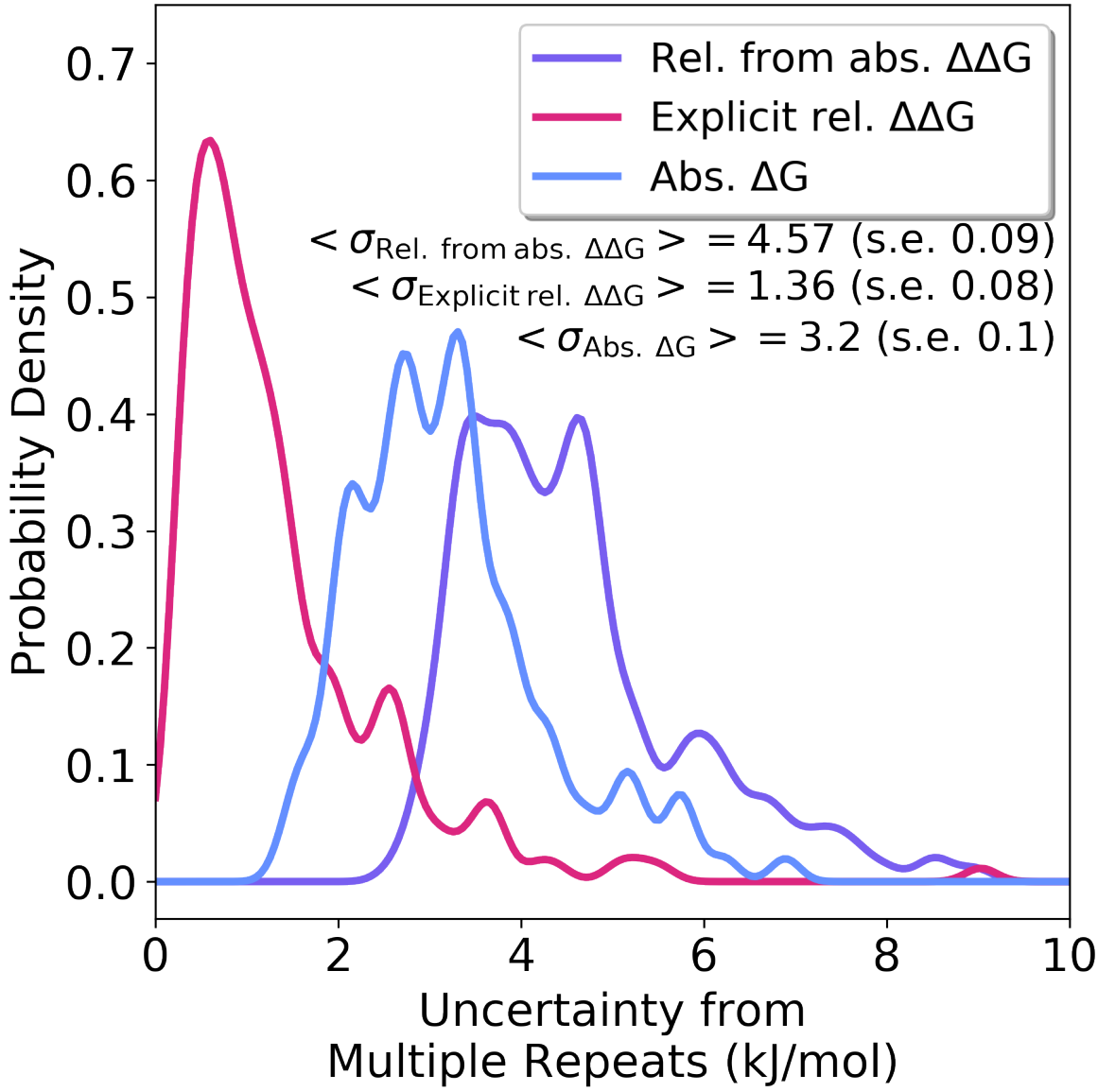

Figure S10: Distributions of uncertainty estimates for absolute and relative binding free energies. Distributions smoothed with a Gaussian kernel density estimator. Mean uncertainties and their standard errors are reported in units of kJ/mol. Computing relative binding free energies directly through alchemical transformations of the ligands results in much smaller uncertainties, as little work is dissipated. For absolute binding free energies, the uncertainties are much larger. Recalculating the relative free energies from absolute ones results in even larger uncertainties due to propagation of errors.

Section 1. Generation of restraints

When the ligand is decoupled from the protein (main text Fig. 7, state B), it is able to explore the whole simulation box. This leads to poor convergence of the binding free energy, as it is difficult to fully sample the movement of the ligand over a large volume. To limit the accessible volume, restraints can be applied to restrict the ligand's translational and rotational degrees of freedom[2]. In this work we followed a rigorous protein-ligand restraining scheme[3], where three atoms of the protein and three of the ligand are connected with harmonic potentials: one for distance, two for angles, and three for dihedrals. These restraints keep the decoupled ligand close to its coupled position and orientation within the binding pocket.

As harmonic potentials translate into Gaussian probability distributions in equilibrium, the protein and ligand atoms the restraints are acting on were selected to ensure a Gaussian distribution for each of the six restrained degrees of freedom in the superpositioned ensemble (main text Fig. 7, state B). The Gaussian approximations for the distributions were fit by computing the mean and variance of the degrees of freedom to be used for restraining (after correcting for periodicity of angles and dihedrals). Atom sets where these ideal Gaussian distributions were too different from the superpositioned ensemble (p-value at or above 0.05 in a Kolmogorov-Smirnov test) were rejected.

Due to an implementation detail within GROMACS[1], an additional condition has to be satisfied by a proper choice of atoms to restrain. As restrained angles are calculated via a cross product of two vectors $\theta_i = \arcsin(|\hat{\mathbf{a}} \times \hat{\mathbf{b}}|)$, crossing the periodic limits of $0 \leq \theta_i < \pi$ during a simulation is not possible leading to non-Gaussian distributions of θ_i near those regions. To prevent this from happening in NEQ morphing simulations, atoms that would produce a restraining potential on an angle θ_i with values $U(\theta_i = 0) < 5kT$ or $U(\theta_i = \pi) < 5kT$ (where k is the Boltzmann constant and T the temperature) were also rejected.

Section 2. Corrections for effects of superpositioning

When a ligand in the active site is coupled to the protein the potential acting between the molecules leads to correlations between their motions, resulting in reduced fluctuations in ligand coordinates. The superpositioning approach for generating the conformations of the decoupled ligand in the active site uses the coupled trajectory as a reference. This transfers some correlations to the generated conformations, limiting the variability of those poses as well (Fig. S11). The presence of correlations can be detected by evaluating the Pearson correlation coefficient between different restrained degrees of freedom after superpositioning (Fig. S12).

The six independent harmonic protein-ligand restraints are only an approximation for the potential required to produce the distribution obtained by superpositioning, which is not the same as the equilibrium distribution produced by the Boresch et al.[3] orthogonal harmonic restraints. Starting the morphing simulations that drive the decoupled state of a ligand in the protein to the coupled state using these conformations under the independent restraints would violate the assumptions of the Crooks theorem. A better approximation for the restraining potential can be obtained by treating the superpositioned distribution as a multivariate normal distribution, a correlated six-dimensional Gaussian:

$$\boldsymbol{\zeta} = [\zeta_0, \zeta_1, \zeta_2, \zeta_3, \zeta_4, \zeta_5] = [r_\mathrm{d}, \theta_\mathrm{A}, \theta_\mathrm{B}, \phi_\mathrm{A}, \phi_\mathrm{B}, \phi_\mathrm{C}] \tag{S1}$$

$$P(\boldsymbol{\zeta}) = \frac{\exp(-\frac{1}{2}(\boldsymbol{\zeta} - \boldsymbol{\mu})^\mathrm{T})\boldsymbol{\Sigma}^{-1}(\boldsymbol{\zeta} - \boldsymbol{\mu}))}{\sqrt{(2\pi)^3 \det \boldsymbol{\Sigma}}} \quad , \tag{S2}$$

where $\boldsymbol{\mu}$ is the vector of restraint equilibrium values (or means of the one dimensional distributions) and $\boldsymbol{\Sigma}$ is the covariance matrix of the restrained degrees of freedom. The restraining potential that can reproduce this can be approximated via Boltzman inversion.

$$U_\mathrm{correlated}(\boldsymbol{\zeta}) = -kT \ln(P(\boldsymbol{\zeta})) \tag{S3}$$

Such a multivariate potential, however, is incompatible with the GROMACS molecular dynamics engine and implementing new potentials into the highly optimized software is beyond the scope of this study. Instead, to correct for the effect of these correlations without modifying the GROMACS source code we developed and tested two decorrelation approaches.

The first approach (Fig. S13 left), called *a priori*, reproduces the equilibrium distributions of the decoupled state expected from the uncorrelated restraints. This is achieved by independently sampling restraint coordinates for each frame from the independent Gaussian distributions. The ligand in each structure of the superpositioned ensemble is then rotated and translated to satisfy those sampled restraint coordinates. Simulations of the alchemical ligand coupling to the rest of the system are then performed starting from the generated ensemble.

The second approach (Fig. S13 right), refered to as *post hoc* and used as part of the superpositioning protocol, performs the alchemical transition simulations starting from the superpositioned configurations, but adjusts the work

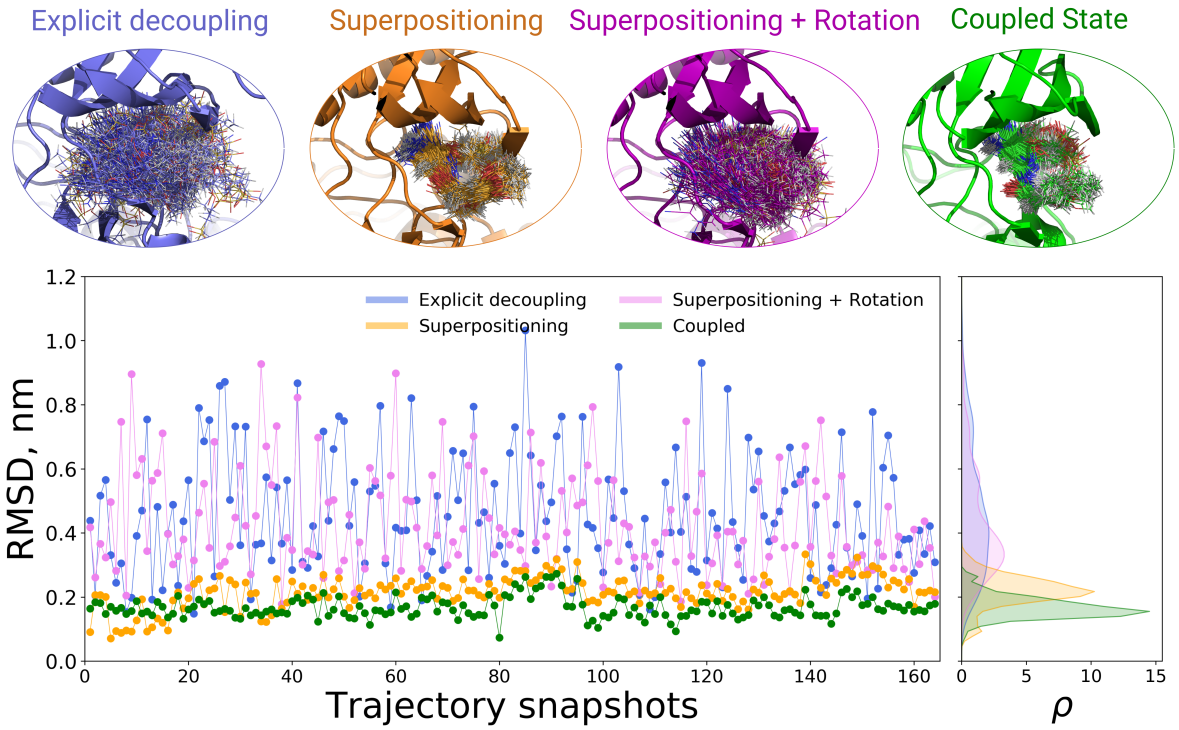

Figure S11: Ensembles of ligand poses in the decoupled and coupled states and their RMSD fluctuations. Decoupled states generated with superpositioning (orange) inherit the limited range of motion and correlations with protein motion present in the coupled state (green). Adding rotational and translational noise consistent with the restraining potential (purple) to the superpositioning ensemble recovers an ensemble close to that of the explicitly simulated decoupled state (blue). Structures and RMSD data shown are limited to a single independent repeat of the 18660-1 ligand[7] with jnk1.

values associated with transitions from the decoupled ligand to its coupled state by adding the free energy difference between the two restraining potentials $\Delta G_{\mathrm{MC}} - \Delta G_{\mathrm{restr}}$. While the free energy imposed by the uncorrelated restraints can be calculated analytically[3], that for the correlated restraints is computed via Monte Carlo integration of $U_{\mathrm{correlated}}(\zeta)$. Fig. S15 shows this method to have a better agreement with the more expensive approach of explicitly simulating the decoupled state.

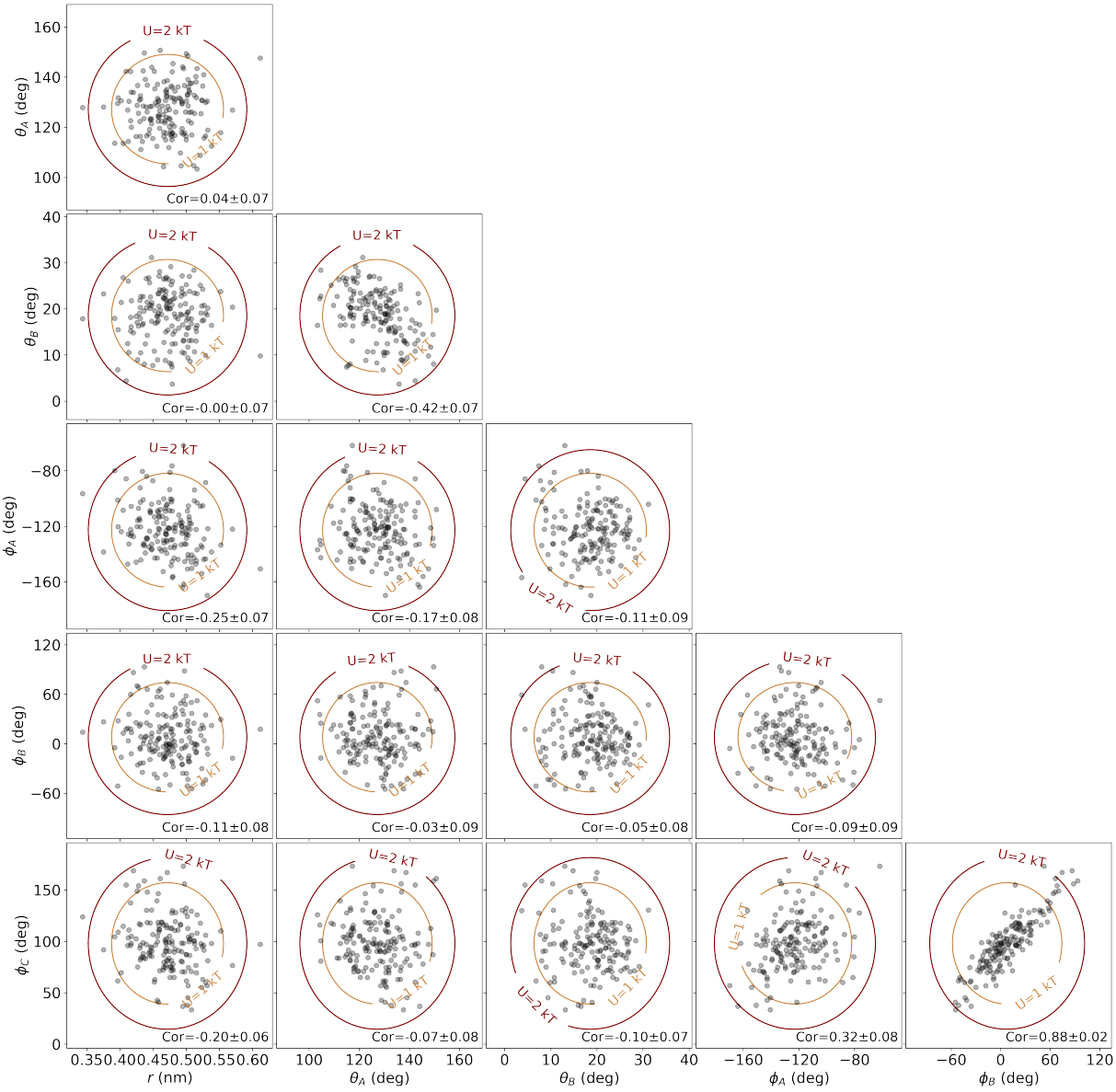

Figure S12: Example of correlated restraint distributions introduced by superpositioning. Restraint coordinates of frames produced by superpositioning are shown as black points while the contour lines represent the strength of the restraining potential. This data comes from the superpositioned trajectory for jnk1 in the apo conformation with the 18660-1 ligand. In this example, the dihedral restraint coordinates ϕ_B and ϕ_C are highly correlated.

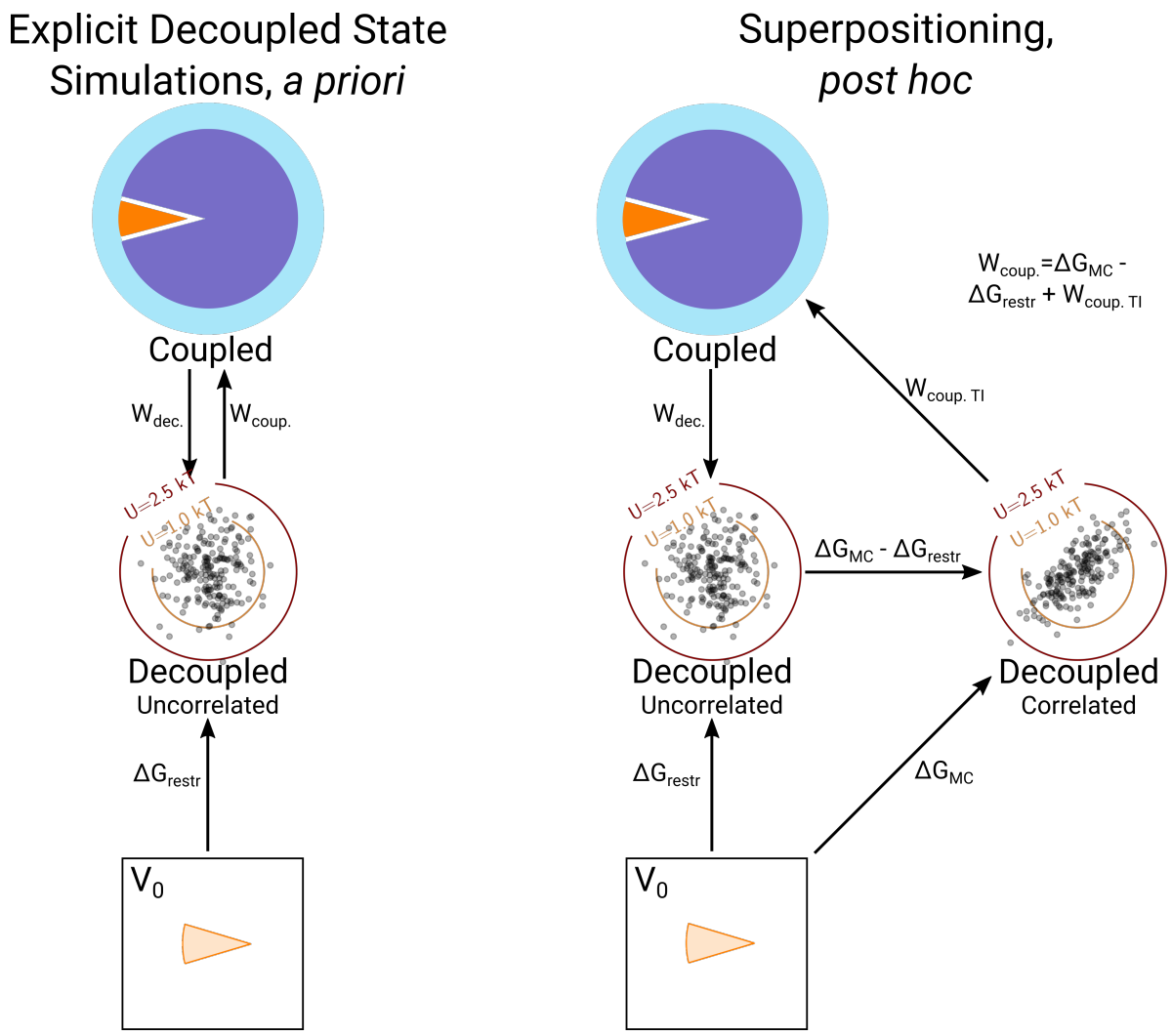

Figure S13: Schemes for decorrelation approaches. While *a priori* decorrelation (left) reproduces the equilibrium distribution of the decoupled end state with the orthogonal restraints, *post hoc* decorrelation (right) corrects for the free energy difference between the ensembles produced by the orthogonal and correlated restraints by modifying the work values of the coupling transition.

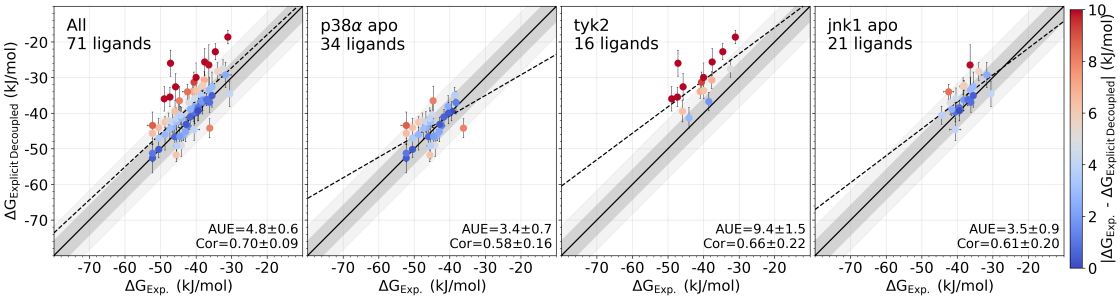

Figure S14: Absolute binding free energies obtained with explicit equilibrium simulations of the decoupled states broken down by system. Dark and light shaded areas represent regions deviating from experiment by at most 1 and 2 kcal/mol.

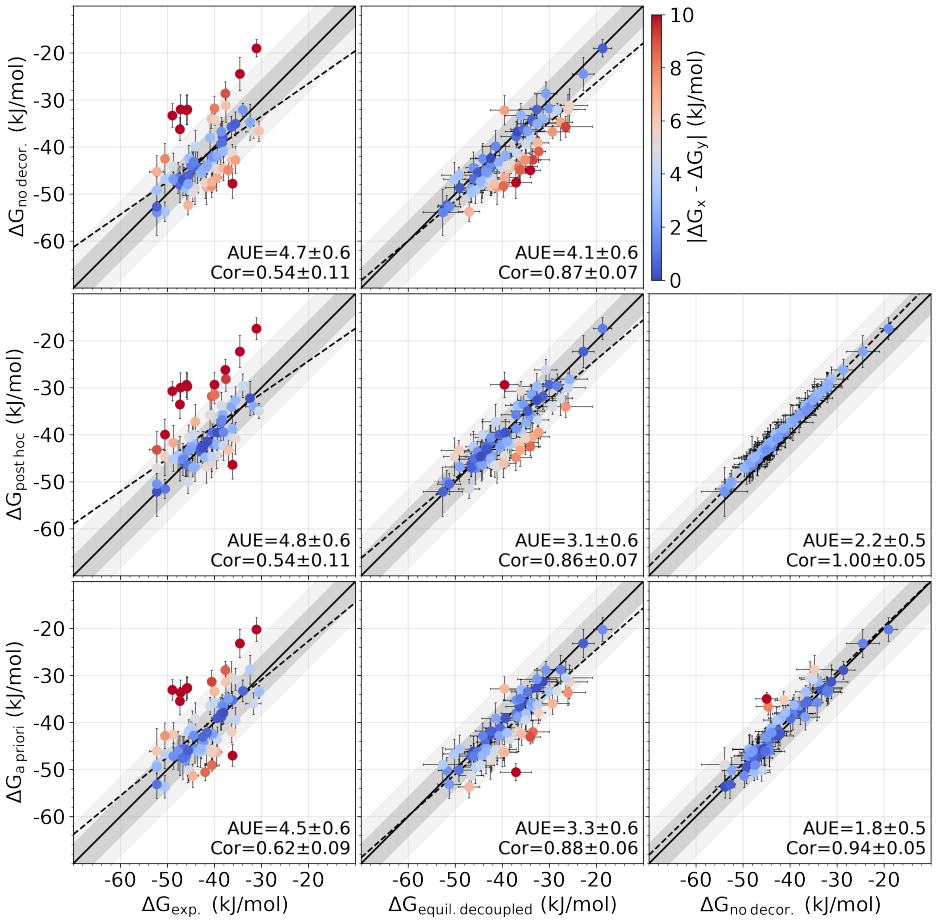

Figure S15: Effects of various decorrelation methods. Explicitly equilibrating the decoupled state provides an exact reference for further comparison. The simplest protocol (top row) involves no corrections for correlation. *Post hoc* decorrelation (middle row) adjusts the coupling work values by the free energy difference between the effects of the correlated and orthogonal restraints. *A priori* decorrelation (bottom row) rotates the ligand in the superpositioned structures to orientations sampled from a distribution that satisfies the orthogonal restraints. Data is shown for the tyk2, p38α and jnk1 systems. Dark and light shaded areas represent regions deviating from experiment by at most 1 and 2 kcal/mol.

protein	ligand	ΔG (kJ/mol) simulation	ΔG (kJ/mol) experimental
tyk2	jmc_30	-29.7 ± 2.7	-45.7 ± 0.7 [11]
tyk2	jmc_28	-29.4 ± 2.7	-45.9 ± 0.7 [11]
tyk2	jmc_27	-30.0 ± 3.1	-47.2 ± 0.7 [11]
tyk2	jmc_23	-30.7 ± 2.1	-48.9 ± 0.7 [11]
tyk2	ejm_55	-33.8 ± 2.8	-38.5 ± 0.7 [10]
tyk2	ejm_54	-37.2 ± 3.7	-44.1 ± 0.7 [10]
tyk2	ejm_50	-28.2 ± 2.6	-37.6 ± 0.7 [10]
tyk2	ejm_49	-32.2 ± 3.6	-32.4 ± 0.7 [10]
tyk2	ejm_48	-26.2 ± 2.2	-37.7 ± 0.7 [10]
tyk2	ejm_47	-31.8 ± 2.7	-40.6 ± 0.7 [10]
tyk2	ejm_46	-33.6 ± 3.0	-47.3 ± 0.7 [10]
tyk2	ejm_45	-29.4 ± 2.5	-40.0 ± 0.7 [10]
tyk2	ejm_44	-17.4 ± 2.3	-31.0 ± 0.7 [10]
tyk2	ejm_43	-22.3 ± 3.5	-34.6 ± 0.7 [10]
tyk2	ejm_42	-35.9 ± 3.6	-40.9 ± 0.7 [10]
tyk2	ejm_31	-31.5 ± 2.0	-39.9 ± 0.7 [10]
jnk1	18660-1	-40.5 ± 2.9	-36.3 ± 0.3 [14]
jnk1	18659-1	-42.0 ± 3.5	-39.4 ± 0.3 [14]
jnk1	18658-1	-39.5 ± 1.8	-40.4 ± 0.4 [14]
jnk1	18652-1	-40.0 ± 2.2	-44.4 ± 0.4 [14]
jnk1	18639-1	-46.0 ± 3.1	-40.5 ± 0.6 [14]
jnk1	18638-1	-44.8 ± 2.8	-42.0 ± 0.4 [14]
jnk1	18637-1	-42.5 ± 2.0	-42.4 ± 1.5 [14]
jnk1	18636-1	-34.0 ± 1.6	-31.7 ± 1.7 [14]
jnk1	18635-1	-35.0 ± 1.6	-30.5 ± 0.4 [14]
jnk1	18634-1	-41.4 ± 2.2	-41.5 ± 0.3 [14]
jnk1	18633-1	-36.9 ± 1.4	-38.4 ± 1.0 [14]
jnk1	18632-1	-42.4 ± 3.1	-38.1 ± 1.5 [14]
jnk1	18631-1	-39.5 ± 2.8	-39.4 ± 1.2 [14]
jnk1	18630-1	-39.5 ± 3.3	-38.1 ± 0.2 [14]
jnk1	18629-1	-38.7 ± 2.9	-36.2 ± 0.4 [14]
jnk1	18628-1	-34.1 ± 2.3	-36.4 ± 0.9 [14]
jnk1	18627-1	-40.8 ± 3.3	-35.5 ± 0.8 [14]
jnk1	18626-1	-43.3 ± 3.5	-37.0 ± 0.6 [14]
jnk1	18625-1	-29.6 ± 2.5	-33.9 ± 0.8 [14]
jnk1	18624-1	-32.7 ± 2.9	-35.5 ± 0.7 [14]
jnk1	17124-1	-44.7 ± 3.2	-40.6 ± 1.6 [14]
p38α	p38a_2aa	-44.4 ± 2.8	-39.5 ± 1.1 [8]
p38α	p38a_3fmk	-51.5 ± 2.0	-50.5 ± 0.9 [8]
p38α	p38a_3fmh	-46.1 ± 2.8	> -45.4 [8]
p38α	p38a_3flz	-41.9 ± 3.2	-40.5 ± 0.9 [8]
p38α	p38a_3fly	-41.7 ± 3.7	-48.7 ± 0.2 [8]
p38α	p38a_3flw	-50.0 ± 2.5	-45.6 ± 0.4 [8]
p38α	p38a_3flq	-42.7 ± 3.4	-42.4 ± 1.0 [8]
p38α	p38a_3fln	-44.6 ± 2.1	-46.4 ± 1.0 [8]
p38α	p38a_2z	-46.8 ± 3.2	-44.1 ± 0.8 [8]
p38α	p38a_2y	-46.8 ± 3.5	-41.6 ± 1.0 [8]
p38α	p38a_2x	-42.6 ± 3.3	-43.6 ± 0.7 [8]
p38α	p38a_2v	-35.8 ± 2.7	-38.2 ± 0.7 [8]
p38α	p38a_2u	-52.1 ± 5.3	-52.2 ± 0.3 [8]
p38α	p38a_2t	-40.0 ± 3.2	-50.5 ± 0.4 [8]
p38α	p38a_2s	-45.1 ± 4.8	-47.3 ± 0.4 [8]
p38α	p38a_2r	-46.4 ± 2.4	-46.0 ± 0.4 [8]
p38α	p38a_2q	-50.5 ± 2.2	-52.2 ± 0.5 [8]
p38α	p38a_2p	-44.4 ± 3.0	-47.7 ± 0.2 [8]
p38α	p38a_2o	-41.7 ± 3.9	-44.7 ± 0.5 [8]
p38α	p38a_2n	-44.7 ± 4.0	-50.5 ± 0.6 [8]

p38α	p38a_2m	-45.2 ± 3.4	-46.7 ± 0.3 [8]
p38α	p38a_2l	-47.1 ± 3.5	-52.2 ± 0.3 [8]
p38α	p38a_2k	-42.6 ± 1.6	-44.9 ± 0.4 [8]
p38α	p38a_2j	-43.6 ± 2.2	-42.9 ± 0.7 [8]
p38α	p38a_2i	-43.1 ± 2.7	-42.8 ± 0.9 [8]
p38α	p38a_2h	-42.0 ± 2.1	-40.4 ± 0.8 [8]
p38α	p38a_2gg	-40.6 ± 2.6	-45.4 ± 0.3 [8]
p38α	p38a_2g	-43.3 ± 2.0	-45.4 ± 1.0 [8]
p38α	p38a_2ff	-45.2 ± 2.6	-48.7 ± 0.4 [8]
p38α	p38a_2f	-46.4 ± 3.1	-36.1 ± 0.9 [8]
p38α	p38a_2ee	-43.2 ± 3.9	-52.2 ± 1.5 [8]
p38α	p38a_2e	-46.5 ± 2.5	-45.6 ± 1.1 [8]
p38α	p38a_2c	-47.0 ± 1.9	-44.6 ± 1.2 [8]
p38α	p38a_2bb	-37.5 ± 3.8	-38.6 ± 0.8 [8]
pde2	43249674	-35.4 ± 4.1	-49.2 ± 0.9 [12]
pde2	50181001	-44.0 ± 4.3	-39.8 ± 0.3 [12]
pde2	50107616	-51.5 ± 5.8	-39.5 ± 1.8 [12]
pde2	49932714	-38.3 ± 4.4	-41.3 ± 0.1 [12]
pde2	49932129	-31.7 ± 2.9	-40.5 ± 1.4 [12]
pde2	49585367	-38.2 ± 4.3	-40.2 ± 1.1 [12]
pde2	49582468	-42.8 ± 5.4	-43.2 ± 0.9 [12]
pde2	49582390	-41.7 ± 4.6	-45.1 ± 1.7 [12]
pde2	49580115	-41.0 ± 2.8	-42.3 ± 1.1 [12]
pde2	49396360	-44.0 ± 3.7	-43.2 ± 0.9 [12]
pde2	49220548	-50.1 ± 3.2	-47.0 ± 1.0 [12]
pde2	49220392	-47.2 ± 6.9	-47.2 ± 1.2 [12]
pde2	49175828	-51.0 ± 5.8	-50.3 ± 2.1 [12]
pde2	49175789	-52.0 ± 5.7	-46.5 ± 1.3 [12]
pde2	49137530	-42.9 ± 5.1	-46.8 ± 1.7 [12]
pde2	49137374	-33.8 ± 3.4	-47.2 ± 1.7 [12]
pde2	49072088	-39.5 ± 6.2	-36.9 ± 0.7 [12]
pde2	48271249	-41.2 ± 5.0	-45.5 ± 0.5 [12]
pde2	48168913	-34.2 ± 3.4	-43.6 ± 1.3 [12]
pde2	48022468	-40.6 ± 5.1	-50.0 ± 1.1 [12]
pde2	48009208	-32.3 ± 3.3	-46.1 ± 1.1 [12]
cdk2	32	-29.9 ± 5.7	-42.1 ± 0.3 [9]
cdk2	31	-34.5 ± 3.5	-41.2 ± 0.3 [9]
cdk2	30	-43.9 ± 2.0	-42.3 ± 0.3 [9]
cdk2	29	-40.8 ± 3.9	-42.8 ± 1.2 [9]
cdk2	28	-36.9 ± 2.1	-47.9 ± 0.1 [9]
cdk2	26	-35.9 ± 2.6	-36.5 ± 0.1 [9]
cdk2	22	-32.6 ± 1.9	-34.0 ± 0.2 [9]
cdk2	21	-31.3 ± 2.4	-33.9 ± 0.3 [9]
cdk2	20	-39.7 ± 3.0	-37.8 ± 0.7 [9]
cdk2	1oiy	-36.5 ± 2.5	-42.6 ± 1.7 [9]
cdk2	1oiu	-40.9 ± 4.3	-39.3 ± 0.5 [9]
cdk2	1oi9	-39.7 ± 2.5	-42.1 ± 0.1 [9]
cdk2	1h1s	-37.0 ± 4.0	-48.6 ± 0.5 [9]
cdk2	1h1r	-34.3 ± 3.1	-29.4 ± 1.7 [9]
cdk2	1h1q	-33.7 ± 2.1	-35.4 ± 0.2 [9]
cdk2	17	-33.0 ± 2.8	-30.9 ± 2.1 [9]
cmet	CHEMBL3402761_1_21	-50.0 ± 5.2	-50.9 ± 0.8 [6]
cmet	CHEMBL3402755_4200_15	-29.3 ± 2.6	-30.4 ± 0.8 [6]
cmet	CHEMBL3402754_40_14	-38.0 ± 3.2	-41.8 ± 0.8 [6]
cmet	CHEMBL3402753_200_13	-32.4 ± 4.0	-37.9 ± 0.8 [6]
cmet	CHEMBL3402752_30000_12	-29.6 ± 3.3	-25.6 ± 0.8 [6]
cmet	CHEMBL3402751_2100_11	-24.5 ± 3.3	-32.1 ± 0.8 [6]
cmet	CHEMBL3402750_400_10	-34.3 ± 2.2	-36.2 ± 0.8 [6]
cmet	CHEMBL3402749_500_9	-34.1 ± 3.6	-35.6 ± 0.8 [6]

cmet	CHEMBL3402748_5300_8	-20.7 ± 3.2	-29.8 ± 0.8 [6]
cmet	CHEMBL3402747_3400_7	-24.8 ± 4.7	-30.9 ± 0.8 [6]
cmet	CHEMBL3402745_200_5	-31.0 ± 3.8	-37.9 ± 0.8 [6]
cmet	CHEMBL3402744_300_4	-30.6 ± 4.3	-36.9 ± 0.8 [6]
galectin	08_ligNH2	-31.2 ± 3.1	-27.8 ± 0.1 [13]
galectin	07_ligOH	-36.1 ± 4.2	-26.0 ± 0.2 [13]
galectin	06_ligPyr	-29.6 ± 3.9	-19.8 ± 2.4 [13]
galectin	05_ligOEt	-28.5 ± 1.5	-22.8 ± 0.3 [13]
galectin	04_ligNMe2	-28.2 ± 2.8	-24.7 ± 0.2 [13]
galectin	03_ligNHMe	-31.5 ± 2.2	-26.6 ± 0.1 [13]
galectin	02_ligOMe	-30.2 ± 3.3	-26.6 ± 0.3 [13]
galectin	01_ligF	-29.7 ± 2.9	-30.7 ± 0.2 [13]

Table S1: Calculated and experimental absolute free energies of binding used for Figure 1 in the main text. Holo crystallographic structures with the ligand removed were used to initialize simulations of the apo state for tyk2 and pde2, while apo crytallographic structures were used for all other proteins. Experimental free energies were obtained from published IC_{50} and K_d values at experimental temperatures. For IC_{50} derived data where no K_m was available (p38α, pde2, and cmet), $K_I \approx IC_{50}$ was used.

protein	pdb id	resolution (Å)	apo or ligand name (if in dataset)	used for modeling initial ligand poses
cdk2	1h1q	2.5	1h1q	yes
cdk2	1h27	2.2	apo	no
jnk1	2gmx	3.5	17124-1	yes
jnk1	3o17	3.0	apo	no
p38α	3fly	3.0	3fly	yes
p38α	1wfc	3.0	apo	no
galectin	5e89	1.5	not in dataset	yes
galectin	3zsl	1.08	apo	no
tyk2	4gih	2.0	ejm_46	yes
cmet	4r1y	2.0	not in dataset	yes
cmet	1r1w	1.8	apo	no
pde2	6ezf	1.5	not in dataset	no
pde2	4htz	2.0	apo	no
pde2	4d08	1.9	49137530	yes
pde2	4d09	2.5	not in dataset	no

Table S2: Summary of the crystallographic structures used.

References

[1] Mark James Abraham, Teemu Murtola, Roland Schulz, Szilárd Páll, Jeremy C. Smith, Berk Hess, and Erik Lindahl. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX*, 1-2:19–25, September 2015. ISSN 2352-7110. doi: 10.1016/j.softx.2015.06.001.

[2] Matteo Aldeghi, Alexander Heifetz, Michael J. Bodkin, Stefan Knapp, and Philip C. Biggin. Predictions of Ligand Selectivity from Absolute Binding Free Energy Calculations. *J. Am. Chem. Soc.*, 139(2):946–957, January 2017. ISSN 0002-7863. doi: 10.1021/jacs.6b11467.

[3] Stefan Boresch, Franz Tettinger, Martin Leitgeb, and Martin Karplus. Absolute Binding Free Energies: A Quantitative Approach for Their Calculation. *J. Phys. Chem. B*, 107(35):9535–9551, September 2003. ISSN 1520-6106. doi: 10.1021/jp0217839.

[4] Rodolfo Briones, Christian Blau, Carsten Kutzner, Bert L. de Groot, and Camilo Aponte-Santamaría. GROmaρs: A GROMACS-Based Toolset to Analyze Density Maps Derived from Molecular Dynamics Simulations. *Biophysical Journal*, 116(1):4–11, January 2019. ISSN 0006-3495. doi: 10.1016/j.bpj.2018.11.3126.

[5] Gavin E. Crooks. Entropy production fluctuation theorem and the nonequilibrium work relation for free energy differences. *Phys. Rev. E*, 60(3):2721–2726, September 1999. ISSN 1063-651X, 1095-3787. doi: 10.1103/PhysRevE.60.2721.

[6] Dieter Dorsch, Oliver Schadt, Frank Stieber, Michael Meyring, Ulrich Grädler, Friedhelm Bladt, Manja Friese-Hamim, Christine Knühl, Ulrich Pehl, and Andree Blaukat. Identification and optimization of pyridazinones as potent and selective c-Met kinase inhibitors. *Bioorganic & Medicinal Chemistry Letters*, 25(7):1597–1602, April 2015. ISSN 0960894X. doi: 10.1016/j.bmcl.2015.02.002.

[7] Vytautas Gapsys, Laura Pérez-Benito, Matteo Aldeghi, Daniel Seeliger, Herman van Vlijmen, Gary Tresadern, and Bert L. de Groot. Large scale relative protein ligand binding affinities using non-equilibrium alchemy. *Chem. Sci.*, 11(4):1140–1152, January 2020. ISSN 2041-6539. doi: 10.1039/C9SC03754C.

[8] David M. Goldstein, Michael Soth, Tobias Gabriel, Nolan Dewdney, Andreas Kuglstatter, Humberto Arzeno, Jeffrey Chen, William Bingenheimer, Stacie A. Dalrymple, James Dunn, Robert Farrell, Sandra Frauchiger, JoAnn La Fargue, Manjiri Ghate, Bradford Graves, Ronald J. Hill, Fujun Li, Renee Litman, Brad Loe, Joel McIntosh, Daniel McWeeney, Eva Papp, Jaehyeon Park, Harlan F. Reese, Richard T. Roberts, David Rotstein, Bong San Pablo, Keshab Sarma, Martin Stahl, Man-Ling Sung, Rebecca T. Suttman, Eric B. Sjogren, Yunchou Tan, Alejandra Trejo, Mary Welch, Paul Weller, Brian R. Wong, and Hasim Zecic. Discovery of 6-(2,4-Difluorophenoxy)-2-[3-hydroxy-1-(2-hydroxyethyl)propylamino]-8-methyl-8 *H* -pyrido[2,3- *d*]pyrimidin-7-one (Pamapimod) and 6-(2,4-Difluorophenoxy)-8-methyl-2-(tetrahydro-2 *H* -pyran-4-ylamino)pyrido[2,3- *d*]pyrimidin-7(8 *H*)-one (R1487) as Orally Bioavailable and Highly Selective Inhibitors of p38α Mitogen-Activated Protein Kinase. *J. Med. Chem.*, 54 (7):2255–2265, April 2011. ISSN 0022-2623, 1520-4804. doi: 10.1021/jm101423y.

[9] Ian R. Hardcastle, Christine E. Arris, Johanne Bentley, F. Thomas Boyle, Yuzhu Chen, Nicola J. Curtin, Jane A. Endicott, Ashleigh E. Gibson, Bernard T. Golding, Roger J. Griffin, Philip Jewsbury, Jerome Menyerol, Veronique Mesguiche, David R. Newell, Martin E. M. Noble, David J. Pratt, Lan-Zhen Wang, and Hayley J. Whitfield. N ^{2} -Substituted *O* ^{6} -Cyclohexylmethylguanine Derivatives: Potent Inhibitors of Cyclin-Dependent Kinases 1 and 2. *J. Med. Chem.*, 47(15):3710–3722, July 2004. ISSN 0022-2623, 1520-4804. doi: 10.1021/jm0311442.

[10] Jun Liang, Vickie Tsui, Anne Van Abbema, Liang Bao, Kathy Barrett, Maureen Beresini, Leo Berezhkovskiy, Wade S. Blair, Christine Chang, James Driscoll, Charles Eigenbrot, Nico Ghilardi, Paul Gibbons, Jason Halladay, Adam Johnson, Pawan Bir Kohli, Yingjie Lai, Marya Liimatta, Priscilla Mantik, Kapil Menghrajani, Jeremy Murray, Amy Sambrone, Yisong Xiao, Steven Shia, Young Shin, Jan Smith, Sue Sohn, Mark Stanley, Mark Ultsch, Birong Zhang, Lawren C. Wu, and Steven Magnuson. Lead identification of novel and selective TYK2 inhibitors. *European Journal of Medicinal Chemistry*, 67:175–187, September 2013. ISSN 02235234. doi: 10.1016/j.ejmech.2013.03.070.

[11] Jun Liang, Anne van Abbema, Mercedesz Balazs, Kathy Barrett, Leo Berezhkovsky, Wade Blair, Christine Chang, Donnie Delarosa, Jason DeVoss, Jim Driscoll, Charles Eigenbrot, Nico Ghilardi, Paul Gibbons, Jason Halladay, Adam Johnson, Pawan Bir Kohli, Yingjie Lai, Yanzhou Liu, Joseph Lyssikatos, Priscilla Mantik, Kapil Menghrajani, Jeremy Murray, Ivan Peng, Amy Sambrone, Steven Shia, Young Shin, Jan Smith, Sue Sohn, Vickie Tsui, Mark Ultsch, Lawren C. Wu, Yisong Xiao, Wenqian Yang, Judy Young, Birong Zhang, Bing-yan Zhu, and Steven Magnuson. Lead Optimization of a 4-Aminopyridine Benzamide Scaffold To Identify Potent, Selective, and Orally Bioavailable TYK2 Inhibitors. *J. Med. Chem.*, 56(11):4521–4536, June 2013. ISSN 0022-2623, 1520-4804. doi: 10.1021/jm400266t.

[12] Laura Pérez-Benito, Henrik Keränen, Herman van Vlijmen, and Gary Tresadern. Predicting Binding Free Energies of PDE2 Inhibitors. The Difficulties of Protein Conformation. *Sci. Rep.*, 8(1):4883, March 2018. ISSN 2045-2322. doi: 10.1038/s41598-018-23039-5.

[13] Kristoffer Peterson. *Molecular Basis for Galectin-Ligand Interactions: Design, Synthesis and Analysis of Ligands*. PhD thesis, Lund University, Faculty of Science, Department of Chemistry, Centre for Analysis and Synthesis, Lund, Sweden, 2018.

[14] Bruce G. Szczepankiewicz, Christi Kosogof, Lissa T. J. Nelson, Gang Liu, Bo Liu, Hongyu Zhao, Michael D. Serby, Zhili Xin, Mei Liu, Rebecca J. Gum, Deanna L. Haasch, Sanyi Wang, Jill E. Clampit, Eric F. Johnson, Thomas H. Lubben, Michael A. Stashko, Edward T. Olejniczak, Chaohong Sun, Sarah A. Dorwin, Kristi Haskins, Cele Abad-Zapatero, Elizabeth H. Fry, Charles W. Hutchins, Hing L. Sham, Cristina M. Rondinone, and James M. Trevillyan. Aminopyridine-Based c-Jun N-Terminal Kinase Inhibitors with Cellular Activity and Minimal Cross-Kinase Activity †. *J. Med. Chem.*, 49(12):3563–3580, June 2006. ISSN 0022-2623, 1520-4804. doi: 10.1021/jm060199b.