

Is structure based drug design ready for selectivity optimization?

Steven K. Albanese^{1,2}, John D. Chodera², Simon Peng³, Robert Abel³, Lingle Wang^{3*}

¹Louis V. Gerstner, Jr. Graduate School of Biomedical Sciences, Memorial Sloan Kettering Cancer Center, New York, NY 10065; ²Computational and Systems Biology Program, Sloan Kettering Institute, Memorial Sloan Kettering Cancer Center, New York, NY 10065; ³Schrödinger, New York, NY 10036

*For correspondence: lingle.wang@schrodinger.com (LW)

Abstract

Alchemical free energy calculations are now widely used to drive or maintain potency in small molecule lead optimization, where the binding affinity to a protein target can be computed—in well-behaved cases—to roughly 1 kcal/mol inaccuracy, which is believed to primarily stem from force field errors. Despite this, the potential to use free energy calculations to drive optimization of compound *selectivity* among two similar targets has been relatively unexplored. In the most optimistic scenario, the similarity of binding sites might lead to a fortuitous cancellation of force field errors and allow selectivity to be predicted more accurately than affinity. Here, we assess the accuracy with which selectivity can be predicted in the context of small molecule kinase inhibitors, considering the very similar binding sites of human kinases CDK2 and CDK9, as well as another series of ligands attempting to achieve selectivity between the more distantly related kinases CDK2 and ERK2. Using a novel Bayesian analysis approach, we separate force field error from statistical error and quantify the correlation in force field errors between selectivity targets. We find that, in the closely related CDK2/CDK9 case, a high correlation in force field errors suggests free energy calculations can have significant impact in aiding chemists in achieving selectivity, while in more distantly related kinases (CDK2/ERK2), limited correlation in force field errors reduces the ability for free energy calculations to aid selectivity optimization. In both cases, the correlation in force field error suggests that longer simulations are beneficial to properly balance statistical error with systematic error to take full advantage of the increase in accuracy in selectivity prediction possible due to fortuitous cancellation of error.

Free energy methods have proven useful in aiding structure-based drug design by driving the optimization or maintenance of potency in lead optimization. Alchemical free energy calculations allow for prediction of ligand binding free energies, including all enthalpic and entropic contributions [1]. Advances in atomistic molecular mechanics forcefields and free energy methodologies [2–5] have allowed free energy methods to reach a level of accuracy sufficient for predicting ligand potencies [6]. Free energy methods have been applied prospectively to develop inhibitors for Tyk2 [7], Syk [8], BACE1 [9], GPCRs [10], and HIV protease [11]. A recent large-scale review found that the use of FEP+ [12] to predicting potency for 92 different projects and 3021 compounds found a median RMSE of 1 kcal/mol [13].

Selectivity is an important consideration in drug design

In addition to maintaining or optimizing potency, free energy methods can be applied to predicting the selectivity of a ligand between two or more targets. Selectivity is an important property to consider in drug development, either in the pursuit of a maximally selective inhibitor [14, 15] or in pursuit a polypharmacological agent [16–20], to avoid on-target toxicity (arising from inhibition of the intended target) [21] and off-target toxicity (arising from inhibition of unintended targets) [22, 23]. In either paradigm, considering the

selectivity of a compound is complicated by the biology of the target. For example, kinases exist as nodes in complex signaling networks [24, 25] with feedback inhibition and cross-talk between pathways. Careful consideration of which off-targets are being inhibited can avoid off-target toxicity due to alleviating feedback inhibition and inadvertently reactivating the targeted pathway [24, 25], or the upregulation of a secondary pathway by alleviation of cross-talk inhibition [26, 27]. Off-target toxicity can also be caused by inhibiting unrelated targets, such as gefitinib, an EGFR inhibitor, inhibiting CYP2D6 [22] and causing hepatotoxicity in lung cancer patients. In a cancer setting, on-target toxicity can be avoided by considering the selectivity for the oncogenic mutant form of the kinase over the wild type form of the kinase [28–30], demonstrated by number of first generation EGFR inhibitors. Selectivity considerations can also lead to beneficial effects: Imatinib, initially developed to target BCR-Abl fusion proteins, is also approved for treating gastrointestinal stromal tumors (GIST) [31] due to its activity against receptor tyrosine kinase KIT.

Use of physical modeling to predict selectivity is relatively unexplored

While predicting selectivity is important for drug discovery, but the utility of free energy methods for predicting this property is relatively unexplored. If there is fortuitous cancellation of errors for closely related systems, free energy methods may be much more accurate than expected given the errors made in predicting the potency for each individual target. The selectivity of Imatinib for Abl kinase over Src [32, 33] and within a family of non-receptor tyrosine kinases [34] has been studied extensively using molecular dynamics and free energy calculations. This work focuses on understanding the role reorganization energy plays in the exquisite selectivity of imatinib for Abl over Src despite high similarity between cocrystallized binding mode and kinase conformations, and does not touch on the evaluation of the accuracy of these methods, or their application to drug discovery on congeneric series of ligands. Previous work predicting the selectivity of three bromodomain inhibitors across the bromodomain family achieved promising accuracy for single target potencies of roughly 1 kcal/mol, but does not explicitly evaluate any selectivity metrics [35] or look at correlation in the errors made for each bromodomain.

Kinases are an interesting and particularly challenging model system for selectivity predictions

Kinases are a useful model system to work with for assessing the utility of free energy calculations to predict selectivity. With the approval of imatinib for the treatment of chronic myelogenous leukemia in 2001, targeted small molecule kinase inhibitors (SMKIs) have become a major class of therapeutics in treating cancer and other diseases. Currently, there are 43 FDA-approved SMKIs [36], and it is estimated that kinase targeted therapies account for as much as 50% of current drug development [37], with many more compounds currently in clinical trials. While there have been a number of successes, the current stable of FDA-approved kinase inhibitors targets only a small number of kinases implicated in disease, and the design of new selective kinase inhibitors remains a significant challenge. Achieving desired selectivity profiles is particularly difficult for kinase targets, making them a system where physical modelling has the potential for a large impact. Achieving selective inhibition of kinases is challenging as there are more than 518 protein kinases [38, 39] with a highly conserved ATP binding site that is targeted by the majority of SMKIs [40]. While kinase inhibitors have been designed to target kinase-specific subpockets and binding modes to achieve selectivity [41–46], previous work has shown that both Type I (binding to the active, DFG-in conformation) and Type II (binding to the inactive, DFG-out conformation) inhibitors display a wide variety of selectivities [47, 48], often exhibiting significant binding a number of other targets in addition to their primary target. Even FDA-approved inhibitors—often the result of extensive drug development programs—bind to a large number of off-target kinases [49]. Kinases are also targets of interest for developing polypharmacological compounds, or inhibitors that are specifically designed to inhibit multiple kinase targets. Resistance to MEK inhibitors in KRAS-mutant lung and colon cancer has been shown to be driven by HER3 upregulation [50], providing rational for dual MEK/ERBB family inhibitors. Similarly, combined MEK and VEGFR1 inhibition has been proposed as a combinatorial approach to treat KRAS-mutant lung cancer [51]. Developing inhibitors with the desired polypharmacology means navigating more complex selectivity profiles. In well-behaved kinase systems, free energy calculations potency predictions have achieved mean unsigned errors of less than 1.0 kcal/mol [7, 12], suggesting that kinases can be computationally tractable as well as clinically interesting.

Assessing the ability of alchemical free energy methods to predict selectivity

We anticipate difficulty in predicting selectivity if the errors in the alchemical free energy calculations for two targets are largely uncorrelated, or even anticorrelated. However, correlation in the forcefield errors of the free energies for the two targets could lead to a fortuitous cancellation of errors in predicting the selectivity between targets, making selectivity predictions *more* accurate than potency predictions. Such correlation could occur because the same chemical elements appear in the ligand and in highly related binding sites. Here, we investigate the magnitude of this correlation (ρ) and the utility of alchemical free energy calculations for the prediction of selectivity, hereafter taken to mean the $\Delta\Delta G$ in binding free energies of the same compound for two targets. We employed state of the art relative free energy calculations [12, 13] to predict the selectivities of two different congeneric ligand series [52, 53], as well as present a simple numerical model to quantify the potential speed up in selectivity optimization expected for different combinations of per target errors and correlation coefficient values. To tease out the effects of a limited number of experimental measurements, we develop a new Bayesian approach to quantify the uncertainty in the correlation coefficient in the predicted change in selectivity on ligand modification, incorporating all sources of uncertainty and correlation in the computation to separate statistical from force field error. We find that in the closely related systems of CDK2 and CDK9, a high correlation of force field errors suggests that free energy methods can have a significant impact on speeding up selectivity optimization. In the more distantly related case (CDK2/ERK2), limited correlation hampers the ability for free energy methods to speed up selectivity optimization.

Methods

Numerical model of selectivity

To model the impact correlation would have on the expected uncertainty for selectivity predictions, $\sigma_{\text{selectivity}}$ was calculated using Equation 1 for 1000 evenly spaced values of the correlation coefficient (ρ) from 0 to 1, for a number of combinations of per target errors (σ_{target1} and σ_{target2})

$$\sigma_{\text{selectivity}} = \sqrt{\sigma_{\text{target1}}^2 + \sigma_{\text{target2}}^2 - 2\rho\sigma_{\text{target1}}\sigma_{\text{target2}}} \quad (1)$$

The speed up in selectivity optimization that could be expected from using free energy calculations of a particular per target error ($\sigma_{\text{selectivity}}$) was quantified as follows using NumPy (v 1.14.2). An original, true distribution for the change in selectivity of 200000000 new compounds proposed with respect to a reference compound was modeled as a normal distribution centered around 0 with a standard deviation of 1 kcal/mol. This assumption was made on the basis that the majority of selectivity is driven by the scaffold, and R group modifications will do little to drive changes in selectivity. The 1 kcal/mol distribution is supported by the standard deviations of the selectivity in the experimental datasets referenced in this work, which are all less than, but close, to 1 kcal/mol.

Each of these proposed compounds were "screened" by a free energy calculation technique with a per target error (σ_{target}) of 1 kcal/mol [12] and a specified correlation coefficient ρ . A $\sigma_{\text{selectivity}}$ was calculated according to Equation 1. The noise of the computational method was modeled as a normal distribution centered around 0 with a standard deviation of $\sigma_{\text{selectivity}}$ and added to the "true" change in selectivity. Any compound predicted to have an improvement in selectivity of 1.4 kcal/mol (1 log unit) would then be made and have its selectivity experimentally measured. The speedup value for each value of ρ is calculated as the proportion of compounds made with a true selectivity gain of 1.4 kcal/mol divided by the proportion of compounds with a 1.4 kcal/mol improvement in the original distribution, where all of the compounds were made.

Finally, this process was repeated for a 100x (2.8 kcal/mol, 2 log unit) selectivity optimization and 50 linearly spaced values of the correlation coefficient (ρ) between 0 and 1, for four values of $\sigma_{\text{selectivity}}$ and 40000000 compounds in the original distribution.

Structure Preparation

Structures from the Shao [52] and Hole [54], and Blake [53] papers were downloaded from the PDB [55], selecting structures with the same co-ligand crystallized. For the Shao dataset, 4BCK (CDK2) and 4BCI (CDK9) were

selected, which have ligand 12c cocrystallized. For the Blake dataset, 5K4J (CDK2) and 5K4I (ERK2) were selected, cocrystallized with ligand 21. The structures were prepared using Schrodinger's Protein Preparation Wizard [56] (release 2017-3). This pipeline modeled in internal loops and missing atoms, added hydrogens at the reported experimental pH (7.0 for the Shao dataset, 7.3 for the Blake dataset) for both the protein and the ligand. All crystal waters were retained. The ligand was assigned protonation and tautomer states using Epik at the experimental pH \pm 2, and hydrogen bonding was optimized using PROPKA at the experimental pH \pm 2. Finally, the entire structure was minimized using OPLS3 with an RMSD cutoff of 0.3Å.

Ligand Pose Generation

Ligands were extracted from the publication entries in the BindingDB as 2D SMILES strings. 3D conformations were generated using LigPrep with OPLS3 [57]. Ionization state was assigned using Epik at experimental pH \pm 2. Stereoisomers were computed by retaining any specified chiralities and varying the rest. The tautomer and ionization state with the lowest epik state penalty was selected for use in the calculation. Ligand poses were generated by first aligning to the co-crystal ligand using the Largest Common Bemis-Murcko scaffold with fuzzy matching (Schrodinger 2017-4). Ligands that were poorly aligned or failed to align were then aligned using Maximum Common Substructure (MCSS). Finally, large R-groups were allowed to sample different conformations using MM-GBSA with a common core restrained. VSGB solvation model was used with the OPLS3 forcefield. No flexible residues were defined for the ligand.

Free Energy Calculations

The FEP+ panel (Maestro release 2017-4) was used to generate perturbation maps. Neutral perturbations were run for 15ns per replica, using an NPT ensemble and water buffer size of 5Å. A GCMC solvation protocol was used to sample buried water molecules in the binding pocket prior to the calculation, which discards any retained crystal waters.

Charge Change Free Energy Calculations

For ligands where a protonation state change was expected to be relevant to binding based on a small state penalty, Jaguar pKa prediction calculations [58] were run to identify protonation state changes with pKas within 1 log unit of the experimental pH. The predicted pKas for one ligand (Shao 12b, 7.84) was within this range. To account for this, a pKa correction was performed. For this ligand, a separate perturbation map containing ligands 12a, 12c, 12b (neutral) and 12b (charged) was run for 30ns per replica using a post-calculation Coulombic charge correction. Each charge change perturbation edge was run with a 150mM NaCl concentration. The pKa correction was performed using Equation 2:

$$\Delta\Delta G_{corrected} = \Delta\Delta G_{uncorrected} - RT \log \left(\frac{10^{pK_a - pH} + 1}{e^{\frac{\Delta G_{neutral} - \Delta G_{charged}}{RT}} * (10^{pK_a - pH} + 1)} \right) \quad (2)$$

$\Delta\Delta G$ for each edge in perturbation map with 12a, 12c and 12b (neutral) was updated using the correction above and merged into the final map.

Statistical Analysis of FEP+ calculations

Each FEP+ calculation has a reported mean unsigned error (MUE) and root mean squared error (RMSE) with a bootstrapped 95% confidence interval. The MUE was calculated according Equation 3, while the RMSE was calculated according to Equation 4.

$$MUE = \frac{\sum_0^n |\Delta G_{calc} - \Delta G_{exp}|}{n} \quad (3)$$

$$RMSE = \frac{\sum_0^n \sqrt{\Delta G_{calc}^2 - \Delta G_{exp}^2}}{n} \quad (4)$$

Each RMSE and MUE is reported with a 95% confidence interval calculated from 10000 replicates of a choose-one-replace bootstrap protocol on the ΔG values reported to account for the finite sample size of the ligands. The code used to bootstrap these values is available on github: <https://github.com/choderalab/selectivity>

Quantification of the correlation coefficient ρ

To quantify ρ , we built a Bayesian graphical model using pymc3 (v. 3.5) [59] and theano (v 1.0.3) [60], which has been made available on Github. For each phase (complex and solvent), the absolute free energy (G) of ligand i was treated as a normal distribution (Equation 5). For each set of calculations, one ligand was chosen as the reference, and pinned to 0, with a standard deviation of 1 kcal/mol in order to improve the efficiency of sampling from the model.

$$G_{i,target}^{phase} = \mathcal{N}(\mu = 0, sd = 25.0 \text{ kcal/mol}) \quad (5)$$

For each edge of the FEP map (ligand $i \rightarrow$ ligand j), there is a contribution from dummy atoms, that was modeled as in Equation 6.

$$c_{i,j} = \mathcal{N}(\mu = 0, sd = 25.0 \text{ kcal/mol}) \quad (6)$$

The model was restrained by including data from the FEP+ calculation.

$$\Delta G_{phase, ij, target}^{BAR} = \mathcal{N}(G_{j,target}^{phase} - G_{i,target}^{phase}, \delta^2 \Delta G_{phase, ij, target}^{BAR}, observed = \Delta G_{phase, ij, target}^{calc}) \quad (7)$$

Where $\delta^2 \Delta G_{phase, ij, target}^{BAR}$ is the reported BAR uncertainty from the calculation, and $\Delta G_{phase, ij, target}^{calc}$ is the BAR estimate of the free energy for the perturbation between ligands i and j in a given phase.

From this, we can calculate the $\Delta \Delta G^{FEP}$ for each edge as in Equation 8:

$$\Delta \Delta G_{target, ij}^{FEP} = \Delta G_{complex, ij, target}^{BAR} - \Delta G_{solvent, ij, target}^{BAR} \quad (8)$$

To model the way an offset is calculated for the ΔG reported by the FEP+ panel in Maestro:

$$offset = \frac{\sum^n G_{i,target}^{complex} - G_{i,target}^{solvent}}{n} - \frac{\sum^n \Delta G_i^{exp}}{n} \quad (9)$$

The offset was added to each ΔG_i^{BAR} to calculate ΔG_i^{sch} .

The experimental binding affinity was treated as a true value ($\Delta G_{i,target}^{true}$) corrupted by experimental uncertainty, which is assumed to be 0.3 kcal/mol [6], with the values reported in the papers ($\Delta G_{i,target}^{obs}$) treated as observations from this distribution (Equation 10)

$$\Delta G_{i,target}^{exp} = \mathcal{N}(\Delta G_{i,target}^{true}, 0.3 \text{ kcal/mol}, observed = \Delta G_{i,target}^{obs}) \quad (10)$$

$\Delta G_{i,target}^{true}$ was assigned a weak normal prior, as in equation 11.

$$\Delta G_{i,target}^{true} = \mathcal{N}(0, 50 \text{ kcal/mol}) \quad (11)$$

The error for a given ligand was calculated as in Equation 12.

$$\epsilon_i = \Delta G_i^{sch} - \Delta G_i^{true} \quad (12)$$

From these ϵ values, we calculated the correlation coefficient, ρ as in Equation 13.

$$\rho = \frac{cov(\epsilon_{target1}, \epsilon_{target2})}{\sigma_{target1} \sigma_{target2}} \quad (13)$$

Where σ is the standard deviation of ϵ . To quantify ρ for the CDK2/ERK2 calculations, the default NUTS sampler with jitter+adapt_diag initialization, 1000 tuning steps, and a target accept probability of 0.8 was used to draw 10000 samples from the model. The CDK2/CDK9 model was sampled 20000 times using default NUTS sampler with jitter+adapt_diag initialization and 3000 tuning steps.

Results

Free energy methods can be used to predict the selectivity of a compound

While ligand potency for a single target is often quantified as a free energy of binding ($\Delta G_{\text{binding}}$), there are a number of different metrics for quantifying the selectivity of a compound [61, 62]. Here, we propose a more granular view of selectivity: the change in free energy of binding for a given ligand between two different targets ($\Delta\Delta G_{\text{selectivity}}$), which can be calculated as in Equation 14. $\Delta\Delta G_{\text{selectivity}}$ is a useful measure of compound selectivity once a single, or small panel, of off-targets have been identified.

$$\Delta\Delta G_{\text{selectivity}} = \Delta G_{\text{binding, target 2}} - \Delta G_{\text{binding, target 1}} \quad (14)$$

To predict the $\Delta\Delta G_{\text{selectivity}}$ of a compound, we developed a protocol that uses a relative free energy calculation (FEP+) [12] to run a map of perturbations between ligands in a congeneric series, as described in depth in the methods section. The calculation is repeated for each target of interest, with identical perturbations (edges) between each ligand (nodes). Each edge represents a relative free energy calculation that quantifies the $\Delta\Delta G$ between the ligands, or nodes. By using provided experimental data, we can convert the $\Delta\Delta G$ from each edge to a single potency value for each value against that target (ΔG_{target}). From this sets of calculations, we can calculate a $\Delta\Delta G_{\text{selectivity}}$ for each ligand given two targets of interest. Previous work shows that FEP+ can achieve an accuracy (σ_{target}) of roughly 1 kcal/mol when predicting potency, which is a combination of systematic forcefield and random statistical error [12]. However, it is possible that the forcefield component of that error may fortuitously cancel when computing $\Delta\Delta G_{\text{selectivity}}$ leading to a selectivity uncertainty ($\sigma_{\text{selectivity}}$) that is lower than would be expected.

Correlation of errors can make selectivity predictions more accurate and speed up ligand optimization

To demonstrate the potential impact correlation has on the uncertainty of selectivity predictions ($\sigma_{\text{selectivity}}$) using alchemical free energy techniques, we created a simple numerical model following equation 1, which takes into account each of the per target errors expected from the methodology as well as the correlation in those errors. As seen in Figure 1A, if the per target errors (σ_1 and σ_2) are the same, $\sigma_{\text{selectivity}}$ approaches 0 as the correlation coefficient (ρ) approaches 1. If the error for the free energy method is not the same, $\sigma_{\text{selectivity}}$ gets smaller but approaches a non-zero value as ρ approaches 1. To quantify the expected speedup in selectivity optimization, we modeled the change in selectivity with respect to a reference compound for a number of compounds a medicinal chemist might suggest as a normal distribution centered around 0 with a standard deviation of 1 kcal/mol (Figure 1B, black curve), reflecting that most proposed changes would not drive large changes in selectivity. Then, suppose that each compound is screened computationally with a method free energy methodology with a per target (σ_{target}) error of 1 kcal/mol, and all compounds predicted to have a 1.4 kcal/mol improvement in selectivity are synthesized and experimentally tested (Figure 1B, colored curves). The fold-change in the proportion of compounds that are made that have a true 1.4 kcal/mol improvement in selectivity compared to the original distribution can be calculated as a surrogate for the expected speedup. For a 1.4 kcal/mol selectivity improvement threshold (1 log unit), a correlation of 0.5 gives an expected speed up of 4.1x, which can be interpreted as 4.1x fewer compounds needing to be made before achieving a 1 log unit improvement in selectivity. This process can be extended for the even more difficult proposition of achieving a 2 log unit improvement in selectivity (Figure 1C), where 200-300x speedups can be expected, depending on σ_{target} for the free energy methodology.

The CDK2 and CDK9 experimental dataset demonstrates the difficulty in achieving selectivity for closely related kinases

To begin quantifying the correlation of errors in free energy predictions for selectivity, we set out to gather datasets that met a number of criteria. We looked for datasets that contained binding affinity data for a number of kinase targets and ligands, as well as having crystal structures for each target with the same co-crystallized ligand. For the CDK2/CDK9 dataset [52], ligand 12c was cocrystallized with CDK2/cylin A (Figure 2A, left) and CDK9/cyclin T (Figure 2B, left), work that was published in a companion paper [54]. In both CDK2 and CDK9, ligand 12c forms relatively few hydrogen bond interactions with the kinase. Each

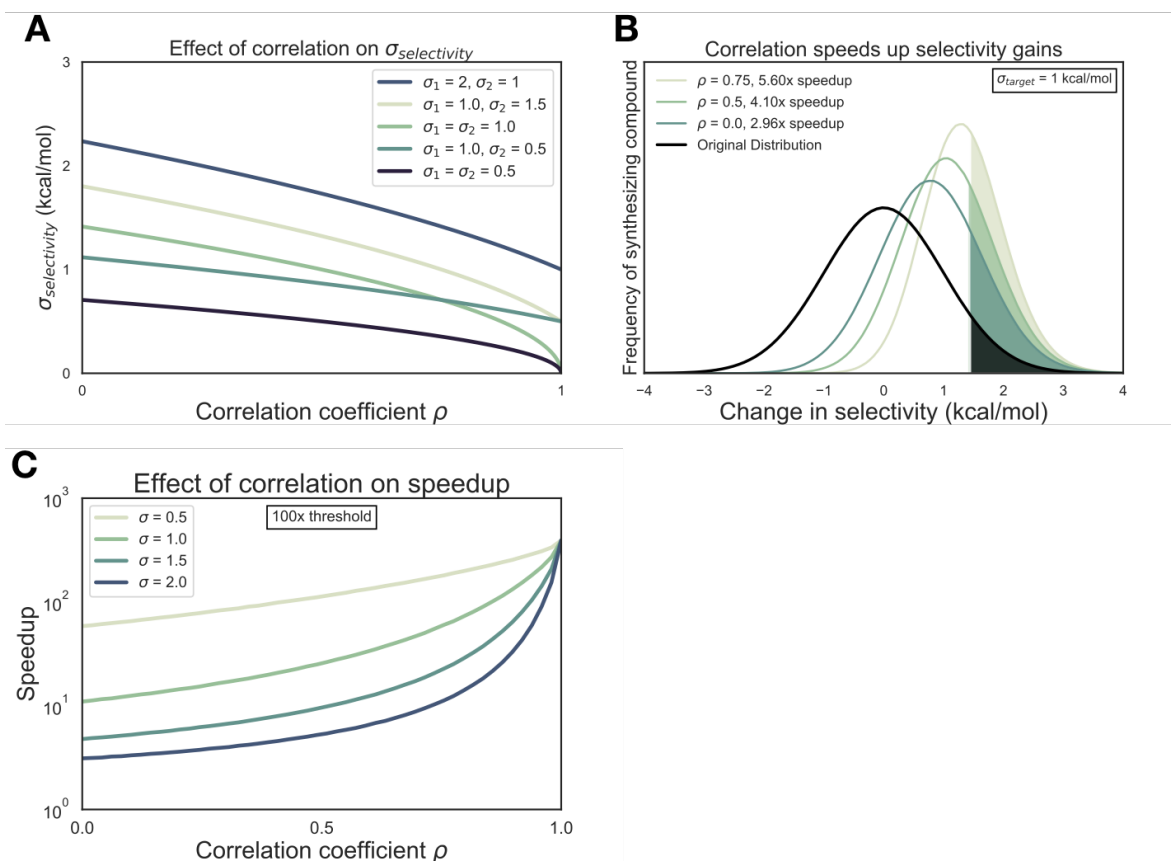


Figure 1. Free energy calculations speed up selectivity optimization

(A) The effect of correlation on expected errors for predicting selectivity ($\sigma_{selectivity}$) in kcal/mol. Each curve represents a different combination of target errors (σ_1 and σ_2). (B) The change in selectivity for molecules proposed by medicinal chemists optimizing a lead candidate can be modeled by a normal distribution centered on 0 with a standard deviation of 1 kcal/mol (black curve). Each green curve corresponds to the distribution of compounds made after screening for a 1 log unit (1.4 kcal/mol) improvement in selectivity with a free energy methodology with a 1 kcal/mol per target error and a particular correlation. The shade region of each curve corresponds to the compounds with a real 1 log unit improvement in selectivity. The speed up is calculated as the ratio of the percentage of compounds made with a real 1 log unit improvement to the percentage of compounds that would be expected in the original distribution. (C) The speedup (y-axis, log scale) expected for 100x (2 log units, 2.8 kcal/mol) selectivity optimization as a function of correlation coefficient ρ . Each curve corresponds to a different σ_{target} value.

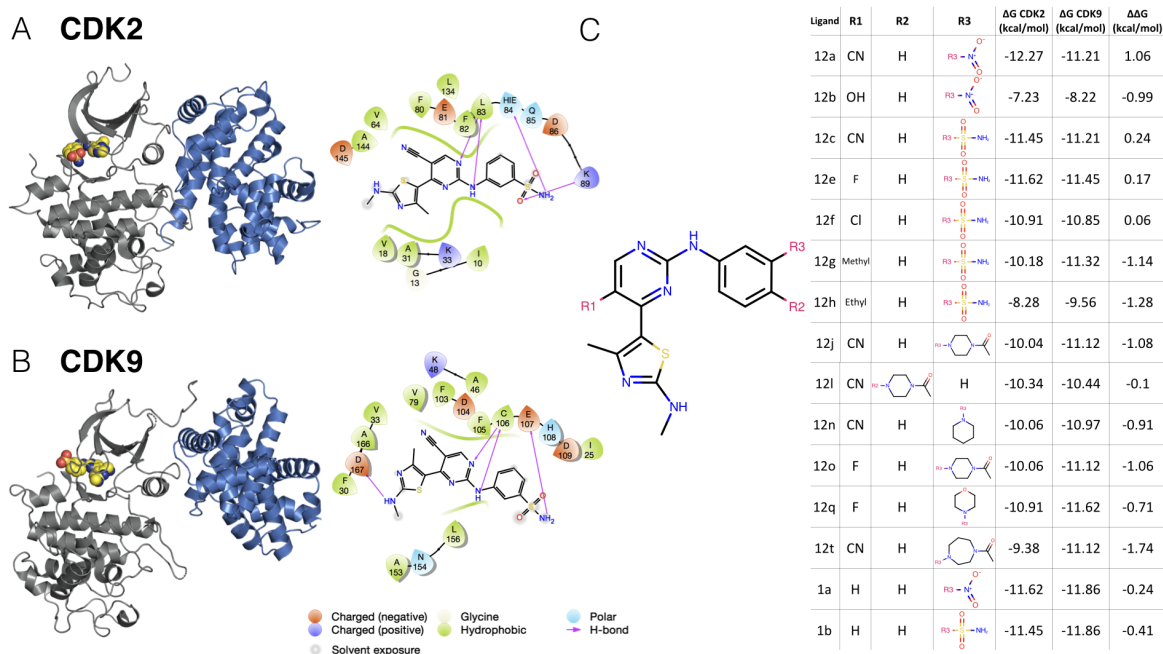


Figure 2. A CDK2/CDK9 selectivity dataset from Shao *et al.*, 2013

(A) (left) Crystal Structure (4BCK)[54] of CDK2 (gray ribbon) bound to ligand 12c (yellow spheres). Cyclin A is shown in blue ribbon (right) 2D ligand interaction map of ligand 12c in the CDK2 binding site. (B) (left) Crystal structure of CDK9 (4BCI)[54] (gray ribbon) bound to ligand 12c (yellow spheres). Cyclin T is shown in blue ribbon. (right) 2D ligand interaction map of ligand 12c in the CDK9 binding site. (C) (left) 2D structure of the common scaffold for all ligands in congeneric ligand series 12 from the publication (right) A table summarizing all R group substitutions as well as the published experimental binding affinities and selectivities[52].

kinase forms a set of hydrogen bonds between the ligand scaffold and a hinge residue (C106 in CDK9 and L83 in CDK2) that is conserved across all of the ligands in this series. CDK9, which has slightly lower affinity for ligand 12c (Figure 2C, right), forms a lone interaction between the sulfonamide of ligand 12c and residue E107. On the other hand, CDK2 forms interactions between the sulfonamide of ligand 12c and residues K89 and H84. The congeneric series of ligands contains a number of challenging perturbations, particularly at substituent point R3 (Figure 2C, left). Ligand 12i also presented a challenging perturbation, moving the 1-(piperazine-1-yl)ethanone from the *meta* to *para* location.

This congeneric series of ligands also highlights two of the challenges of working from publicly available data. First, the dynamic range of selectivity is incredibly narrow, with a mean $\Delta\Delta G_{\text{selectivity}}$ (CDK9 - CDK2) of only -0.65 kcal/mol, and a standard deviation of 0.88 kcal/mol. Additionally, experimental uncertainties are not reported for the experimental measurements. Thus, for this and subsequent sets of ligands, the experimental uncertainty is assumed to be 0.3 kcal/mol based on previous work done to summarize uncertainty in experimental data [6, 63].

The CDK2 and ERK2 dataset achieves higher levels of selectivity for more distantly related kinases. The CDK2/ERK2 dataset from Blake *et al.*, 2016 also met the criteria described above. Crystal structures for both CDK2 (Figure 3A, top) and ERK2 (Figure 3B, top) were available with ligand 22 co-crystallized. Of note, CDK2 was not crystallized with cyclin A, despite cyclin A being included in the affinity assay reported in the paper [53]. CDK2 adopts a DFG-in conformation with the α C helix rotated out, away from the ATP binding site and breaking the conserved salt bridge between K33 and E51 (Supp. Figure 1A), indicative of an inactive kinase [43, 64]. By comparison, the CDK2 structure from the CDK2/CDK9 dataset adopts a DFG-in conformation with the α C helix rotated in, forming the ionic bond between K33 and E51 indicative of an active kinase, due to allosteric activation by cyclin A. While missing cyclins have caused problems for free

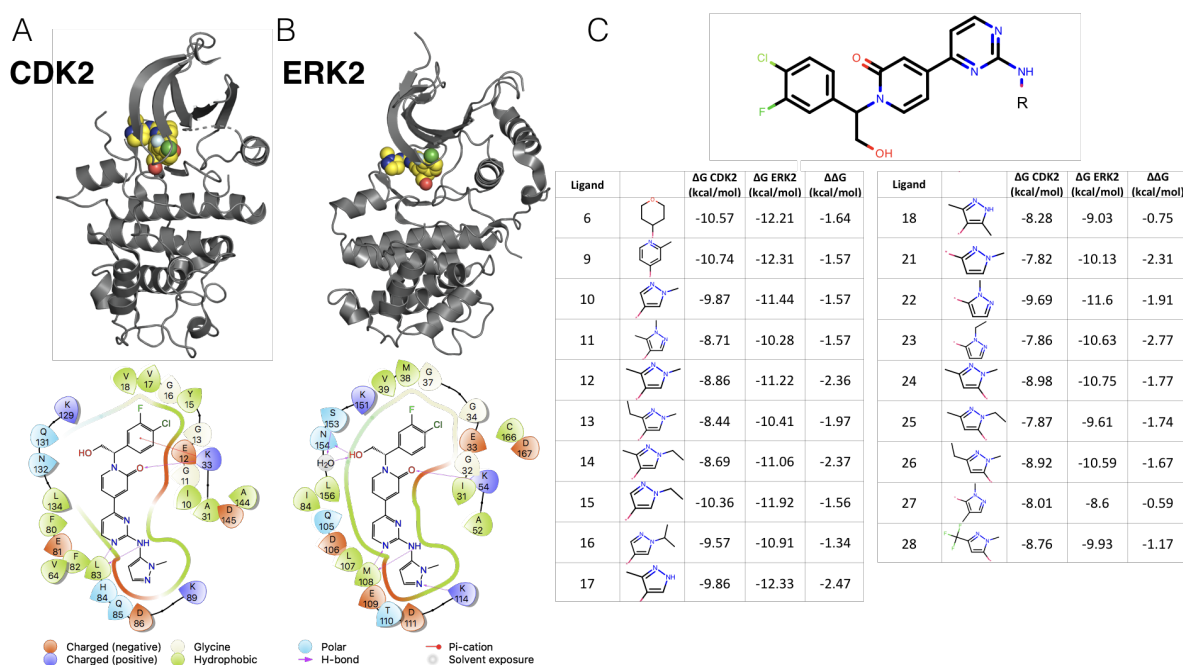


Figure 3. CDK2 and ERK2 selectivity dataset from Blake et al., 2016

(A) (top) Crystal structure of CDK2 (5K4J) shown in gray cartoon and ligand 22 shown in yellow spheres. (bot) 2D interaction map of ligand 22 in the binding pocket of CDK2 (B) (top) Crystal structure of ERK2 (5K4I) shown in gray cartoon with ligand 22 shown in yellow spheres. (bot) 2D interaction map of ligand 22 in the binding pocket of ERK2. (C) (top) Common scaffold for all of the ligands in the Blake dataset, with R denoting attachment side for substitutions. (bot) Table showing R group substitutions and experimentally measured binding affinities and selectivities. Ligand numbers correspond to those used in publication.

energy calculations in prior work, it is possible that the fully active conformation contributes equally to binding affinity for all of the ligands in the series, and the high accuracy of the potency predictions (Figure 4, top left) is the result of fortuitous cancellation of errors. The binding mode for this series is similar between both kinases. There is a set of conserved hydrogens bonds between the scaffold of the ligand and the backbone of one of the hinge residues (L83 for CDK2 and M108 for ERK2). The conserved lysine (K33 for CDK2 and K54 for ERK2), normally involved in the formation of a ionic bond with the α C helix, forms a hydrogen bond with the scaffold (Figure 4A and 4B, bottom) in both CDK2 and ERK2. However, in the ERK2 structure, the hydroxyl engages a crystallographic water as well as N154 in a hydrogen bond network that is not present in the CDK2 structure. The congeneric ligand series features a single subsituent point, with the R groups exposed to the solvent. This helps explain the extremely narrow distribution of selectivities, with a mean selectivity of -1.74 kcal/mol (ERK2 - CDK2) and standard deviation of 0.56 kcal/mol. This suggests that the selectivity is largely driven by the scaffold and unaffected by the R group substitutions.

FEP+ calculations show accurate potency predictions for ERK2/CDK2 and larger errors for CDK2/CDK9. The FEP+ predictions of single target potencies (ΔG) showed good accuracy for the CDK2 and ERK2 dataset (Figure 4, top), with an RMSE of $0.37^{0.57}_{0.16}$ and $0.53^{0.81}_{0.22}$ kcal/mol, respectively. All of the CDK2 and ERK2 potencies were predicted within 1 log unit of the experimental value. The selectivity ($\Delta\Delta G_{\text{selectivity}}$) predictions show an RMSE of $0.81^{1.26}_{0.37}$ kcal/mol, with all of the predictions falling within 1 log unit of the experimental values (Figure 4, top right panel). Despite the high accuracy of the predictions, the narrow dynamic range and high uncertainty from experiment and calculation obscures any signal in the data. The CDK2 and CDK9 datasets show higher errors in the potency predictions, with an RMSE of $1.39^{2.05}_{0.58}$ and $1.71^{2.61}_{0.61}$ kcal/mol respectively. There are a number of outliers that fall outside of 1 log unit from the experimental value. While the higher per target errors make predicting potency more difficult, the selectivity predictions show a much lower RMSE

of $0.74^{1.25}_{0.31}$ kcal/mol. This suggests that some correlation in the error is leading to fortuitous cancellation of systematic error, leading to more accurate than expected predictions of $\Delta\Delta G_{selectivity}^*$.

Correlation of forcefield errors accelerates selectivity optimization

To quantify the correlation coefficient (ρ) of the forcefield errors in our calculations, we built a Bayesian graphical model to separate the forcefield error from the statistical error, as described in the methods section. Briefly, we modeled the absolute free energy (G) of each ligand in each phase (complex and solvent) as in equation 5. The model was chained to the FEP+ calculations by providing the $\Delta G_{phase,ij,target}^{calc}$ as observed data, as in equation 7. As in equation, the experimental data was modeled as a normal distribution centered around the true free energy of binding ($\Delta G_{i,target}^{true}$) corrupted by experimental error, which is assumed to be 0.3 kcal/mol from previous work done to quantify the uncertainty in publicly available data [6]. The reported IC50 values from each dataset were treated as data observations (Equation 10) and the $\Delta G_{i,target}^{true}$ was assigned a weak normal prior (Equation 11). The correlation coefficient was calculated for each sample according to equation 12. The correlation coefficient ρ for the CDK2/ERK2 calculations was quantified to be $0.5^{0.33}_{-0.23}$, indicating that the errors are largely uncorrelated between ERK2 and CDK2 (Figure 5A, right). The joint marginal distribution of the error (ϵ) for each target is symmetric, which is expected for cases in which ρ is 0 (Supp. Figure 2). Despite the weak correlation in errors, the high per target accuracy of these calculations should have a 2-3x speed up for 1 log unit selectivity optimization, and a 20-30x speed up for 2 log unit selectivity optimization (Figure 5A, right). The CDK2/CDK9 calculations show strong evidence of correlation, with a correlation coefficient of $0.70^{0.82}_{0.57}$ (Figure 5B, right). The joint marginal distribution of errors is strongly diagonal, which is expected based on the value for ρ (Figure 5B, left). The high correlation in errors leads to a speed up of 4-5 for 1 log unit selectivity optimization and 30-40x for 2 log unit selectivity optimization (Figure 5B, right), despite the much higher per target errors. Quantifying ρ for these calculations enables estimation of $\sigma_{selectivity}^*$, which is useful for estimating expected error for prospective studies, where the experimental values for $\Delta\Delta G_{selectivity}^*$ are not yet known. Based on the distribution quantified for ρ , the expected $\sigma_{selectivity}^*$ for the CDK2/CDK9 calculations is between 0.76 and 1.16 kcal/mol (Supp. Figure 3), which is in good agreement with the bootstrapped RMSE (Figure 4, bottom). For the CDK2/ERK2 calculations, $\sigma_{selectivity}^*$ is expected to fall between 0.82 kcal/mol and 1.10 kcal/mol (Supp. Figure 3), which is also in good agreement with the bootstrapped RMSE (Figure 4, top).

Discussion and Conclusions

We have demonstrated, using a simple numerical model, the impact that free energy calculations with even weakly correlated errors can have on speeding up the optimization of selectivity in small molecule kinase inhibitors. While the expected speed up is dependent on the per target error of the method (σ_{target}), the speedup is also highly dependent on the correlation of errors made for both targets. Unsurprisingly, free energy methods have greater impact as the threshold for selectivity optimization goes from 10x to 100x. While 100x selectivity optimization is difficult to achieve, the expected benefit from free energy calculations is also quite high, with 1 and 2 order of magnitude speedups possible. To quantify the correlation of errors in two example systems, we gathered experimental data for two congeneric ligand series with experimental data for CDK2 and ERK2, as well as CDK2 and CDK9. These datasets, which had crystal structures for both targets with the same ligand co-crystallized, exemplify the difficulty in predicting selectivity. The dynamic range of selectivity for both systems is incredibly narrow, with most of the perturbations not having a major impact on the overall selectivity achieved. Further, the data was reported with unreliable experimental uncertainties, which makes quantifying the errors made by the free energy calculations difficult. This issue is common when considering selectivity, as many kinase-oriented high throughput screens are carried out at a single concentration and not highly quantitative.

discuss quantification of rho (is there a way to know a priori what rho might be?)

discussion of the expected speedup and sigma based on the quantification of rho

extensions include: separating statistical error (unless this gets added into the results section)

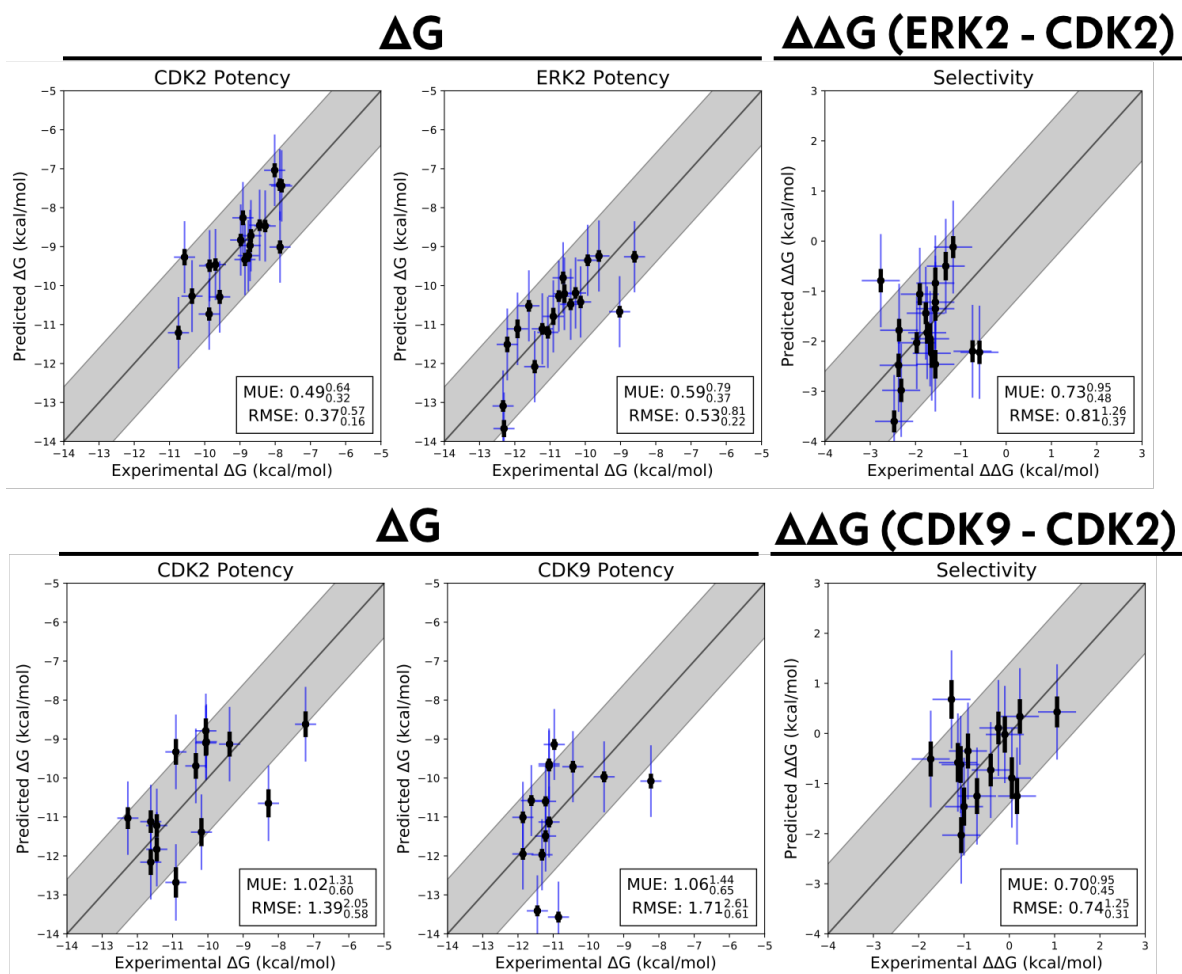


Figure 4. Relative free energy calculations can accurately predict potency, but show larger errors for selectivity predictions.

Single target potencies and selectivities for CDK2/ERK2 from the Blake datasets (*top*), and CDK2/CDK9 (*bottom*) from the Shao datasets. The experimental values are shown on the X-axis and calculated values on the Y-axis. Each data point corresponds to a ligand for a given target. All values are shown in units of kcal/mol. The horizontal error bars show the assumed experimental uncertainty of 0.3 kcal/mol[6]. To better highlight outliers that are unlikely due simply to forcefield errors, we presume the forcefield error ($\sigma_{FF} \approx 0.9$ kcal mol⁻¹ [4]) also behaves as a random error. We show the total estimated statistical and forcefield error ($\sqrt{\sigma_{FF}^2 + \sigma_{calc}^2}$) as vertical blue error bars. The black vertical error bars correspond to the statistical error (σ_{calc}). The black line indicates agreement between calculation and experiment, while the gray shaded region represent 1.36 kcal/mol (or 1 log unit) error. The MUE and RMSE are shown on each plot with bootstrapped 95% confidence intervals.

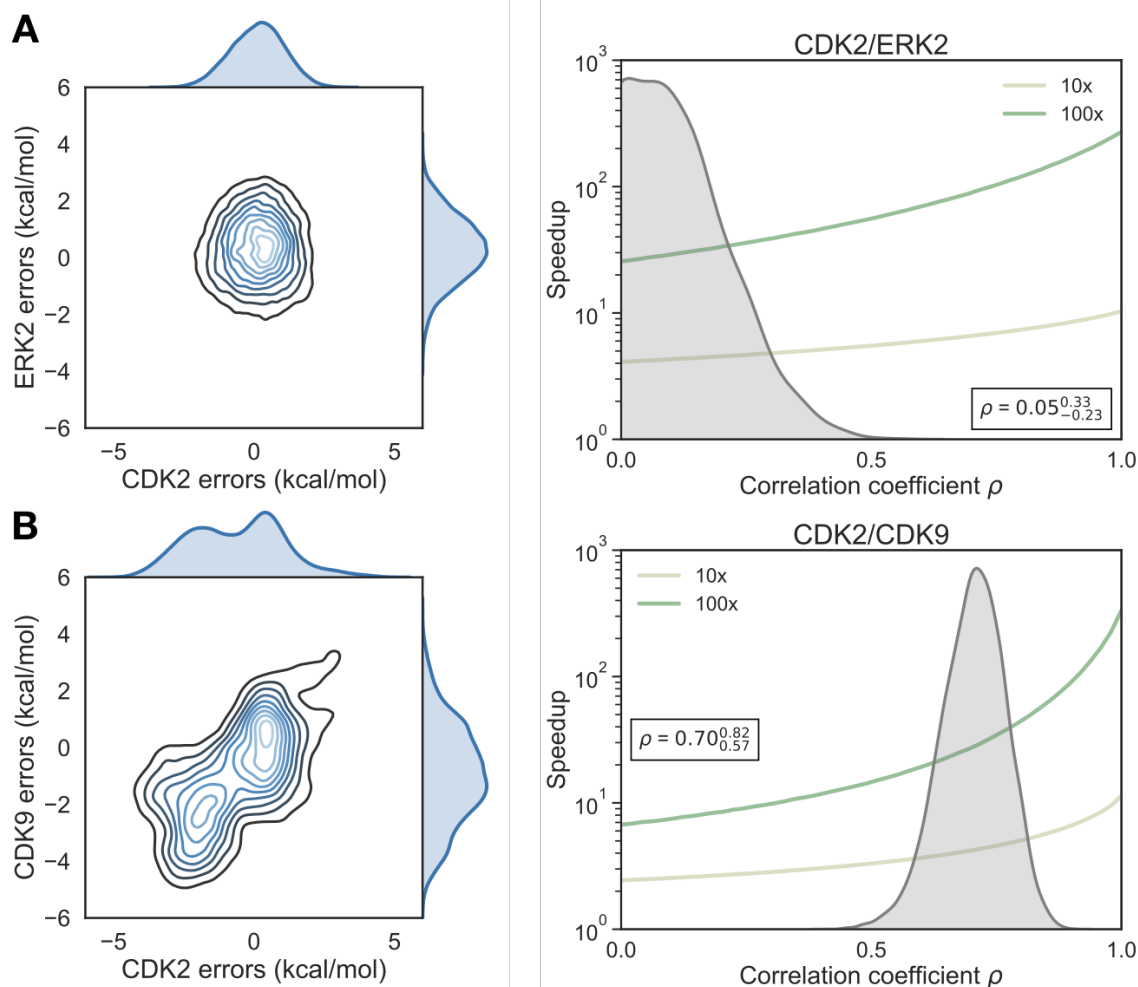


Figure 5. Correlation in selectivity prediction errors can be used to accelerate selectivity optimization

(A) (left) The joint posterior distribution of the prediction errors for CDK2 (X-axis) and ERK2 (Y-axis) from the Bayesian graphical model. (right) Speedup in selectivity optimization (Y-axis) as a function of correlation coefficient (X-axis). The posterior marginal distribution of the correlation coefficient (ρ) is shown in gray, while the expected speed up is shown for 100x (green curve) and 10x (yellow curve) selectivity optimization. The inserted box shows the mean and 95% confidence interval for the correlation coefficient. (B) (left) The same as above, with CDK2 (X-axis) and CDK9 (Y-axis). (right) As above, for the CDK2/CDK9 calculations.

Acknowledgments

JDC and SKA acknowledge support from NIH grant R01 121505. JDC acknowledges partial support from NIH grant P30 CA008748.

To be filled out soon

Disclosures

JDC was a member of the Scientific Advisory Board for Schrödinger, LLC during part of this study. JDC is a current member of the Scientific Advisory Board of OpenEye Scientific Software. The Chodera laboratory receives or has received funding from multiple sources, including the National Institutes of Health, the National Science Foundation, the Parker Institute for Cancer Immunotherapy, Relay Therapeutics, Entasis Therapeutics, Silicon Therapeutics, EMD Serono (Merck KGaA), AstraZeneca, XtalPi, the Molecular Sciences Software Institute, the Starr Cancer Consortium, the Open Force Field Consortium, Cycle for Survival, a Louis V. Gerstner Young Investigator Award, and the Sloan Kettering Institute. A complete funding history for the Chodera lab can be found at <http://choderalab.org/funding>

Author Contributions

Conceptualization: SKA, LW, RA, JDC

Methodology: SKA, LW, JDC

Investigation: SKA, SP

Writing – Original Draft: SKA

Writing – Review & Editing: SKA, JDC

Funding Acquisition: RA, JDC

Resources: LW, JDC

Supervision: LW, JDC

References

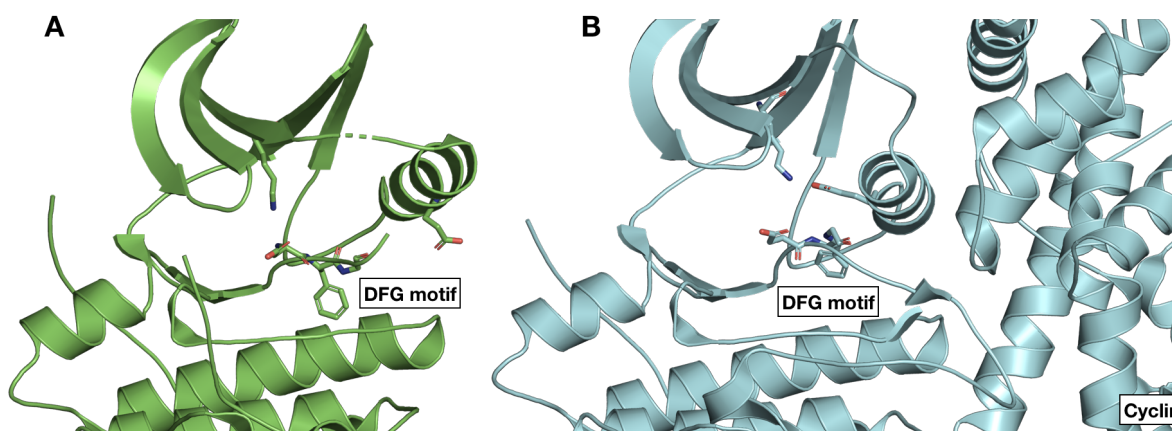
- [1] Chodera JD, Mobley DL, Shirts MR, Dixon RW, Branson K, Pande VS. Alchemical free energy methods for drug discovery: progress and challenges. *Curr Opin Struct Biol.* 2011 Apr; 21(2):150–160.
- [2] Huang J, MacKerell AD. CHARMM36 All-Atom Additive Protein Force Field: Validation Based on Comparison to NMR Data. *J Comput Chem.* 2013 Sep; 34(25):2135–2145. doi: 10.1002/jcc.23354.
- [3] Maier JA, Martinez C, Kasavajhala K, Wickstrom L, Hauser KE, Simmerling C. ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *J Chem Theory Comput.* 2015 Aug; 11(8):3696–3713. doi: 10.1021/acs.jctc.5b00255.
- [4] Harder E, Damm W, Maple J, Wu C, Reboul M, Xiang JY, Wang L, Lupyan D, Dahlgren MK, Knight JL, Kaus JW, Cerutti DS, Krilov G, Jorgensen WL, Abel R, Friesner RA. OPLS3: A Force Field Providing Broad Coverage of Drug-like Small Molecules and Proteins. *J Chem Theory Comput.* 2016 Jan; 12(1):281–296. doi: 10.1021/acs.jctc.5b00864.
- [5] Cournia Z, Allen B, Sherman W. Relative Binding Free Energy Calculations in Drug Discovery: Recent Advances and Practical Considerations. *Journal of chemical information and modeling.* 2017 Dec; 57(12):2911–2937.
- [6] Brown SP, Muchmore SW, Hajduk PJ. Healthy Skepticism: Assessing Realistic Model Performance. *Drug Discov Today.* 2009; 14(7):420 – 427. doi: <http://dx.doi.org/10.1016/j.drudis.2009.01.012>.
- [7] Abel R, Mondal S, Masse C, Greenwood J, Harriman G, Ashwell MA, Bhat S, Wester R, Frye L, Kapeller R, Friesner RA. Accelerating drug discovery through tight integration of expert molecular design and predictive scoring. *Curr Opin Struct Biol.* 2017 Apr; 43(Supplement C):38–44.
- [8] Lovering F, Aevazelis C, Chang J, Dehnhardt C, Fitz L, Han S, Janz K, Lee J, Kaila N, McDonald J, Moore W, Moretto A, Papaioannou N, Richard D, Ryan MS, Wan ZK, Thorarensen A. Imidazotriazines: Spleen Tyrosine Kinase (Syk) Inhibitors Identified by Free-Energy Perturbation (FEP). *ChemMedChem.* 2016 Jan; 11(2):217–233.
- [9] Ciordia M, Pérez-Benito L, Delgado F, Trabanco AA, Tresadern G. Application of Free Energy Perturbation for the Design of BACE1 Inhibitors. *Journal of chemical information and modeling.* 2016 Sep; 56(9):1856–1871.
- [10] Lenselink EB, Louvel J, Forti AF, van Veldhoven JPD, de Vries H, Mulder-Krieger T, McRobb FM, Negri A, Goose J, Abel R, van Vlijmen HWT, Wang L, Harder E, Sherman W, IJzerman AP, Beuming T. Predicting Binding Affinities for GPCR Ligands Using Free-Energy Perturbation. *ACS omega.* 2016 Aug; 1(2):293–304.
- [11] Jorgensen WL. Computer-aided discovery of anti-HIV agents. *Bioorganic & medicinal chemistry.* 2016 Oct; 24(20):4768–4778.
- [12] Wang L, Wu Y, Deng Y, Kim B, Pierce L, Krilov G, Lupyan D, Robinson S, Dahlgren MK, Greenwood J, Romero DL, Masse C, Knight JL, Steinbrecher T, Beuming T, Damm W, Harder E, Sherman W, Brewer M, Wester R, et al. Accurate and Reliable Prediction of Relative Ligand Binding Potency in Prospective Drug Discovery by Way of a Modern Free-Energy Calculation Protocol and Force Field. *J Am Chem Soc.* 2015 Feb; 137(7):2695–2703. doi: 10.1021/ja512751q.
- [13] Abel R, Wang L, Harder ED, Berne BJ, Friesner RA. Advancing Drug Discovery through Enhanced Free Energy Calculations. *Accounts of chemical research.* 2017 Jul; 50(7):1625–1632.
- [14] Zhang J, Yang PL, Gray NS. Targeting cancer with small molecule kinase inhibitors. *Nat Rev Cancer.* 2009 Jan; 9(1):28–39.
- [15] Huggins DJ, Sherman W, Tidor B. Rational approaches to improving selectivity in drug design. *J Med Chem.* 2012 Feb; 55(4):1424–1444.
- [16] Fan QW, Cheng CK, Nicolaides TP, Hackett CS, Knight ZA, Shokat KM, Weiss WA. A dual phosphoinositide-3-kinase alpha/mTOR inhibitor cooperates with blockade of epidermal growth factor receptor in PTEN-mutant glioma. *Cancer Res.* 2007 Sep; 67(17):7960–7965.
- [17] Apsel B, Blair JA, Gonzalez B, Nazif TM, Feldman ME, Aizenstein B, Hoffman R, Williams RL, Shokat KM, Knight ZA. Targeted polypharmacology: discovery of dual inhibitors of tyrosine and phosphoinositide kinases. *Nat Chem Biol.* 2008 Nov; 4(11):691–699.
- [18] Knight ZA, Lin H, Shokat KM. Targeting the Cancer Kinome through Polypharmacology. *Nat Rev Cancer.* 2010; 10(2):130.
- [19] Hopkins AL, Mason JS, Overington JP. Can we rationally design promiscuous drugs? *Curr Opin Struct Biol.* 2006 Feb; 16(1):127–136.

- [20] **Hopkins AL.** Network pharmacology: the next paradigm in drug discovery. *Nat Chem Biol.* 2008 Nov; 4(11):682–690.
- [21] **Rudmann DG.** On-target and off-target-based toxicologic effects. *Toxicol Pathol.* 2013 Feb; 41(2):310–314.
- [22] **Kijima T**, Shimizu T, Nonen S, Furukawa M, Otani Y, Minami T, Takahashi R, Hirata H, Nagatomo I, Takeda Y, Kida H, Goya S, Fujio Y, Azuma J, Tachibana I, Kawase I. Safe and successful treatment with erlotinib after gefitinib-induced hepatotoxicity: difference in metabolism as a possible mechanism. *J Clin Oncol.* 2011 Jul; 29(19):e588–90.
- [23] **Liu S**, Kurzrock R. Toxicity of targeted therapy: Implications for response and impact of genetic polymorphisms. *Cancer Treat Rev.* 2014 Aug; 40(7):883–891.
- [24] **Mendoza MC**, Er EE, Blenis J. The Ras-ERK and PI3K-mTOR pathways: cross-talk and compensation. *Trends Biochem Sci.* 2011 Jun; 36(6):320–328.
- [25] **Tricker EM**, Xu C, Uddin S, Capelletti M, Ercan D, Ogino A, Pratilas CA, Rosen N, Gray NS, Wong KK, Jänne PA. Combined EGFR/MEK Inhibition Prevents the Emergence of Resistance in EGFR-Mutant Lung Cancer. *Cancer Discov.* 2015 Sep; 5(9):960–971.
- [26] **Bailey ST**, Zhou B, Damrauer JS, Krishnan B, Wilson HL, Smith AM, Li M, Yeh JJ, Kim WY. mTOR Inhibition Induces Compensatory, Therapeutically Targetable MEK Activation in Renal Cell Carcinoma. *PLoS One.* 2014 Sep; 9(9):e104413.
- [27] **Chandarlapaty S**, Sawai A, Scaltriti M, Rodrik-Outmezguine V, Grbovic-Huezo O, Serra V, Majumder PK, Baselga J, Rosen N. AKT Inhibition Relieves Feedback Suppression of Receptor Tyrosine Kinase Expression and Activity. *Cancer Cell.* 2011 Jan; 19(1):58–71. doi: 10.1016/j.ccr.2010.10.031.
- [28] **Pao W**, Miller V, Zakowski M, Doherty J, Politi K, Sarkaria I, Singh B, Heelan R, Rusch V, Fulton L, Mardis E, Kupfer D, Wilson R, Kris M, Varmus H. EGF receptor gene mutations are common in lung cancers from “never smokers” and are associated with sensitivity of tumors to gefitinib and erlotinib. *Proceedings of the National Academy of Sciences.* 2004 Sep; 101(36):13306–13311.
- [29] **Kim Y**, Li Z, Apetri M, Luo B, Settleman JE, Anderson KS. Temporal resolution of autophosphorylation for normal and oncogenic forms of EGFR and differential effects of gefitinib. *Biochemistry.* 2012 Jun; 51(25):5212–5222.
- [30] **Juchum M**, Günther M, Laufer SA. Fighting Cancer Drug Resistance: Opportunities and Challenges for Mutation-Specific EGFR Inhibitors. *Drug Resist Updat.* 2015 May; 20:12–28. doi: 10.1016/j.drug.2015.05.002.
- [31] **Din OS**, Woll PJ. Treatment of gastrointestinal stromal tumor: focus on imatinib mesylate. *Ther Clin Risk Manag.* 2008 Feb; 4(1):149–162.
- [32] **Lin YL**, Meng Y, Jiang W, Roux B. Explaining why Gleevec is a specific and potent inhibitor of Abl kinase. *Proc Natl Acad Sci U S A.* 2013 Jan; 110(5):1664–1669.
- [33] **Lin YL**, Meng Y, Huang L, Roux B. Computational Study of Gleevec and G6G Reveals Molecular Determinants of Kinase Inhibitor Selectivity. *J Am Chem Soc.* 2014 Oct; 136(42):14753–14762.
- [34] **Lin YL**, Roux B. Computational Analysis of the Binding Specificity of Gleevec to Abl, c-Kit, Lck, and c-Src Tyrosine Kinases. *J Am Chem Soc.* 2013 Oct; 135(39):14741–14753.
- [35] **Aldeghi M**, Heifetz A, Bodkin MJ, Knapp S, Biggin PC. Predictions of Ligand Selectivity from Absolute Binding Free Energy Calculations. *J Am Chem Soc.* 2017 Jan; 139(2):946–957.
- [36] **Robert Roskoski Jr.** USFDA Approved Protein Kinase Inhibitors. . 2017; <http://www.brimr.org/PKI/PKIs.htm>, updated 3 May 2017.
- [37] **Santos R**, Ursu O, Gaulton A, Bento AP, Donadi RS, Bologa CG, Karlsson A, Al-Lazikani B, Hersey A, Oprea TI, Overington JP. A Comprehensive Map of Molecular Drug Targets. *Nat Rev Drug Discov.* 2016 Dec; 16(1):19–34. doi: 10.1038/nrd.2016.230.
- [38] **Volkamer A**, Eid S, Turk S, Jaeger S, Rippmann F, Fulle S. Pocketome of human kinases: prioritizing the ATP binding sites of (yet) untapped protein kinases for drug discovery. *J Chem Inf Model.* 2015 Mar; 55(3):538–549.
- [39] **Manning G**, Whyte DB, Martinez R, Hunter T, Sudarsanam S. The Protein Kinase Complement of the Human Genome. *Science.* 2002 Dec; 298(5600):1912–1934.
- [40] **Wu P**, Nielsen TE, Clausen MH. FDA-approved small-molecule kinase inhibitors. *Trends Pharmacol Sci.* 2015 Jul; 36(7):422–439.

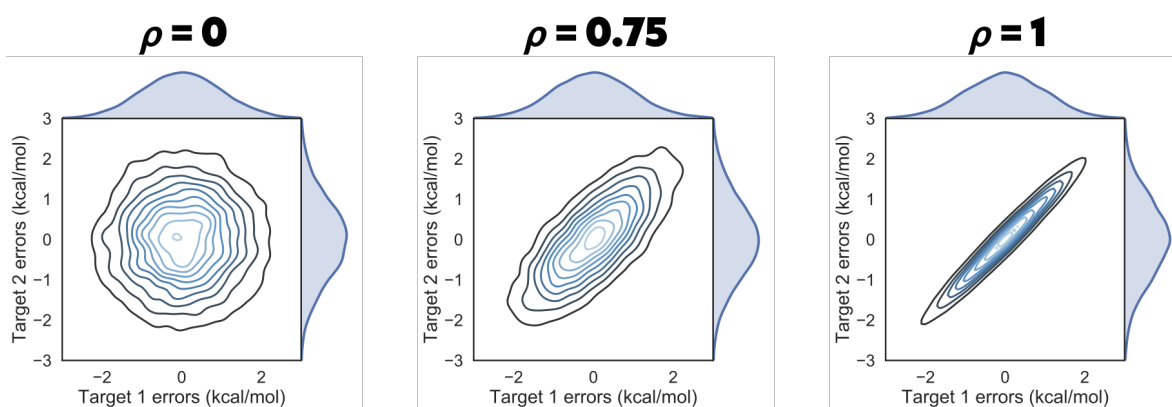
- [41] **Cowan-Jacob SW**, Fendrich G, Floersheimer A, Furet P, Liebetanz J, Rummel G, Rheinberger P, Centeleghe M, Fabbro D, Manley PW, IUCr. Structural biology contributions to the discovery of drugs to treat chronic myelogenous leukaemia. *Acta Crystallogr D Biol Crystallogr*. 2007 Jan; 63(1):80–93.
- [42] **Seeliger MA**, Nagar B, Frank F, Cao X, Henderson MN, Kuriyan J. c-Src Binds to the Cancer Drug Imatinib with an Inactive Abl/c-Kit Conformation and a Distributed Thermodynamic Penalty. *Structure*. 2007 Mar; 15(3):299–311.
- [43] **Huse M**, Kuriyan J. The conformational plasticity of protein kinases. *Cell*. 2002 Jan; 109(3):275–282.
- [44] **Harrison SC**. Variation on an Src-like theme. *Cell*. 2003 Mar; 112(6):737–740.
- [45] **Volkamer A**, Eid S, Turk S, Rippmann F, Fulle S. Identification and Visualization of Kinase-Specific Subpockets. *J Chem Inf Model*. 2016 Feb; 56(2):335–346.
- [46] **Christmann-Franck S**, van Westen GJP, Papadatos G, Beltran Escudie F, Roberts A, Overington JP, Domine D. Unprecedentedly Large-Scale Kinase Inhibitor Set Enabling the Accurate Prediction of Compound–Kinase Activities: A Way toward Selective Promiscuity by Design? *Journal of chemical information and modeling*. 2016 Sep; 56(9):1654–1675.
- [47] **Anastassiadis T**, Deacon SW, Devarajan K, Ma H, Peterson JR. Comprehensive assay of kinase catalytic activity reveals features of kinase inhibitor selectivity. *Nat Biotechnol*. 2011 Nov; 29(11):1039–1045.
- [48] **Davis MI**, Hunt JP, Herrgard S, Ciceri P, Wodicka LM, Pallares G, Hocker M, Treiber DK, Zarrinkar PP. Comprehensive Analysis of Kinase Inhibitor Selectivity. *Nat Biotechnol*. 2011 Oct; 29(11):1046–1051. doi: [10.1038/nbt.1990](https://doi.org/10.1038/nbt.1990).
- [49] **Klaeger S**, Heinzlmeier S, Wilhelm M, Polzer H, Vick B, Koenig PA, Reinecke M, Ruprecht B, Petzoldt S, Meng C, Zecha J, Reiter K, Qiao H, Helm D, Koch H, Schoof M, Canevari G, Casale E, Depaolini SR, Feuchtinger A, et al. The target landscape of clinical kinase drugs. *Science*. 2017 Dec; 358(6367).
- [50] **Sun C**, Hobor S, Bertotti A, Zecchin D, Huang S, Galimi F, Cottino F, Prahallad A, Grenrum W, Tzani A, Schlicker A, Wessels LFA, Smit EF, Thunnissen E, Halonen P, Lieftink C, Beijersbergen RL, Di Nicolantonio F, Bardelli A, Trusolino L, et al. Intrinsic resistance to MEK inhibition in KRAS mutant lung and colon cancer through transcriptional induction of ERBB3. *CellReports*. 2014 Apr; 7(1):86–93.
- [51] **Manchado E**, Weissmueller S, Morris JP, Chen CC, Wullenkord R, Lujambio A, de Stanchina E, Poirier JT, Gainor JF, Corcoran RB, Engelman JA, Rudin CM, Rosen N, Lowe SW. A combinatorial strategy for treating KRAS-mutant lung cancer. *Nature*. 2016 Jun; 534(7609):647–651.
- [52] **Shao H**, Shi S, Huang S, Hole AJ, Abbas AY, Baumli S, Liu X, Lam F, Foley DW, Fischer PM, Noble M, Endicott JA, Pepper C, Wang S. Substituted 4-(Thiazol-5-yl)-2-(phenylamino)pyrimidines Are Highly Active CDK9 Inhibitors: Synthesis, X-ray Crystal Structures, Structure–Activity Relationship, and Anticancer Activities. *J Med Chem*. 2013 Feb; 56(3):640–659.
- [53] **Blake JF**, Burkard M, Chan J, Chen H, Chou KJ, Diaz D, Dudley DA, Gaudino JJ, Gould SE, Grina J, Hunsaker T, Liu L, Martinson M, Moreno D, Mueller L, Orr C, Pacheco P, Qin A, Rasor K, Ren L, et al. Discovery of (S)-1-(1-(4-Chloro-3-fluorophenyl)-2-hydroxyethyl)-4-(2-((1-methyl-1H-pyrazol-5-yl)amino)pyrimidin-4-yl)pyridin-2(1H)-one (GDC-0994), an Extracellular Signal-Regulated Kinase 1/2 (ERK1/2) Inhibitor in Early Clinical Development. *J Med Chem*. 2016 Jun; 59(12):5650–5660.
- [54] **Hole AJ**, Baumli S, Shao H, Shi S, Huang S, Pepper C, Fischer PM, Wang S, Endicott JA, Noble ME. Comparative Structural and Functional Studies of 4-(Thiazol-5-yl)-2-(phenylamino)pyrimidine-5-carbonitrile CDK9 Inhibitors Suggest the Basis for Isotype Selectivity. *J Med Chem*. 2013 Feb; 56(3):660–670.
- [55] **Berman HM**, Battistuz T, Bhat TN, Bluhm WF, Bourne PE, Burkhardt K, Feng Z, Gilliland GL, Iype L, Jain S, Fagan P, Marvin J, Padilla D, Ravichandran V, Schneider B, Thanki N, Weissig H, Westbrook JD, Zardecki C. The Protein Data Bank. *Acta Crystallogr D Biol Crystallogr*. 2002 Jun; 58(Pt 6):899–907.
- [56] **Sastry GM**, Adzhigirey M, Day T, Annabhimoju R, Sherman W. Protein and ligand preparation: parameters, protocols, and influence on virtual screening enrichments. *J Comput Aided Mol Des*. 2013 Mar; 27(3):221–234.
- [57] **Harder E**, Damm W, Maple J, Wu C, Reboul M, Xiang JY, Wang L, Lupyan D, Dahlgren MK, Knight JL, Kaus JW, Cerutti DS, Krilov G, Jorgensen WL, Abel R, Friesner RA. OPLS3: A Force Field Providing Broad Coverage of Drug-like Small Molecules and Proteins. *J Chem Theory Comput*. 2016 Jan; 12(1):281–296.
- [58] **Bochevarov AD**, Harder E, Hughes TF, Greenwood JR, Braden DA, Philipp DM, Rinaldo D, Halls MD, Zhang J, Friesner RA. Jaguar: a high-performance quantum chemistry software program with strengths in life and materials sciences. *Int J Quantum Chem*. 2013; 113(18):2110–2142.

- 504 [59] **Salvatier J**, Wiecki TV, Fonnesbeck C. Probabilistic programming in Python using PyMC3. *PeerJ Computer Science*.
505 2016; 2:e55.
- 506 [60] **Al-Rfou R**, Alain G, Almahairi A, Angermueller C, Bahdanau D, Ballas N, Bastien F, Bayer J, Belikov A, Belopolsky A,
507 Bengio Y, Bergeron A, Bergstra J, Bisson V, Blecher Snyder J, Bouchard N, Boulanger-Lewandowski N, Bouthillier X,
508 de Brébisson A, Breuleux O, et al. Theano: A Python framework for fast computation of mathematical expressions.
509 arXiv e-prints. 2016 May; abs/1605.02688. <http://arxiv.org/abs/1605.02688>.
- 510 [61] **Bosc N**, Meyer C, Bonnet P. The use of novel selectivity metrics in kinase research. *BMC bioinformatics*. 2017 Jan;
511 18(1):17.
- 512 [62] **Cheng AC**, Eksterowicz J, Geuns-Meyer S, Sun Y. Analysis of kinase inhibitor selectivity using a thermodynamics-based
513 partition index. *J Med Chem*. 2010 Jun; 53(11):4502–4510.
- 514 [63] **Hauser K**, Negron C, Albanese SK, Ray S, Steinbrecher T, Abel R, Chodera JD, Wang L. Predicting resistance of clinical
515 Abl mutations to targeted kinase inhibitors using alchemical free-energy calculations. *Communications Biology*. 2018
516 Jun; 1(1):70.
- 517 [64] **Hari SB**, Merritt EA, Maly DJ. Sequence determinants of a specific inactive protein kinase conformation. *Chemistry &*
518 *biology*. 2013 Jun; 20(6):806–815.
- 519 [65] **Hu J**, Ahuja LG, Meharena HS, Kannan N, Kornev AP, Taylor SS, Shaw AS. Kinase regulation by hydrophobic spine
520 assembly in cancer. *Molecular and cellular biology*. 2015 Jan; 35(1):264–276.

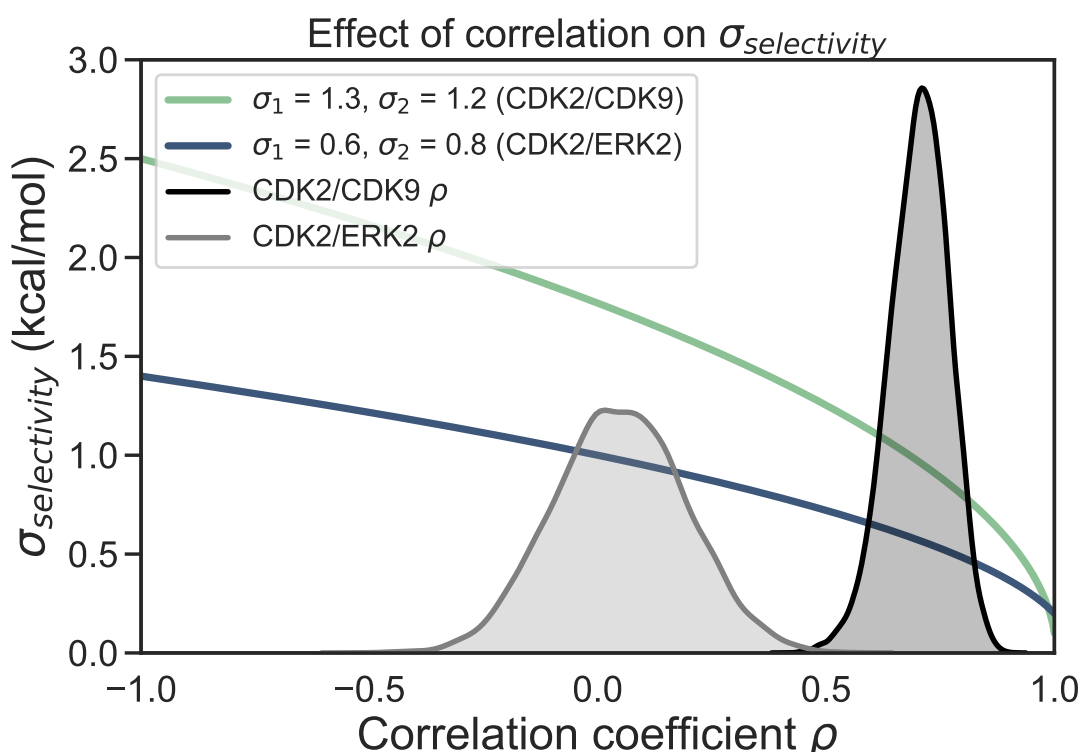
521 Supplemental Information

**Supplemental Figure. 1. CDK2 adopts an inactive conformation in the crystal structure used for the CDK2/ERK2 calculations**

(A) CDK2 (5K4J) adopts an inactive conformation in the absence of its cyclin. The DFG motif is in a DFG-in conformation, with the α C helix rotated outwards, breaking the salt bridge between K33 and E51 (Uniprot numbering) that is typically a marker of an active conformation. Notably, the Phe in the DFG motif does not completely form the hydrophobic spine due to the rotation of the α C helix [65] (B) The CDK2 structure used for the CDK2/CDK9 calculations (4BCK) contains cyclin A and adopts a DFG-in/ α C helix-in conformation that forms the salt bridge between K33 and E51. This is typically indicative of a fully active kinase [43, 64].

**Supplemental Figure. 2. Correlation coefficient ρ controls the shape of the joint marginal distribution of errors**

As ρ increases, the joint marginal distribution of errors become more diagonal. Each panel shows 10000 samples drawn from a multivariate normal distribution centered around 0 kcal/mol, where the per target error was set to 1 kcal/mol and ρ to the value indicated in bold over the plot.



Supplemental Figure. 3. Correlation reduces the expected error for selectivity predictions

As correlation coefficient ρ increases, $\sigma_{selectivity}$ decreases. The intersection between CDK2/CDK9 $\sigma_{selectivity}$ (green curve) and ρ (black distribution) indicates the range of expected $\sigma_{selectivity}$ values. The intersection for CDK2/ERK $\sigma_{selectivity}$ (blue curve) and ρ (gray distribution) suggests the expected $\sigma_{selectivity}$ range for that set of calculations.