

PREDICTING SELECTIVITY AND THE  
FUNCTIONAL IMPACT OF CLINICAL CANCER  
MUTATIONS FOR SMALL MOLECULE KINASE  
INHIBITORS USING PHYSICAL MODELING

by

Steven K. Albanese

A Dissertation

Presented to the Faculty of the Louis V. Gerstner Jr.,  
Graduate School of Biomedical Sciences,  
Memorial Sloan Kettering Cancer Center  
in Partial Fulfillment of the Requirements for the Degree of  
Doctor of Philosophy

New York, NY

February 2019

---

John D. Chodera, PhD  
Dissertation Mentor

---

Date

© 2019 Steven K. Albanese

ALL RIGHTS RESERVED

## ABSTRACT

Small molecule kinase inhibitors have become a major focus of drug development for treating cancer, which accounted for 610,000 deaths and 1.7 million diagnoses in the United States in 2018 alone. Currently, there are 44 FDA approved small-molecule kinase inhibitors. The dominant paradigm for designing such inhibitors has been to optimize maximally selective ligands for a single target. Unfortunately, many such inhibitors fail in clinical trials due to a lack of efficacy and clinical safety. Tumors can evade inhibitors through multiple routes of resistance, including upregulation of a second kinase, mutations in the target kinase, or amplification of the target kinase. On the other hand, toxicity arises from on-target inhibition of the wild type kinase or off-target effects of promiscuous small molecules or their metabolites. ATP-competitive kinase inhibitors have great potential for promiscuity, as there are over 520 kinases in the human kinome that each bind a common substrate, ATP. Further, advances in sequencing technology have enabled the generation of datasets of disease associated alterations rich in missense mutations in kinases. While this technology has been particularly transformative in the field of oncology, where many patients are treated with kinase targeted therapies, most kinase missense mutations are rare, making it difficult to assess their functional impact. Physical modeling can provide a route for predicting small molecule kinase inhibitor selectivity, and the impact that missense mutations have on kinase structure and inhibitor binding. To assess the utility of free energy calculations for predicting selectivity, we performed relative free energy calculations on publicly available congeneric series of ligands on multiple kinase targets. We built a Bayesian graphical model to quantitate the correlation of errors for a given ligand on both target, to interrogate whether any fortuitous

cancellation of errors makes selectivity predictions more accurate than expected. To understand the functional impact of mutations on protein kinase structure and inhibitor binding, we performed massively parallel molecular dynamics simulations of clinically observed hyperactivating mTOR mutations and alchemical free energy calculations on clinical Abl mutations. To rigorously test our predictions, we have developed a panel of kinase expression constructs (now available through AddGene) appropriate for automated, high-throughput expression protocol in *E. coli*. We have demonstrated the utility of these constructs for engineering and expressing clinically observed missense mutations, testing a panel of 96 mutations in Src and Abl kinases gathered from publicly available cancer genomics datasets as well as the MSK-IMPACT clinical sequencing panel, and further expressed a separate panel of 95 clinically relevant Abl mutations. We measured the binding free energies for these clinically relevant Abl mutations for a panel of FDA-approved small molecule kinase inhibitors. Using this dataset as a benchmark, we tested the sensitivity and accuracy of absolute free energy calculations to predict the impact of mutations on inhibitor binding. Taken together, this work provides an assessment of physical modeling for predicting selectivity and resistance in drug design of small molecule kinase inhibitors.

## **BIOGRAPHICAL SKETCH**

Your biosketch goes here. Make sure it sits inside the brackets.

This document is dedicated to my father, whose unwavering support and love  
inspired me and kept me going. He is terribly missed.

## **ACKNOWLEDGEMENTS**

Your acknowledgements go here. Make sure it sits inside the brackets.

# Table of Contents

Biographical Sketch . . . . .	v
Dedication . . . . .	vi
Acknowledgements . . . . .	vii
Table of Contents . . . . .	viii
List of Tables . . . . .	xi
List of Figures . . . . .	xii
<b>1 Introduction</b>	<b>1</b>
<b>2 Predicting the selectivity of small molecule kinase inhibitors</b>	<b>2</b>
2.1 Introduction . . . . .	2
2.1.1 Free energy methods can aid structure-based drug design . . . . .	2
2.1.2 Selectivity is an important consideration in drug design . . . . .	2
2.1.3 Use of physical modeling to predict selectivity is relatively unexplored . . . . .	3
2.1.4 Kinases are an interesting and particularly challenging model system for selectivity predictions . . . . .	4
2.1.5 Assessing the ability of alchemical free energy methods to predict selectivity . . . . .	5
2.2 Methods . . . . .	7
2.2.1 Numerical model of selectivity . . . . .	7
2.2.2 Structure Preparation . . . . .	8
2.2.3 Ligand Pose Generation . . . . .	9
2.2.4 Free Energy Calculations . . . . .	9
2.2.5 Charge Change Free Energy Calculations . . . . .	10
2.2.6 Statistical Analysis of FEP+ calculations . . . . .	10
2.2.7 Quantification of the correlation coefficient $\rho$ . . . . .	11
2.3 Results . . . . .	13
2.3.1 Free energy methods can be used to predict the selectivity of a compound . . . . .	13
2.3.2 Correlation of errors can make selectivity predictions more accurate and speed up ligand optimization . . . . .	14
2.3.3 The CDK2 and CDK9 experimental dataset demonstrates the difficulty in achieving selectivity for closely related kinases . . . . .	17
2.3.4 The CDK2 and ERK2 dataset achieves higher levels of selectivity for more distantly related kinases . . . . .	20

2.3.5	FEP+ calculations show accurate potency predictions for ERK2/CDK2 and larger errors for CDK2/CDK9 . . . . .	23
2.3.6	Free energy calculation errors are correlated, accelerating selectivity optimization . . . . .	25
2.4	Discussion and Conclusions . . . . .	28
<b>3</b>	<b>Understanding the functional impact of mTOR clinical kinase mutations using physical modeling</b>	<b>33</b>
3.1	Introduction . . . . .	33
3.1.1	mTOR forms the catalytic core of protein complexes that control a number of cellular processes . . . . .	33
3.1.2	mTOR is the targeted by inhibitors with two distinct mechanisms of action . . . . .	35
3.1.3	mTOR signaling is dysregulated in cancer by hyperactivating missense mutations . . . . .	39
3.1.4	Using physical modeling to understand the functional impact of mTOR mutations . . . . .	41
3.2	Methods . . . . .	42
3.2.1	Molecular dynamics simulations . . . . .	42
3.2.2	Contact Map Analysis . . . . .	44
3.2.3	Mean Contact formation over time . . . . .	44
3.2.4	Alchemical free energy calculations . . . . .	45
3.3	Results . . . . .	47
3.3.1	Missense mutations perturb the structure of mTOR kinase domain . . . . .	47
3.3.2	Missense mutations do not appear to shift the formation of an active kinase domain . . . . .	50
3.3.3	Missense mutations do not disrupt the formation of inhibitory salt bridges between the kinase and FAT domains . . . . .	52
3.3.4	Free energy calculations show promise in predicting impact of mutations on small molecule and ATP affinity . . . . .	54
3.4	Conclusions . . . . .	57
<b>4</b>	<b>Predicting the impact of clinically-observed mutations using physical modeling and biophysical experiments</b>	<b>58</b>
4.1	Introduction . . . . .	58
4.1.1	The long tail of rare kinase mutations frustrates prediction of drug resistance . . . . .	59
4.1.2	Alchemical free-energy methods can predict inhibitor binding affinities . . . . .	61
4.1.3	Alchemical approaches can predict the impact of protein mutations on free energy . . . . .	61
4.1.4	Assessing the potential for physical modeling to predict resistance to FDA-approved TKIs . . . . .	62
4.1.5	A benchmark of $\Delta pIC_{50}$ s for predicting mutational resistance . . . . .	63

4.1.6	Summary of experimental data that is available and what needs to be done to make better information available . . . . .	68
4.1.7	Open-source YANK/Amber forcefield stuff . . . . .	68

# List of Tables

4.1 Public $\Delta\text{pIC}_{50}$ datasets for 144 Abl kinase mutations and eight tyrosine kinase inhibitors (TKIs) with corresponding wild-type co-crystal structures . . . . .	65
---	----

# List of Figures

2.1	Free energy calculations speed up selectivity optimization . . . . .	16
2.2	A CDK2/CDK9 selectivity dataset from Shao et al., 2013 . . . . .	19
2.3	CDK2 and ERK2 selectivity dataset from Blake et al., 2016 . . . . .	22
2.4	Relative free energy calculations can accurately predict potency, but show larger errors for selectivity predictions. . . . .	24
2.5	Correlation in selectivity prediction errors can be used to accelerate selectivity optimization . . . . .	27
2.6	CDK2 adopts an inactive conformation in the crystal structure used for the CDK2/ERK2 calculations . . . . .	30
2.7	Correlation coefficient $\rho$ controls the shape of the joint marginal distribution of errors . . . . .	31
2.8	Correlation reduces the expected error for selectivity predictions . . .	32
3.1	mTOR is an atypical kinase with a number of regulatory domains . .	34
3.2	mTORC1 integrates signaling from a number of inputs . . . . .	38
3.3	Hyperactivating mTOR missense mutations have been observed in cancer . . . . .	40
3.4	Missense mutations can perturb local structure . . . . .	49
3.5	Missense mutations do not shift the population of the active conformation on a common activation order parameter . . . . .	51
3.6	Missense mutations in kinase domain do not disrupt interactions between the kinase and FAT domains . . . . .	53
3.7	Free energy calculations identify potential resistance mutations . .	56
4.1	Relative alchemical free-energy calculations can be used to predict affinity changes of FDA-approved selective kinase inhibitors arising from clinically-identified mutations in their targets of therapy. . . .	60
4.2	Cross-comparison of the experimentally measured effects that mutations in Abl kinase have on ligand binding, performed by different labs. . . . .	67

**CHAPTER 1**

**INTRODUCTION**

# CHAPTER 2

## PREDICTING THE SELECTIVITY OF SMALL MOLECULE KINASE INHIBITORS

### 2.1 Introduction

#### 2.1.1 Free energy methods can aid structure-based drug design

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].

#### 2.1.2 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 polypharmac-

logical 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.

### **2.1.3 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.

#### **2.1.4 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 to 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.

### 2.1.5 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 ( $\rho$ ) 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.

## 2.2 Methods

### 2.2.1 Numerical model of selectivity

To model the impact correlation would have on the expected uncertainty for selectivity predictions,  $\sigma_{selectivity}$  was calculated using Equation 2.1 for 1000 evenly spaced values of the correlation coefficient ( $\rho$ ) from 0 to 1, for a number of combinations of per target errors ( $\sigma_{target1}$  and  $\sigma_{target2}$ )

$$\sigma_{selectivity} = \sqrt{\sigma_{target1}^2 + \sigma_{target2}^2 - 2\rho\sigma_{target1}\sigma_{target2}} \quad (2.1)$$

The speed up in selectivity optimization that could be expected from using free energy calculations of a particular per target error ( $\sigma_{selectivity}$ ) was quantify as follows using NumPy (v 1.14.2). An original, true distribution for the changed in selectivity of 200000000 new compounds proposed with respected 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 ( $\sigma_{target}$ ) of 1 kcal/mol [12] and a specified correlation coefficient  $\rho$ . A  $\sigma_{selectivity}$  was calculated according to Equation 2.1. The noise of the computational method was modeled as a normal distribution centered around 0 with a standard deviation of  $\sigma_{selectivity}$  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 experimental measured. The speedup value for each value of  $\rho$  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 ( $\rho$ ) between 0 and 1, for four values of  $\sigma_{selectivity}$  and 40000000 compounds in the original distribution.

### 2.2.2 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 0.3.

### **2.2.3 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.

### **2.2.4 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 5. 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.

## 2.2.5 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 2.2:

$$\Delta\Delta G_{corrected} = \Delta\Delta G_{uncorrected} - RT \log \left( \frac{10^{pK_a - pH} + 1}{e^{\frac{\Delta G_{neutral} - \Delta G_{charged}}{RT}} * (10^{pK_a - pH} + 1)} \right) \quad (2.2)$$

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

## 2.2.6 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 2.3, while the RMSE was calculated according to Equation 2.4.

$$MUE = \frac{\sum_0^n |\Delta G_{calc} - \Delta G_{exp}|}{n} \quad (2.3)$$

$$\text{RMSE} = \frac{\sum_0^n \sqrt{\Delta G_{\text{calc}}^2 - \Delta G_{\text{exp}}^2}}{n} \quad (2.4)$$

Each RMSE and MUE is reported with a 95% confidence interval calculated from 10000 replicates of a choose-one-replace bootstrap protocol on the  $\Delta G$  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>

### 2.2.7 Quantification of the correlation coefficient $\rho$

To quantify  $\rho$ , 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 2.5). 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.

$$G_{i,\text{target}}^{\text{phase}} = \mathcal{N}(\mu = 0, \text{sd} = 25.0 \text{ kcal/mol}) \quad (2.5)$$

For each edge of the FEP map (ligand  $i - j$  ligand  $j$ ), there is a contribution from dummy atoms, that was modeled as in Equation 2.6.

$$c_{i,j} = \mathcal{N}(\mu = 0, \text{sd} = 25.0 \text{ kcal/mol}) \quad (2.6)$$

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

$$\Delta G_{phase, ij, target}^{BAR} = \mathcal{N}(G_{j,target}^{phase} - G_{i,target}^{phase}, \delta^2 \Delta G_{phase, ij, target}^{BAR}, observed = \Delta G_{phase, ij, target}^{calc}) \quad (2.7)$$

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 2.8:

$$\Delta\Delta G_{target, ij}^{FEP} = \Delta G_{complex, ij, target}^{BAR} - \Delta G_{solvent, ij, target}^{BAR} \quad (2.8)$$

To model the way an offset is calculated for the  $\Delta G$  reported by the FEP+ panel in Maestro:

$$offset = \frac{\sum^n G_{i,target}^{complex} - G_{i,target}^{solvent}}{n} - \frac{\sum^n \Delta G_i^{exp}}{n} \quad (2.9)$$

The offset was added to each  $\Delta G_i^{BAR}$  to calculate  $\Delta G_i^{sch}$ .

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 2.10)

$$\Delta G_{i,target}^{exp} = \mathcal{N}(\Delta G_{i,target}^{true}, 0.3 \text{ kcal/mol}, observed = \Delta G_{i,target}^{obs}) \quad (2.10)$$

$\Delta G_{i,target}^{true}$  was assigned a weak normal prior, as in equation 2.11.

$$\Delta G_{i,target}^{true} = \mathcal{N}(0, 50 \text{ kcal/mol}) \quad (2.11)$$

The error for a given ligand was calculated as in Equation 2.12.

$$\epsilon_i = \Delta G_i^{sch} - \Delta G_i^{true} \quad (2.12)$$

From these  $\epsilon$  values, we calculated the correlation coefficient,  $\rho$  as in Equation 2.13.

$$\rho = \frac{cov(\epsilon_{target1}, \epsilon_{target2})}{\sigma_{target1}\sigma_{target2}} \quad (2.13)$$

Where  $\sigma$  is the standard deviation of  $\epsilon$ . To quantify  $\rho$  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.

## 2.3 Results

### 2.3.1 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 ( $\Delta G_{binding}$ ), 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

$(\Delta\Delta G_{selectivity})$ , which can be calculated as in Equation 2.14.  $\Delta\Delta G_{selectivity}$  is a useful measure of compound selectivity once a single, or small panel, of off-targets have been identified.

$$\Delta\Delta G_{selectivity} = \Delta G_{binding, target 2} - \Delta G_{binding, target 1} \quad (2.14)$$

To predict the  $\Delta\Delta G_{selectivity}$  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 ( $\Delta G_{target}$ ). From this set of calculations, we can calculate a  $\Delta\Delta G_{selectivity}$  for each ligand given two targets of interest. Previous work shows that FEP+ can achieve an accuracy ( $\sigma_{target}$ ) 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  $\Delta\Delta G_{selectivity}$ , leading to a selectivity uncertainty ( $\sigma_{selectivity}$ ) that is lower than would be expected.

### 2.3.2 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 ( $\sigma_{selectivity}$ ) using alchemical free energy techniques, we created

a simple numerical model following equation 2.1, 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 2.1A, if the per target errors ( $\sigma_1$  and  $\sigma_2$ ) are the same,  $\sigma_{selectivity}$  approaches 0 as the correlation coefficient ( $\rho$ ) approaches 1. If the error for the free energy method is not the same,  $\sigma_{selectivity}$  gets smaller but approaches a non-zero value as  $\rho$  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 2.1B, 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 ( $\sigma_{target}$ ) 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 2.1B, 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 2.1C), where 200-300x speedups can be expected, depending on  $\sigma_{target}$  for the free energy methodology.

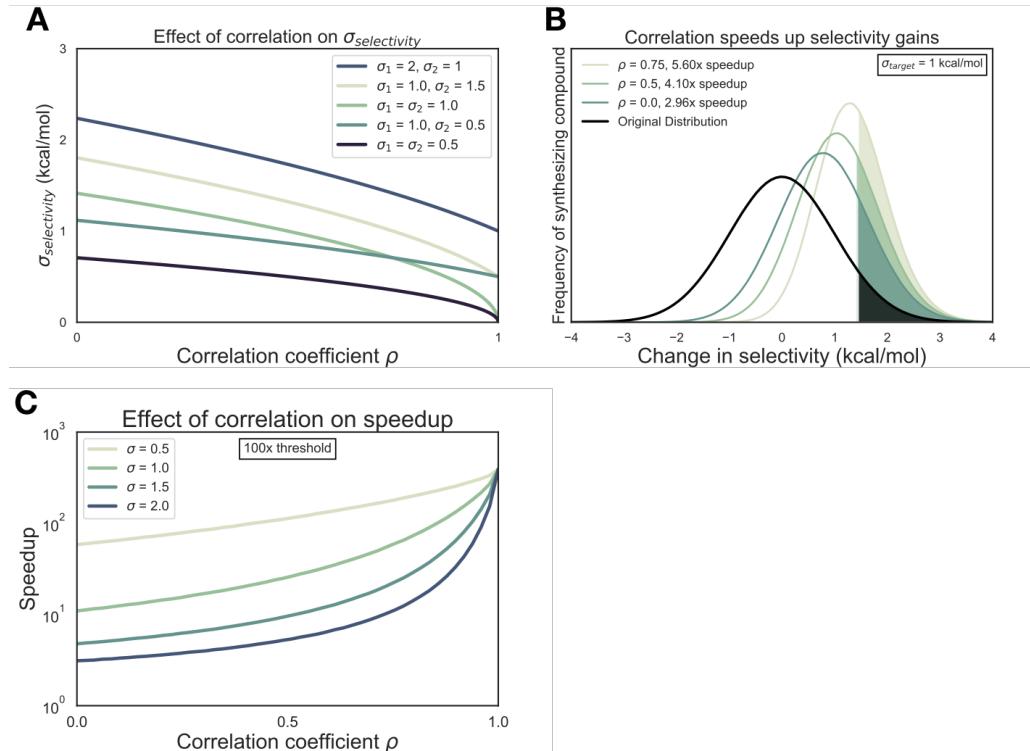


Figure 2.1: Free energy calculations speed up selectivity optimization (A) The effect of correlation on expected errors for predicting selectivity ( $\sigma_{selectivity}$ ) in kcal/mol. Each curve represents a different combination of target errors ( $\sigma_1$  and  $\sigma_2$ ). (B) The change in 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  $\rho$ . Each curve corresponds to a different  $\sigma_{target}$  value.

### 2.3.3 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/cyclin A (Figure 2.2A, left) and CDK9/cyclin T (Figure 2.2B, 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 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 2.2C, 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 2.2C, 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  $\Delta\Delta G_{selectivity}$  (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].

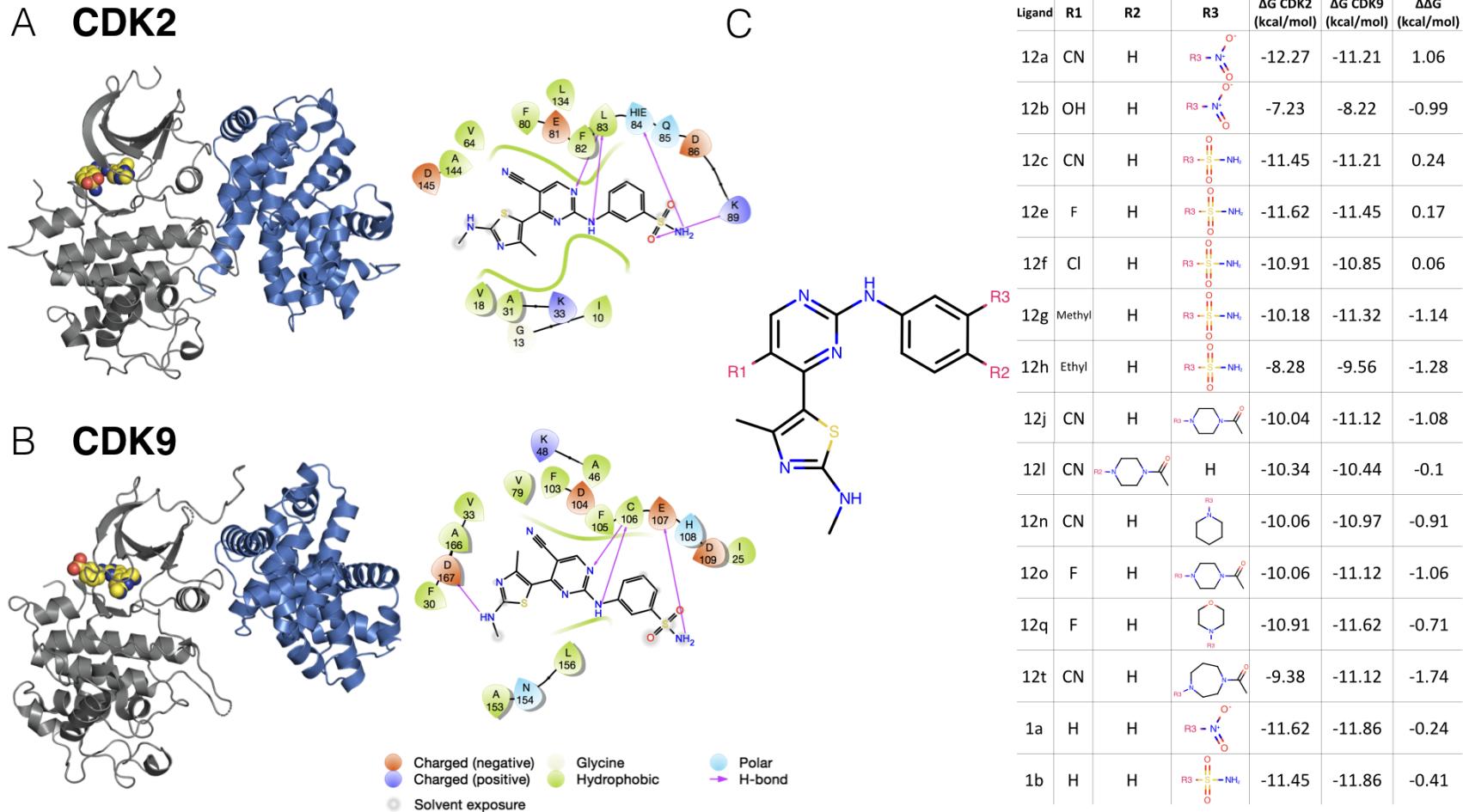


Figure 2.2: A CDK2/CDK9 selectivity dataset from Shao *et al.*, 2013 (A) (left) Crystal Structure (4BCK)[54] of CDK2 (gray ribbon) bound to ligand 12c (yellow spheres). Cyclin A is shown in blue ribbon (right) 2D ligand interaction map of ligand 12c in the CDK2 binding site. (B) (left) Crystal structure 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].

### 2.3.4 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 2.3A, top) and ERK2 (Figure 2.3B, 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  $\alpha$ C helix rotated out, away from the ATP binding site and breaking the conserved salt bridge between K33 and E51 (Supp. Figure 2.6A), indicative of an inactive kinase [43, 64]. By comparison, the CDK2 structure from the CDK2/CDK9 dataset adopts a DFG-in conformation with the  $\alpha$ 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 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 2.4, 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  $\alpha$ C helix, forms a hydrogen bond with the scaffold (Figure 2.4A and 2.4B, 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 congenic ligand series features a single substituent 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

I probably need to either remove the reference to how closely related the kinases are, or come up with a way to quantify that - SKA

is there a good citation for this?

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.

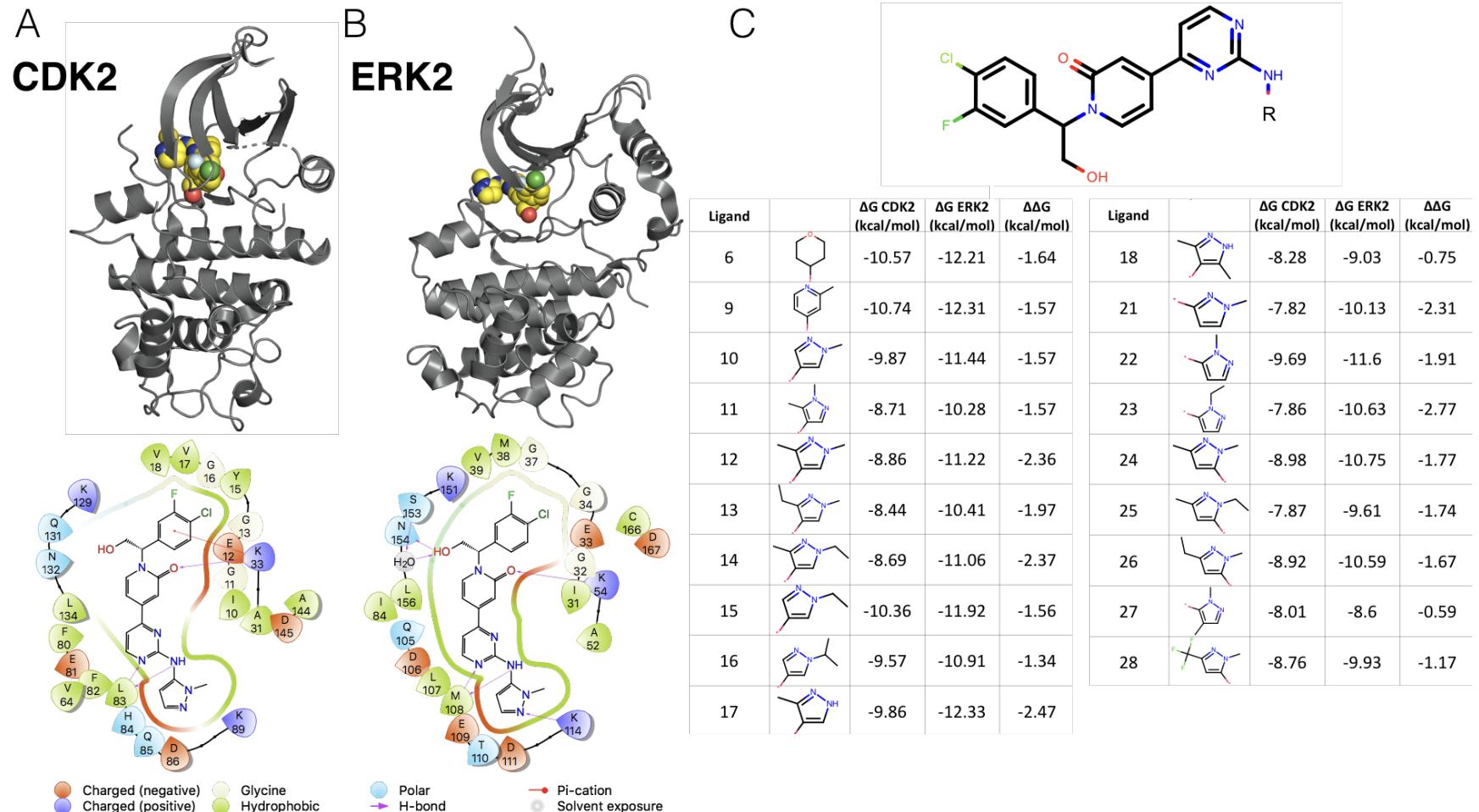


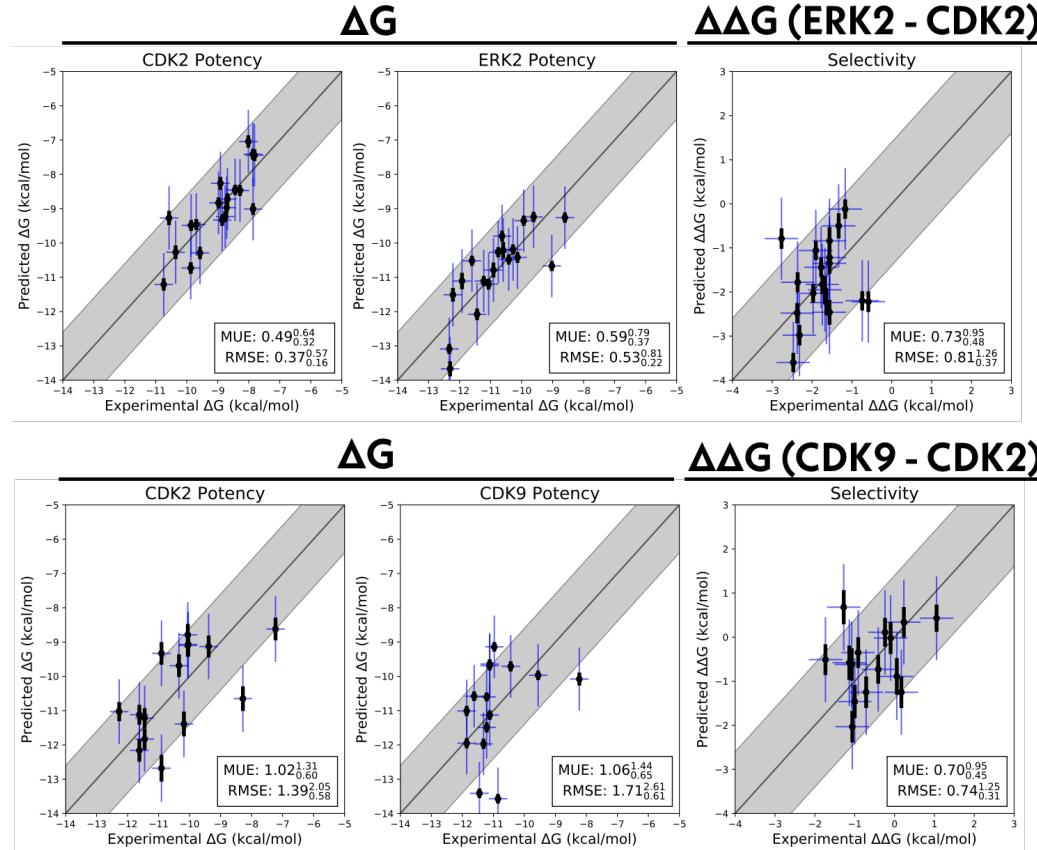
Figure 2.3: CDK2 and ERK2 selectivity dataset from Blake et al., 2016

(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.

### 2.3.5 FEP+ calculations show accurate potency predictions for ERK2/CDK2 and larger errors for CDK2/CDK9

The FEP+ predictions of single target potencies ( $\Delta G$ ) showed good accuracy for the CDK2 and ERK2 dataset (Figure 2.4, 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. Despite the high accuracy for the single target potency predictions, the selectivity ( $\Delta\Delta G_{selectivity}$ ) 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 2.4, 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  $\Delta\Delta G_{selectivity}$ .

this will change  
if we take off the  
error bars



**Figure 2.4: 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<sup>-1</sup> [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 ( $\sigma_{calc}$ ). 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.

## 2.3.6 Free energy calculation errors are correlated, accelerating selectivity optimization

To quantify the correlation coefficient ( $\rho$ ) of the errors in our calculations, we built a Bayesian graphical model, 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 2.5. The model was chained to the FEP+ calculations by providing the  $\Delta G_{phase,ij,target}^{calc}$  as observed data, as in equation 2.7. 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 2.10) and the  $\Delta G_{i,target}^{true}$  was assigned a weak normal prior (Equation 2.11). The correlation coefficient was calculated for each sample according to equation 2.12. The correlation coefficient  $\rho$  for the CDK2/ERK2 calculations was quantified to be  $0.5_{-0.23}^{0.33}$ , indicating that the errors are largely uncorrelated between ERK2 and CDK2 (Figure 2.5A, right). The joint marginal distribution of the error ( $\epsilon$ ) for each target is symmetric, which is expected for cases in which  $\rho$  is 0 (Supp. Figure 2.7). 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 2.5A, right). The CDK2/CDK9 calculations show strong evidence of correlation, with a correlation coefficient of  $0.70_{-0.57}^{0.82}$  (Figure 2.5B, right). The joint marginal distribution of errors is strongly diagonal, which is expected based on the value for  $\rho$  (Figure 2.5B, 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 2.5B, right), despite the much higher per target

errors. Quantifying  $\rho$  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  $\rho$ , the expected  $\sigma_{selectivity}$  for the CDK2/CDK9 calculations is between 0.76 and 1.16 kcal/mol (Supp. Figure 2.8), which is in good agreement with the bootstrapped RMSE (Figure 2.4, bottom). For the CDK2/ERK2 calculations,  $\sigma_{selectivity}$  is expected to fall between 0.82 kcal/mol and 1.10 kcal/mol (Supp. Figure 2.8), which is also in good agreement with the bootstrapped RMSE (Figure 2.4, top).

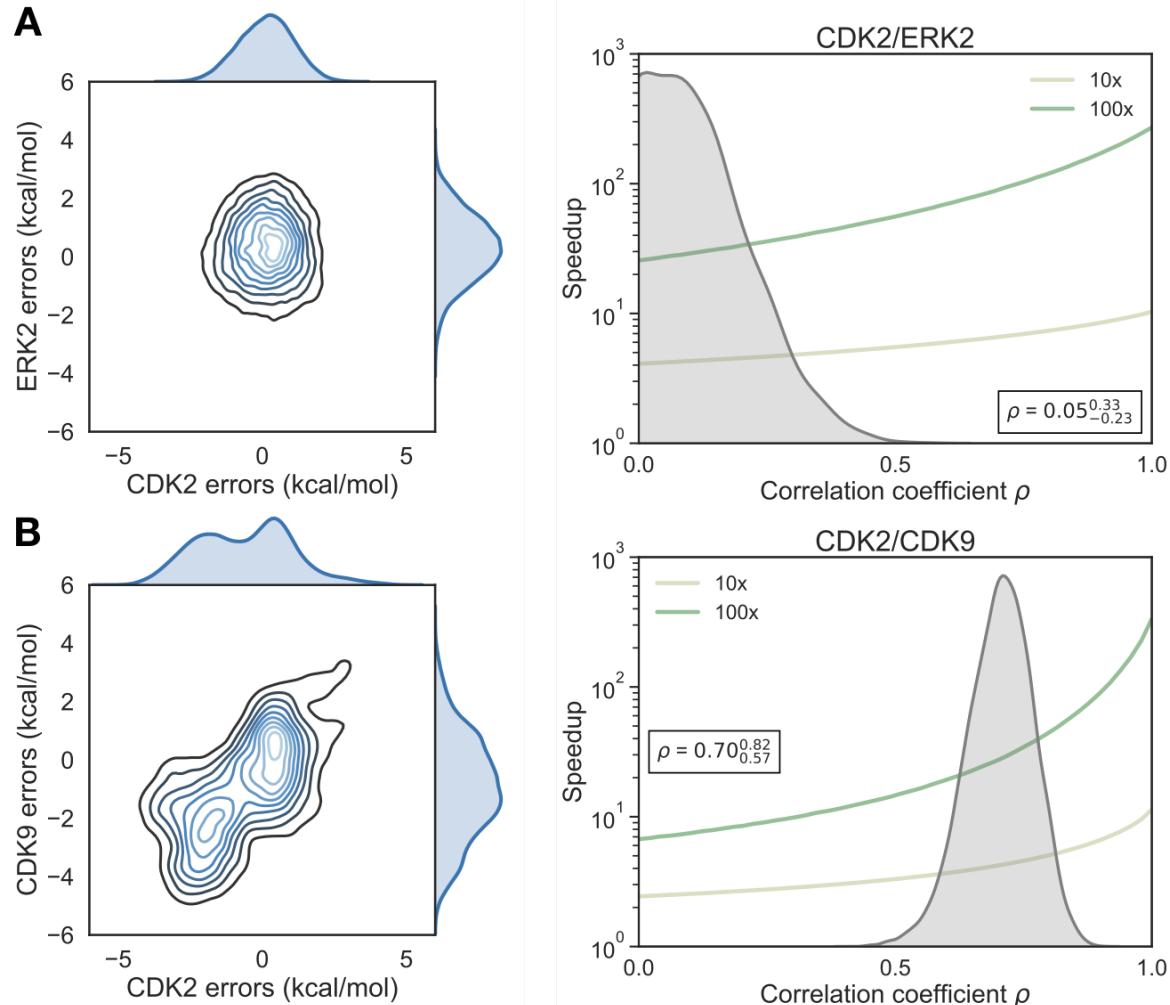


Figure 2.5: Correlation in selectivity prediction errors can be used to accelerate selectivity optimization

(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 ( $\rho$ ) 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.

## 2.4 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 ( $\sigma_{target}$ ), 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, are 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. Furhter, 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. Work is being done to increase the availability of publicly available, high quality biophysical data for selectivity measurements, which will benefit future work on predicting selectivity using physical models and machine learning techniques.

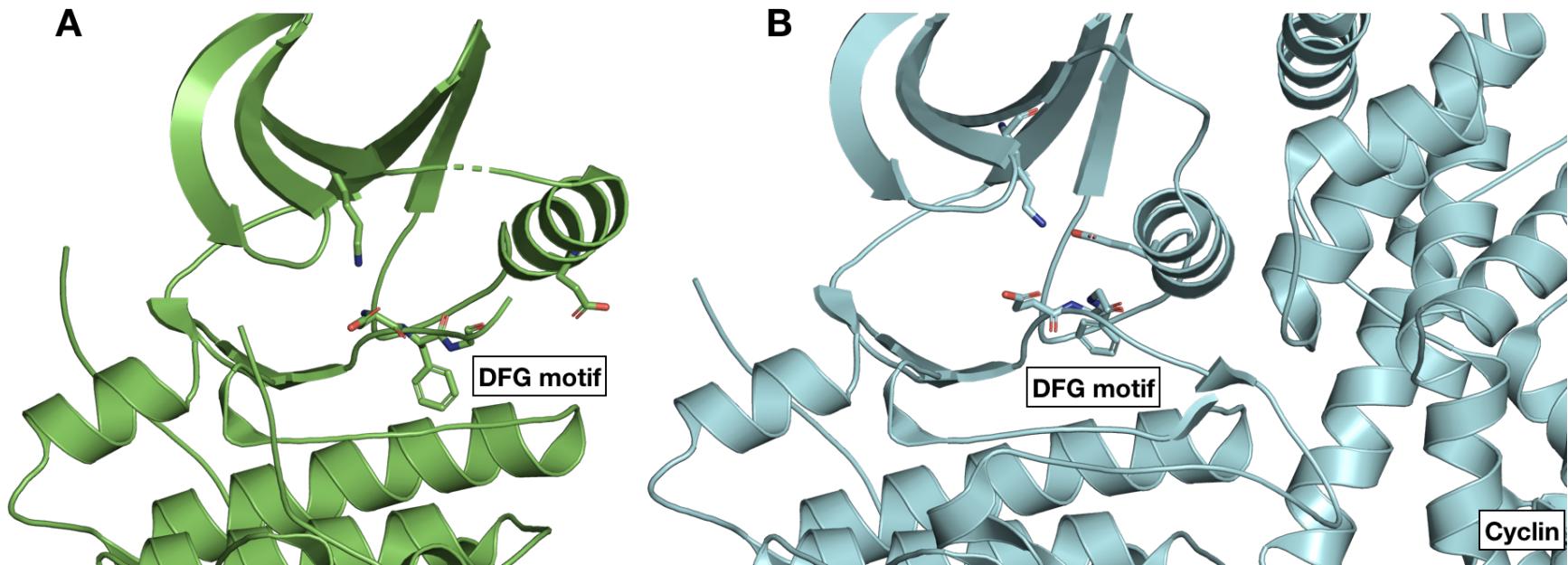
insert a discussion of outliers in the FEP+ calculations here?

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)

Do we want to mention protein mutation FEP+ to enable cycle closure analysis for certain highly related targets?



**Figure 2.6: 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-out conformation, with the  $\alpha$ 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  $\alpha$ C helix [65] (B) The CDK2 structure used for the CDK2/CDK9 calculations (4BCK) contains cyclin A and adopts a DFG-in/ $\alpha$ C helix-in conformation that forms the salt bridge between K33 and E51. This is typically indicative of a fully active kinase [43, 64].

IE

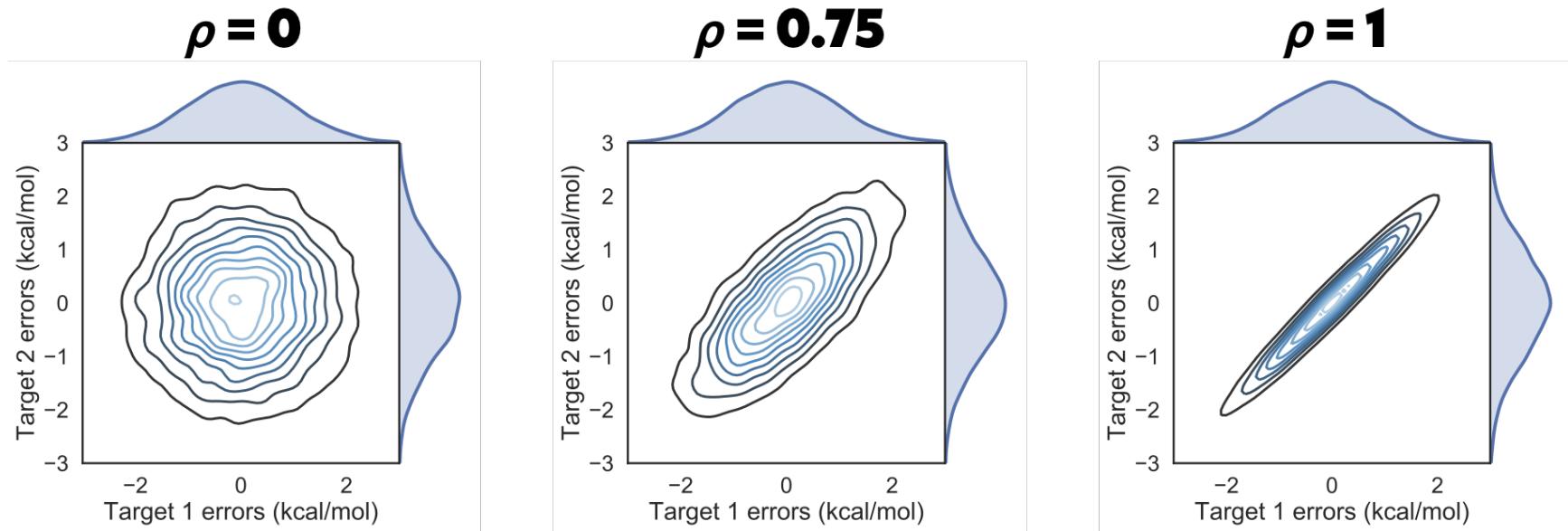
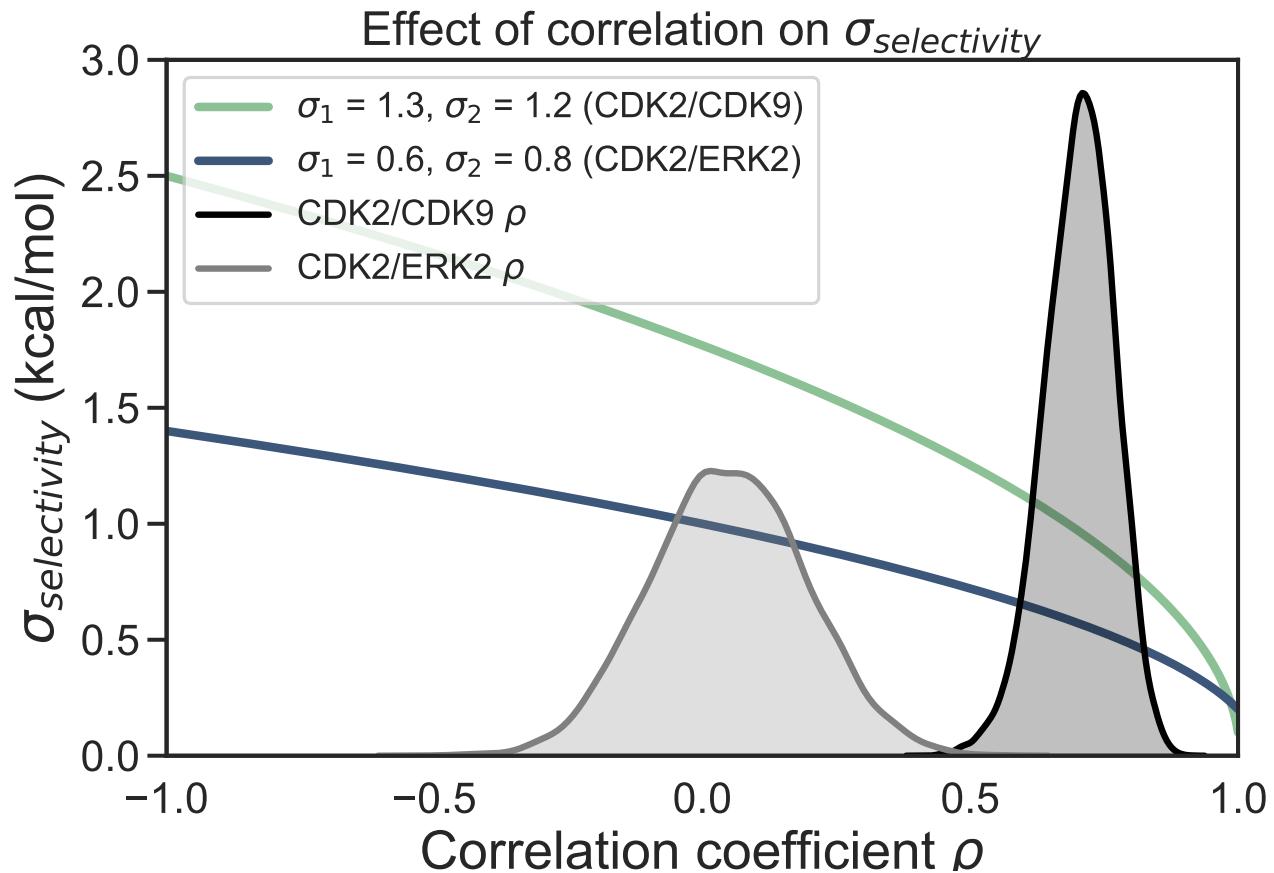


Figure 2.7: **Correlation coefficient  $\rho$  controls the shape of the joint marginal distribution of errors** As  $\rho$  increases, the joint marginal distribution of errors become more diagonal. Each panel shows 10000 samples drawn from a multivariate normal distribution centered around 0 kcal/mol, where the per target error was set to 1 kcal/mol and  $\rho$  to the value indicated in bold over the plot.



**Figure 2.8: Correlation reduces the expected error for selectivity predictions** As corelation coefficient  $\rho$  increases,  $\sigma_{selectivity}$  decreases. The intersection between CDK2/CDK9  $\sigma_{selectivity}$  (green curve) and  $\rho$  (black distribution) indicates the range of expected  $\sigma_{selectivity}$  values. The intersection for CDK2/ERK  $\sigma_{selectivity}$  (blue curve) and  $\rho$  (gray distribution) suggests the expected  $\sigma_{selectivity}$  range for that set of calculations.

CHAPTER 3

**UNDERSTANDING THE FUNCTIONAL IMPACT OF MTOR CLINICAL  
KINASE MUTATIONS USING PHYSICAL MODELING**

### **3.1 Introduction**

#### **3.1.1 mTOR forms the catalytic core of protein complexes that control a number of cellular processes**

mTOR (mammalian target of rapamycin) is a serine-threonine kinase that controls a number of cellular processes [66, 67] by integrating signaling from the MAPK and PI3K pathways [24]. mTOR forms the catalytic core of heteromeric protein complexes, mTORC1 and mTORC2, that differ in the regulatory proteins that decorate the kinase. mTORC1 is defined by RAPTOR [68, 69] and PRAS40 [70], while mTORC2 is characterized by RICTOR [71]. Each complex contains a number of shared regulatory proteins, such as mLST8 [72] and DEPTOR [73]. mTOR itself is an atypical kinase, and includes a number of insertions that deviate from the canonical kinase fold. The C-terminal fragment crystal structure (Figure 3.1) shows that the kinase domain is hugged by a FAT domain [74], forming a C-shaped clamp around the kinase domain. The FAT domain forms a number of regulatory salt bridges that have been implicated in controlling the activity of the kinase [74]. The FK506-rapamycin-binding (FRB) domain hangs over the opening cleft to the active site [74] (Figure 3.1). The FRB domain is exposed in mTORC1, while it is thought to be occluded and inaccessible to FKBP12 in mTORC2 [75]. Complex activity is hypothesized to be regulated via restricted access to the active site by architectural elements as well as rapalog-mediated recruitment of FKBP to the FRB domain [76].

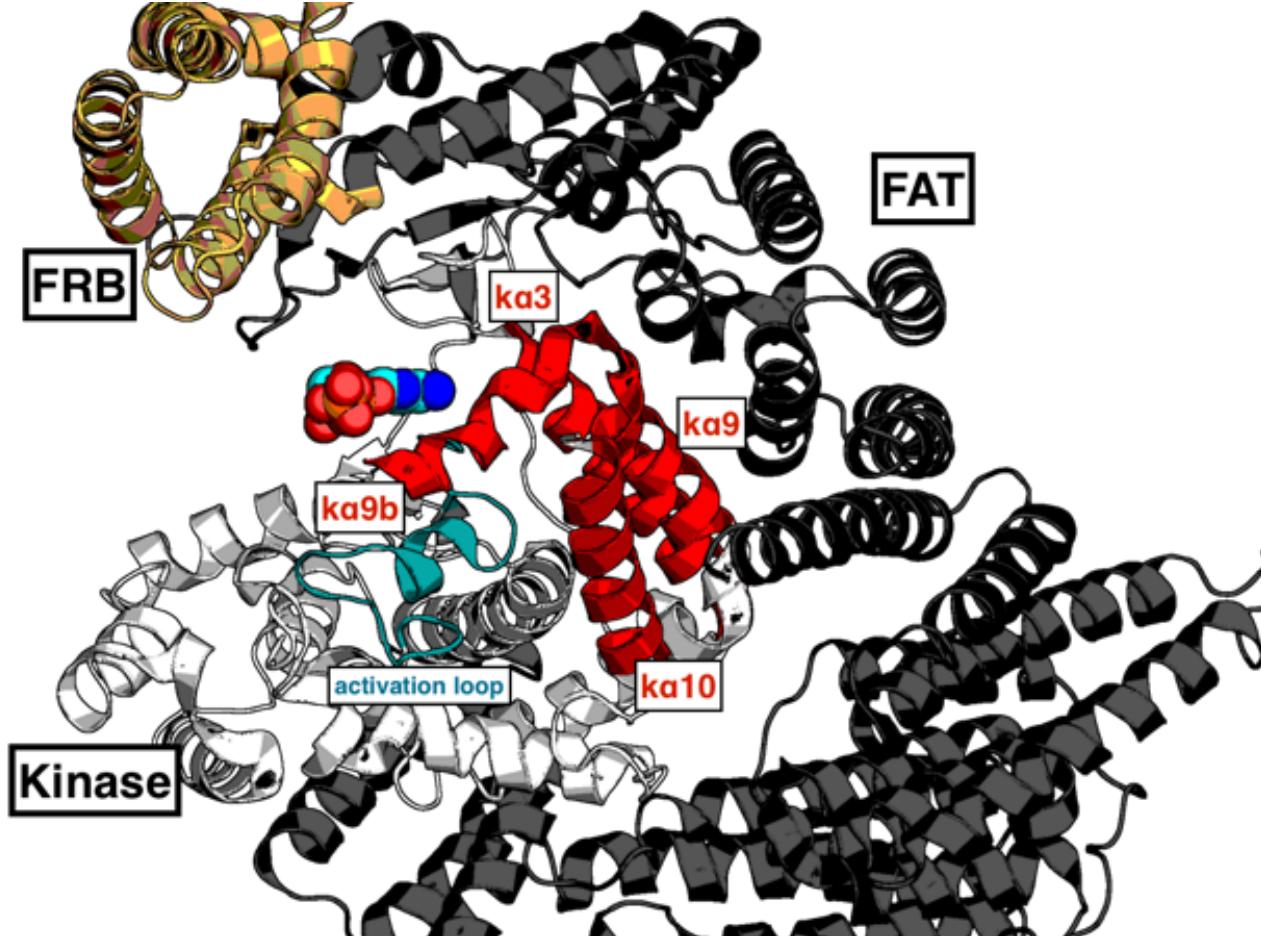


Figure 3.1: **mTOR is an atypical kinase with a number of regulatory domains** mTOR (PDBID: 4JSV) is shown as a cartoon diagram. The kinase domain (white) has a number of structural features highlighted, such as the activation loop (teal) as well as a network of regulatory  $\alpha$  helices (red) in the active site. The FAT domain (black) clamps around the kinase domain and forms a number of salt bridge and hydrophobic contacts with the kinase domain. The FRB domain (gold) hangs over the active site clef and is the site of rapalog binding and rapalog-mediated FBKBP recruitment. This figure reprinted with permission of James Hsieh

mTORC1 controls a number of biological processes by integrating multiple upstream signals (Figure 3.2). Tuberous sclerosis complex 1/2 (TCS1/2) integrates growth factor signaling from the PI3K/AKT pathway [77, 78] and DNA damage from AMPK [79]. TSC1/2 acts as GTPase activating protein (GAP) for RHEB [80], which binds to and activates mTORC1 at the lysosomal surface. mTORC1 localization is controlled by the second major signal it integrates: amino acid availability. Upon stimulation by amino acids, the Ragulator [81] recruits inactive mTORC1 from the cytosol to the lysosomal surface [82–84], enabling interaction with RHEB. Once active, mTORC1 controls cell growth and protein synthesis by phosphorylating and activating S6K at T389 and inactivating 4EBP1 via a phosphorylation at S65 [66, 85]. mTORC1 also activates SREBP1/2, which regulates lipid biosynthesis [86]. mTORC1 activation inhibits autophagy through via ULK1. Recent work suggests that mTORC1 can also control the biophysical properties of the cytoplasm by regulating crowding [87], impacting the rate of diffusion and expanding the already diverse array of essential processes that mTOR controls.

### 3.1.2 mTOR is targeted by inhibitors with two distinct mechanisms of action

Due to the central role mTOR plays in a wide array of biological processes, it has emerged as the target of extensive drug discovery programs. There are two distinct classes of inhibitors developed to target mTOR: ATP-competitive inhibitors and allosteric inhibitors [88, 89]. Temsirolimus [90] and Evirolimus [91] are FDA-approved inhibitors [36] that are thought to inhibit mTOR through recruiting FKBP family members, such as FKBP12, to the FRB domain [92]. Evidence suggests FKBP12 recruitment may inhibit mTORC1 activity by inducing a conformational

change that prevents S6K binding [93]. Further incubation leads to destabilization and disassembly of the protein complex, which is consistent with the time-dependent nature of rapalog inhibition of mTORC1 mediated 4EBP1 phosphorylation [93]. The rapalogs are potent inhibitors of mTORC1, while mTORC2 requires chronic rapalog treatment and is highly dependent on FKBP expression levels [94]. A cryo-EM structure of TORC2 suggests that the FRB domain is inaccessible in TORC2, explaining the relative insensitivity of mTORC2 to rapalog inhibition [75]. Rapalogs have also been shown to only partially inhibit mTORC1, often failing to reduce 4EBP phosphorylation levels [67], despite efficacy at preventing S6K phosphorylation. There are an array of ATP-competitive inhibitors developed to target mTOR and other members of the PI3K family of kinases, such as MLN0128 [95, 96], INK-228 (TAK-128) [97], AZD8055 [98], PKI-587 [99], and BEZ-235 [100]. ATP-competitive inhibitors exhibit a wide range of selectivity, with many inhibiting mTORC1, mTORC2, and multiple isoforms of PI3K. ATP competitive inhibitors target the catalytic activity of mTOR and more fully inhibit mTORC1 and mTORC2 phosphorylation of downstream targets [67]. While many of these inhibitors have struggled clinically due to modest clinical benefits and high toxicity leading to low tolerability [101], there is considerable excitement about developing mTOR inhibitors to treat a number of diseases. ATP-competitive inhibitor BEZ-235 has shown promising results in preclinical models of lapatinib-resistant PI3K hyperactivated breast cancer [102]. Dual targeted catalytic inhibitors that target both PI3K and mTOR have shown success by abrogating reactivation of Akt, which is caused by the alleviation of negative feedback due to long term treatment with mTOR inhibitors [103]. For example, a dual-inhibitor of mTORC and PI3K $\alpha$ , PI-103, showed promising activity against glioma xenografts [104].

mTOR inhibitors show the most promise in treating diseases with mTOR pathway alterations pathway [105]. Despite initial success, resistance mutations in the

FRB or kinase domain have been observed [106]. Work has also been done to combine the two mechanisms of action in Rapalink-1, which tethers rapamycin and MLN0128 [107] with a flexible linker and overcome such acquired resistance. Rapalink also shows improvements over ATP-competitive inhibitors or rapalogs outside of the context of resistance. In glioblastoma models, ATP-competitive inhibitor MLN0128 and allosteric inhibitor rapamycin showed poor *in vivo* efficacy. In contrast, RapaLink-1, by overcoming poor residence time and potency, was able to durably inhibit mTORC1 and cross the blood-brain barrier [108].

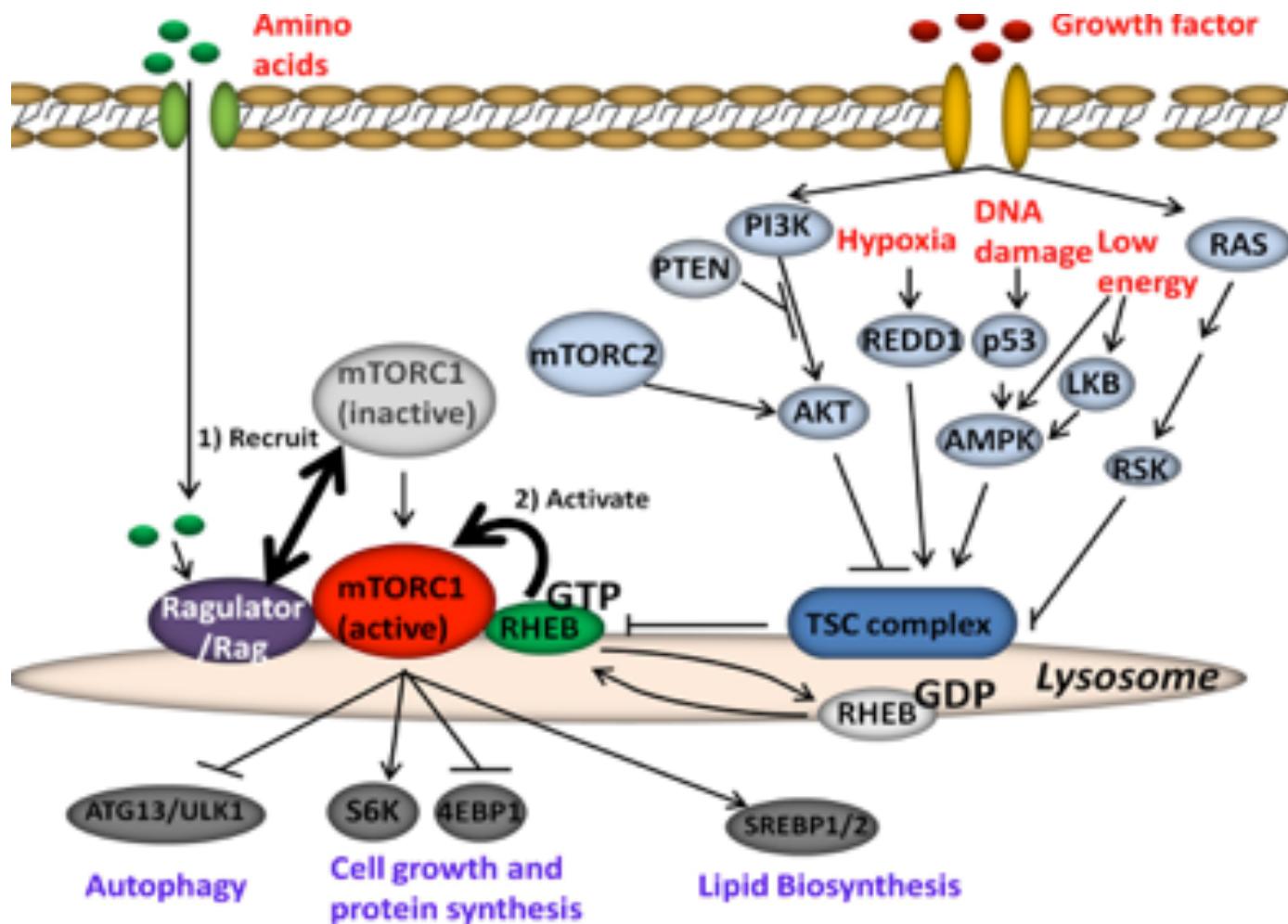


Figure 3.2: mTORC1 integrates signaling from a number of inputs. The signaling pathway of mTOR involves integrating signaling from the growth factors, amino acids, hypoxia, and DNA damage. Integrating such signals controls the localization and activity of mTORC1, which controls autophagy, cell growth, and macromolecule synthesis through phosphorylation of downstream targets. This figure courtesy of James Hsieh

### **3.1.3 mTOR signaling is dysregulated in cancer by hyperactivating missense mutations**

mTOR pathway alterations have been observed in a wide array of cancer types [109] and extensively characterized. Less well studied are missense mutations in mTOR itself. An exceptional responder in a phase 1 clinical trial of pazopanib and everolimus with metastatic urothelial carcinoma lead to the identification of two missense mutations in mTOR [105], E2014K and E2419K. Subsequent work identified 33 MTOR mutations gathered from publicaly available tumor sequencing data, and found that a number of them activated the mTOR pathway and conferred sensitivity to mTOR inhibitors when engineered into various cancer cell lines *in vitro* [110]. While these missense mutations are pulled from an array of different cancer types, mTOR missense mutations have been observed in about 15% of clear cell renal cell carcinoma [111]. Characterization of these mutations revealed that not only do many of these mutations hyperactivate mTOR (Figure 3.3), but they can be grouped into complementation groups by determining which double mutations cause hyperactivation at levels higher than the constituent single mutants alone [112]. This suggests that these mutations can activate mTOR through different, seemingly complementary mechanisms.

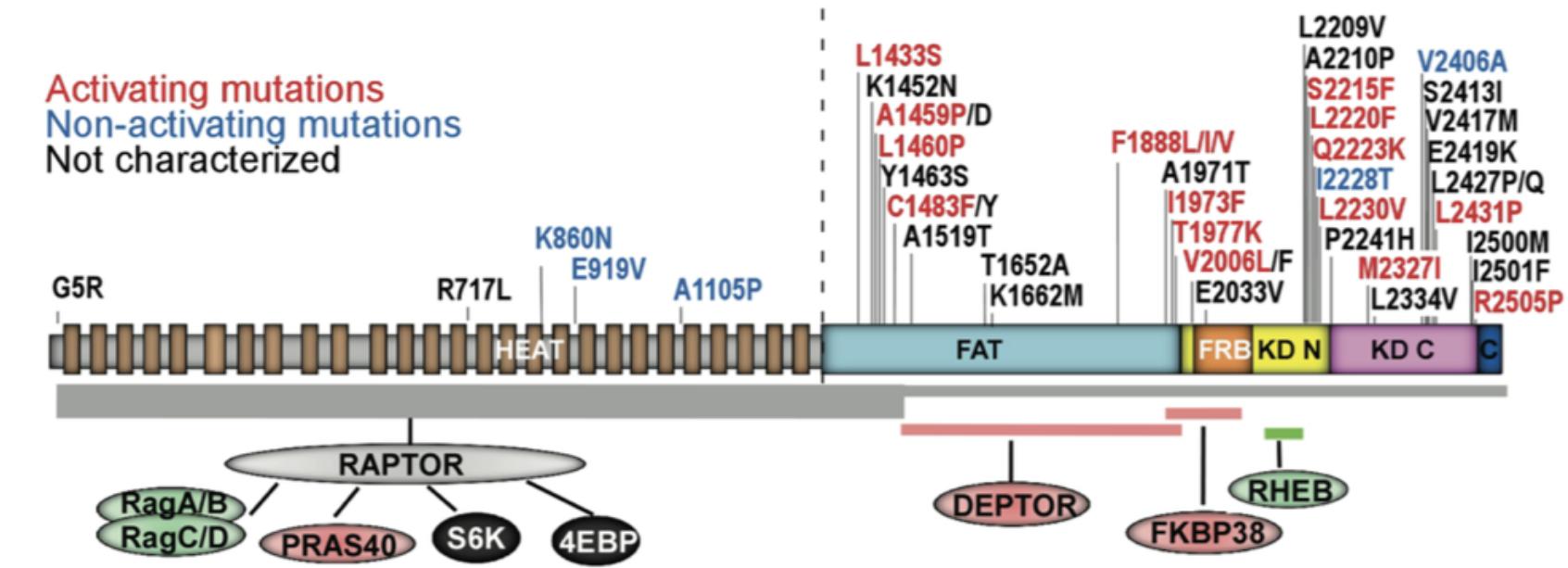


Figure 3.3: **Hyperactivating mTOR missense mutations have been observed in cancer** Diagram shows the domain structure of mTOR, its regulatory interaction partners (negative regulators in pink, positive regulators in green, and a dual-role regulator in gray), and the substrates of mTORC1 complex. The positions within mTOR that are involved in the interaction with the regulatory partners are highlighted below the domain structure. The thickness of the horizontal bar of RAPTOR-mTOR interaction indicates the relative binding affinity. mTOR missense mutations derived from ccRCC are mapped and color coded to summarize their respective effects on mTORC1 signaling (activating mutations in red). KD N, kinase domain N lobe; KD C, kinase domain C lobe. **This figure reprinted with permission of James Hsieh and the Journal of Clinical investigation [112].**

### **3.1.4 Using physical modeling to understand the functional impact of mTOR mutations**

To begin to understand the impact of these missense mutations at an atomistic level, we performed massively parallel molecular dynamics [113] using the computing resource Folding@Home [114]. Molecular Dynamics simulations have been used previously to understand the mechanism of oncogenic and resistance mutations on the structure and activation of EGFR [115, 116]. They have also been applied to understanding missense mutations in p53 [117], CLIC2 [118], opsin [119], and a host of oncogenes and tumor suppressors[120]. Using the previously solved crystal structure [74], we built models for 45 single, and 190 double mutations, far more than could be crystallized individually [112]. We analyzed the simulations for changes in contact formation, as well as changes in order parameters for activation mined from other kinase studies and previous work on mTOR biochemistry. We also piloted alchemical free energy calculations [121] on a number clinically observed mutants, to compute physical testable properties such as change in affinity for ATP-competitive inhibitors and ATP itself. In doing so, we identify promising candidates for resistance mutations to an ATP competitive inhibitor, and lay the ground work for future studies on the application of these methods to studying the functional impact of mutations on inhibitor binding. Taken together, this work forms the beginning of a comprehensive analysis of the impact of these mutations on structure and small molecule binding, which has implications for the treatment of patients with these mutations.

## 3.2 Methods

### 3.2.1 Molecular dynamics simulations

This work was performed and previously described in reference [112]. The canonical wild-type mTOR sequence for the UniProt-annotated PI3K/PI4K domain span (residues 2182-2516) was modeled onto the X-ray structure of mTOR from chain A of RCSB entry 4JSN using the Ensembler automated simulation setup tool [122] with default parameters. A second set of simulations was performed using the full length, C-terminal fragment of the mTOR crystal structure (Uniprot sequence residues 1376-2549) from RCSB entry 4JSN, hereafter called the mTOR kinase+FAT simulations. This sequence was modeled onto the crystal structure using Ensembler with default parameters.

All residues were assigned default protonation states typical of pH 7.4. The AMBER 99SB-ILDN [123] forcefield was used for the protein along with the TIP3P solvent model [124] with neutralizing monovalent Na<sup>+</sup> or Cl<sup>-</sup> counterions. The resulting simulation box had 80,983 atoms. The OpenMM 6.2 simulation package [125] was used for all minimization, equilibration, and production simulations. Equilibration simulations utilized Langevin dynamics with a timestep of 2 fs and collision rate of 20/ps, along with a Monte Carlo barostat with molecular scaling and update interval of 50 steps, with temperature and pressure control set to 300 K and 1 atm. Particle-mesh Ewald (PME) with default parameters was used for long-range electrostatic treatment, direct-space and Lennard-Jones interactions were truncated at 9 Å, and a long-range dispersion correction was employed. Bonds to hydrogen were constrained using CCMA [126] using the default tolerance of 1e-5, and waters were rigidly constrained using SETTLE [127]. The

Ensembler package [122] handles energy minimization and refinement in implicit solvent followed by a short minimization and equilibration step in explicit solvent prior to production simulations. Production simulations of the wild-type kinase domain and full length C-terminal fragment were run on Folding@home [114] using a simulation core based on OpenMM 6.2 and the same simulation parameters, with the exception of a reduced collision rate of 1/ps. The structure obtained after 589.5 ns—which had relaxed much of the initial loop-modeling-induced structural artifacts—was used as a starting model for further modeling of mutations and subsequent production simulations. To model mTOR mutants and wild-type behavior, PDBFixer v1.2 [128], part of the Omnia molecular simulation suite, was used to generate mutant versions of the mTOR kinase domain and full length C-terminal fragment using the relaxed wild-type structure. Subsequent simulation steps utilized reaction-field electrostatics with a cutoff of 10A in place of PME to allow longer trajectories to be generated. Proteins were resolved in TIP3P water with NaCl counterions to neutralize the system and produce an environment of approximately 150 mM NaCl to using a padding of 11A around the kinase, resulting in systems of approximately 81K atoms. Langevin dynamics with a collision rate of 5/ps was used for subsequent production simulations of mutant and wild-type kinase domains, which also employed a simulation core based on OpenMM 6.2 on Folding@home. Simulation boxes were energy minimized with the OpenMM LocalEnergyMinimizer facility before subjecting them to subsequent dynamics. 20 replicate simulations of each mutant simulation box were run on Folding@home, with each replicate receiving a unique random number seed ensuring rapid decorrelation of trajectories. Each of the trajectories were 501 nanoseconds of simulation. The initial 100 ns of each simulation were discarded and subsequent simulation data was analyzed for structural alternations indicative of rapid mutation-induced conformational changes.

### 3.2.2 Contact Map Analysis

Conformational changes were detected by generating a contact map, which shows the net change in the probability of forming a contact between a pair of residues from wild-type to mutant. To calculate these probabilities, mdtraj [129] was first used to calculate the distance between every residue pair based on the closest heavy atoms in each frame of the simulations after 100 ns. Using 5 angstroms as the threshold at which a contact was formed, the number of frames in which a contact was formed was divided by the total number of frames for each simulation, giving a probability for each residue pair in that simulation which could be averaged over the number of replicates per mutant. The wild-type probability was subtracted from the mutant, giving a net change in probability to form a contact for each residue pair in the protein. The structural images were generated using PyMOL to visually inspect areas of interest identified by the contact map over the course of each simulation.

### 3.2.3 Mean Contact formation over time

The fractional contact formation was calculated for 20 replicate 500ns trajectories. Analysis was carried out using 20ns sliding window chunks, with a single frame step (2fs) between each window. Contact cutoff at 4Å for closest heavy atom. Plotted is the mean  $\pm$  SEM, calculated as in Equation 3.2

$$\sigma = \sqrt{\frac{\sum^n(X_i - \mu_x)^2}{n - 1}} \quad (3.1)$$

$$\text{SEM} = \frac{\sigma}{\sqrt{n}} \quad (3.2)$$

Where  $\sigma$  in Equation 3.1 is the standard deviation,  $X_i$  fractional contact for a frame  $i$ ,  $\mu$  is the mean fractional contact for a 20ns window, and n is the total number of frames in that sliding window.

### 3.2.4 Alchemical free energy calculations

#### Structure and Ligand Preparation

Structures for mTOR (4JSP [74] (ATP) and 4JSX [74] (AZD8055)) were downloaded from the PDB [55]. Models were prepared from chain A of each structure using Schrödinger's PrepWizard (2016-1) [56], keeping only the canonical wild-type mTOR sequence for the UniProt-annotated PI3K/PI4K domain span (residues 2182-2516). The FAT and FRB domains were removed from the structure. All other chains were deleted. PrepWizard was used to add in hydrogens at pH 7.4 for both protein residues and the cocrystallized ligand. The protonation state of the cocrystallized ligand was assigned the lowest energy state using Epik at pH  $7.4 \pm 2$ . Hydrogen bonding was optimized using PROPKA at pH  $7.4 \pm 2$ . Each of the structures was minimized using OPLS3 [4] and an RMSD convergence cutoff of 0.3Å. The missing loops, due to their large size, were not modeled in, and were capped by PrepWizard.

ATP and AZD8055 were prepared for docking using Schrödinger's LigPrep (2016-1). 3D structures were generated using OPL3, and ionization state determined by Epik at pH  $7.4 \pm 2$ . All other settings were left on default. The lowest Epik state penalty state was selected for each ligand.

## Docking

ATP and AZD8055 were docked into 4JSP and 4JSX, respectively, using Schrödinger's GLIDE (2016-1) [130–132]. The receptor grid was generated centered on the cocrystallized ligand using the default settings of the Receptor Grid Generation panel. None of the rotatable groups were allowed to rotate to improve computational efficiency. The ligands were docked into these receptor grids using the extra precision (XP) protocol. Ligand sampling was set to flexible, allowing for nitrogen inversions and sampling of different ring conformations. Epik state penalties were added to the docking score, although only one input for each ligand was used. A post docking minimization was performed on the top 5 poses for each ligand. The docking protocol was set to write out only the best pose for each ligand, on the basis of the docking score after minimization.

## Protein mutation FEP+

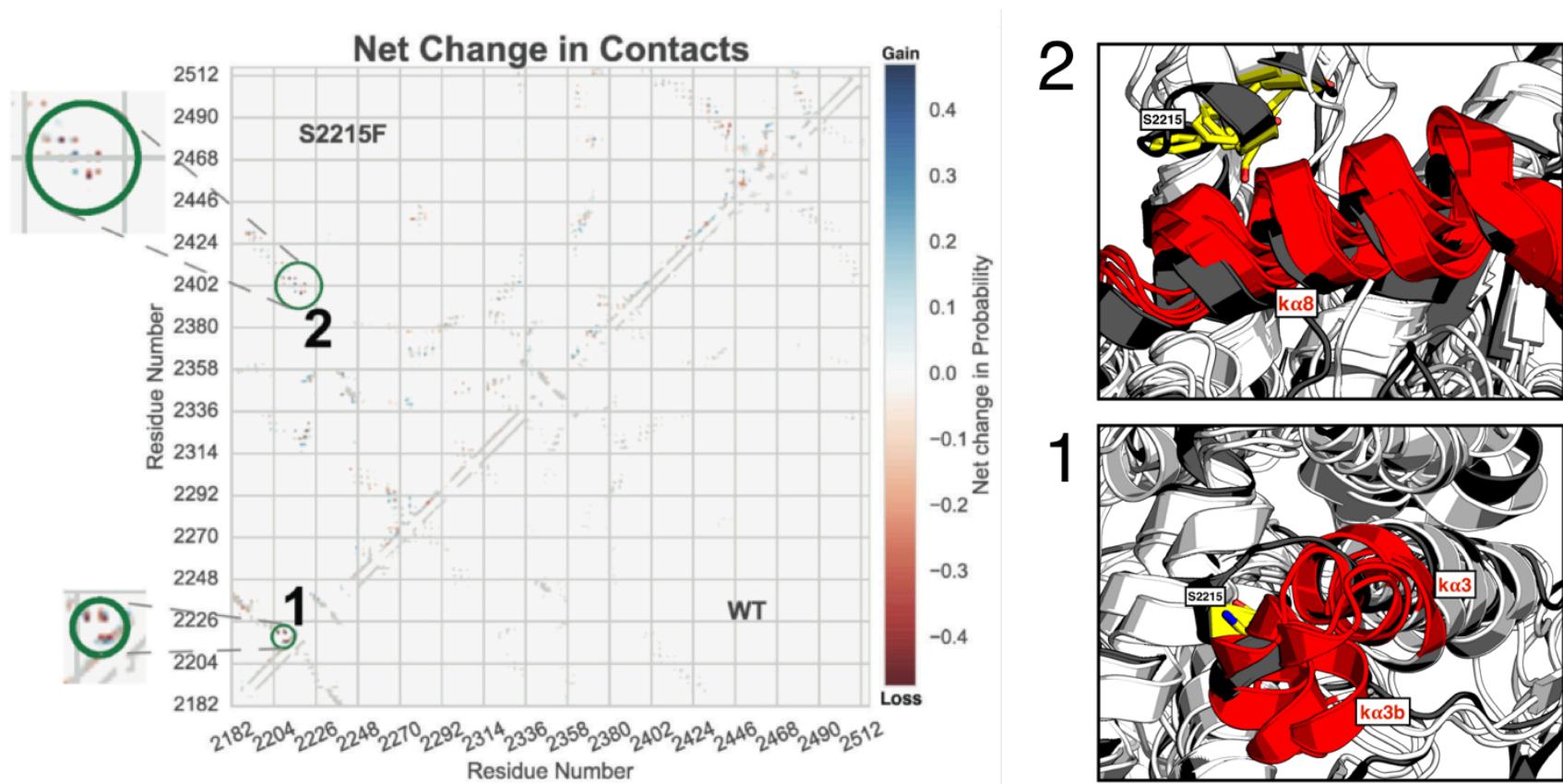
Protein mutation FEP+ [13, 63, 133] was used to calculate  $\Delta\Delta G_{mutation}$  for both ATP and AZD8055 structures for each mutation. The models generated above were parameterized using the default OPLS3 forcefield [57] that shipped with Maestro 2016-1. SPC parameters were used for the water model [134]. The FEP+ panel offers an automated workflow with only requires an input structure and specified mutations. The protocol carried out as described in reference [63], with only single replicates performed. The reported uncertainties in each  $\Delta\Delta G_{mutation}$  are the BAR uncertainty estimates [135, 136].

### 3.3 Results

#### 3.3.1 Missense mutations perturb the structure of mTOR kinase domain

To begin understanding the types of rearrangements missense mutations can induce in the local structure of mTOR, we employed contact map analysis to identify regions in which we detected changes in the formation of contacts from the wild type to mutant simulations. This analysis is described in depth in the methods section. Briefly, we first calculated the distance between every residue pair based on the closest heavy atoms in each frame of the simulations after 100 ns. For each residue pair, we calculated the probability of forming a contact across all of the simulation data and replicas, using 5 Å as the threshold at which a contact was formed. The wild-type probability was subtracted from the mutant, giving a net change in probability to form a contact for each residue pair in the protein. This analysis was performed only for the kinase domain simulations, as the full contact map for the full length C terminal fragment was too computationally intensive to compute for the 190 mutant simulations. Using these maps, it is possible to identify regions in which the mutant is causing a conformational change, both locally near the mutant and more distantly through some allosteric mechanism. As an example of the 45 different mutant kinase domain simulations, we show the kinase domain simulations of S2215F, a highly activating and recurrent missense mutation [? ]. S2215F shows evidence of a rapid, mutation-induced local rearrangement, causing disruption of the  $\alpha$ -helix in which it occurs (Figure 3.4, region 1). This unfolding is evident in many of the replicates, shown in red in the region 1 panel highlighted by Figure 3.4. S2215F, shown in yellow in Figure 3.4, also causes unwinding and

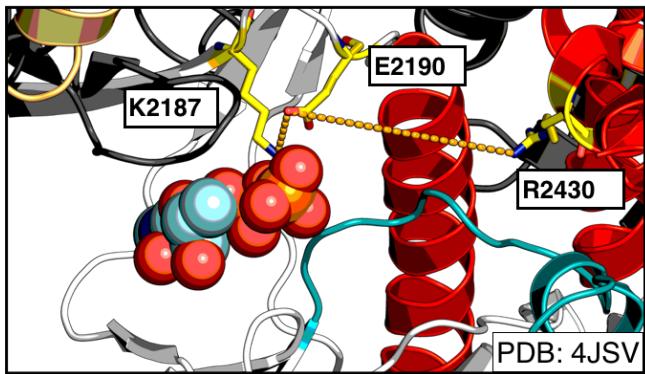
loosening of  $\alpha$ -helix  $k\alpha 8$ , despite this secondary structural element being far from where the mutant occurs (Figure 3.4, region 2).



**Figure 3.4: Missense mutations can perturb local structure** **Left** Contact map showing the difference in probability of forming a contact between WT and mutant S2215F for the kinase domain simulations. **Right** Regions one and two highlighted in contact map, showing a structural perturbation in indicated helices. Starting structure is shown in gray, the residues indicated in the contact map are shown in red and residue 2215 is shown in yellow. All trajectories started from PDB: 4JSN. This left panel of this figure is a modified version of a figure that appears in [112]. Reprinted with permission of James Hsieh and the Journal of Clinical Investigation

### 3.3.2 Missense mutations do not appear to shift the formation of an active kinase domain

Despite the promise of being able to identify structural rearrangements in an automated, high-throughput fashion, contact map analysis is difficult to use to explain mechanistically how these mutants are activating. Seeking a more mechanistic understanding of how these mutants are activating, we looked at a common order parameter for activation from previous work on kinases [137]. In most kinases, there is a highly conserved lysine residue that coordinates ATP when it is bound in the active site. In mTOR, this lysine has been identified as K2187 [74]. To test whether mutations are activating mTOR by shifting the kinase domain into a more active conformation, we looked at the contact formation over time between residues K2187-E2190 (indicative of an active kinase), shown in Figure 3.5 in red, and E2190-R2430, shown in Figure 3.5 in blue. We hypothesized that formation of the E2190-R2430 contact suggests an inactive kinase domain, as the  $\alpha$ C would need to rotate out and away from the active site, which has been observed in inactive kinases in previous work (Figure 3.5, inset panel) [137, 138]. Show on the right in Figure 3.5 is a representative panel of the activating mutants from the kinase domain simulations, as well as the control wild type. While some of the mutants seem less stable than the wild type, none of them appear to shift the distribution of the kinase domain further towards formation of the K2187-E2190 contact, indicative of activation. This is likely because the mTOR kinase domain adopts an active conformation in all of the available crystal structures, and the simulations largely stay in the local minima of the crystal structure. This is further supported by the hypothesis that the mTOR kinase domain is constitutively active, and is regulated by restriction of substrate access [66, 67, 74].



**Indicative of active kinase**  
**Indicative of inactive kinase**

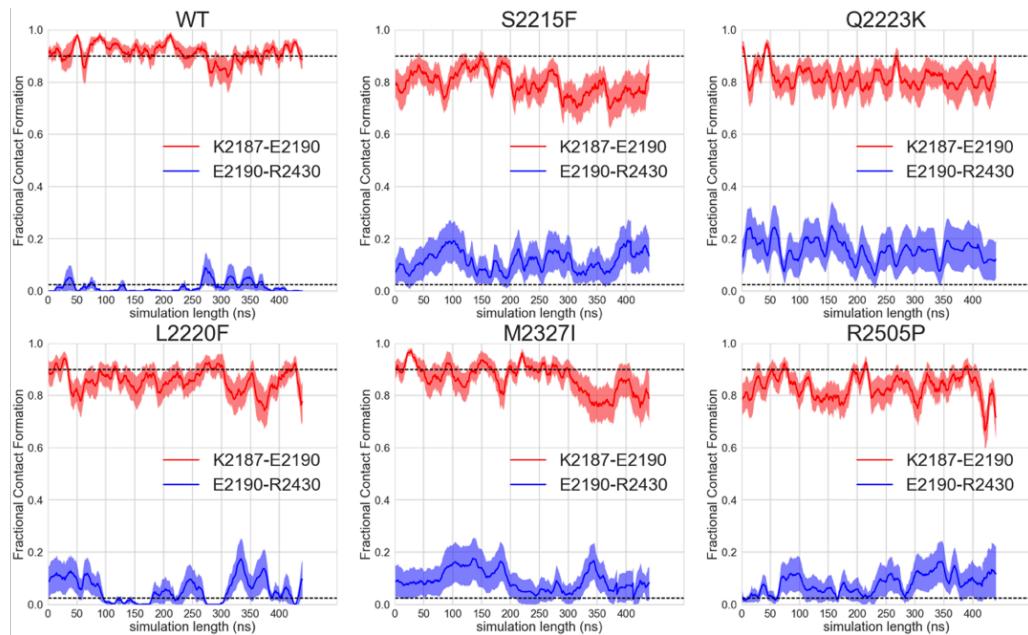
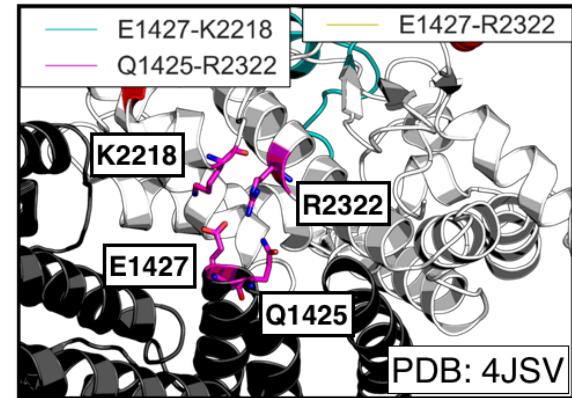
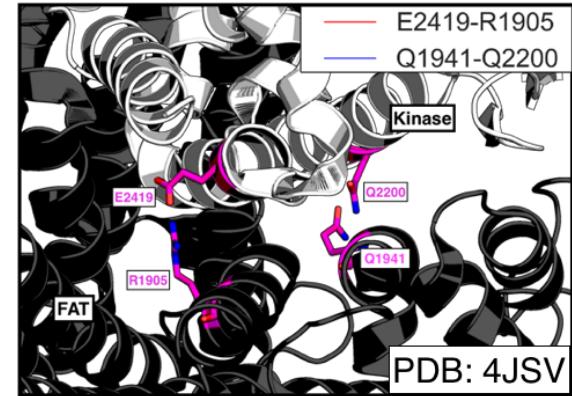
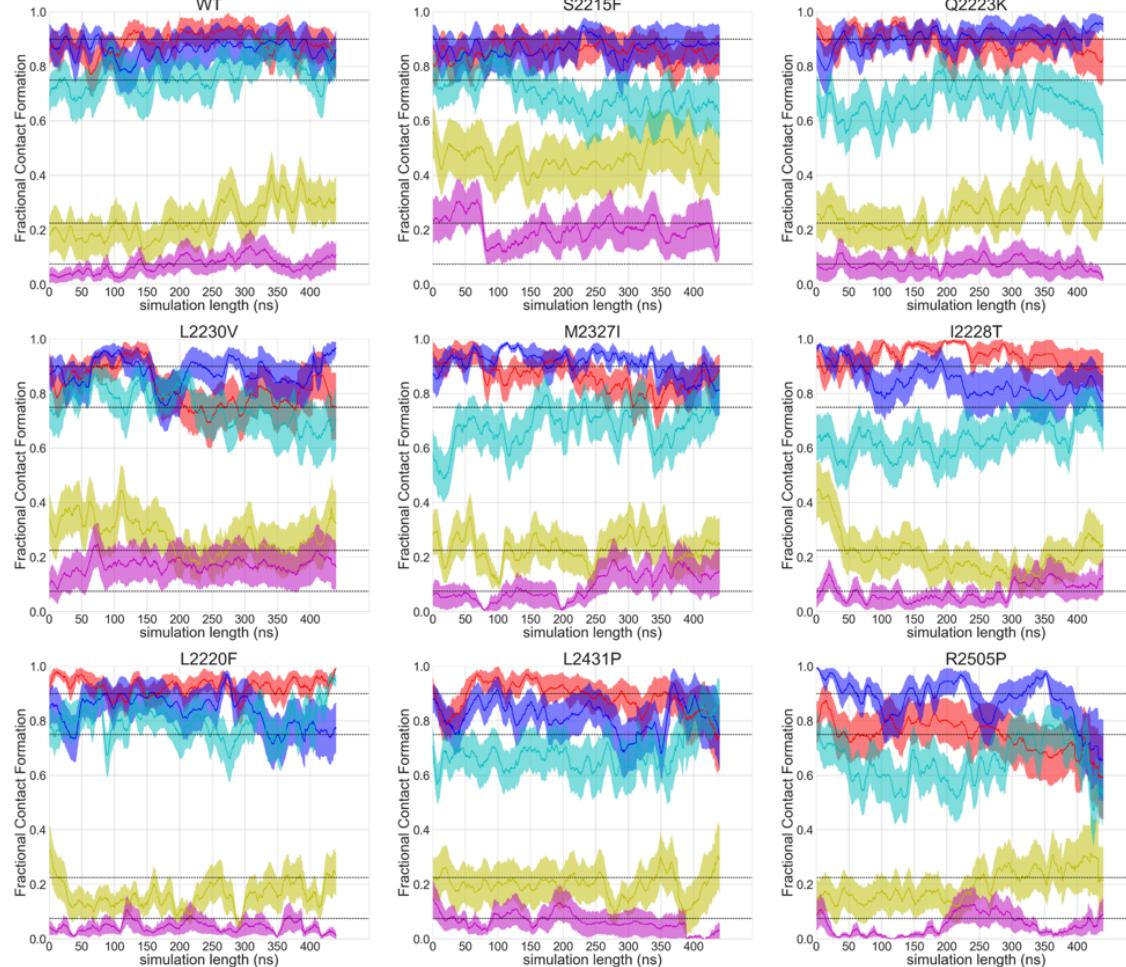


Figure 3.5: Missense mutations do not shift the population of the active conformation on a common activation order parameter  
**Right** Fractional contact formation analysis for 20 500ns trajectories, analyzed in 20ns sliding window chunks. Contacts cutoff at 4 for closest heavy atom. Plotted is the mean  $\pm$  SEM. The dashed lines represent the proportion roughly populated in the WT simulations.  
**Left** Illustration of the bond formed between K2187 and E2190 (red) or E2190 and R2430 (blue). ATP is shown in spheres. The kinase domain (white) is shown with the activation loop (teal) and the  $\alpha$  helices (red). The FRB domain is shown in gold.

### 3.3.3 Missense mutations do not disrupt the formation of inhibitory salt bridges between the kinase and FAT domains

Another hypothesis for how the missense mutations may be hyperactivating looked at the role the FAT domain, which clamps around the kinase domain, plays in regulating the activity of mTOR. In previous work, point mutations that disrupted the formation of the salt bridge between E2419 and R1905 (Figure 3.6, red) activated TOR signaling in yeast and mammalian cell lines [139]. From this observation, we hypothesized that stable contacts formed between the kinase domain and FAT domain have a negative regulatory role on the activity of mTOR, and that disruption of these bonds through some allosteric mechanism would allow an oncogenic mutant to activate mTOR. Another set of contacts, Gln1941 to Gln2200 (Figure 3.6, blue) is conserved across many mTOR orthologs [74], suggesting a potential regulatory role. We also looked at several other salt bridges in the vicinity of these contacts: E1147 and K2218 (cyan), Q1425 and R2322 (magenta), and E1427 and R2322 (yellow). Shown in Figure 3.6 is a representative selection of mutations from the kinase domain, demonstrating that none of the salt bridges showed a prominent change from the wild type to mutant simulations. Of note, R2505P appears to have a disruption in the formation of the Q1941 and Q2200, E1147 and K2218, and E2419 and R1905 contacts, indicating that this may be a possible mechanism of activation for this mutant. However, this is difficult to say with certainty without functional experimental data or additional computation to confirm this result.



**Figure 3.6: Missense mutations in kinase domain do not disrupt interactions between the kinase and FAT domains** **Left** Fractional contact formation analysis for 10 500ns trajectories, analyzed in 20ns sliding window chunks. Contacts cutoff at 4 for closest heavy atom. Plotted is the mean  $\pm$  SEM. The dashed lines represent the proportion roughly populated in the WT simulations. **Right** Illustration of the distances being plotted between E2419 and R1905 (red), Q1941 and 2200 (blue), E1147 and K2218 (cyan), Q1425 and R2322 (magenta), or E1427 and R2322 (yellow). The kinase domain (white) interacts with the FAT domain (black) through salt bridges and hydrogen bonds formed by the highlighted residues (magenta)

### 3.3.4 Free energy calculations show promise in predicting impact of mutations on small molecule and ATP affinity

The preceding work on trying to understand the functional impact of mutations using traditional molecular dynamics highlights one of the key challenges of the application of physical modeling to this question: the need to compute meaningful, physically testable quantities. Alchemical free energy calculations have been previously used to predict the impact of mutations on protein-protein binding [140] and protein thermostabilities [141]. Based on this work, we set out to see if free energy calculations could be used to identify clinically observed mutations that might cause resistance to an ATP-competitive inhibitor. Here, we propose a model of resistance mutants that cause a decrease in affinity for an ATP-competitive inhibitor while maintaining a basic level of ATP affinity (Figure 3.7, left panels). An increase in ATP affinity could be a mechanism of resistance by making it more difficult for ATP-competitive inhibitors to out-compete ATP, as well as a mechanism of activation. To explore the impact of mutations on affinity for these compounds, we built models of mTOR bound to ATP (Figure 3.7, top left) and AZD8055 (Figure 3.7, bottom left). Using protein mutation FEP+ [63, 141], we ran preliminary calculations for a small panel of mutations from the MSKCC-IMPACT assay [142]. Three showed greater than 1-log unit, or 1.4 kcal/mol, decrease in affinity for AZD8055 (Figure 3.7, blue bars). W2239C, a missense mutation observed in a patient with cervical squamous cell carcinoma that occurred together with a known activating mutation S2215F [143, 144], showed the greatest decrease in affinity (shown as a positive  $\Delta\Delta G_{mutation}$  in Figure 3.7). While we were not able to confirm via prediction that ATP-affinity is not maintained, this result suggests that treatment with an ATP-competitive inhibitor targeting mTOR may be less effective than expected. M2345V and Y2225C are also both predicted to de-

crease the affinity for AZD8055. M2345V was observed in a cell line from a patient with cancer of unknown primary tissue and co-occurred with a TSC2 E75K mutation, which is likely to be loss of function [145]. Y2225C was observed in a patient with tubular stomach adenocarcinoma, who also had a slight copy number gain in both MTOR and DEPTOR [146]. All three of these mutations are exceedingly rare and occur in tumors with relatively high mutation burden. These preliminary results suggest that free energy calculations can potentially aid in understanding or teasing apart the functional impact of rare mutations in complex tumors.

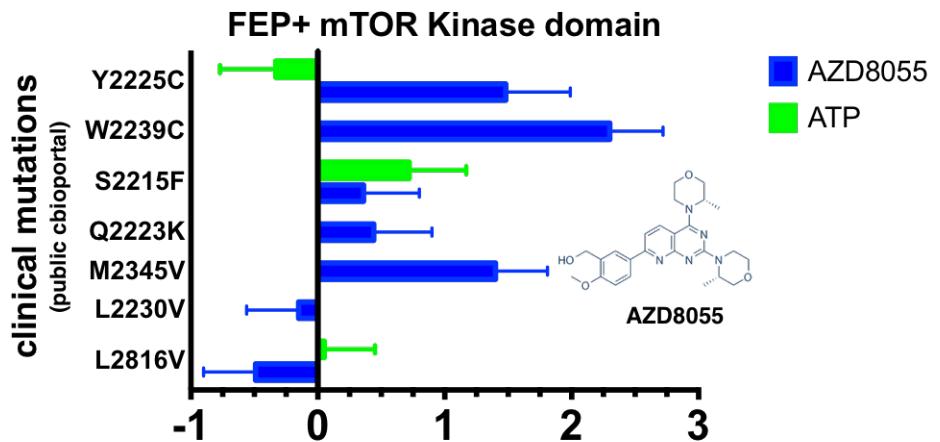
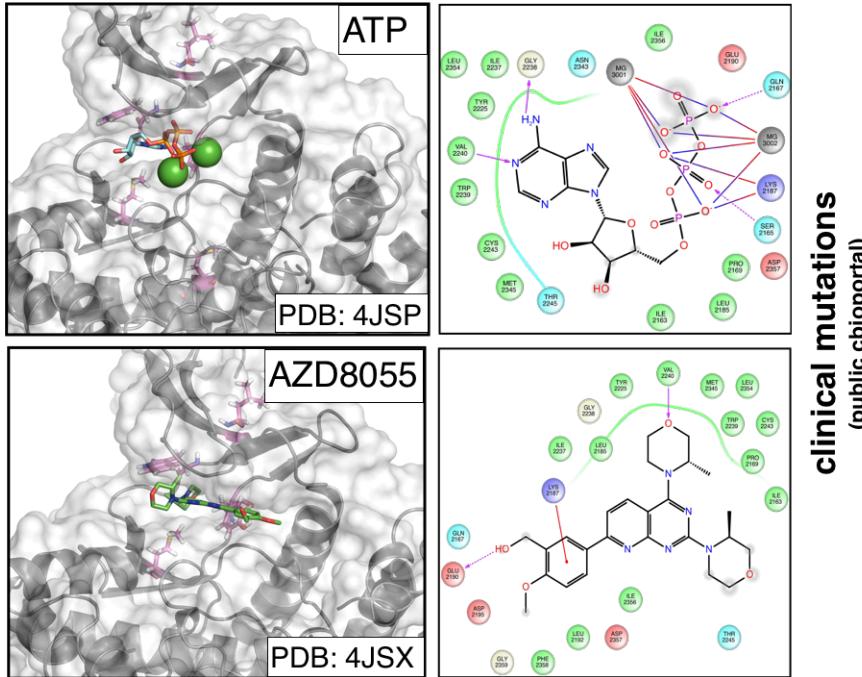


Figure 3.7: Free energy calculations identify potential resistance mutations **Left** Structures and 2D interactions maps of ATP (top, PDBID: 4JSP) and AZD8055 (bottom, PDBID: 4JSX) docked to the kinase domain of mTOR. Magnesium ions are shown as green spheres. **Right**  $\Delta\Delta G_{mutation}$  (x-axis) calculated for a number of clinically observed mTOR mutations(y-axis) [142] for ATP (green) and AZD8055 (blue). Error bars correspond to the BAR uncertainty estimate.

### **3.4 Conclusions**

# CHAPTER 4

## PREDICTING THE IMPACT OF CLINICALLY-OBSERVED MUTATIONS USING PHYSICAL MODELING AND BIOPHYSICAL EXPERIMENTS

### 4.1 Introduction

Targeted kinase inhibitors are a major therapeutic class in the treatment of cancer. A total of 44 selective small molecule kinase inhibitors have now been approved by the FDA [36], including 34 approved to treat cancer, and perhaps 50% of all current drugs in development target kinases [37]. Despite the success of selective inhibitors, the emergence of drug resistance remains a challenge in the treatment of cancer [147–154] and has motivated the development of second- and then third-generation inhibitors aimed at overcoming recurrent resistance mutations [30, 155–158].

While a number of drug resistance mechanisms have been identified in cancer (e.g., induction of splice variants [159], or alleviation of feedback [27]), inherent or acquired missense mutations in the kinase domain of the target of therapy are a major form of resistance to tyrosine kinase inhibitors (TKI) [18, 154, 160]. Oncology is entering a new era with major cancer centers now deep sequencing tumors to reveal genetic alterations that may render subclonal populations susceptible or resistant to targeted inhibitors [142], but the use of this information in precision medicine has lagged behind. It would be of enormous value in clinical practice if an oncologist could reliably ascertain whether these mutations render the target of therapy resistant or susceptible to available inhibitors; such tools would facilitate the enrollment of patients in mechanism-based basket trials [161, 162], help prioritize candidate compounds for clinical trials, and aid the development

of next-generation inhibitors.

### 4.1.1 The long tail of rare kinase mutations frustrates prediction of drug resistance

While some cancer missense mutations are highly recurrent and have been characterized clinically or biochemically, a long tail of rare mutations collectively accounts for the majority of clinically observed missense mutations (Figure ??a), leaving clinicians and researchers without knowledge of whether these uncharacterized mutations might lead to resistance. While rules-based and machine learning schemes are still being assessed in oncology contexts, work in predicting drug response to microbial resistance has shown that rare mutations present a significant challenge to approaches that seek to predict resistance to therapy [163]. Clinical cancer mutations may impact drug response through a variety of mechanisms by altering kinase activity, ATP affinity, substrate specificities, and the ability to participate in regulatory interactions, compounding the difficulties associated with limited datasets that machine learning approaches face. In parallel with computational approaches, high-throughput experimental techniques such as MITE-Seq [164] have been developed to assess the impact of point mutations on drug response. However, the complexity of defining selection schemes that reliably correlate with *in vivo* drug effectiveness and long turn-around times might limit their ability to rapidly and reliably impact clinical decision-making.

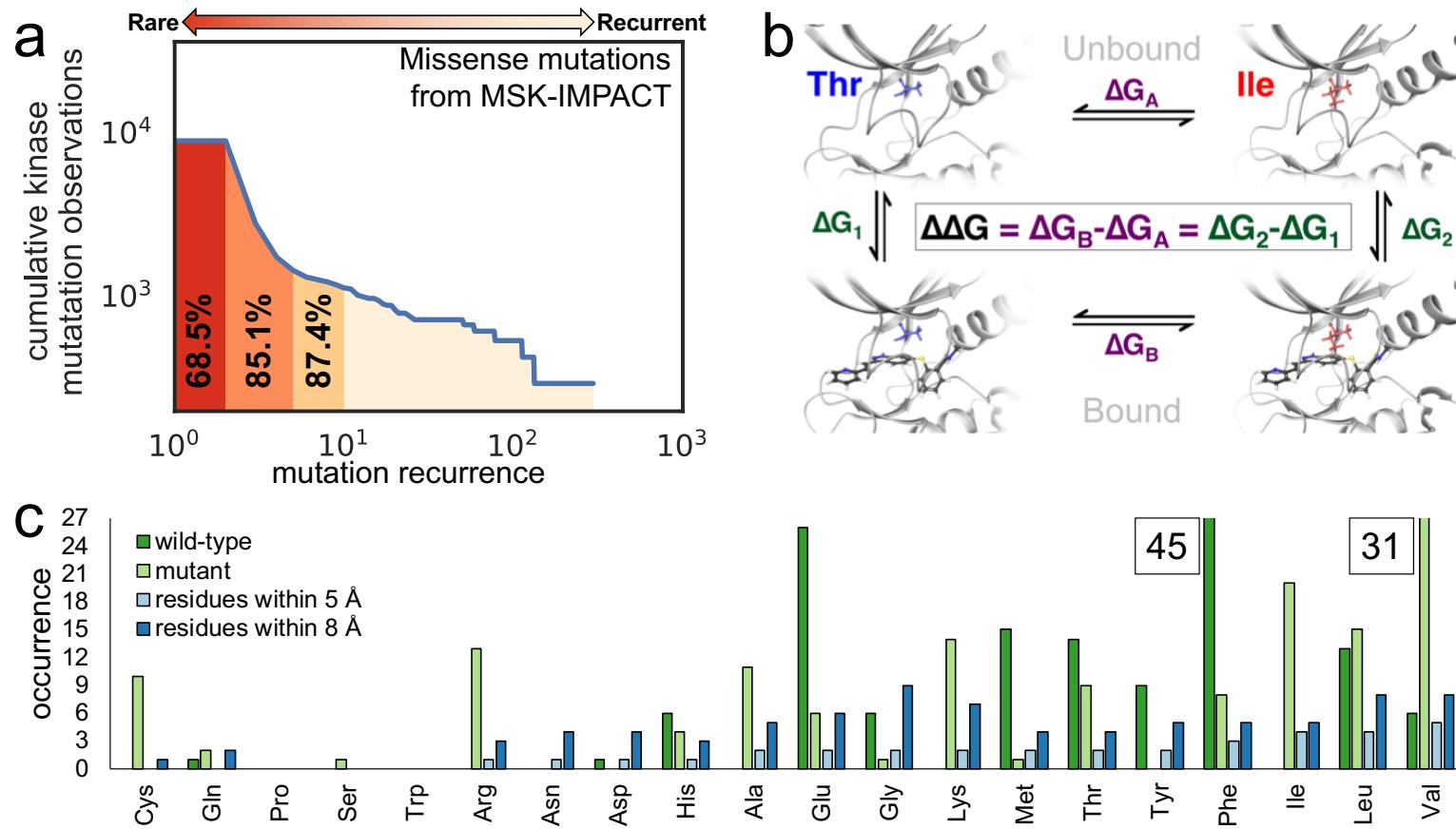


Figure 4.1: Relative alchemical free-energy calculations can be used to predict affinity changes of FDA-approved selective kinase inhibitors arising from clinically-identified mutations in their targets of therapy. (a) Missense mutation statistics derived from 10,336 patient samples subjected to Memorial Sloan Kettering-Integrated Mutation Profiling of Actionable Cancer Targets (MSK-IMPACT) deep sequencing panel [142] show that 68.5% of missense kinase mutations in cancer patients have never been observed previously, while 87.4% have been observed no more than ten times; the vast majority of clinically observed missense kinase mutations are unique to each patient. (b) To compute the impact of a clinical point mutation on inhibitor binding free energy, a thermodynamic cycle can be used to relate the free energy of the wild-type and mutant kinase in the absence (top) and presence (bottom) of the inhibitor. (c) Summary of mutations studied in this work. Frequency of the wild-type (dark green) and mutant (green) residues for the 144 clinically-identified Abl mutations used in this study (see [Table 4.1](#) for data sources). Also shown is the frequency of residues within 5 (light blue) and 8 (blue) of the binding pocket. The ordering of residues along the x-axis corresponds to the increasing occurrence of residues within 5 of the binding pocket. The number of wild-type Phe residues (n=45) and mutant Val residues (n=31) exceeded the limits of the y-axis.

## 4.1.2 Alchemical free-energy methods can predict inhibitor binding affinities

Physics-based approaches could be complementary to machine-learning and experimental techniques in predicting changes in TKI affinity due to mutations with few or no prior clinical observations. Modern atomistic molecular mechanics forcefields such as OPLS3 [4], CHARMM [2], and AMBER FF14SB [3] have reached a sufficient level of maturity to enable the accurate and reliable prediction of receptor-ligand binding free energy. Alchemical free-energy methods permit receptor-ligand binding energies to be computed rigorously, including all relevant entropic and enthalpic contributions [121]. Encouragingly, kinase:inhibitor binding affinities have been predicted using alchemical free-energy methods with mean unsigned errors of  $1.0 \text{ kcal mol}^{-1}$  for CDK2, JNK1, p38, and Tyk2 [12, 165]. Beyond kinases, alchemical approaches have predicted the binding affinity of BRD4 inhibitors with mean absolute errors of  $0.6 \text{ kcal mol}^{-1}$  [166]. Alchemical methods have also been observed to have good accuracy ( $0.6 \text{ kcal mol}^{-1}$  mean unsigned error for Tyk2 tyrosine kinase) in the prediction of relative free energies for ligand transformations within a complex whose receptor geometry was generated using a homology model [167].

## 4.1.3 Alchemical approaches can predict the impact of protein mutations on free energy

Alchemical free-energy calculations have also been used to predict the impact of mutations on protein-protein binding [140] and protein thermostabilities [141]. Recent work has found that protein mutations can be predicted to be stabilizing or

destabilizing with a classification accuracy of 71% across ten proteins and 62 mutations [168]. The impact of Gly to D-Ala mutations on protein stability was predicted using an alchemical approach with a similar level of accuracy [169]. Recently, one study has hinted at the potential utility of alchemical free-energy calculations in oncology by predicting the impact of a single clinical mutation on the binding free energies of the TKIs dasatinib and RL45 [170].

#### 4.1.4 Assessing the potential for physical modeling to predict resistance to FDA-approved TKIs

Here, we ask whether physical modeling techniques may be useful in predicting whether clinically-identified kinase mutations lead to drug resistance or drug sensitivity. We perform state-of-the-art relative alchemical free-energy calculations using FEP+ [12], recently demonstrated to achieve sufficiently good accuracy to drive the design of small-molecule inhibitors for a broad range of targets during lead optimization [12, 121, 165, 171], to calculate the effect of point mutation on the binding free energy between the inhibitor and the kinase receptor (Figure 4.1b). We compare this approach against a fast but approximate physical modeling method implemented in Prime [172] (an MM-GBSA approach) in which an implicit solvent model is used to assess the change in minimized interaction energy of the ligand with the mutant and wild-type kinase. We consider whether these methods can predict a ten-fold reduction in inhibitor affinity (corresponding to a binding free energy change of  $1.36 \text{ kcal mol}^{-1}$ ) to assess baseline utility. As a benchmark, we compile a set of reliable inhibitor  $\Delta\text{pIC}_{50}$  data for 144 clinically-identified mutants of the human kinase Abl, an important oncology target dysregulated in cancers like chronic myelogenous leukemia (CML), for which six [36] FDA-approved TKIs

are available. While  $\Delta pIC_{50}$  can approximate a dissociation constant  $\Delta K_D$ , other processes contributing to changes in cell viability might affect  $IC_{50}$  in ways that are not accounted for by a traditional binding experiment, motivating a quantitative comparison between  $\Delta pIC_{50}$  and  $\Delta K_D$ . The results of this benchmark demonstrate the potential for FEP+ to predict the impact that mutations in Abl kinase have on drug binding, and a classification accuracy of  $88^{93}_{82}\%$  (for all statistical metrics reported in this paper, the 95% confidence intervals (CI) is shown in the form of  $(x_{lower}^{upper})$ ), an RMSE of  $1.07^{1.26}_{0.89}$  kcal mol $^{-1}$ , and an MUE of  $0.79^{0.92}_{0.67}$  kcal mol $^{-1}$  was achieved.

#### 4.1.5 A benchmark of $\Delta pIC_{50}$ s for predicting mutational resistance

To construct a benchmark evaluation dataset, we compiled a total of 144  $\Delta pIC_{50}$  measurements of Abl:TKI affinities, summarized in Table 4.1, taking care to ensure all measurements for an individual TKI were reported in the same study from experiments run under identical conditions. 131  $\Delta pIC_{50}$  measurements were available across the six TKIs with available co-crystal structures with wild-type Abl—26 for axitinib and 21 for bosutinib, dasatinib, imatinib, nilotinib, and ponatinib. 13  $\Delta pIC_{50}$  measurements were available for the two TKIs for which docking was necessary to generate Abl:TKI structures—7 for erlotinib and 6 for gefitinib. For added diversity, this set includes TKIs for which Abl is not the primary target—axitinib, erlotinib, and gefitinib. All mutations in this benchmark dataset have been clinically-observed (Table 4.1). Due to the change in bond topology required by mutations involving proline, which is not currently supported by the FEP+ technology for protein residue mutations, the three mutations H396P (axitinib,

gefitinib, erlotinib) were excluded from our assessment. As single point mutations were highly represented in the Memorial Sloan Kettering-Integrated Mutation Profiling of Actionable Cancer Targets (MSK-IMPACT) study analyzed in Figure 4.1a, we excluded double mutations from this work. However, the impact of mutations from multiple sites can potentially be modeled by sequentially mutating each site and this will be addressed in future work.

Table 4.1: Public  $\Delta\text{pIC}_{50}$  datasets for 144 Abl kinase mutations and eight tyrosine kinase inhibitors (TKIs) with corresponding wild-type co-crystal structures used in this study

TKI	$N_{\text{mut}}$			PDB	$ \Delta G_{\text{max}} - \Delta G_{\text{min}} $	(kcal mol <sup>-1</sup> ) Source	(kcal mol <sup>-1</sup> ) $\Delta G_{\text{WT}}$
		R	S				
axitinib	26	0	26	4wa9	2.05	[173]	-8.35
bosutinib	21	4	17	3ue4	2.79	[174]	-9.81
dasatinib	21	5	16	4xey	5.08	[174]	-11.94
imatinib	21	5	16	1opj	2.16	[174]	-9.19
nilotinib	21	4	17	3cs9	3.88	[174]	-10.74
ponatinib	21	0	21	3oxz	1.00	[174]	-11.70
<b>subtotal</b>	<b>131</b>	<b>18</b>	<b>113</b>				
erlotinib	7	1	6	Dock 3ue4	1.73	[48]	-9.77
gefitinib	6	0	6	Dock 3ue4	1.79	[48]	-8.84
<b>total</b>	<b>144</b>	<b>19</b>	<b>125</b>				

$N_{\text{mut}}$ : Total number of mutants for which  $\Delta\text{pIC}_{50}$  data was available.

Number of Resistant, Susceptible mutants using 10-fold affinity change threshold.

PDB: Source PDB ID, or Dock to 3ue4, which used 3ue4 as the receptor for Glide-SP docking inhibitors without co-crystal structure.

$\Delta G_{\text{WT}}$ : Binding free energy of inhibitor to wild-type Abl, as estimated from IC<sub>50</sub> data.

Experimental  $\Delta pIC_{50}$  measurements for wild-type and mutant Abl were converted to  $\Delta\Delta G$  in order to make direct comparisons between physics-based models and experiment. However, computation of experimental uncertainties were required to understand the degree to which differences between predictions and experimental data were significant. Since experimental error estimates for measured  $IC_{50}$ s were not available for the data in [Table 4.1](#), we compared that data to other sources that have published  $IC_{50}$ s for the same mutations in the presence of the same TKIs ([Figure 2.2a,b,c](#)). Cross-comparison of 97 experimentally measured  $\Delta\Delta G$ s derived from cell viability assay  $IC_{50}$  data led to an estimate of experimental variability of  $0.32_{0.28}^{0.36}$  kcal mol $^{-1}$  root-mean square error (RMSE) that described the expected repeatability of the measurements. Because multiple factors influence the  $IC_{50}$  aside from direct effects on the binding affinity—the focus of this study—we also compared  $\Delta\Delta G$ s derived from  $\Delta pIC_{50}$ s with those derived from binding affinity measurements ( $\Delta K_d$ ) for which data for a limited set of 27 mutations was available (Figure ??d); the larger computed RMSE of  $0.81_{0.59}^{1.04}$  kcal mol $^{-1}$  represents an estimate of the lower bound of the RMSE to the  $IC_{50}$ -derived  $\Delta\Delta G$ s that we might hope to achieve with FEP+ or Prime, which were performed using non-phosphorylated models, when comparing sample statistics directly. In comparing 31 mutations for which phosphorylated and non-phosphorylated  $\Delta K_d$ s were available, we found a strong correlation between the  $\Delta\Delta G$ s derived from those data ( $r=0.94$ , Supplementary Figure ??); the statistics of that comparison are similar to those of the inter-lab variability comparison.

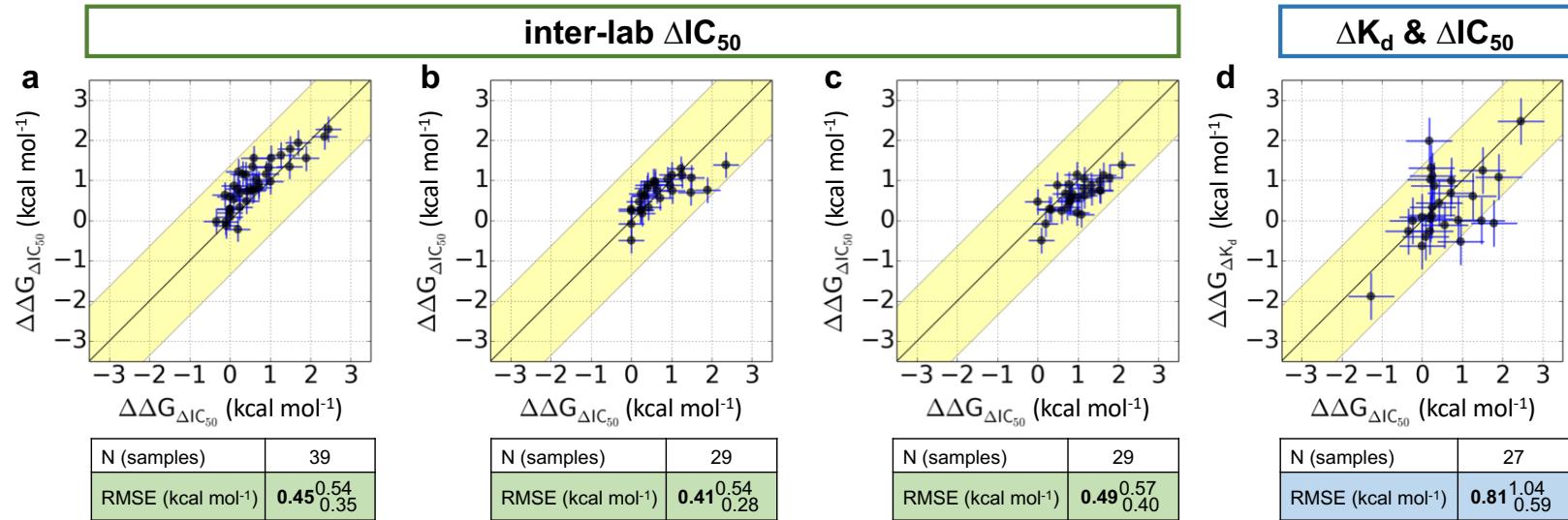


Figure 4.2: **Cross-comparison of the experimentally measured effects that mutations in Abl kinase have on ligand binding, performed by different labs.**  $\Delta\text{G}$  was computed from publicly available  $\Delta\text{pIC}_{50}$  or  $\Delta\text{p}K_d$  measurements and these values of  $\Delta\Delta\text{G}$  were then plotted and the RMSE between them reported. (a)  $\Delta\text{pIC}_{50}$  measurements (X-axis) from [174] compared with  $\Delta\text{pIC}_{50}$  measurements (Y-axis) from [175]. (b)  $\Delta\text{pIC}_{50}$  measurements (X-axis) from [174] compared with  $\Delta\text{pIC}_{50}$  measurements (Y-axis) from [176]. (c)  $\Delta\text{pIC}_{50}$  measurements (X-axis) from [175] compared with  $\Delta\text{pIC}_{50}$  measurements (Y-axis) from [176]. (d)  $\Delta\text{pIC}_{50}$  measurements (X-axis) from [174] compared with  $\Delta\text{p}K_d$  measurements (Y-axis) from [48] using non-phosphorylated Abl kinase. Scatter plot error bars in a,b, and c are  $\pm$ standard error (SE) taken from the combined 97 inter-lab  $\Delta\Delta\text{Gs}$  derived from the  $\Delta\text{pIC}_{50}$  measurements, which was  $0.32_{0.28}^{0.36}$ ; the RMSE was  $0.45_{0.39}^{0.51}$  kcal mol<sup>-1</sup>. Scatter plot error bars in d are the  $\pm$ standard error (SE) of  $\Delta\Delta\text{Gs}$  derived from  $\Delta\text{pIC}_{50}$  and  $\Delta\text{p}K_d$  from a set of 27 mutations, which is  $0.58_{0.42}^{0.74}$  kcal mol<sup>-1</sup>; the RMSE was  $0.81_{0.59}^{1.04}$  kcal mol<sup>-1</sup>.

**4.1.6 Summary of experimental data that is available and what  
needs to be done to make better information available**

**4.1.7 Open-source YANK/Amber forcefield stuff**

## BIBLIOGRAPHY

- [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](https://doi.org/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](https://doi.org/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](https://doi.org/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](https://doi.org/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, Pratillas 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](https://doi.org/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. EGFR 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.drup.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, Heinzelmeir 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, Grernrum 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. *Cell Reports*. 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 Isootype 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 61):899–907.
- [56] Sastry GM, Adzhigirey M, Day T, Annabhimmoju 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.
- [59] Salvatier J, Wiecki TV, Fonnesbeck C. Probabilistic programming in Python using PyMC3. *PeerJ Computer Science*. 2016; 2:e55.
- [60] Al-Rfou R, Alain G, Almahairi A, Angermueller C, Bahdanau D, Ballas N, Bastien F, Bayer J, Belikov A, Belopolsky A, Bengio Y, Bergeron A, Bergstra J, Bisson V, Bleecher Snyder J, Bouchard N, Boulanger-Lewandowski N, Bouthillier X, de Brébisson A, Breuleux O, et al. Theano: A Python framework for fast computation of mathematical expressions. *arXiv e-prints*. 2016 May; abs/1605.02688. <http://arxiv.org/abs/1605.02688>.
- [61] Bosc N, Meyer C, Bonnet P. The use of novel selectivity metrics in kinase research. *BMC bioinformatics*. 2017 Jan; 18(1):17.
- [62] Cheng AC, Eksterowicz J, Geuns-Meyer S, Sun Y. Analysis of kinase inhibitor selectivity using a thermodynamics-based partition index. *J Med Chem*. 2010 Jun; 53(11):4502–4510.
- [63] Hauser K, Negron C, Albanese SK, Ray S, Steinbrecher T, Abel R, Chodera JD, Wang L. Predicting resistance of clinical Abl mutations to targeted kinase inhibitors using alchemical free-energy calculations. *Communications Biology*. 2018 Jun; 1(1):70.
- [64] Hari SB, Merritt EA, Maly DJ. Sequence determinants of a specific inactive protein kinase conformation. *Chemistry & biology*. 2013 Jun; 20(6):806–815.
- [65] Hu J, Ahuja LG, Meharena HS, Kannan N, Kornev AP, Taylor SS, Shaw AS. Kinase regulation by hydrophobic spine assembly in cancer. *Molecular and cellular biology*. 2015 Jan; 35(1):264–276.
- [66] Laplante M, Sabatini DM. mTOR signaling in growth control and disease. *Cell*. 2012 Apr; 149(2):274–293.
- [67] Saxton RA, Sabatini DM. mTOR Signaling in Growth, Metabolism, and Disease. *Cell*. 2017 Mar; 168(6):960–976.
- [68] Kim DH, Sarbassov DD, Ali SM, King JE, Latek RR, Erdjument-Bromage H, Tempst P, Sabatini DM. mTOR interacts with raptor to form a nutrient-sensitive complex that signals to the cell growth machinery. *Cell*. 2002 Jul; 110(2):163–175.

- [69] Hara K, Maruki Y, Long X, Yoshino Ki, Oshiro N, Hidayat S, Tokunaga C, Avruch J, Yonezawa K. Raptor, a binding partner of target of rapamycin (TOR), mediates TOR action. *Cell*. 2002 Jul; 110(2):177–189.
- [70] Yang H, Jiang X, Li B, Yang HJ, Miller M, Yang A, Dhar A, Pavletich NP. Mechanisms of mTORC1 activation by RHEB and inhibition by PRAS40. *Nature*. 2017 Dec; 552(7685):368–373.
- [71] Sarbassov DD, Ali SM, Kim DH, Guertin DA, Latek RR, Erdjument-Bromage H, Tempst P, Sabatini DM. Rictor, a novel binding partner of mTOR, defines a rapamycin-insensitive and raptor-independent pathway that regulates the cytoskeleton. *Current biology : CB*. 2004 Jul; 14(14):1296–1302.
- [72] Bar-Peled L, Sabatini DM. SnapShot: mTORC1 signaling at the lysosomal surface. *Cell*. 2012 Dec; 151(6):1390–1390.e1.
- [73] Peterson TR, Laplante M, Thoreen CC, Sancak Y, Kang SA, Kuehl WM, Gray NS, Sabatini DM. DEPTOR is an mTOR inhibitor frequently overexpressed in multiple myeloma cells and required for their survival. *Cell*. 2009 May; 137(5):873–886.
- [74] Yang H, Rudge DG, Koos JD, Vaidalingam B, Yang HJ, Pavletich NP. mTOR kinase structure, mechanism and regulation. *Nature*. 2013 May; 497(7448):217–223.
- [75] Gaubitz C, Oliveira TM, Prouteau M, Leitner A, Karuppasamy M, Konstantinidou G, Rispal D, Eltschinger S, Robinson GC, Thore S, Aebersold R, Schaffitzel C, Loe with R. Molecular Basis of the Rapamycin Insensitivity of Target Of Rapamycin Complex 2. *Molecular cell*. 2015 Jun; 58(6):977–988.
- [76] Aylett CHS, Sauer E, Imseng S, Boehringer D, Hall MN, Ban N, Maier T. Architecture of human mTOR complex 1. *Science*. 2016 Jan; 351(6268):48–52.
- [77] Inoki K, Li Y, Zhu T, Wu J, Guan KL. TSC2 is phosphorylated and inhibited by Akt and suppresses mTOR signalling. *Nature cell biology*. 2002 Sep; 4(9):648–657.
- [78] Manning BD, Tee AR, Logsdon MN, Blenis J, Cantley LC. Identification of the tuberous sclerosis complex-2 tumor suppressor gene product tuberin as a target of the phosphoinositide 3-kinase/akt pathway. *Molecular cell*. 2002 Jul; 10(1):151–162.
- [79] Jones RG, Plas DR, Kubek S, Buzzai M, Mu J, Xu Y, Birnbaum MJ, Thompson CB. AMP-activated protein kinase induces a p53-dependent metabolic checkpoint. *Molecular cell*. 2005 Apr; 18(3):283–293.
- [80] Inoki K, Li Y, Xu T, Guan KL. Rheb GTPase is a direct target of TSC2 GAP activity and regulates mTOR signaling. *Genes & development*. 2003 Aug; 17(15):1829–1834.
- [81] Bar-Peled L, Schweitzer LD, Zoncu R, Sabatini DM. Ragulator is a GEF for the rag GTPases that signal amino acid levels to mTORC1. *Cell*. 2012 Sep; 150(6):1196–1208.
- [82] Kim E, Goraksha-Hicks P, Li L, Neufeld TP, Guan KL. Regulation of TORC1 by Rag GTPases in nutrient response. *Nature cell biology*. 2008 Aug; 10(8):935–945.
- [83] Sancak Y, Bar-Peled L, Zoncu R, Markhard AL, Nada S, Sabatini DM. Ragulator-Rag complex targets mTORC1 to the lysosomal surface and is necessary for its activation by amino acids. *Cell*. 2010 Apr; 141(2):290–303.
- [84] Efeyan A, Zoncu R, Sabatini DM. Amino acids and mTORC1: from lysosomes to disease. *Trends in molecular medicine*. 2012 Sep; 18(9):524–533.

- [85] Hay N, Sonenberg N. Upstream and downstream of mTOR. *Genes & development*. 2004 Aug; 18(16):1926–1945.
- [86] Lamming DW, Sabatini DM. A Central role for mTOR in lipid homeostasis. *Cell metabolism*. 2013 Oct; 18(4):465–469.
- [87] Delarue M, Brittingham GP, Pfeffer S, Surovtsev IV, Pinglay S, Kennedy KJ, Schaffer M, Gutierrez JL, Sang D, Poterewicz G, Chung JK, Plitzko JM, Groves JT, Jacobs-Wagner C, Engel BD, Holt LJ. mTORC1 Controls Phase Separation and the Biophysical Properties of the Cytoplasm by Tuning Crowding. *Cell*. 2018 Jun; .
- [88] Ballou LM, Lin RZ. Rapamycin and mTOR kinase inhibitors. *Journal of chemical biology*. 2008 Nov; 1(1-4):27–36.
- [89] Lamming DW, Ye L, Sabatini DM, Baur JA. Rapalogs and mTOR inhibitors as anti-aging therapeutics. *The Journal of clinical investigation*. 2013 Mar; 123(3):980–989.
- [90] Hudes G, Carducci M, Tomczak P, Dutcher J, Figlin R, Kapoor A, Staroslawska E, Sosman J, McDermott D, Bodrogi I, Kovacevic Ž, Lesovoy V, Schmidt-Wolf IGH, Barbarash O, Gokmen E, O'Toole T, Lustgarten S, Moore L, Motzer RJ, Global ARCC Trial. Temsirolimus, interferon alfa, or both for advanced renal-cell carcinoma. *The New England journal of medicine*. 2007 May; 356(22):2271–2281.
- [91] Motzer RJ, Escudier B, Oudard S, Hutson TE, Porta C, Bracarda S, Grünwald V, Thompson JA, Figlin RA, Hollaender N, Urbanowitz G, Berg WJ, Kay A, Lebwohl D, Ravaud A, RECORD-1 Study Group. Efficacy of everolimus in advanced renal cell carcinoma: a double-blind, randomised, placebo-controlled phase III trial. *Lancet (London, England)*. 2008 Aug; 372(9637):449–456.
- [92] Hausch F, Kozany C, Theodoropoulou M, Fabian AK. FKBP<sub>s</sub> and the Akt/mTOR pathway. *Cell cycle (Georgetown, Tex)*. 2013 Aug; 12(15):2366–2370.
- [93] Yip CK, Murata K, Walz T, Sabatini DM, Kang SA. Structure of the human mTOR complex I and its implications for rapamycin inhibition. *Molecular cell*. 2010 Jun; 38(5):768–774.
- [94] Schreiber KH, Ortiz D, Academia EC, Anies AC, Liao CY, Kennedy BK. Rapamycin-mediated mTORC2 inhibition is determined by the relative expression of FK506-binding proteins. *Aging cell*. 2015 Apr; 14(2):265–273.
- [95] Slotkin EK, Patwardhan PP, Vasudeva SD, de Stanchina E, Tap WD, Schwartz GK. MLN0128, an ATP-competitive mTOR kinase inhibitor with potent in vitro and in vivo antitumor activity, as potential therapy for bone and soft-tissue sarcoma. *Molecular cancer therapeutics*. 2015 Feb; 14(2):395–406.
- [96] Hassan B, Akcakanat A, Sangai T, Evans KW, Adkins F, Eterovic AK, Zhao H, Chen K, Chen H, Do KA, Xie SM, Holder AM, Naing A, Mills GB, Meric-Bernstam F. Catalytic mTOR inhibitors can overcome intrinsic and acquired resistance to allosteric mTOR inhibitors. *Oncotarget*. 2014 Sep; 5(18):8544–8557.
- [97] García-García C, Ibrahim YH, Serra V, Calvo MT, Guzmán M, Grueso J, Aura C, Pérez J, Jessen K, Liu Y, Rommel C, Taberner J, Baselga J, Scaltriti M. Dual mTORC1/2 and HER2 blockade results in antitumor activity in preclinical models of breast cancer resistant to anti-HER2 therapy. *Clinical cancer research : an official journal of the American Association for Cancer Research*. 2012 May; 18(9):2603–2612.

- [98] Chresta CM, Davies BR, Hickson I, Harding T, Cosulich S, Critchlow SE, Vincent JP, Ellston R, Jones D, Sini P, James D, Howard Z, Dudley P, Hughes G, Smith L, Maguire S, Hummersone M, Malagu K, Menear K, Jenkins R, et al. AZD8055 is a potent, selective, and orally bioavailable ATP-competitive mammalian target of rapamycin kinase inhibitor with *in vitro* and *in vivo* antitumor activity. *Cancer research*. 2010 Jan; 70(1):288–298.
- [99] Mallon R, Feldberg LR, Lucas J, Chaudhary I, Dehnhardt C, Santos ED, Chen Z, dos Santos O, Ayral-Kaloustian S, Venkatesan A, Hollander I. Antitumor efficacy of PKI-587, a highly potent dual PI3K/mTOR kinase inhibitor. *Clinical cancer research : an official journal of the American Association for Cancer Research*. 2011 May; 17(10):3193–3203.
- [100] Mukherjee B, Tomimatsu N, Amacherla K, Camacho CV, Pichamoorthy N, Burma S. The Dual PI3K/mTOR Inhibitor NVP-BEZ235 Is a Potent Inhibitor of ATM- and DNA-PKCs-Mediated DNA Damage Responses. *Neoplasia*. 2012 Jan; 14(1):34–IN8.
- [101] Pongas G, Fojo T. BEZ235: When Promising Science Meets Clinical Reality. *The oncologist*. 2016 Sep; 21(9):1033–1034.
- [102] Eichhorn PJA, Gili M, Scaltriti M, Serra V, Guzman M, Nijkamp W, Beijersbergen RL, Valero V, Seoane J, Bernards R, Baselga J. Phosphatidylinositol 3-kinase hyperactivation results in lapatinib resistance that is reversed by the mTOR/phosphatidylinositol 3-kinase inhibitor NVP-BEZ235. *Cancer research*. 2008 Nov; 68(22):9221–9230.
- [103] Rodrik-Outmezguine VS, Chandarlapaty S, Pagano NC, Poulikakos PI, Scaltriti M, Moskatek E, Baselga J, Guichard S, Rosen N. mTOR kinase inhibition causes feedback-dependent biphasic regulation of AKT signaling. *Cancer discovery*. 2011 Aug; 1(3):248–259.
- [104] Fan QW, Knight ZA, Goldenberg DD, Yu W, Mostov KE, Stokoe D, Shokat KM, Weiss WA. A dual PI3 kinase/mTOR inhibitor reveals emergent efficacy in glioma. *Cancer Cell*. 2006 May; 9(5):341–349.
- [105] Wagle N, Grabiner BC, Van Allen EM, Hodis E, Jacobus S, Supko JG, Stewart M, Choueiri TK, Gandhi L, Cleary JM, Elfiky AA, Taplin ME, Stack EC, Signoretti S, Loda M, Shapiro GI, Sabatini DM, Lander ES, Gabriel SB, Kantoff PW, et al. Activating mTOR mutations in a patient with an extraordinary response on a phase I trial of everolimus and pazopanib. *Cancer discovery*. 2014 May; 4(5):546–553.
- [106] Wagle Nikhil, Grabiner Brian C, Van Allen Eliezer M, Amin-Mansour Ali, Taylor-Weiner Amaro, Rosenberg Mara, Gray Nathanael, Barletta Justine A, Guo Yanan, Swanson Scott J, Ruan Daniel T, Hanna Glenn J, Haddad Robert I, Getz Gad, Kwiatkowski David J, Carter Scott L, Sabatini David M, Jänne Pasi A, Garraway Levi A, Lorch Jochen H. Response and acquired resistance to everolimus in anaplastic thyroid cancer. *The New England journal of medicine*. 2014 Oct; 371(15):1426–1433.
- [107] Rodrik-Outmezguine VS, Okaniwa M, Yao Z, Novotny CJ, McWhirter C, Banaji A, Won H, Wong W, Berger M, de Stanchina E, Barratt DG, Cosulich S, Klinowska T, Rosen N, Shokat KM. Overcoming mTOR resistance mutations with a new-generation mTOR inhibitor. *Nature*. 2016 Jun; 534(7606):272–276.
- [108] Fan Q, Aksoy O, Wong RA, IlkhaniZadeh S, Novotny CJ, Gustafson WC, Truong AYQ, Cayanan G, Simonds EF, Haas-Kogan D, Phillips JJ, Nicolaides T, Okaniwa M,

- Shokat KM, Weiss WA. A Kinase Inhibitor Targeted to mTORC1 Drives Regression in Glioblastoma. *Cancer Cell*. 2017 Mar; 31(3):424–435.
- [109] Guertin DA, Sabatini DM. Defining the role of mTOR in cancer. *Cancer Cell*. 2007 Jul; 12(1):9–22.
- [110] Grabiner BC, Nardi V, Birsoy K, Possemato R, Shen K, Sinha S, Jordan A, Beck AH, Sabatini DM. A diverse array of cancer-associated MTOR mutations are hyperactivating and can predict rapamycin sensitivity. *Cancer discovery*. 2014 May; 4(5):554–563.
- [111] Cancer Genome Atlas Research Network. Comprehensive molecular characterization of clear cell renal cell carcinoma. *Nature*. 2013 Jul; 499(7456):43–49.
- [112] Xu J, Pham CG, Albanese SK, Dong Y, Oyama T, Lee CH, Rodrik-Outmezguine V, Yao Z, Han S, Chen D, Parton DL, Chodera JD, Rosen N, Cheng EH, Hsieh JJ. Mechanistically distinct cancer-associated mTOR activation clusters predict sensitivity to rapamycin. *The Journal of clinical investigation*. 2016 Sep; 126(9):3526–3540.
- [113] Salsbury FR. Molecular dynamics simulations of protein dynamics and their relevance to drug discovery. *Current opinion in pharmacology*. 2010 Dec; 10(6):738–744.
- [114] Shirts, M, Pande, V S. COMPUTING: Screen Savers of the World Unite! *Science*. 2000 Dec; 290(5498):1903–1904.
- [115] Shan Y, Eastwood MP, Zhang X, Kim ET, Arkhipov A, Dror RO, Jumper J, Kuriyan J, Shaw DE. Oncogenic mutations counteract intrinsic disorder in the EGFR kinase and promote receptor dimerization. *Cell*. 2012 May; 149(4):860–870.
- [116] Sutto L, Gervasio FL. Effects of oncogenic mutations on the conformational free-energy landscape of EGFR kinase. *Proceedings of the National Academy of Sciences of the United States of America*. 2013 Jun; 110(26):10616–10621.
- [117] Demir Ö, Baronio R, Salehi F, Wassman CD, Hall L, Hatfield GW, Chamberlin R, Kaiser P, Lathrop RH, Amaro RE. Ensemble-based computational approach discriminates functional activity of p53 cancer and rescue mutants. *PLoS Computational Biology*. 2011 Oct; 7(10):e1002238.
- [118] Witham S, Takano K, Schwartz C, Alexov E. A missense mutation in CLIC2 associated with intellectual disability is predicted by in silico modeling to affect protein stability and dynamics. *Proteins*. 2011 Aug; 79(8):2444–2454.
- [119] Tsukamoto H, Farrens DL. A constitutively activating mutation alters the dynamics and energetics of a key conformational change in a ligand-free G protein-coupled receptor. *The Journal of biological chemistry*. 2013 Sep; 288(39):28207–28216.
- [120] Stehr H, Jang SHJ, Duarte JM, Wierling C, Lehrach H, Lappe M, Lange BMH. The structural impact of cancer-associated missense mutations in oncogenes and tumor suppressors. *Molecular cancer*. 2011 May; 10(1):54.
- [121] 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. doi: [10.1016/j.sbi.2011.01.011](https://doi.org/10.1016/j.sbi.2011.01.011).
- [122] Parton DL, Grinaway PB, Hanson SM, Beauchamp KA, Chodera JD. Ensembler: Enabling High-Throughput Molecular Simulations at the Superfamily Scale. *PLoS computational biology*. 2016 Jun; 12(6):e1004728.

- [123] Lindorff-Larsen K, Piana S, Palmo K, Maragakis P, Klepeis JL, Dror RO, Shaw DE. Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins*. 2010 Jun; 78(8):1950–1958.
- [124] Jorgensen, William L, Chandrasekhar, Jayaraman, Madura, Jeffry D, Impey, Roger W, Klein, Michael L. Comparison of simple potential functions for simulating liquid water. *The Journal of chemical physics*. 1998 Aug; 79(2):926–935.
- [125] Eastman, Peter, Swails, Jason, Chodera, John D, McGibbon, Robert T, Zhao, Yutong, Beauchamp, Kyle A, Wang, Lee-Ping, Simmonett, Andrew C, Harrigan, Matthew P, Stern, Chaya D, Wiewiora, Rafal P, Brooks, Bernard R, Pande, Vijay S. OpenMM 7: Rapid development of high performance algorithms for molecular dynamics. *PLoS Computational Biology*. 2017 Jul; 13(7):e1005659.
- [126] Eastman P, Pande VS. CCMA: A Robust, Parallelizable Constraint Method for Molecular Simulations. *Journal of chemical theory and computation*. 2010 Feb; 6(2):434–437.
- [127] Miyamoto S, Kollman PA. Settle: An analytical version of the SHAKE and RATTLE algorithm for rigid water models. *Journal of Computational Chemistry*. 1992; 13(8):952–962.
- [128] Eastman P, Friedrichs MS, Chodera JD, Radmer RJ, Bruns CM, Ku JP, Beauchamp KA, Lane TJ, Wang LP, Shukla D, Tye T, Houston M, Stich T, Klein C, Shirts MR, Pande VS. OpenMM 4: A Reusable, Extensible, Hardware Independent Library for High Performance Molecular Simulation. *Journal of Chemical Theory and Computation*. 2013 Jan; 9(1):461–469.
- [129] McGibbon, Robert T, Beauchamp, Kyle A, Harrigan, Matthew P, Klein, Christoph, Swails, Jason M, Hernández, Carlos X, Schwantes, Christian R, Wang, Lee-Ping, Lane, Thomas J, Pande, Vijay S. MDTraj: A Modern Open Library for the Analysis of Molecular Dynamics Trajectories. *Biophysical journal*. 2015 Oct; 109(8):1528–1532.
- [130] Friesner RA, Banks JL, Murphy RB, Halgren TA, Klicic JJ, Mainz DT, Repasky MP, Knoll EH, Shelley M, Perry JK, Shaw DE, Francis P, Shenkin PS. Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. *Journal of medicinal chemistry*. 2004 Mar; 47(7):1739–1749.
- [131] Halgren TA, Murphy RB, Friesner RA, Beard HS, Frye LL, Pollard WT, Banks JL. Glide: a new approach for rapid, accurate docking and scoring. 2. Enrichment factors in database screening. *Journal of medicinal chemistry*. 2004 Mar; 47(7):1750–1759.
- [132] Friesner RA, Murphy RB, Repasky MP, Frye LL, Greenwood JR, Halgren TA, Sanschagrin PC, Mainz DT. Extra precision glide: docking and scoring incorporating a model of hydrophobic enclosure for protein-ligand complexes. *Journal of medicinal chemistry*. 2006 Oct; 49(21):6177–6196.
- [133] 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.
- [134] Berendsen HJC, Postma JPM, van Gunsteren WF, Hermans J. Interaction Models for Water in Relation to Protein Hydration. In: *Intermolecular Forces* Dordrecht: Springer, Dordrecht; 1981.p. 331–342.

- [135] Bennett CH. Efficient estimation of free energy differences from Monte Carlo data. *Journal of Computational Physics*. 1976 Oct; 22(2):245–268.
- [136] Shirts MR, Bair E, Hooker G, Pande VS. Equilibrium Free Energies from Nonequilibrium Measurements Using Maximum-Likelihood Methods. *Physical Review Letters*. 2003; 91(14).
- [137] Shukla D, Meng Y, Roux B, Pande VS. Activation pathway of Src kinase reveals intermediate states as targets for drug design. *Nature communications*. 2014; 5:3397.
- [138] Jura N, Zhang X, Endres NF, Seeliger MA, Schindler T, Kuriyan J. Catalytic Control in the EGF Receptor and Its Connection to General Kinase Regulatory Mechanisms. *Molecular cell*. 2011 Apr; 42(1):9–22.
- [139] Urano J, Sato T, Matsuo T, Otsubo Y, Yamamoto M, Tamanoi F. Point mutations in TOR confer Rheb-independent growth in fission yeast and nutrient-independent mammalian TOR signaling in mammalian cells. *Proceedings of the National Academy of Sciences*. 2007 Feb; 104(9):3514–3519.
- [140] Clark AJ, Gindin T, Zhang B, Wang L, Abel R, Murret CS, Xu F, Bao A, Lu NJ, Zhou T, et al. Free Energy Perturbation Calculation of Relative Binding Free Energy between Broadly Neutralizing Antibodies and the gp120 Glycoprotein of HIV-1. *Journal of molecular biology*. 2017; 429(7):930–947.
- [141] Steinbrecher T, Zhu C, Wang L, Abel R, Negron C, Pearlman D, Feyfant E, Duan J, Sherman W. Predicting the Effect of Amino Acid Single-Point Mutations on Protein Stability—Large-Scale Validation of MD-Based Relative Free Energy Calculations. *Journal of molecular biology*. 2017; 429(7):948–963.
- [142] Zehir A, Benayed R, Shah RH, Syed A, Middha S, Kim HR, Srinivasan P, Gao J, Chakravarty D, Devlin SM, Hellmann MD, Barron DA, Schram AM, Hameed M, Dogan S, Ross DS, Hechtman JF, DeLair DF, Yao J, Mandelker DL, et al. Mutational Landscape of Metastatic Cancer Revealed from Prospective Clinical Sequencing of 10,000 Patients. *Nat Med*. 2017 May; 23(6):703–713. doi: [10.1038/nm.4333](https://doi.org/10.1038/nm.4333).
- [143] Cerami E, Gao J, Dogrusoz U, Gross BE, Sumer SO, Aksoy BA, Jacobsen A, Byrne CJ, Heuer ML, Larsson E, Antipin Y, Reva B, Goldberg AP, Sander C, Schultz N. The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. *Cancer discovery*. 2012 May; 2(5):401–404.
- [144] Gao J, Aksoy BA, Dogrusoz U, Dresdner G, Gross B, Sumer SO, Sun Y, Jacobsen A, Sinha R, Larsson E, Cerami E, Sander C, Schultz N. Integrative Analysis of Complex Cancer Genomics and Clinical Profiles Using the cBioPortal. *Science signaling*. 2013 Apr; 6(269):pl1–pl1.
- [145] Barretina J, Caponigro G, Stransky N, Venkatesan K, Margolin AA, Kim S, Wilson CJ, Lehár J, Kryukov GV, Sonkin D, Reddy A, Liu M, Murray L, Berger MF, Monahan JE, Morais P, Meltzer J, Korejwa A, Jané-Valbuena J, Mapa FA, et al. The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. *Nature*. 2012 Mar; 483(7391):603–607.
- [146] Cancer Genome Atlas Research Network, Analysis Working Group: Asan University, BC Cancer Agency, Brigham and Women's Hospital, Broad Institute, Brown University, Case Western Reserve University, Dana-Farber Cancer Institute, Duke University, Greater Poland Cancer Centre, Harvard Medical School, Institute for

- Systems Biology, KU Leuven, Mayo Clinic, Memorial Sloan Kettering Cancer Center, National Cancer Institute, Nationwide Children's Hospital, Stanford University, University of Alabama, University of Michigan, et al. Integrated genomic characterization of oesophageal carcinoma. *Nature*. 2017 Jan; 541(7636):169–175.
- [147] Shah NP, Nicoll JM, Nagar B, Gorre ME, Paquette RL, Kuriyan J, Sawyers CL. Multiple BCR-ABL Kinase Domain Mutations Confer Polyclonal Resistance to the Tyrosine Kinase Inhibitor Imatinib (ST1571) in Chronic Phase and Blast Crisis Chronic Myeloid Leukemia. *Cancer Cell*. 2002 Aug; 2(2):117–125.
  - [148] Buczek M, Escudier B, Bartnik E, Szczylik C, Czarnecka A. Resistance to tyrosine kinase inhibitors in clear cell renal cell carcinoma: From the patient's bed to molecular mechanisms. *Biochimica et Biophysica Acta (BBA) - Reviews on Cancer*. 2014; 1845(1):31 – 41. <http://www.sciencedirect.com/science/article/pii/S0304419X13000437>, doi: <https://doi.org/10.1016/j.bbcan.2013.10.001>.
  - [149] Huang L, Fu L. Mechanisms of Resistance to EGFR Tyrosine Kinase Inhibitors. *Acta Pharm Sin B*. 2015; 5(5):390–401.
  - [150] Meyer SC, Levine RL. Molecular Pathways: Molecular Basis for Sensitivity and Resistance to JAK Kinase Inhibitors. *Clin Cancer Res*. 2014; 20(8):2051–2059. doi: [10.1158/1078-0432.CCR-13-0279](https://doi.org/10.1158/1078-0432.CCR-13-0279).
  - [151] Davare MA, Vellore NA, Wagner JP, Eide CA, Goodman JR, Drilon A, Deininger MW, O'Hare T, Druker BJ. Structural Insight into Selectivity and Resistance Profiles of ROS1 Tyrosine Kinase Inhibitors. *Proc Natl Acad Sci*. 2015; 112(39):E5381–E5390. doi: [10.1073/pnas.1515281112](https://doi.org/10.1073/pnas.1515281112).
  - [152] Van Allen EM, Wagle N, Sucker A, Treacy DJ, Johannessen CM, Goetz EM, Place CS, Taylor-Weiner A, Whittaker S, Kryukov GV, Hodis E, Rosenberg M, McKenna A, Cibulskis K, Farlow D, Zimmer L, Hillen U, Gutzmer R, Goldinger SM, Ugurel S, et al. The Genetic Landscape of Clinical Resistance to RAF Inhibition in Metastatic Melanoma. *Cancer Discov*. 2014; 4(1):94–109. doi: [10.1158/2159-8290.CD-13-0617](https://doi.org/10.1158/2159-8290.CD-13-0617).
  - [153] Rani S, Corcoran C, Shiels L, Germano S, Breslin S, Madden S, McDermott MS, Browne BC, OttextquoterightDonovan N, Crown J, Gogarty M, Byrne AT, OttextquoterightDriscoll L, Neuromedin U: A Candidate Biomarker and Therapeutic Target to Predict and Overcome Resistance to HER-Tyrosine Kinase Inhibitors. *Cancer Res*. 2014; 74(14):3821–3833. doi: [10.1158/0008-5472.CAN-13-2053](https://doi.org/10.1158/0008-5472.CAN-13-2053).
  - [154] Holohan C, Van Schaeybroeck S, Longley DB, Johnston PG. Cancer Drug Resistance: An Evolving Paradigm. *Nat Rev Cancer*. 2013 Sep; 13(10):714–726. doi: [10.1038/nrc3599](https://doi.org/10.1038/nrc3599).
  - [155] Weisberg E, Manley PW, Cowan-Jacob SW, Hochhaus A, Griffin JD. Second Generation Inhibitors of BCR-ABL for the Treatment of Imatinib-Resistant Chronic Myeloid Leukaemia. *Nat Rev Cancer*. 2007 May; 7(5):345–356. doi: [10.1038/nrc2126](https://doi.org/10.1038/nrc2126).
  - [156] Y Lu X, Cai Q, Ding K. Recent Developments in the Third Generation Inhibitors of Bcr-Abl for Overriding T315I Mutation. *Curr Med Chem*. 2011 May; 18(14):2146–2157. doi: [10.2174/092986711795656135](https://doi.org/10.2174/092986711795656135).
  - [157] Song Z, Wang M, Zhang A. Alectinib: A Novel Second Generation Anaplastic Lymphoma Kinase (ALK) Inhibitor for Overcoming Clinically-Acquired Resistance. *Acta Pharm Sin B*. 2015 Jan; 5(1):34–37. doi: [10.1016/j.apsb.2014.12.007](https://doi.org/10.1016/j.apsb.2014.12.007).

- [158] Neel DS, Bivona TG. Resistance Is Futile: Overcoming Resistance to Targeted Therapies in Lung Adenocarcinoma. *Npj Precis Oncol.* 2017 Dec; 1(1). doi: 10.1038/s41698-017-0007-0.
- [159] Gruber F, Hjorth-Hansen H, Mikkola I, Stenke L, TA J. A Novel BCR-ABL Splice Isoform Is Associated with the L248V Mutation in CML Patients with Acquired Resistance to Imatinib. *Leuk Off J Leuk Soc Am Leuk Res Fund UK.* 2006 Dec; 20:2057–60.
- [160] Housman G, Byler S, Heerboth S, Lapinska K, Longacre M, Snyder N, Sarkar S. Drug Resistance in Cancer: An Overview. *Cancers.* 2014 Sep; 6(3):1769–1792. doi: 10.3390/cancers6031769.
- [161] Redig AJ, Jänne PA. Basket Trials and the Evolution of Clinical Trial Design in an Era of Genomic Medicine. American Society of Clinical Oncology; 2015.
- [162] Hyman DM, Taylor BS, Baselga J. Implementing Genome-Driven Oncology. *Cell.* 2017 Feb; 168(4):584–599. doi: 10.1016/j.cell.2016.12.015.
- [163] Pesesky MW, Hussain T, Wallace M, Patel S, Andleeb S, Burnham CAD, Dantas G. Evaluation of Machine Learning and Rules-Based Approaches for Predicting Antimicrobial Resistance Profiles in Gram-Negative Bacilli from Whole Genome Sequence Data. *Front Microbiol.* 2016 Nov; 7. doi: 10.3389/fmicb.2016.01887.
- [164] Melnikov A, Rogov P, Wang L, Gnrke A, Mikkelsen TS. Comprehensive Mutational Scanning of a Kinase *in Vivo* Reveals Substrate-Dependent Fitness Landscapes. *Nucleic Acids Res.* 2014 Aug; 42(14):e112–e112. doi: 10.1093/nar/gku511.
- [165] Abel R, Mondal S, Masse C, Greenwood J, Harriman G, Ashwell MA, Bhat S, Wester R, Frye L, Kapeller R, et al. Accelerating drug discovery through tight integration of expert molecular design and predictive scoring. *Current opinion in structural biology.* 2017; 43:38–44.
- [166] Aldeghi M, Heifetz A, Bodkin MJ, Knapp S, Biggin PC. Accurate Calculation of the Absolute Free Energy of Binding for Drug Molecules. *Chem Sci.* 2016; 7(1):207–218. doi: 10.1039/C5SC02678D.
- [167] Cappel D, Hall ML, Lenselink EB, Beuming T, Qi J, Bradner J, Sherman W. Relative Binding Free Energy Calculations Applied to Protein Homology Models. *J Chem Inf Model.* 2016; 56(12):2388–2400. doi: 10.1021/acs.jcim.6b00362.
- [168] Ford MC, Babaoglu K. Examining the Feasibility of Using Free Energy Perturbation (FEP+) in Predicting Protein Stability. *J Chem Inf Model.* 2017 Jun; 57(6):1276–1285. doi: 10.1021/acs.jcim.7b00002.
- [169] Zou J, Song B, Simmerling C, Raleigh D. Experimental and Computational Analysis of Protein Stabilization by Gly-to- D -Ala Substitution: A Convolution of Native State and Unfolded State Effects. *J Am Chem Soc.* 2016 Dec; 138(48):15682–15689. doi: 10.1021/jacs.6b09511.
- [170] Mondal J, Tiwary P, Berne BJ. How a Kinase Inhibitor Withstands Gatekeeper Residue Mutations. *J Am Chem Soc.* 2016; 138(13):4608–4615. doi: 10.1021/jacs.6b01232.

- [171] 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. doi: [10.1002/cmdc.201500333](https://doi.org/10.1002/cmdc.201500333).
- [172] Rapp C, Kalyanaraman C, Schiffmiller A, Schoenbrun EL, Jacobson MP. A Molecular Mechanics Approach to Modeling Protein–Ligand Interactions: Relative Binding Affinities in Congeneric Series. *J Chem Inf Model*. 2011 Sep; 51(9):2082–2089. doi: [10.1021/ci200033n](https://doi.org/10.1021/ci200033n).
- [173] Pemovska T, Johnson E, Kontro M, Repasky GA, Chen J, Wells P, Cronin CN, McTigue M, Kallioniemi O, Porkka K, Murray BW, Wennerberg K. Axitinib Effectively Inhibits BCR-ABL1(T315I) with a Distinct Binding Conformation. *Nature*. 2015 Feb; 519(7541):102–105. doi: [10.1038/nature14119](https://doi.org/10.1038/nature14119).
- [174] Schrock A, Chen TH, Clackson T, Rivera VM. Comprehensive Analysis Of The In Vitro Potency Of Ponatinib, and All Other Approved BCR-ABL Tyrosine Kinase Inhibitors (TKIs), Against a Panel Of Single and Compound BCR-ABL Mutants. *Blood*. 2013; 122(21):3992–3992.
- [175] Soverini S, Colarossi S, Gnani A, Rosti G, Castagnetti F, Poerio A, Iacobucci I, Amabile M, Abruzzese E, Orlandi E, Radaelli F, Ciccone F, Tribelli M, di Lorenzo R, Caracciolo C, Izzo B, Pane F, Saglio G, Baccarani M, Martinelli G. Contribution of ABL Kinase Domain Mutations to Imatinib Resistance in Different Subsets of Philadelphia-Positive Patients: By the GIMEMA Working Party on Chronic Myeloid Leukemia. *Clinical Cancer Research*. 2006; 12(24):7374–7379. <http://clincancerres.aacrjournals.org/content/12/24/7374>, doi: [10.1158/1078-0432.CCR-06-1516](https://doi.org/10.1158/1078-0432.CCR-06-1516).
- [176] O'Hare T, Eide CA, Deininger MW. Bcr-Abl kinase domain mutations, drug resistance, and the road to a cure for chronic myeloid leukemia. *Blood*. 2007; 110(7):2242–2249.