

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

May 2019

John D. Chodera, PhD
Dissertation Mentor

Date

© 2019 Steven K. Albanese
ALL RIGHTS RESERVED

To my father, whose unwavering support and love inspired me and kept me going. He is terribly missed.

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 more than 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 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

Steven was born to Eileen and Joseph Albanese in Wayne, New Jersey in 1991, where he would spend the rest of his childhood. Steven graduated from Wayne Hills High School in 2010, and began college at Cornell University that fall. Steven would graduate *cum laude* with a Bachelor of Arts in Cell Biology and Chemistry in 2014. During his time at Cornell, Steven completed an honors thesis in the laboratory of Holger Sondermann in the Department of Molecular Medicine, studying the molecular binding partners of Atlastin, a commonly mutated gene in spastic paraplegia. Steven spent two years working as a study group leader for Intro Biology with the Biology Scholars Program, which served students from underrepresented minorities or the first in their family to attend college. In July 2014, Steven began his first year at Louis V. Gerstner Jr. Graduate School of Biomedical Sciences, achieving a life-long goal of working at Memorial Sloan Kettering Cancer Center. During his time at GSK, Steven served as co-chair of the student council, participated in numerous recruiting trips for the school, and piloted the internship program.

ACKNOWLEDGEMENTS

There are so many people to whom I owe so much for their help on my journey to producing this dissertation, that I hardly know where to begin. It has been a long and fulfilling journey, and I am so grateful to everyone that I have met, worked with, and spent time with along the way. I will do my best to remember everyone, and beg forgiveness from anyone whom I have mistakenly left out.

I should start with my parents, whose love and support has enabled everything that I have accomplished. My mom, who instilled in me the importance and joy of learning, who always pushed me to be the best that I could be, and who fought to give me the best chance to succeed. My dad, who taught me the value of a hard day's work, family, and laughter while he was with us. I miss him terribly. Without their loving guidance, I would not be the person I am today. To my sister Caitie, who has been my best friend and biggest fan, and to my crazy Aunt Maureen and Uncle Bob: thank you for always being there for me.

I would like to thank the following people from my time at Cornell, not in order of importance but order of chronology: Jeff McCaffrey, my undergraduate biology advisor; Colleen Kearns, the research advisor in the Office of Undergraduate Biology at the time; Holger Sondermann, my undergraduate thesis advisor who took a chance on a freshman looking for a summer job and introduced me to protein structure and biophysics; Debashree Chatterjee and John O'Donnell, the graduate students in the Sondermann lab that taught me all of the basics of wetlab biology; and Neil Lewis, my friend and mentor who helped me navigate the graduate school application process. The time I spent on the Hill was a truly formative experience, where I discovered my love for research and developed the courage to chase down fields of study I could never have dreamed of prior.

They say it takes a village to raise a child, and it certainly took a village to shepherd me through graduate school. I have been so fortunate to be surrounded by

a wonderful community of administrators and friends during my time at Gerstner Sloan Kettering. Dean emeritus Ken Marians, Maria Torres, Ivan Gerena, and Iwona Abramek - the original GSK office crew. Thank you for all of the help, fun and encouragement during my early years in graduate school. I cannot overstate how important it was to have so many supportive members of the graduate school to rely on as I acclimated to a new city and program. Thank you to the current administrators at GSK: Dean Michael Overholtzer, Associate Deans Linda Burnley and Thomas Magaldi, David McDonagh, Stacy De La Cruz and Julie Masen - I hope you know how much the students appreciate your hard work! My classmates, Michelle Riegmann, Chris Hulton, and Ben Tischler in particular, who made the core course and first year in general so much fun.

Finally, I have been truly humbled to work alongside an amazing array of scientists, many of whom made substantial contributions to the work in this dissertation and deserve my gratitude: John Chodera, for being an outstanding mentor, who has guided me and taught me so much throughout my graduate career; Patrick Grinaway, for teaching me about statistical mechanics and coding; Andrea Rizzi, Josh Fass, and Mehtap Isik, whose help and advice I appreciate so much; Sonya Hanson and Danny Parton, who have contributed greatly to this work; Lucelenie Rodríguez-Laureano and Erin Grundy, for being excellent lab technicians and a joy to work with; the rest of the Chodera Lab group, whose enthusiasm for science and willingness to help created an excellent environment to learn and grow as a scientist. And to the collaborators whose expertise made my science better and gave me the opportunity to work on exciting problems: Michael Kharas and Gerard Minuesa; James Hsieh and Jianing Xu; Lingle Wang and Kevin Hauser; Robert Benezra and Paulina Wojnarowicz; Andy Intlekofer; Andrea Volkamer; Nicholas Levinson; and Marcus Seeliger.

My thesis committee, consisting of Christina Leslie (chair), John Chodera (advisor), Robert Abel, Sarat Chandarlapaty and Daniel Heller, deserves special thanks for reviewing this dissertation as well as providing sound scientific and career advice over the last few years. The text of Chapter 2 contains work in preparation for submission. Chapter 3 contains material published in *Communications Biology*. In this chapter, Kevin Hauser primarily carried out the calculations, under the supervision of Lingle Wang and John Chodera. I was fortunate to be involved in shaping the experiments and identifying sources of data, as well as helping Kevin write and edit the paper, and analyze some of the outliers. Chapter 4 contains work that was published in *Biochemistry*, and was the culmination of years of work by a huge number of people. Daniel Parton wrote the structural informatics code and helped plan some of the experiments in the paper; Scott Gradia and Chris Jeans carried out the high throughput expression at the MacroLab at UC Berkeley; Nicholas Levinson and Markus Seeliger helped interpret the data and lent their considerable expertise with kinase structural biology to the effort; Mehtap Işık and Lucelenie Rodríguez-Laureano designed the biophysical assays with me and carried them out; John Chodera helped write and edit the manuscript with me.

Table of Contents

List of Tables	xiii
List of Figures	xiv
1 Introduction	1
1.1 Perspective	1
1.2 Synopsis	5
2 Predicting the selectivity of small molecule kinase inhibitors	7
2.1 Gloss	7
2.2 Abstract	10
2.3 Introduction	11
2.3.1 Free energy methods can aid structure-based drug design	11
2.3.2 Selectivity is an important consideration in drug design	11
2.3.3 The use of physical modeling to predict selectivity is relatively unexplored	12
2.3.4 Kinases are an important and particularly challenging model system for selectivity predictions	13
2.3.5 The correlation coefficient measures how useful predictions are in achieving selectivity	14
2.4 Results	16
2.4.1 Alchemical free energy methods can be used to predict compound selectivity	16
2.4.2 Correlation in force field errors can significantly enhance accuracy of selectivity predictions	18
2.4.3 Poor selectivity is achieved for the closely related kinases CDK2/CDK9	22
2.4.4 Greater selectivity is achieved for more distantly related kinases CDK2/ERK2	25
2.4.5 FEP+ calculations show smaller than expected errors for ΔS predictions	28
2.4.6 Correlation of forcefield errors accelerates selectivity optimization	31
2.4.7 Expending more effort to reduce statistical error can be beneficial in selectivity optimization	34
2.5 Discussion and Conclusions	37
2.6 Methods	41

2.6.1	Numerical model of selectivity optimization speedup	41
2.6.2	Numerical model of impact of statistical error on selectivity optimization	42
2.6.3	Binding Site Similarity analysis	43
2.6.4	Extracting the binding free energy ΔG from reported experimental data	44
2.6.5	Structure Preparation	45
2.6.6	Free Energy Calculations	46
2.6.7	Statistical Analysis of FEP+ calculations	47
2.6.8	Quantification of the correlation coefficient ρ	48
2.6.9	Calculating the marginal distribution of speedup	50
2.7	Acknowledgments	50
2.8	Funding	51
2.9	Disclosures	51
2.10	Author Contributions	52
3	Predicting the impact of clinically-observed kinase mutations using physical modeling	53
3.1	Gloss	53
3.2	Abstract	57
3.3	Introduction	57
3.3.1	The long tail of rare kinase mutations frustrates prediction of drug resistance	58
3.3.2	Alchemical free-energy methods can predict inhibitor binding affinities	61
3.3.3	Alchemical approaches can predict the impact of protein mutations on free energy	61
3.3.4	Assessing the potential for physical modeling to predict resistance to FDA-approved TKIs	62
3.4	Results	63
3.4.1	A benchmark of ΔpIC_{50} s for predicting mutational resistance	63
3.4.2	Most mutations do not significantly reduce TKI potency	69
3.4.3	FEP+ predicts affinity changes for clinical Abl mutants	72
3.4.4	FEP+ accurately classifies affinity changes for Abl mutants	76
3.4.5	How reliant are classification results on choice of cutoff?	77
3.4.6	Bayesian analysis can estimate the true error	79
3.4.7	How transferable is FEP+ across the six TKIs?	80
3.4.8	Understanding the origin of mispredictions	83
3.4.9	How strongly is accuracy affected for docked TKIs?	84
3.5	Discussion and Conclusions	88
3.6	Methods	95
3.6.1	System preparation	95
3.6.2	Force field parameter assignment	100
3.6.3	Prime (MM-GBSA)	100
3.6.4	Alchemical free energy perturbation calculations using FEP+	101

3.6.5	Obtaining $\Delta\Delta G$ from ΔpIC_{50} benchmark set data	103
3.6.6	Assessing prediction performance	105
3.6.7	Data availability	109
3.6.8	Code availability	109
3.7	Acknowledgments	109
3.8	Author Contributions	110
3.9	Competing Interests	110
4	Enabling high-throughput biophysical experiments on clinically- observed mutations	111
4.1	Gloss	111
4.2	Abstract	114
4.3	Introduction	115
4.4	Results	119
4.4.1	Construct boundary choice impacts Abl kinase domain ex- pression	119
4.4.2	Screen of 96 kinases finds 52 with useful levels of automated <i>E. coli</i> expression	123
4.4.3	High-expressing kinases are folded with a well-formed ATP binding site	129
4.4.4	Fluorescence-based thermostability assay	131
4.4.5	ATP-competitive inhibitor binding fluorescence assay	132
4.4.6	Expressing clinically-derived Src and Abl mutants	135
4.5	Discussion	141
4.6	Methods	143
4.6.1	Semi-automated selection of kinase construct sequences for <i>E. coli</i> expression	143
4.6.2	Mutagenesis protocol	147
4.6.3	Expression testing	148
4.6.4	Fluorescence-based thermostability assay	149
4.6.5	ATP-competitive inhibitor binding fluorescence assay	151
4.6.6	Large Scale expression and purification protocol for MK14	154
4.7	Author Contributions	155
4.8	Acknowledgments	156
5	Conclusion	157
	Appendices	160
A	Supplemental Figures from Chapter 2	161
	Bibliography	172

List of Tables

3.1	Public ΔpIC_{50} datasets for 144 Abl kinase mutations and eight tyrosine kinase inhibitors (TKIs) with corresponding wild-type co-crystal structures used in this study	65
3.2	$\Delta\Delta G$ data derived from publicly available ΔpIC_{50} measurements and sources of mutation clinical-observation	70
3.3	Summary of FEP+ and Prime statistics in predicting mutational resistance or sensitivity to FDA-approved TKI	90
3.4	Summary of the preparation of the 6 Abl:TKI co-crystal structure complexes	98
4.1	Kinase domain constructs with yields $>2 \mu\text{g}/\text{mL}$ culture for 96-kinase expression screen.	126
4.2	Expression yields for engineered clinical missense mutants of Abl kinase domains with yields $> 2 \mu\text{g}/\text{mL}$ culture.	136
4.3	Expression yields for engineered clinical missense mutants of Src kinase domains with yields $> 2 \mu\text{g}/\text{mL}$ culture.	137
A.1	The CDK2 and CDK9 binding sites are more similar than the CDK2 and ERK2 binding sites	171

List of Figures

2.1	Free energy calculations can accelerate selectivity optimization. . . .	21
2.2	A CDK2/CDK9 selectivity dataset.	24
2.3	A CDK2/ERK2 selectivity set	27
2.4	Selectivity predictions suggest correlation in forcefield error.	30
2.5	Correlation in force field errors between targets can significantly accelerate selectivity optimization.	33
2.6	Reducing statistical uncertainty when force field error correlation is high improves optimization speedups	36
3.1	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
3.2	Cross-comparison of the experimentally measured effects that mutations in Abl kinase have on ligand binding, performed by different labs.	67
3.3	Comparison of 31 mutations for which phosphorylated and non-phosphorylated ΔK_d s were available.	68
3.4	Comparison of experimentally-measured binding free-energy changes ($\Delta\Delta G$) for 131 clinically observed mutations and 6 selective kinase inhibitors for which co-crystal structures of wild-type kinase with inhibitor are available.	74
3.4	Figure caption	75
3.5	TKI-by-TKI truth tables with increasingly large classification cutoffs.	78
3.6	Physical modeling accuracy in computing the impact of clinical Abl mutations on selective inhibitor binding.	81
3.6	Figure caption	82
3.7	Predicting resistance mutations using FEP+ for inhibitors for which co-crystal structures with wild-type kinase are not available.	87
4.1	Abl kinase domain construct expression screen illustrates high sensitivity to construct boundaries.	121
4.2	Expression yields of Abl kinase domain constructs for all constructs with detectable expression.	122
4.3	Kinome wide search for expressible kinases.	125
4.4	Fluorescence-based thermostability assay demonstrates many high-expressing kinases are well-folded.	130

4.5	Fluorescence emission spectra as a function of the fluorescent ATP-competitive kinase inhibitor bosutinib demonstrates the presence of a well-formed ATP binding pocket.	134
4.6	Expression yields for engineered clinically-derived Src and Abl mis-sense mutants	140
A.1	CDK2 adopts an inactive conformation in the crystal structure used for the CDK2/ERK2 calculations	162
A.2	Correlation coefficient ρ controls the shape of the joint marginal distribution of errors	163
A.3	Correlation reduces the expected error for selectivity predictions	164
A.4	Each replicate of the CDK2/CDK9 calculations yields a consistent RMSE and MUE	165
A.5	Each replicate of the CDK2/CDK9 calculations yields consistent errors and correlation coefficient	166
A.6	Each replicate of the CDK2/ERK2 calculations yields a consistent RMSE and MUE	167
A.7	Each replicate of the CDK2/ERK2 calculations yields consistent errors and correlation coefficient	168
A.8	The pooled replicates show good agreement in predictions for individual ligands	169
A.9	The standard deviation for each edge is smaller than the estimated cycle closure uncertainties	170

CHAPTER 1

INTRODUCTION

1.1 Perspective

In 2018, cancer accounted for 610,000 deaths and 1.7 million diagnoses in the United States alone [1]. Since the FDA approval of imatinib in 2001, therapeutics targeting kinases now account for over 50% of current cancer drug discovery and close to 30% of total drug discovery efforts [2], with 44 FDA approved small molecule kinase inhibitors (SMKIs) on the market [3]. However, there has been a decrease in productivity using current design strategies, with many drugs failing in late stage clinical trials. By the time a drug fails in Phase III, a typical pharmaceutical company has spent 12 years and almost \$1 billion on development [4]. SMKIs can fail late in the development pipeline for two main reasons: safety issues or lack of efficacy. Tumors have multiple routes to resistance, including target amplification [5, 6], effectively increasing the amount of drug required to get the same level of inhibition. Inhibitor resistance occurs through the presence or upregulation of a redundant pathway [7, 8], mutation of the target kinase [9, 10], activation of downstream kinases [11], or relief of feedback inhibition [12]. On-target toxicity, from inhibition of wild type kinase, can cause efficacy issues by limiting the maximally tolerated dose (MTD). Safety issues arise from adverse events due off-target toxicity, such as gefitinib inhibiting CYP2D68 and causing hepatotoxicity in lung cancer patients, or from the on-target toxicity of inhibiting the wild type kinase, such as EGFR inhibitors causing skin rash and ocular toxicities [13, 14].

Each kinase inhibitor has a certain selectivity profile, or group of biological targets a molecule binds to and inhibits strongly enough to produce a phenotype.

Kinase inhibitors have potential for a great diversity of selectivity—the number of targets a molecule binds to below a certain K_d threshold. There are more than 520 members of the human kinome [15], each with a highly similar, druggable ATP-binding site [16–20], giving inhibitors targeted to them huge potential for promiscuity, like staurosporine, which inhibits a large percentage of the kinome with high affinity. Even FDA approved drugs have a wide range of selectivities. In a 2011 paper, Davis *et al.*, characterized the interaction of 72 known kinase inhibitors against a panel of 442 kinases [21] using a competitive binding assay. Of the 72 compounds screened, 70% had a K_d less than 3 μM for more than 10% of the 442 distinct kinases screened. While this study confirmed that type II inhibitors, SMKIs that bind an active site adjacent pocket exposed in the ‘DFG-out’ conformation, are more likely to be selective than Type I inhibitors, those that can bind to either the ‘DFG-out’ or ‘DFG-in’ conformation, it also found that there are several type II inhibitors that have low selectivity. Conversely, several Type I inhibitors exhibited a high level of selectivity. This suggests that either binding mode is a viable option when seeking to design a selective inhibitor. Additionally, 17 of the 72 compounds bound to fewer than 5 off-target kinases with affinity comparable to their primary target and also had a K_d less than 3 μM for less than 10% of the assayed kinases. This suggests that it is possible to design compounds that are selective for multiple targets, a strategy that has been suggested as a possible design paradigm termed targeted polypharmacology [11, 22–24].

Current design efforts focus on achieving maximal selectivity for a single target by improving a weak inhibitor through analogue synthesis [25], which is not always rational, or through structure-informed design [25, 26], which is difficult because kinases exist as nodes in complex signaling networks [27, 28], with feedback inhibition and pathway cross-talk muddling the relationship between binding and signaling. This complicates the notion of inhibiting a single kinase to shut-

down a pathway, as alleviating negative feedback can lead to re-activation of the target pathway [27, 29], or lead to activation of a secondary pathway previously regulated by inhibitory cross-talk [27, 30]. Further, tumors can easily evade inhibition [11] by mutating the target to ablate inhibitor binding; mutating downstream effectors to bypass the inhibited node in the pathway; up-regulating a redundant signaling pathway or branch; or up-regulating the target kinase to increase the amount of drug needed for efficacious inhibition to occur. Rationally designing SMKIs for a given selectivity profile, or to overcome resistance will be essential to continue making advances in the treatment of cancer. Selectivity can be improved by antitargeting kinases closely related to the desired therapeutic target, such as positively designing an inhibitor for EGFR while antitargeting HER2. Multitarget design could also reduce on-target toxicity, thereby improving the therapeutic window, by targeting the oncogenic mutated kinase and antitargeting the wild type kinase. This could improve upon the success of certain EGFR inhibitors such as gefitinib and erlotinib [31–33] or aid the development of second generation inhibitors for use in treating patients with clinically-acquired resistance mutations [10, 34, 35], such as the ALK inhibitor alectinib [36]

A further factor complicating the design of small molecule kinase inhibitors is the proliferation of mutations observed in the clinic. Next-generation sequencing has enabled generation of massive datasets rich with missense alterations in kinases observed directly in the clinic [37–39], and has been particularly transformative in the field of oncology. While this technology has drastically advanced our understanding of disease, it presents a problem for rational drug design. While some missense mutations are highly recurrent and have been characterized clinically or biochemically, a long tail of rare mutations accounts for the majority of clinically observed mutations, leaving clinicians, researchers, and chemists without an understanding of whether these mutations might be activating and drive

cancer, or cause resistance to a previously-developed compound [40]. Understanding the impact of these mutations will enable the development of next-generation inhibitors that overcome resistance and have lower-levels of on-target toxicity by sparing the wild-type form of the kinase in favor of the oncogenic mutant kinase.

Physical modeling can be used to characterize the impact of these rare mutations, as well as enable the design of inhibitor selectivity. Molecular dynamics have been successfully applied to a number of drug discovery projects [41], such as for identifying allosteric binding sites [42] or enabling more accurate virtual docking screens [43] through the relaxed complex scheme (RCS) [42, 44]. While these methods have led to successful development of inhibitors for HIV integrase [45] and FKBP [42] as well as some kinase [46], docking scoring algorithms focus on enrichment [47] and do not correlate well with ligand binding affinity [48]. Alchemical free energy calculations allow for prediction of ligand binding free energies, including all enthalpic and entropic contributions [49–51]. Advances in atomistic molecular mechanics forcefields and free energy methodologies [52–55] have allowed free energy methods to reach high levels of accuracy for predicting ligand potencies [56]. Free energy methods have been applied prospectively to develop inhibitors for Tyk2 [57], Syk [58], BACE1 [59], GPCRs [60], and HIV protease [61]. While predicting affinity with free energy calculations has been well-studied, the utility of these methods for predicting selectivity is understudied. Early work on using free energy calculations to understand the selectivity of imatinib for Abl over Src [62, 63] and predicting the selectivity of three inhibitors within the bromodomain family [64] promises that these methods may be useful for designing the selectivity of inhibitors.

Physical models can also be used to understand the impact of clinical mutations on the structure and function of proteins, enabling the design and selection

of inhibitors for the era of personalized medicine. Molecular Dynamics simulations have been applied to understanding the mechanism of mutations on the activity [65, 66] and ligand binding [67] of EGFR. They have also been applied to understanding missense mutations in p53 [68], CLIC2 [69], and opsin [70]. Alchemical free energy calculations have been applied to several widescale studies of protein thermostability [71–73] as well as understanding how gatekeeper mutations impact inhibitor binding [74].

As will be discussed at length in this work, physical modeling can enable the predicting of selectivity and understanding the impact mutations have on protein structure, function, and inhibitor binding. These methods can greatly aid the design of small molecule kinase inhibitors that will address the complex biology of kinases and overcome resistance, both of which limit the benefit patients receive from current compounds.

1.2 Synopsis

This thesis is organized as follows. Chapter 2 contains a manuscript presenting a study of the utility of free energy methods to predict selectivity, using small molecule kinase inhibitors as a clinically-relevant and particularly challenging test system. It will present a numerical model for the speedup that can be expected when optimizing the selectivity of an inhibitor using free energy methods. Chapter 3 presents work from a manuscript published in *Communications Biology* that uses alchemical free energy methods to calculate the impact of clinical mutations on inhibitor binding affinities for Abl kinase. This work highlights the benefit of using physical models to compute physically-meaningful and testable quantities, as well as highlights some of the challenges and limitations in using publicly

available data to benchmark free energy calculations. Finally, Chapter 4 contains a manuscript published in *Biochemistry* that presents work done to enable the generation of high-quality binding affinity data, that is critical for testing the accuracy and utility of free energy methods in drug discovery.

CHAPTER 2
PREDICTING THE SELECTIVITY OF SMALL MOLECULE KINASE
INHIBITORS

2.1 Gloss

The work in this chapter is in preparation, and will be submitted as follows.

Is structure based drug design ready for selectivity optimization?

Steven K. Albanese^{1,2}, John D. Chodera², Andrea Volkamer³, Simon Peng⁴, Robert Abel⁴, Lingle Wang^{4,*}

¹ *Louis V. Gerstner, Jr. Graduate School of Biomedical Sciences, Memorial Sloan Kettering Cancer Center, New York, NY 10065*

² *Computational and Systems Biology Program, Sloan Kettering Institute, Memorial Sloan Kettering Cancer Center, New York, NY 10065*

³ *Charité – Universitätsmedizin Berlin, Charitéplatz 1, 10117 Berlin*

⁴ *Schrödinger, New York, NY 10036*

** Corresponding Author*

This chapter contains work that I started while doing a summer internship at Schrödinger under the direction of Lingle Wang and finished Gerstner Sloan Kettering under the direction of John Chodera. We were interested in whether selectivity could be predicted using alchemical free energy methods. Selectivity is an important property to consider in drug development, when designing either an inhibitor that is maximally selective [25, 26] or polypharmacological [11, 22–24, 75] agent engineered to bind to multiple targets. Controlling selectivity can help avoid off-target toxicity [14, 76], such as in the case of staurosporine, a toxin that binds to most of the human kinome [21]. Designing with selectivity in mind can be useful in avoiding on-target toxicity by selectively targeting disease mutations [13] in the tumor and sparing the wild type form of the target in normal tissue.

Despite the importance of considering selectivity in drug design, most of the prior work in the field has focused on the utility of physical modeling for predicting potency. Previous work had estimated the per target forcefield error for OPLS3, the forcefield used in this study, at roughly 1.0 kcal/mol [54] when predicting potency. However, it was unknown how much this error might be correlated, and therefore cancel, when making predictions for two targets that have the same small molecule ligands and varying degrees of similarity in the binding site. Should this error be uncorrelated and behave randomly, the selectivity predictions would be *worse* than potency predictions. Even a small degree of correlation in the forcefield error could make selectivity predictions *more* accurate than expected. Molecular dynamics and free energy calculations have been used to extensively explore the mechanism underlying the selectivity of imatinib for Abl kinase over Src [62, 63] and within a family of non-receptor tyrosine kinases [77], termed reorganization energy. This work evaluate neither the accuracy of these methods, nor their application to drug discovery on congeneric series of ligands. The most extensive prior work in this field evaluated the use of absolute free energy methods for predicting the selectivity of three inhibitors across the bromodomain family [64]. These methods achieved promising accuracy for single target potencies of roughly 1 kcal/mol for well behaved systems, but did not explicitly evaluate any selectivity metrics or consider correlation in the forcefield errors made for each bromodomain.

In this work, we present an analysis of the correlation for two sets of kinase selectivity calculations using a Bayesian graphical model to separate the forcefield error from statistical error. One pair of kinases is closely related; one pair is more distantly related. We present models to demonstrate the impact of free energy calculations on selectivity optimization when considering differing levels of forcefield error, forcefield correlation, and statistical error. This work was

designed by Lingle Wang, Robert Abel, John Chodera and me. Simon Peng helped me identify key datasets. Andrea Volkamer performed the multiple sequence alignments to quantify how closely related the kinase binding sites are. Lingle Wang and I set up and ran the free energy calculations. John Chodera and I designed and tested the Bayesian graphical model. I wrote the original draft, and edited it with John Chodera, Lingle Wang and Robert Abel.

The results in this chapter suggest that physical modeling can enhance the design of selective inhibitors, and point to a number of promising future avenues of study. The data sets in this chapter are very limited, with only a handful of small molecules and a limited dynamic range of selectivity each. A future study would benefit greatly from access to the type of prospective data sets commonly generated in the pharmaceutical industry, which can contain up to thousands of compounds. A larger data set would allow for a more accurate quantification of the typical forcefield error correlation for the systems involved, especially if larger small molecule modifications are included. Additional future work could explore the role the relatedness of both targets plays in driving forcefield error correlation, to develop a heuristic to estimate correlation coefficient before running calculations. CDK2, CDK9 and ERK2 share a roughly 50-60% sequence identity in the binding site. A larger scale study looking at more kinase pairs, or pairing protein targets from different families with different binding modes, would provide insight into what level of correlation could be expected *a priori*. Future work could also focus on how forcefield error correlation changes for different forcefields. Beyond understanding the correlation in forcefields, there is much work to be done in exploring how best to combine physical modeling and other types of data, including the wealth of structural and high-throughput biophysical data available for kinases, with machine learning to predict selectivity and polypharmacology.

2.2 Abstract

Alchemical free energy calculations are now widely used to drive or maintain potency in small molecule lead optimization, where the binding affinity to a protein target can be computed—in well-behaved cases—to roughly 1 kcal/mol inaccuracy, which is believed to primarily stem from force field errors. Despite this, the potential to use free energy calculations to drive optimization of compound *selectivity* among two similar targets has been relatively unexplored. In the most optimistic scenario, the similarity of binding sites might lead to a fortuitous cancellation of force field errors and allow selectivity to be predicted more accurately than affinity. Here, we assess the accuracy with which selectivity can be predicted in the context of small molecule kinase inhibitors, considering the very similar binding sites of human kinases CDK2 and CDK9, as well as another series of ligands attempting to achieve selectivity between the more distantly related kinases CDK2 and ERK2. Using a Bayesian analysis approach, we separate force field error from statistical error and quantify the correlation in force field errors between selectivity targets. We find that, in the closely related CDK2/CDK9 case, a high correlation in force field errors suggests free energy calculations can have significant impact in aiding chemists in achieving selectivity, while in more distantly related kinases (CDK2/ERK2), the correlation in force field errors suggests fortuitous cancellation may even occur between systems that are not closely related. In both cases, the correlation in force field error suggests that longer simulations are beneficial to properly balance statistical error with systematic error to take full advantage of the increase in accuracy in selectivity prediction possible due to fortuitous cancellation of error.

2.3 Introduction

2.3.1 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 the prediction of ligand binding free energies, including all enthalpic and entropic contributions [49]. Advances in atomistic molecular mechanics force fields and free energy methodologies [52–55] have allowed free energy methods to reach a level of accuracy sufficient for predicting ligand potencies [56]. These methods have been applied prospectively to develop inhibitors for Tyk2 [57], Syk [58], BACE1 [59], GPCRs [60], and HIV protease [61]. A recent large-scale review of the use of FEP+ [78] to predict potency for 92 different projects and 3 021 compounds determined that predicted binding free energies had a median root mean squared error (RMSE) of 1.0 kcal/mol [79].

2.3.2 Selectivity is an important consideration in drug design

In addition to potency, selectivity is an important property to consider in drug development, either in the pursuit of an inhibitor that is maximally selective [25, 26] or possesses a desired polypharmacology [11, 22–24, 75]. Controlling selectivity can be useful not only in avoiding off-target toxicity (arising from inhibition of unintended targets) [14, 76], but also in avoiding on-target toxicity (arising from inhibition of the intended target) by selectively targeting disease mutations [13]. 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 [27, 28] 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 [27, 28] or the upregulation of a secondary pathway by alleviation of cross-talk inhibition [12, 30]. Off-target toxicity can also be caused by inhibiting unrelated targets, such as gefitinib, an EGFR inhibitor, inhibiting CYP2D6 [76] 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 [80–82], exemplified by a number of first generation EGFR inhibitors. Selective binding to multiple kinases can also lead to beneficial effects: Imatinib, initially developed to target BCR-Abl fusion proteins, is also approved for treating gastrointestinal stromal tumors (GIST) [83] due to its activity against receptor tyrosine kinase KIT.

2.3.3 The use of physical modeling to predict selectivity is relatively unexplored

While engineering compound selectivity is important for drug discovery, the utility of free energy methods for predicting this selectivity with the aim of reducing the number of compounds that must be synthesized to achieve a desired selectivity profile has been relatively unexplored. If there is fortuitous cancellation of systematic (force field) 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. Molecular dynamics and free

energy calculations have been used extensively to investigate the biophysical origins of the selectivity of imatinib for Abl kinase over Src [62, 63] and within a family of non-receptor tyrosine kinases [77]. This work focused on understanding the role reorganization energy plays in the exquisite selectivity of imatinib for Abl over the highly related Src despite high similarity between the cocrystallized binding mode and kinase conformations, and touches on neither the evaluation of the accuracy of these methods nor 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 potency of roughly 1 kcal/mol, but does not explicitly evaluate any selectivity metrics [64] or quantify the correlation in the errors made in predicting affinities for each bromodomain.

2.3.4 Kinases are an important 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 inhibitor selectivity in a drug discovery context. 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 [3], and it is estimated that kinase targeted therapies account for as much as 50% of current drug development [84], 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 selective inhibition of kinases is quite challenging, as there are more than 518 protein kinases [15, 85] sharing a highly conserved ATP binding site that is targeted by the majority of SMKIs [86]. While kinase inhibitors have been designed to target kinase-specific sub-pockets and binding modes to achieve selectivity [16–19, 87, 88], 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 are capable of achieving a range of selectivities [21, 89], 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 [90]. 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 [91], providing a rationale 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 [92]. Developing inhibitors with a desired polypharmacology means navigating more complex selectivity profiles, presenting a problem where physical modeling has the potential to dramatically speed up drug discovery.

2.3.5 The correlation coefficient measures how useful predictions are in achieving selectivity

Since the prediction of selectivity depends on predicting affinities to two or more targets (or relative affinities between pairs of related molecules), a spectrum of possibilities exists for how accurately selectivity can be predicted even when the error in predicting individual target affinities is fixed. In well-behaved kinase systems, for example, free energy calculation potency predictions have achieved mean unsigned errors of less than 1.0 kcal/mol [57, 78], believed to arise predominantly from systematic force field errors [54]. In the best-case scenario, correlation in the force field errors for predicting the interactions of a given ligand with two related protein targets might exactly cancel out, allowing selectivity to be predicted much more accurately than potency. On the other hand, if the force field error acts like an uncorrelated random error between two protein targets because the details of the interactions are different, predictions of selectivity will be *less accurate* than potency predictions. Real-world systems are likely to fall somewhere between these two extremes, and quantifying the *degree* to which error in multiple protein targets is correlated, its implications for the use of free energy calculations for prioritizing synthesis in the pursuit of selectivity, the ramifications for optimal calculation protocols, and rough guidelines governing which systems we might expect good selectivity prediction is the primary focus of this work.

In particular, in this work, we investigate the magnitude of the correlation (ρ) in predicted binding free energy differences between compounds to two different targets ($\Delta\Delta G$), assessing the utility of alchemical free energy calculations for the prediction of selectivity. We employ state of the art relative free energy

calculations [78, 79] to predict the selectivities of two different congeneric ligand series [93, 94], and construct simple numerical models that allow us to quantify the potential speed up in selectivity optimization expected for different combinations of per target force field errors and correlation coefficients. To make a realistic assessment of our confidence in this correlation coefficient derived from 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. Even in the more distantly related case (CDK2/ERK2), correlation in the force field errors allows free energy calculations to speed up selectivity optimization, suggesting that these methodologies can impact drug discovery even when comparing systems that are not closely related. We also present a model of the impact of per target statistical error at different levels of force field error correlation, suggesting that it is worthwhile to expend more effort sampling in systems with high correlation.

2.4 Results

2.4.1 Alchemical free energy methods can be used to predict compound selectivity

While the potency of a ligand i for a single target is often quantified as a free energy of binding ($\Delta G_{i,\text{target}}$), there are a number of different metrics for quantifying compound selectivity [95, 96]. Here, we consider the selectivity S_i between one target and another (an *antitarget*) as the difference in free energy of binding for a given ligand i between the two,

$$S_i \equiv \Delta G_{i,\text{target 2}} - \Delta G_{i,\text{target 1}} \quad (2.1)$$

While in the optimization of potency we are concerned with $\Delta\Delta G_{ij} \equiv \Delta G_j - \Delta G_i$, the relative free energy of binding of ligands i and j to a single target, in the optimization of selectivity, we are concerned with $\Delta S_{ij} \equiv S_j - S_i$, which reflects the change in selectivity between ligand i and a related ligand j ,

$$\begin{aligned} \Delta S_{ij} &\equiv S_j - S_i && (2.2) \\ &= (\Delta G_{j,\text{target 2}} - \Delta G_{j,\text{target 1}}) - (\Delta G_{i,\text{target 2}} - \Delta G_{i,\text{target 1}}) \\ &= \Delta\Delta G_{ij,\text{target 2}} - \Delta\Delta G_{ij,\text{target 1}} \end{aligned}$$

To predict the change in selectivity, ΔS_{ij} , between two related compounds, we developed a protocol that uses a relative free energy calculation (FEP+) [78] to construct a map of alchemical perturbations between ligands in a congeneric series, as described in detail in the Methods. The calculation is repeated for each target of interest, with identical perturbations (edges) between each ligand (nodes). Each edge represents a relative alchemical free energy calculation that

quantifies the $\Delta\Delta G$ between the ligands (nodes) for a single target. From these calculations, we can calculate a ΔS for each pair of ligands between two targets of interest.

Previous work shows that FEP+ can achieve an accuracy (σ_{target}) of roughly 1 kcal/mol in potency prediction, which reflects a combination of systematic force field error and random statistical error [78]. However, it is possible that the force field component of that error (σ_{ff}) may fortuitously cancel when computing ΔS_{ij} , resulting in the force field contribution to the selectivity error ($\sigma_{\text{selectivity}}$) being significantly lower than its contribution to potency error.

If we presume that the force field errors for both targets are distributed according to a bivariate normal distribution with correlation coefficient ρ quantifying the *degree* of correlation (with $\rho = 0$ denoting no correlation and $\rho = 1$ denoting perfect correlation), we can model the error in predicting the S_i as $\sigma_{\text{selectivity}}$ as,

$$\sigma_{\text{selectivity}} = \sqrt{\sigma_{\text{ff},1}^2 + \sigma_{\text{ff},2}^2 - 2\rho \sigma_{\text{ff},1} \sigma_{\text{ff},2}} \quad (2.3)$$

As we shall see below, the quantitative value of the correlation coefficient ρ has important ramifications for the accuracy with which selectivity can be predicted.

2.4.2 Correlation in force field errors can significantly enhance accuracy of selectivity predictions

To demonstrate the potential impact the correlation coefficient ρ in potency prediction error between two targets has on the force field error of selectivity predictions ($\sigma_{\text{selectivity}}$) using alchemical free energy techniques, we created a simple numerical model following Equation 2.3 which takes into account each of

the per-target force field errors ($\sigma_{\text{ff},1}^2, \sigma_{\text{ff},2}^2$) expected from the methodology as well as the correlation in those errors. As seen in Figure 2.1A, if the per target force field errors are the same ($\sigma_{\text{ff},1} = \sigma_{\text{ff},2}$), $\sigma_{\text{selectivity}}$ approaches 0 as the correlation coefficient ρ approaches 1. If the error for the free energy method is not the same ($\sigma_{\text{ff},1} \neq \sigma_{\text{ff},2}$), $\sigma_{\text{selectivity}}$ gets smaller but approaches a non-zero value as ρ approaches 1.

To quantify the expected reduction in number of compounds that must be synthesized to achieve a desired selectivity threshold (hereafter referred to as the *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 the notion that most proposed modifications would not drive large changes in selectivity. We assume a synthetic chemist's proposals function as a unit normal distribution based on data-driven estimates from an Abbott Laboratories data set [97]

Further suppose that each compound is evaluated computationally with a free energy methodology that has a per-target force field error (σ_{ff}) of 1 kcal/mol, where we presume sufficient computational effort has been expended to make statistical error negligible. All compounds predicted to have a 1.4 kcal/mol improvement in selectivity (10 \times in ratio of affinities, or 1 log₁₀ unit) are synthesized and experimentally tested (Figure 2.1B, colored curves), using an experimental technique with perfect measurement accuracy. 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 this 1.4 kcal/mol selectivity improvement threshold, a correlation coefficient ρ of 0.5 gives an

expected speed up of 4.1×, which can be interpreted as needing to make 4.1x fewer compounds to achieve a tenfold improvement in selectivity. This process can be extended for the even more difficult proposition of achieving a hundredfold improvement in selectivity (Figure 2.1C), where 200–300× speedups can be expected, depending on σ_{ff} for the free energy methodology.

While the speedup values illustrate the potential utility of free energy methods for a drug discovery team optimizing the selectivity of a compound series, it is important to acknowledge that observed speedups for a real drug discovery project will be dependent on the details of how the free energy methods are used by the project team. All of the scoring accuracies under consideration here should allow for immediate identification of 1-log-unit-improved selective compounds in a single round of chemistry, provided a sufficient number of molecules are scored and a sufficiently stringent synthesis rule is used [98, 99].

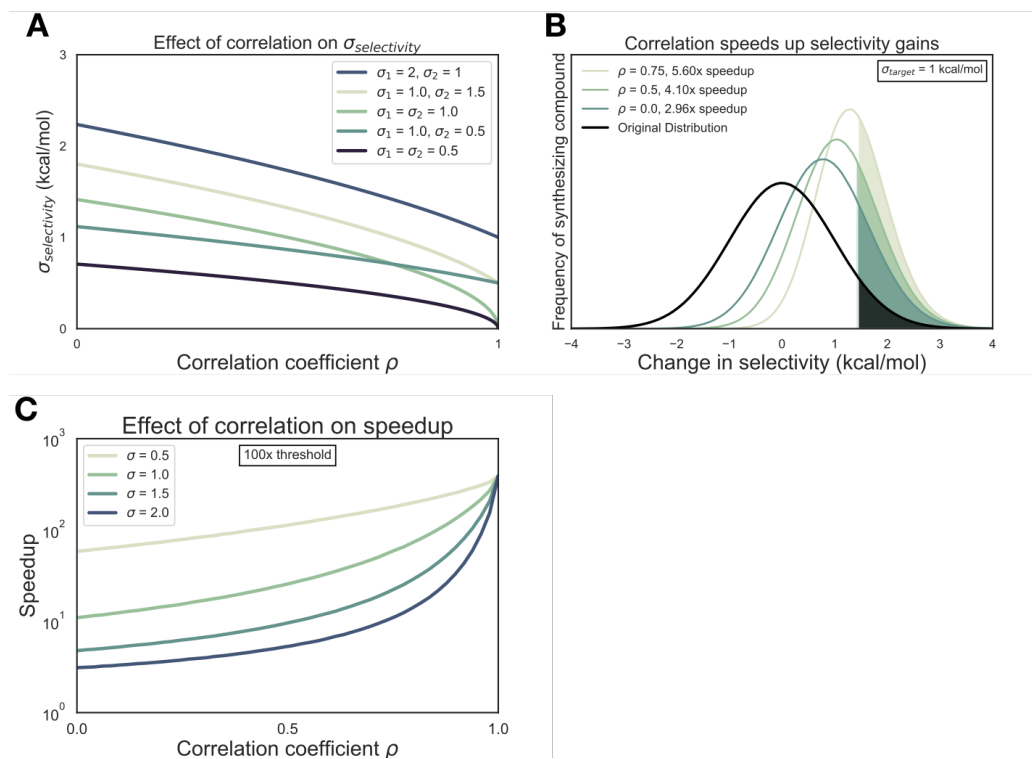


Figure 2.1: **Free energy calculations can accelerate selectivity optimization.** (A) The effect of correlation on expected errors for predicting selectivity ($\sigma_{\text{selectivity}}$) in kcal/mol. Each curve represents a different combination of per target force field errors ($\sigma_{\text{ff},1}$ and $\sigma_{\text{ff},2}$). (B) The change in selectivity for molecules proposed by medicinal chemists optimizing a lead candidate can be modeled by a normal distribution centered on zero with a standard deviation of 1 kcal/mol (black curve), which is consistent with the standard deviation of selectivities observed in the experimental data presented later in this work. Each green curve corresponds to the distribution of compounds made after screening for a 1 log₁₀ unit (1.4 kcal/mol) improvement in selectivity with a free energy methodology with a 1 kcal/mol per target force field error and a particular correlation, in the regime of infinite error where statistical error is zero. The shaded region of each curve corresponds to the compounds with a real 1 log₁₀ unit improvement in selectivity. The speedup is calculated as the ratio of the percentage of compounds made with a real 1 log₁₀ unit improvement to the percentage of compounds that would be expected in the original distribution. (C) The speedup (y-axis, log scale) expected for 100× (2 log₁₀ units, or 2.8 kcal/mol) selectivity optimization as a function of correlation coefficient ρ . Each curve corresponds to a different value of σ_{ff} .

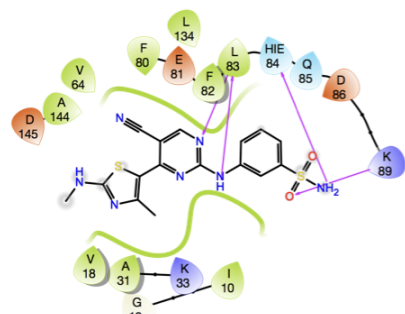
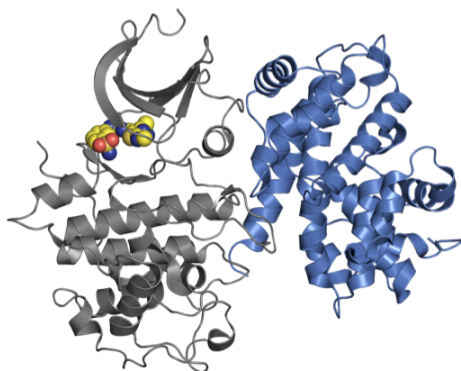
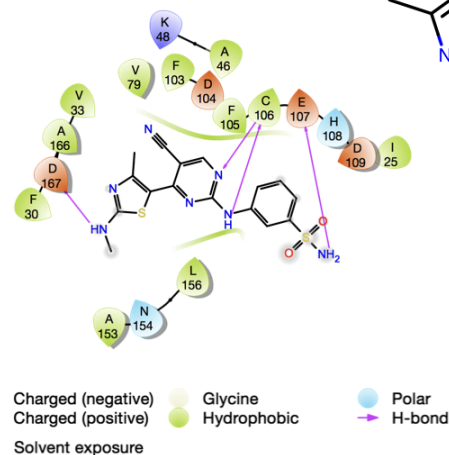
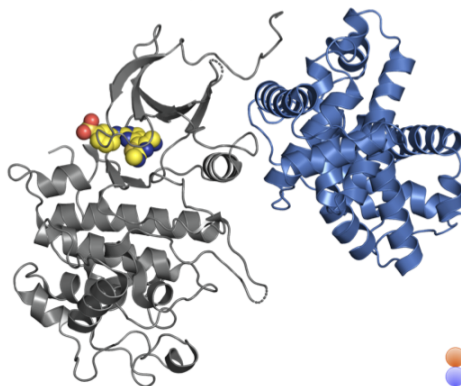
2.4.3 Poor selectivity is achieved for the closely related kinases

CDK2/CDK9

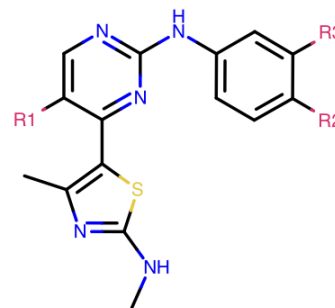
To assess the correlation of errors in free energy predictions for selectivity, we set out to gather data sets that met a number of criteria. We searched for data sets that contained binding affinity data for a number of kinase targets and ligands, as well as having crystal structures for each target with the same co-crystallized ligand.

The first data set we used contained data for a congeneric series of ligands with experimental data for CDK2 and CDK9, with the goal of potently inhibiting CDK9 and sparing CDK2. Based on a multiple sequence alignment of the 85 binding site residues identified in the kinase–ligand interaction fingerprints and structure (KLIFS) database [100, 101], CDK2 and CDK9 share 57% sequence identity (Supp. Table A.1). For this CDK2/CDK9 data set [93], ligand 12c was cocrystallized with CDK2/cylin A (Figure 2.2A, left) and CDK9/cyclin T (Figure 2.2B, left), work that was published in a companion paper [102]. 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 S (CDK9 - CDK2) of only -0.65 kcal/mol, and a standard deviation of 0.88 kcal/mol; the total dynamic range of this data set is 2.8 kcal/mol. Second, experimental uncertainties are not reported for the experimental measurements. This data set reported K_i values calculated from measured IC_{50} , using the K_m (ATP) for CDK2 and CDK9 and [ATP] from the assay. 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. While K_i values are reported, these values are derived from IC_{50} measurements. A number of studies report on the reproducibility of intra-lab IC_{50} measurements. These values range from as low as 0.22 kcal/mol [40], from public data, to as high as 0.4 kcal/mol [56], which was estimated from internal data at Abbott Laboratories. The assumed value of 0.3 kcal/mol falls within this range, and agrees well with the uncertainty reported from Novartis for two different ligand series [103].

A **CDK2**B **CDK9**

C



Ligand	R1	R2	R3	ΔG CDK2 (kcal/mol)	ΔG CDK9 (kcal/mol)	$\Delta\Delta G$ (kcal/mol)
12a	CN	H		-12.27	-11.21	1.06
12b	OH	H		-7.23	-8.22	-0.99
12c	CN	H		-11.45	-11.21	0.24
12e	F	H		-11.62	-11.45	0.17
12f	Cl	H		-10.91	-10.85	0.06
12g	Methyl	H		-10.18	-11.32	-1.14
12h	Ethyl	H		-8.28	-9.56	-1.28
12j	CN	H		-10.04	-11.12	-1.08
12l	CN		H	-10.34	-10.44	-0.1
12n	CN	H		-10.06	-10.97	-0.91
12o	F	H		-10.06	-11.12	-1.06
12q	F	H		-10.91	-11.62	-0.71
12t	CN	H		-9.38	-11.12	-1.74
1a	H	H		-11.62	-11.86	-0.24
1b	H	H		-11.45	-11.86	-0.41

Figure 2.2: **A CDK2/CDK9 selectivity dataset.** Experimental IC_{50} data for a congeneric series of compounds binding to CDK2 and CDK9 was extracted from Shao et al. [93]. (A) (left) Crystal Structure (4BCK) [102] 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)[102] (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 [93], derived from the reported K_i as described in the methods section.

2.4.4 Greater selectivity is achieved for more distantly related kinases CDK2/ERK2

The CDK2/ERK2 data set from Blake et al. [94] also met the criteria described above, with the goal of developing a potent ERK2 inhibitor. Based on a multiple sequence alignment of the KLIFs binding site residues [100, 101], CDK2 and ERK2 share 52% sequence identity (Table A.1), making them slightly less closely related than CDK2 and CDK9. Crystal structures for both CDK2 (Figure 2.3A, top) and ERK2 (Figure 2.3B, top) were available with ligand 22 (according to the manuscript numbering scheme) co-crystallized. Of note, CDK2 was not crystallized with cyclin A, despite cyclin A being included in the affinity assay reported in the paper [94].

CDK2 adopts a DFG-in conformation with the α C helix rotated out, away from the ATP binding site and breaking the conserved salt bridge between K33 and E51 (Supplementary Figure A.1A), indicative of an inactive kinase [18, 104]. By comparison, the CDK2 structure from the CDK2/CDK9 data set adopts a DFG-in conformation with the α C helix rotated in, forming the ionic bond between K33 and E51 indicative of an active kinase, due to allosteric activation by cyclin A. While missing cyclins have caused problems for free 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 hydrogen bonds between the scaffold of the ligand and the backbone of one of the hinge residues (L83 for CDK2 and M108 for ERK2). The conserved

lysine (K33 for CDK2 and K54 for ERK2), normally involved in the formation of a ionic bond with the α C helix, forms a hydrogen bond with the scaffold (Figure 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 congeneric ligand series features a single solvent-exposed substituent. This helps explain the extremely narrow distribution of selectivities, with a mean selectivity of -1.74 kcal/mol (ERK2 - CDK2) and standard deviation of 0.56 kcal/mol; the total dynamic range of this data set is 2.2 kcal/mol. While the small standard deviation suggests that selectivity is difficult to drive with R-group substitution, the total dynamic range demonstrates that R-group substitutions can provide significant selectivity optimization.

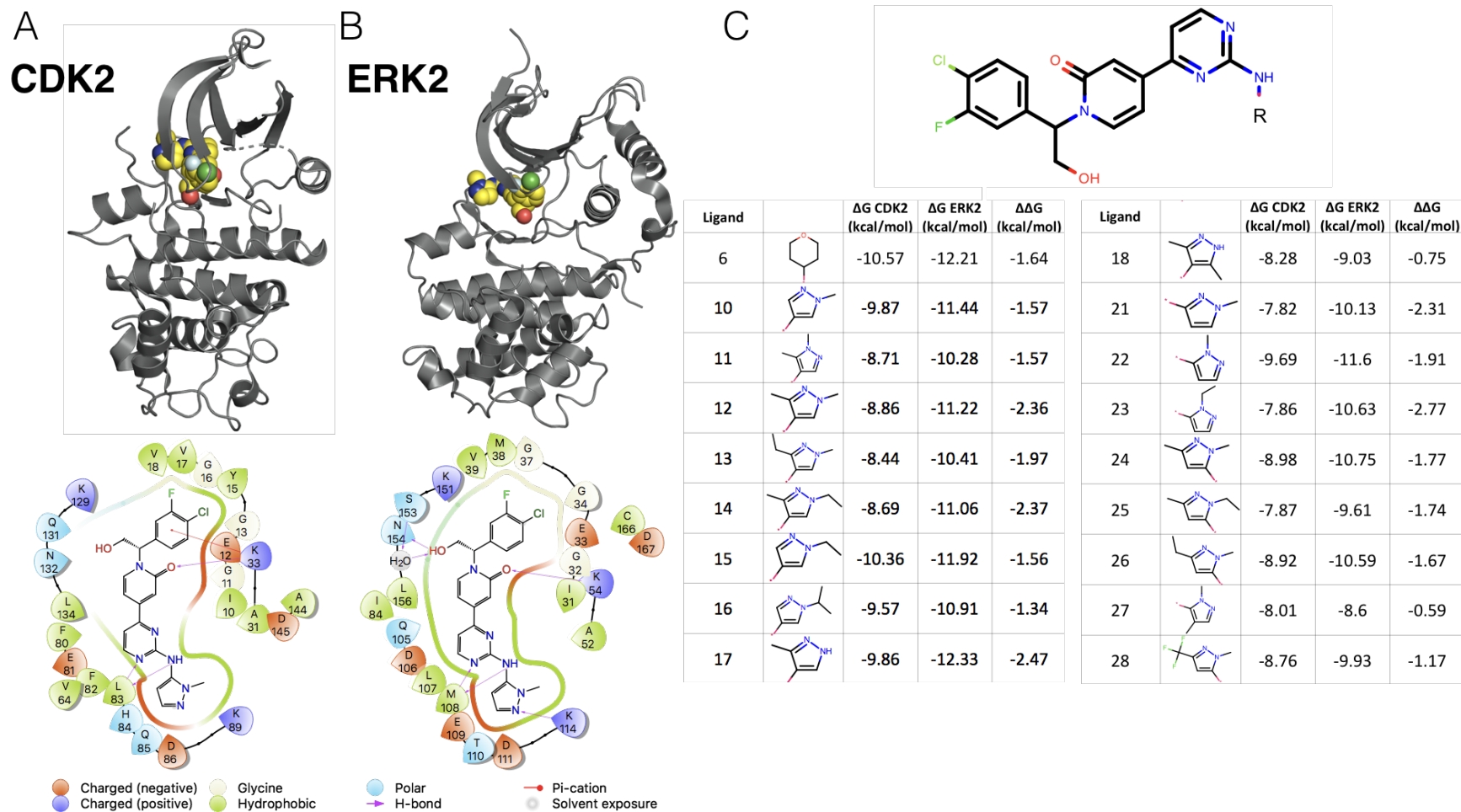


Figure 2.3: **A CDK2/ERK2 selectivity set**

(A) (*top*) Crystal structure of CDK2 (5K4J) shown in gray cartoon and ligand 22 shown in yellow spheres. (*bottom*) 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. (*bottom*) 2D interaction map of ligand 22 in the binding pocket of ERK2. (C) (*top*) Common scaffold for all of the ligands in the Blake data set, with R denoting attachment side for substitutions. (*bottom*) Table showing R group substitutions and experimentally measured binding affinities and selectivities, derived from the IC_{50} values as described in the methods section. Ligand numbers correspond to those used in publication.

2.4.5 FEP+ calculations show smaller than expected errors for ΔS predictions

The FEP+ predictions of the relative free energy of binding between ligands i and j for each target ($\Delta\Delta G_{ij,\text{target}}$) showed good accuracy for the CDK2 and ERK2 data set (Figure 2.4, top). For each ligand i , $\Delta\Delta G_{ij,\text{target}}$ is defined where j is a reference compound, such that $\Delta\Delta G_{ij,\text{target}}$ for the reference compound is 0 kcal/mol.

$$\Delta\Delta G_{ij,\text{target}} = \Delta G_{i,\text{target}} - \Delta G_{\text{reference},\text{target}} \quad (2.4)$$

The reference compounds (Compound 6 for CDK2/ERK2 and Compound 1a for CDK2/CDK9) were selected because they were the compounds from which the synthetic studies were started. Replicate 1 of the CDK2/ERK2 calculations is shown on the bottom of Figure 2.4, with an RMSE of $0.95^{1.25}_{0.62}$ and $0.97^{1.23}_{0.71}$ kcal/mol, respectively. All of the CDK2 and ERK2 $\Delta\Delta G_{ij,\text{target}}$ s were predicted within 1 log unit of the experimental value. The change in selectivity (ΔS) predictions show an RMSE of $1.41^{1.75}_{1.07}$ kcal/mol, with all the confidence intervals of the predictions falling within 1 log unit of the experimental values (Figure 2.4, top right panel). This was consistent across all three replicates of the calculations (Supp. Figure A.6). This consistency across replicates holds true at the individual ligand level as well (Supp. Figure A.8). Despite the low RMSE for the selectivity predictions, the narrow dynamic range and high uncertainty from experiment and calculation makes it difficult to determine which compounds are more selective than others.

Replicate 1 of the CDK2/CDK9 calculations are shown in the top panel of Figure 2.4. The CDK2 and CDK9 data sets show higher errors in $\Delta\Delta G_{ij,\text{target}}$ predictions, with an RMSE of $1.15^{1.58}_{0.67}$ and $2.10^{2.63}_{1.47}$ kcal/mol respectively. There are

a number of outliers that fall outside of 1 log unit from the experimental value for CDK9. While the higher per target errors make predicting potency more difficult, the selectivity predictions show a lower than expected RMSE of $1.37_{1.03}^{1.66}$ kcal/mol. This suggests that some correlation in the error is leading to fortuitous cancellation of the force field error, leading to more accurate than expected predictions of ΔS . These results were consistent across all three replicates of the calculation (Supp. Figure [A.4](#)) as well as each individual ligand (Supp. Figure [A.8](#)).

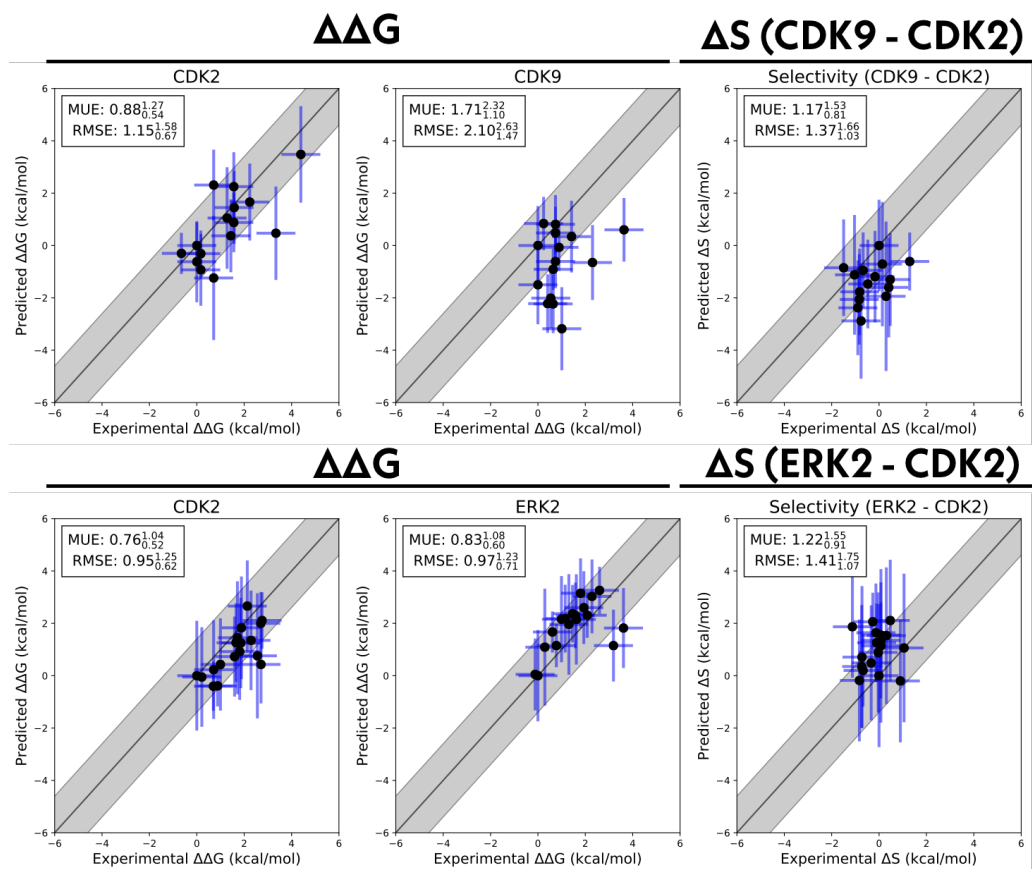


Figure 2.4: Selectivity predictions suggest correlation in forcefield error

$\Delta\Delta G_{i,j,\text{target}}$ and ΔS predictions for CDK2/ERK2 from the Blake data sets (*top*), and CDK2/CDK9 (*bottom*) from the Shao data sets. The experimental values are shown on the X-axis and calculated values on the Y-axis. Each data point corresponds to a transformation between a ligand i to a set reference ligand j for a given target. All values are shown in units of kcal/mol. The horizontal error bars show a 95% CI for the $\delta\Delta G_{ij}^{\text{exp}}$ based on the assumed uncertainty of 0.3 kcal/mol [56, 103] for each ΔG_i^{exp} . We show the 95% CI based on the estimated statistical error (σ_{stat}) as vertical blue error bars. For selectivity, the errors were propagated under the assumption that they were completely uncorrelated. 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.

2.4.6 Correlation of forcefield errors accelerates selectivity optimization

To quantify the correlation coefficient (ρ) of the force field error between targets, we built a Bayesian graphical model to estimate the true model error and quantify our confidence in estimates of ρ (described in depth in Methods). Briefly, we modeled the absolute free energy (G) of each ligand in each thermodynamic phase (ligand-in-complex and ligand-in-solvent, with G determined to an arbitrary additive constant for each phase) as in Equation 2.14. The model was chained to the FEP+ calculations by providing the $\Delta G_{phase,ij,target}^{calc}$ for each edge from the FEP+ maps (where j is now not necessarily the reference compound) as observed data, as in Equation 2.16. As in Equation 2.18, 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 [56]. ΔG values derived from reported IC_{50} s or K_i s, as described in the methods section, were treated as data observations (Equation 2.18) and the $\Delta G_{i,target}^{true}$ was assigned a weak normal prior (Equation 2.19).

The correlation coefficient was calculated for each sample according to equation 2.20. The correlation coefficient ρ for replicate 1 of the CDK2/ERK2 calculations was quantified to be $0.49_{0.27}^{0.68}$, indicating that the errors are correlated between ERK2 and CDK2 (Figure 2.5A, right), which was consistent with the distributions for ρ in replicates 2 and 3 (Supp. Figure A.7). The joint marginal distribution of the error (ϵ) for each target is more diagonal than symmetric, which is expected for cases in which ρ is 0.5 (Supp. Figure A.2). In addition to correlation in the force field errors, the high per target accuracy of these

calculations allow for a predicted 2–3x speed up for 1 \log_{10} unit selectivity optimization, and a 20–50x speed up for 2 \log_{10} unit selectivity optimization (Figure 2.5A, right), in the regime of infinite sampling effort where there is no statistical error.

The CDK2/CDK9 calculations show strong evidence of correlation, with a correlation coefficient of $0.72_{0.58}^{0.83}$ (Figure 2.5B, right) for replicate 1. The rest of the replicates showed strong agreement (Supp. Figure A.5). The joint marginal distribution of errors is strongly diagonal, which is expected based on the value for ρ (Figure 2.5B, left). The high correlation in errors leads to a speed up of 2–3 for 1 \log_{10} unit selectivity optimization and 30–40x for 2 \log_{10} unit selectivity optimization (Figure 2.5B, right), despite the much higher per target RMSE than the CDK2/ERK2 case.

Quantifying ρ for these calculations enables estimation of the force field error in the selectivity predictions, $\sigma_{\text{selectivity}}$. This is useful for estimating expected error for prospective studies, where the experimental values for S are not yet known. Based on the distribution quantified for ρ , the expected $\sigma_{\text{selectivity}}$ for the CDK2/CDK9 calculations is $1.18_{0.95}^{1.38}$ kcal/mol (Supp. Figure A.3), which is in good agreement with the bootstrapped RMSE (Figure 2.4, bottom). For the CDK2/ERK2 calculations, $\sigma_{\text{selectivity}}$ is $0.96_{0.75}^{1.14}$ (Supp. Figure A.3), which is also in good agreement with the bootstrapped RMSE (Figure 2.4, top).

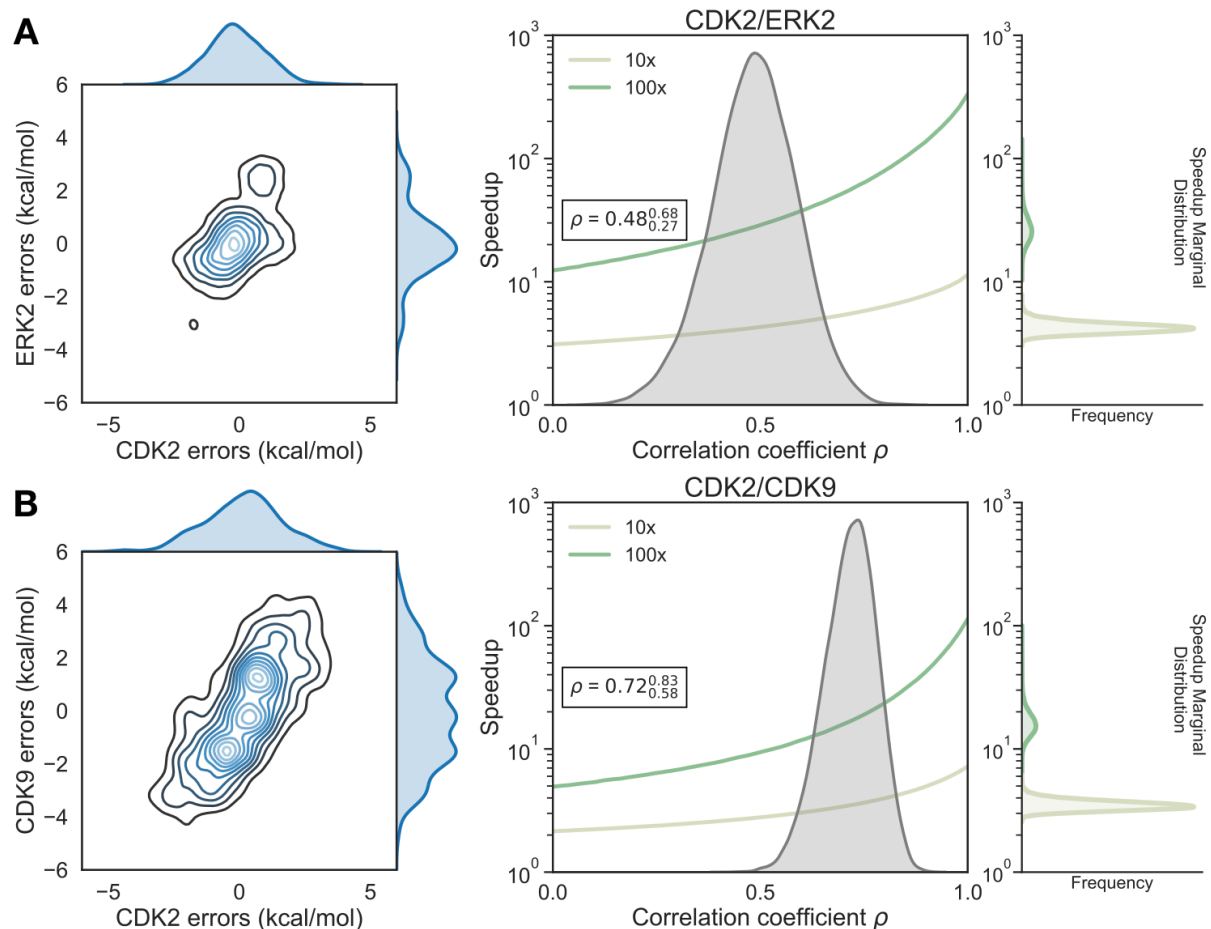


Figure 2.5: **Correlation in force field errors between targets can significantly accelerate selectivity optimization.** (A, left) The joint posterior distribution of the prediction errors for the more distantly related CDK2 (x-axis) and ERK2 (y-axis) from the Bayesian graphical model. (A, right) Speedup in selectivity optimization (x-axis) as a function of correlation coefficient (x-axis). The posterior marginal distribution of the correlation coefficient (ρ) is shown in gray, while the expected speed up is shown for 100 \times (green curve) and 10 \times (yellow curve) selectivity optimization. The inserted box shows the mean and 95% confidence interval for the correlation coefficient. The marginal distribution of speedup is shown on the right side of the plot for both 100 \times (green) and 10 \times (yellow) selectivity optimizations. (B) As above, but for the more closely related CDK2/CDK9 selectivity data set.

2.4.7 Expending more effort to reduce statistical error can be beneficial in selectivity optimization

Up to this point, we have considered only force field error in quantifying the speedup free energy calculations can enable for selectivity optimization, by assuming the statistical error for each target is zero – that we are in the regime of infinite sampling. To begin understanding how statistical error impacts this speedup, we modified the model of speedup by additionally considering the per target statistical error (σ_{stat}), which we define in Equation 2.7 such that at the baseline effort, N , σ_{stat} is 0.2 kcal/mol. In this definition, it takes 4× the sampling, or effort, to reduce statistical error by a factor of 2×. We assume that statistical error is uncorrelated when propagating to two targets, and that σ_{ff} is \approx 1.0 kcal/mol for both targets [40, 54]. As shown in Figure 2.6, expending effort to reduce σ_{stat} when ρ is less than 0.5 does not change the expected speedup for the 100× selectivity threshold in meaningful way, suggesting that it is not worth running calculations longer than the default protocol in this case. However, when $\rho > 0.5$, the curves do start to separate, particularly the 1/4×, 1×, and 4× effort curves. This suggests that when the correlation is high, running longer calculations can net improvements in selectivity optimization speed. Interestingly, the 16×, 48×, and ∞ effort curves do not differ greatly from the 4× effort curve, indicating that there are diminishing returns to running longer calculations.

In order to understand what the current statistical error is for our calculations, we performed three replicates of our calculations, and calculated the standard deviation of the cycle closure $\Delta\Delta G$ for each edge of the map, and compared that value to the cycle closure errors reported for each edge (Supp. Figure A.9). In general, the standard deviation suggests that the statistical error for our

calculations is between 0.1 and 0.3 kcal/mol. While this does not agree well with the cycle closure error (Supp. Figure [A.9](#)), the high variation of the cycle closure errors between replicates of each edge suggest that the standard deviation is a more reliable estimate of the statistical error for these calculations.

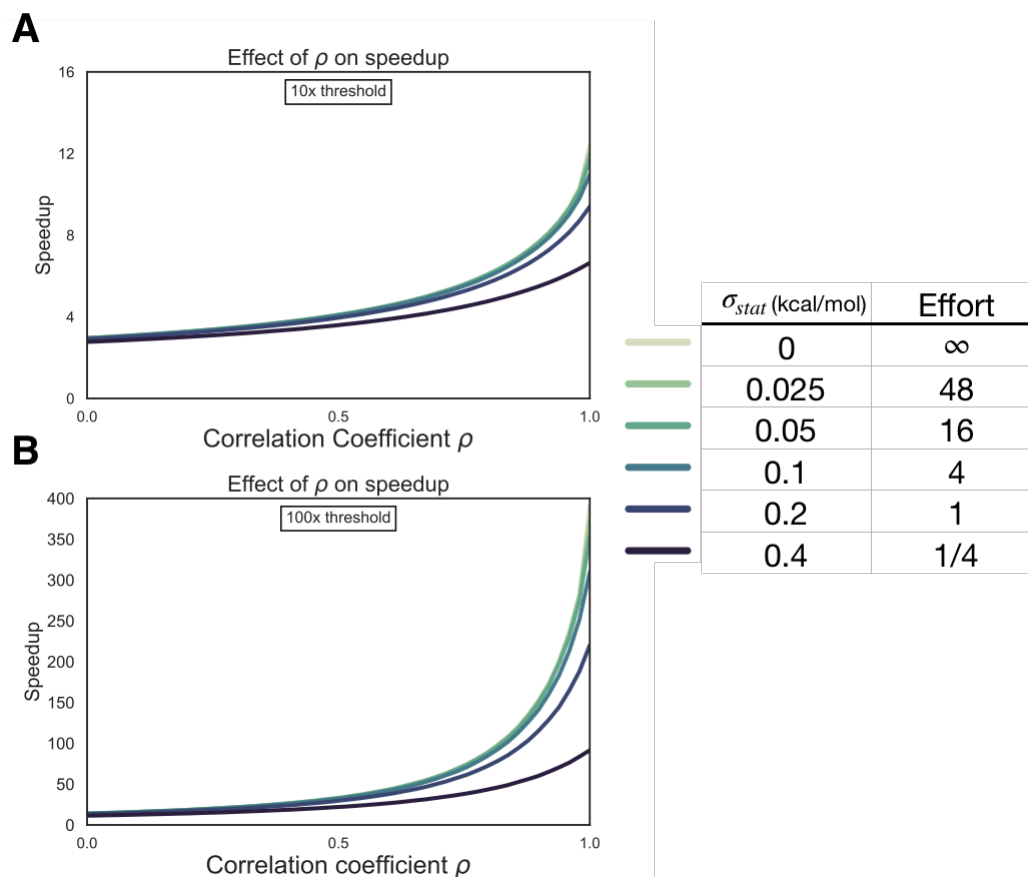


Figure 2.6: Reducing statistical uncertainty when force field error correlation is high improves optimization speedups

(left) The speedup in selectivity (Y-axis) as a function of correlation coefficient (X-axis). Each curve represents a different per target statistical error (σ_{stat}) for 10 \times (1 \log_{10} unit) (A) and 100 \times (2 \log_{10} unit) (B) thresholds (right) Table with the (σ_{stat}), kcal/mol) corresponding to each curve on the left and a rough estimate of the generic amount of computational effort it would take to achieve that statistical uncertainty.

2.5 Discussion and Conclusions

***S* is a useful metric for selectivity in lead optimization**

There are a number of different metrics for quantifying the selectivity of a compound [95], which look at selectivity from different views depending on the information trying to be conveyed. One of the earliest metrics was the standard selectivity score, which conveyed the number of inhibited kinase targets in a broad scale assay divided by the total number of kinases in the assay [105]. The Gini coefficient is a method that does not rely on any threshold, but is highly sensitive to experimental conditions because it is dependent on percent inhibition [106]. Other metrics take a thermodynamic approach to kinase selectivity and are suitable for smaller panel screens [107, 108]. Here, we propose a more granular, thermodynamic view of selectivity that is easy to use free energy methods to calculate: the change in free energy of binding for a given ligand between two different targets (S). S is a useful metric of selectivity in lead optimization once a single, or small panel, of off-targets have been identified and the goal is to use physical modeling to either improve or maintain selectivity within a lead series.

Forcefield error correlation can accelerate selectivity optimization

We have demonstrated, using a simple numerical model, the impact that free energy calculations with even weakly correlated force field 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 force field error of the method (σ_{ff}), 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 10 \times to 100 \times . While 100 \times

selectivity optimization is difficult to achieve, the expected benefit from free energy calculations is also quite high, with speedups of one or two orders of magnitude possible.

Two pairs of kinase test systems suggest forcefield errors can be correlated

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 data sets, which had crystal structures for both targets with the same ligand co-crystallized, exemplify the difficulty in predicting selectivity. The dynamic range of selectivity for both systems is incredibly narrow, with most of the perturbations not having a major impact on the overall selectivity achieved. Further, the data was reported with unreliable experimental uncertainties, which makes quantifying the errors made by the free energy calculations difficult. This issue is common when considering selectivity, as many kinase-oriented high throughput screens are carried out at a single concentration and not highly quantitative.

The CDK9 calculations contained a significant outlier, compound 12h, that drove much of the prediction error for that set. Compound 12e (R1 = F) is the most potent against CDK9 of the compounds in with a sulfonamide at R3 (Figure 2.2). The addition of a single methyl group decreases the potency against CDK9 (compound 12g) and while only slightly changing the affinity for CDK2. However, adding on another methyl group (compound 12h) results in an order of magnitude decrease in K_i for both CDK9 and CDK2. Crystal structures for both kinases show that R1 points into a pocket formed by the backbone, and the sidechains of a Valine and Phenylalanine. While ethyl at R1 in compound 12h is bulkier, the magnitude of the decrease in affinity for both kinases is larger than

might be expected, given that the pocket suggests an ethyl group would be well accommodated in terms of fit and the hydrophobicity of the sidechains. For both kinases, the free energy calculations predict that this addition should *improve* the potency, suggesting that it is possible that the model is missing a chemical detail that might explain the trend seen in the experimental data. We expect that these types of errors, which would be troubling when predicting potency alone, will drive the correlation of forcefield errors and fortuitously cancel.

Despite CDK2 and ERK2 being more distantly related than CDK2 and CDK9, the calculated correlation in the force field error suggests that fortuitous cancellation of errors may be applicable in a wider range of scenarios than closely related kinases within the same family.

Reducing statistical error is beneficial when forcefield errors are correlated

We built a numerical model of the impact of statistical error in the context of different levels of force field error correlation, in order to better understand if there are situations where it is beneficial to expend more effort running longer calculations to minimize statistical error and get improved speedup in selectivity optimization. Our results suggest that unless the correlation $\rho > 0.5$ for the two targets of interest, there is not much benefit in running longer calculations.

However, when the force field error is reduced by correlation, longer calculations can help realize large increases in speedup. Keeping a running quantification of ρ for free energy calculations as compounds are made and the predictions can be tested will allow for decisions to be made about whether running longer calculations is worthwhile. It will also allow for an estimate of $\sigma_{selectivity}$, which is useful for estimating expected force field error for prospective predictions.

Importantly, we expect that correlation will be protocol dependent and changes to

the way the system is modeled are expected to change the observed correlation in the force field error.

Larger data sets with a wide range of protein targets will enable future work

The data sets gathered here were limited by the total number of compounds, the small dynamic range for selectivity (S), and the lack of reliable experimental uncertainties. The small size of the data set makes it difficult to draw broad conclusions about the correlation in forcefield errors. Understanding the degree of correlation *a priori* based on structural similarity requires study on a larger range of targets than the two pairs presented in this study. A larger data set that contained many protein targets, crystal structures and quantitative binding affinity data would be ideal to draw conclusions about the broader prevalence of forcefield error correlation.

This work demonstrates that correlation in the force field errors can allow free energy calculations to facilitate significant speedups in selectivity optimization for drug discovery projects. This is particularly important in kinase systems, where considering multiple targets is an important part of the development process. The results suggest that free energy calculations can be particularly helpful in the design of kinase polypharmacological agents, especially in cases where there is high correlation in the force field errors between multiple targets.

2.6 Methods

2.6.1 Numerical model of selectivity optimization speedup

To model the impact correlation of force field error would have on the expected uncertainty for selectivity predictions in the regime of infinite sampling and zero statistical error, $\sigma_{\text{selectivity}}$ was calculated using Equation 2.5 for 1000 evenly spaced values of the correlation coefficient (ρ) from 0 to 1, for a number of combinations of per target force field errors ($\sigma_{\text{ff},1}$ and $\sigma_{\text{ff},2}$)

$$\sigma_{\text{selectivity}} = \sqrt{\sigma_{\text{ff},1}^2 + \sigma_{\text{ff},2}^2 - 2\rho \sigma_{\text{ff},1} \sigma_{\text{ff},2}} \quad (2.5)$$

The speed up in selectivity optimization that could be expected from using free energy calculations of a particular per target error ($\sigma_{\text{selectivity}}$) was quantified as follows using NumPy (v 1.14.2). An original, true distribution for the change in selectivity of 200 000 000 new compounds proposed with respect to a reference compound was modeled as a normal distribution centered around 0 with a standard deviation of 1 kcal/mol. This assumption was made on the basis that the majority of selectivity is driven by the scaffold, and R group modifications will do little to drive changes in selectivity. The 1 kcal/mol distribution is supported by the standard deviations of the selectivity in the experimental data sets referenced in this work, which are all less than, but close, to 1 kcal/mol.

In this model, we suppose that each of proposed compound is triaged by a free energy calculation and only proposed compounds predicted to increase selectivity by $\Delta S \geq 1.4$ kcal/mol (1 \log_{10} unit) would be synthesized. Based on reported estimates in the literature, we presume that relative free energy calculations have a per-target force field error $\sigma_{\text{ff}} \approx 1$ kcal/mol [54], and explore the impact of the correlation coefficient ρ governing the correlation of these

predictions between targets. The standard error in predicted selectivity, $\sigma_{\text{selectivity}}$, is then given by Equation 2.5, resulting in the error in predicted change in selectivity ΔS modeled as a normal distribution centered around 0 with a standard deviation of $\sigma_{\text{selectivity}}$ and added to the "true" ΔS ,

$$\Delta S_{\text{compound}} = \mathcal{N}_{\text{true}}(\mu = 0, \sigma^2 = 1) + \mathcal{N}_{\text{force field}}(\mu = 0, \sigma_{\text{selectivity}}^2(\rho)) \quad (2.6)$$

We ignore the potential complication of finite experimental error in this thought experiment, presuming the experimental uncertainty is sufficiently small as to be negligible.

The *speedup* in synthesizing molecules that reach this 10 \times selectivity gain is calculated, as a function of ρ , is then the ratio of the number of compounds that exceed the selectivity threshold in the case that molecules predicted to fall below this threshold by free energy calculations were triaged and not synthesized, divided by the number of compounds that exceeded the selectivity threshold without the benefit of free energy triage. This process was repeated for a 100 \times (2.8 kcal/mol, 2 log₁₀ unit) selectivity optimization and 50 linearly spaced values of the correlation coefficient (ρ) between 0 and 1, for four values of $\sigma_{\text{selectivity}}$, using a sample size of 4×10^7 compounds.

2.6.2 Numerical model of impact of statistical error on selectivity optimization

To model the impact of finite statistical error in the alchemical free energy calculations, a similar scheme was used with the following modifications: Each proposed compound was triaged by a free energy calculation with a per target force field error (σ_{ff}) of 1.0 kcal/mol [54] and a specified correlation coefficient ρ .

A $\sigma_{\text{selectivity}}$ was calculated according to Equation 2.5. Additionally, a per target statistical error (σ_{stat}) was considered,

$$\sigma_{\text{stat}} = \sqrt{\frac{2\sigma^2}{N}} \quad (2.7)$$

Where N is the relative effort put into running sampling the calculation and σ is such that when $N = 1$, $\sigma_{\text{stat}} = 0.2$ kcal/mol. The statistical error was propagated assuming it was uncorrelated, as independent sets of calculations are used for each target. This gives an updated model for the error in predicted change in selectivity ΔS . The force field and statistical errors were modeled as Gaussian noise added to the true distribution,

$$\Delta S_{\text{compound}} = \mathcal{N}_{\text{true}}(\mu = 0, \sigma^2 = 1) + \mathcal{N}_{\text{force field}}(\mu = 0, \sigma_{\text{selectivity}}^2(\rho)) + \mathcal{N}_{\text{stat}}(\mu = 0, \sigma_{\text{stat}}^2) \quad (2.8)$$

Any compound predicted to have an improvement in selectivity of above the threshold (either 1.4 /kcal/mol (1 \log_{10} units) or 2.8 kcal/mol (2 \log_{10} units)) would then be made and have its selectivity experimental measured, using an experimental method with perfect accuracy. The speedup value for each value of ρ is calculated as previously described.

2.6.3 Binding Site Similarity analysis

Binding site similarity analysis was performed using a multiple sequence alignment of the residues in the binding site. Binding site definitions for each kinase were taken from the 85 residues defined by the KLIFs database [100]. The

scores presented in Table A.1 are the pair wise identity scores for each pair of kinase.

2.6.4 Extracting the binding free energy ΔG from reported experimental data

K_i values were derived from IC_{50} measurements reported for the ERK2/CDK2 data set (Figure 2.3), assuming Michaelis-Menten binding kinetics for an ATP-competitive inhibitor,

$$IC_{50} = \frac{K_i}{1 + \frac{[S_0]}{K_m}} \quad (2.9)$$

Where the Michaelis-Menten constant for ATP (K_m (ATP)) is much larger than the initial concentration of ATP, S_0 , so that $IC_{50} \approx K_i$.

These K_i values were then used to calculate a ΔG (Equation 2.10),

$$\Delta G = -k_B T \ln K_i \quad (2.10)$$

Here, k_B is the Boltzmann constant and T is absolute temperature (taken to be room temperature, $T \sim 300\text{K}$).

For the CDK2/CDK9 data set, the authors note that the assumption K_m (ATP) $\gg S_0$ does *not* hold, and report K_i s derived from their IC_{50} measurements using the K_m (ATP) for each kinase, as well as the S_0 from their assay. These values were then converted to ΔG using Equation 2.10.

For both data sets, these derived ΔG were used to calculate $\Delta\Delta G$ between ligands for each kinase target.

As mentioned above, the assumption that K_m (ATP) $\gg S_0$ may not always hold, and changes in IC_{50} may be driven by factors other than changes in ligand binding affinity. However, these assumptions have been used successfully to estimate relative free energies previously [40, 109]. Further, data was taken from the same lab and assay for each target. This should minimize errors arising from differing assay conditions that typically complicates the conversion of IC_{50} to free energy of binding.

2.6.5 Structure Preparation

Structures from the Shao [93] (CDK2/CDK9), Hole [102] (CDK2/CDK9), and Blake [94] (CDK2/ERK2) papers were downloaded from the PDB [110], selecting structures with the same co-ligand crystallized.

For the Shao (CDK2/CDK9) data set, PDB IDs 4BCK (CDK2) and 4BCI (CDK9) were selected, which have ligand 12c cocrystallized. For the Blake data set (ERK2/CDK2), 5K4J (CDK2) and 5K4I (ERK2) were selected, cocrystallized with ligand 21. The structures were prepared using Schrodinger's Protein Preparation Wizard [111] (Maestro, Release 2017-3). This pipeline modeled in internal loops and missing atoms, added hydrogens at the reported experimental pH (7.0 for the Shao data set, 7.3 for the Blake data set) 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.

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 [54]. Ionization state was assigned using Epik at experimental $\text{pH}\pm 2$. 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. Any ligands predicted to have a positive or negative charge in its lowest Epik state penalty was excluded, with the exception of Compound 9 from the Blake data set. This ligand was predicted to have a +1 charge for its lowest state penalty state. The neutral form the ligand was include for the sake of cycle closure in the FEP+ map, but was ignored for the sake any analysis afterwards. Ligand poses were generated by first aligning to the co-crystal ligand using the Largest Common Bemis-Murcko scaffold with fuzzy matching (Maestro, Release 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 force field. No flexible residues were defined for the ligand.

2.6.6 Free Energy Calculations

The FEP+ panel (Maestro, Release 2017-4) was used to generate perturbation maps. FEP+ calculations were run using the FEP+ panel from Maestro release 2018-3, using the parameters from the version of OPLS3e that shipped with the 2018-3 release. Any missing ligand torsions were fit using the automated FFbuilder protocol [57]. Custom charges were assigned using the OPLS3e force field using input geometries, according to the automated FEP+ workflow in

Maestro Release 2018-3. Neutral perturbations were run for 15 ns per replica, using an NPT ensemble and water buffer size of 5. The SPC water model was used. 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.

2.6.7 Statistical Analysis of FEP+ calculations

To quantify the overall errors in the FEP+ calculations, we computed the mean unsigned error (MUE),

$$MUE = \frac{\sum_0^n |\Delta\Delta G_{\text{calc}} - \Delta\Delta G_{\text{exp}}|}{n} \quad (2.11)$$

and the root mean squared error (RMSE)

$$RMSE = \sqrt{\frac{\sum_0^n (\Delta\Delta G_{\text{calc}} - \Delta\Delta G_{\text{exp}})^2}{n}} \quad (2.12)$$

MUE and RMSE were computed for the $\Delta\Delta G_{\text{binding}}$ for each ligand with respect to a reference compound, which was pinned to $\Delta\Delta G_{\text{binding}} = 0$ kcal/mol. For the CDK2/CDK9 data set, compound 1a was used as the reference compound, as it was the first compound from which the others in the series were derived. For the CDK2/ERK2 data set, compound 6 was used as the reference compound, since it was the compound from which the investigation was launch. A metabolite of compound 6 (not included in the data set here) was used as the starting compound from which the rest were derived. To account for the finite ligand sample size, we used 10 000 replicates of bootstrapping with replacement to estimate 95% confidence intervals. The code used to bootstrap these values is available on GitHub [<https://github.com/choderalab/selectivity>].

2.6.8 Quantification of the correlation coefficient ρ

To quantify ρ , we built a Bayesian graphical model using pymc3 3.5 [112] and theano 1.0.3 [113]. All code for this model is available on GitHub [<https://github.com/choderalab/selectivity>].

For each phase (complex and solvent), the prior for the absolute free energy (G) of ligand i (up to an arbitrary additive constant for each thermodynamic phase, ligand-in-complex or ligand-insolvent), was treated as a normal distribution (Equation 2.14).

$$G_{i,\text{target}}^{\text{phase}} = \mathcal{N}(\mu = 0, \sigma = 25.0 \text{ kcal/mol}) \quad (2.13)$$

To improve sampling efficiency, for each phase, one ligand was chosen as the reference, and pinned to an absolute free energy of $G = 0$, with a standard deviation of 1 kcal/mol.

$$G_{1,\text{target}}^{\text{phase}} = \mathcal{N}(\mu = 0, \sigma = 1.0 \text{ kcal/mol}) \quad (2.14)$$

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 2.15. Note that here, unlike was done in Figure 2.4, ligand j is not necessarily a reference compound.

$$c_{i,j} = \mathcal{N}(\mu = 0, \sigma = 25.0 \text{ kcal/mol}) \quad (2.15)$$

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

$$\Delta G_{\text{phase}, ij, \text{target}}^{\text{BAR}} = \mathcal{N}(G_{j,\text{target}}^{\text{phase}} - G_{i,\text{target}}^{\text{phase}}, \delta^2 \Delta G_{\text{phase}, ij, \text{target}}^{\text{BAR}}, \text{observed} = \Delta G_{\text{phase}, ij, \text{target}}^{\text{calc}}) \quad (2.16)$$

where $\delta^2 \Delta G_{\text{phase}, ij, \text{target}}^{\text{BAR}}$ is the reported BAR uncertainty from the calculation, and $\Delta G_{\text{phase}, ij, \text{target}}^{\text{calc}}$ 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 $\Delta G_{i, \text{target}}^{\text{FEP}}$ for each ligand and target,

$$\Delta G_{i, \text{target}}^{\text{FEP}} = G_{i, \text{target}}^{\text{complex}} - G_{i, \text{target}}^{\text{solvent}} \quad (2.17)$$

From $\Delta G_{i, \text{target}}^{\text{FEP}}$ we calculated $\Delta \Delta G_{ij, \text{target}}^{\text{FEP}}$ for each pair of ligands, filtering out pairs where i and j are the same ligand and where the reciprocal was already included.

The experimental binding affinity was treated as a true value ($\Delta G_{i, \text{target}}^{\text{true}}$) corrupted by experimental uncertainty, which is assumed to be 0.3 kcal/mol [56]. There are a number of studies that report on the reproducibility and uncertainty of intra-lab IC50 measurements, ranging from as small as 0.22 kcal/mol [40] to as high as 0.4 kcal/mol [56]. The assumed value falls within this range and is in good agreement with the uncertainty reported from multiple replicate measurements in internal data sets at Novartis [103].

The values reported in the papers ($\Delta G_{i, \text{target}}^{\text{obs}}$) were treated as observations from this distribution (Equation 2.18),

$$\Delta G_{i, \text{target}}^{\text{exp}} = \mathcal{N}(\mu = \Delta G_{i, \text{target}}^{\text{true}}, \sigma = 0.3 \text{ kcal/mol, observed} = \Delta G_{i, \text{target}}^{\text{obs}}) \quad (2.18)$$

$\Delta G_{i, \text{target}}^{\text{true}}$ was assigned a weak normal prior, as in Equation 2.19,

$$\Delta G_{i, \text{target}}^{\text{true}} = \mathcal{N}(\mu = 0, \sigma = 50 \text{ kcal/mol}) \quad (2.19)$$

$\Delta \Delta G_{ij, \text{target}}^{\text{true}}$ for each pair of ligands was calculated from $\Delta G_{i, \text{target}}^{\text{true}}$ filtering out pairs where i and j are the same ligand and where the reciprocal was already included as above.

The error for a given ligand was calculated as

$$\epsilon_{i, \text{target}} = \Delta \Delta G_{ij, \text{target}}^{\text{FEP}} - \Delta \Delta G_{ij, \text{target}}^{\text{true}} \quad (2.20)$$

From these ϵ values, we calculated the correlation coefficient, ρ , from the sampled errors for the finite set of molecules for which measurements were available,

$$\rho = \frac{\text{COV}(\epsilon_{\text{target1}}, \epsilon_{\text{target2}})}{\sigma_{\text{target1}} \sigma_{\text{target2}}} \quad (2.21)$$

where σ is the standard deviation of ϵ .

To quantify ρ from these calculations, the default NUTS sampler with `jitter+adapt_diag` initialization, 3 000 tuning steps, and the default target accept probability was used to draw 20 000 samples from the model.

2.6.9 Calculating the marginal distribution of speedup

To quantify the expected speedup from the calculations we ran, we utilized 10^4 replicates of the scheme detailed above to calculate the speedup given parameters ρ , $\sigma_{ff,1}$, and $\sigma_{ff,2}$, in the regime of infinite effort and zero statistical error. Using Numpy 1.14.2, ρ was drawn from a normal distribution with the mean and standard deviation from the posterior distribution of ρ from the Bayesian Graphical model. The per-target force field errors, $\sigma_{ff,1}$ and $\sigma_{ff,2}$, were estimated from the mean of the absolute value of $\epsilon_{\text{target1}}$ and $\epsilon_{\text{target2}}$, which are the magnitude of errors from the Bayesian graphical model. $\sigma_{\text{selectivity}}$ was calculated using Equation 2.3. 10^6 molecules were proposed from true normal distribution, as above. The error of the computational method was modeled as in Equation 2.6.

2.7 Acknowledgments

The authors are grateful to Patrick Grinaway (ORCID: [0000-0002-9762-4201](https://orcid.org/0000-0002-9762-4201)) for useful discussions about Bayesian statistics and Mehtap Işık (ORCID:

[0000-0002-6789-952X](#)) for useful discussion about kinase inhibitor protonation states. SKA is grateful to Haoyu S. Yu, Wei Chen, and Dmitry Lupyan for advice on running FEP+ calculations.

2.8 Funding

Research reported in this publication was supported by the National Institute for General Medical Sciences of the National Institutes of Health under award numbers R01GM121505 and P30CA008748. SKA acknowledges financial support from Schrödinger and the Sloan Kettering Institute. JDC acknowledges financial support from Cycle for Survival and the Sloan Kettering Institute.

2.9 Disclosures

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

2.10 Author Contributions

Conceptualization: SKA, LW, RA, JDC; Methodology: SKA, LW, JDC; Formal Analysis: SKA, JDC, LW; Data Curation: SKA, SP; Investigation: SKA, SP, AV; Writing – Original Draft: SKA; Writing – Review & Editing: SKA, JDC, LW, AV; Visualization: SKA, JDC, LW; Supervision: LW, JDC, RA; Project Administration: SKA, LW, JDC, RA; Funding Acquisition: RA, JDC; Resources: LW, JDC

CHAPTER 3

PREDICTING THE IMPACT OF CLINICALLY-OBSERVED KINASE MUTATIONS USING PHYSICAL MODELING

3.1 Gloss

The work in this chapter was published in *Communications Biology* as follows:

Predicting resistance of clinical Abl mutations to targeted kinase inhibitors using alchemical free-energy calculations

Kevin Hauser¹, Christopher Negron¹, Steven K. Albanese^{2,3}, Soumya Ray, Thomas Steinbrecher⁴, Robert Abel³, John D. Chodera², Lingle Wang^{3,*}

¹ *Schrödinger, New York, NY 10036*

² *Louis V. Gerstner, Jr. Graduate School of Biomedical Sciences, Memorial Sloan Kettering Cancer Center, New York, NY 10065*

³ *Computational and Systems Biology Program, Sloan Kettering Institute, Memorial Sloan Kettering Cancer Center, New York, NY 10065*

³ *Schrödinger, GmbH, Q7 23, 68161 Mannheim, Germany*

** Corresponding Author*

This material is reproduced with permission under the Creative Commons license: <https://creativecommons.org/licenses/by/4.0/>. Some supplementary figures and tables have not been reproduced. They have been replaced in the text with references to the *Communications Biology* source.

This work represents a step towards addressing the feasibility of using alchemical free energy calculations to predict the impact of clinical kinase mutations on small molecule binding. Next-generation sequencing has revolutionized our understanding of cancer, and lead to the generation of massive datasets. While these data sets have provided numerous insights into cancer biology, they present the challenge of interpreting and understanding vast amounts of data. While many highly recurrent mutations have been extensively characterized through structural, biochemical and *in vivo* methods, the majority of these mutations are

rare, occurring fewer than 10 times [40]. They often occur in the context of genetically complex tumors with many rearrangements and mutations [38], which makes generating good disease models difficult. Thus, it is important to develop methods that can describe the impact of individual mutations on properties of interest, such as inhibitor binding, regulatory state, protein stability, and protein conformation. Here, we focus on the use of physical modeling to predict the impact of missense mutations on kinase inhibitor binding. Understanding the impact of these mutations on inhibitor binding is particularly interesting for kinase systems. There are over 43 FDA-approved inhibitors that are currently used to treat patients [3]. While initial responses are typically promising, resistance and disease progression arise in most cases [9, 80, 114]. Developing effective inhibitors to overcome this resistance has been a the focus of extensive drug development programs, leading to the generation of numerous second- and third-generation inhibitors [115–117] for a many different kinase targets.

However, this work has typically relied on data from clinical trials, where patients how have stopped responding to treatment are investigated for potential mutations or changes in cell signaling pathways that may give rise to resistance. While this approach informs a huge number of discoveries, the patients presenting with resistance are often left without good treatment alternatives. This paradigm was flipped in recent work on TRK kinase inhibitors targeting TRK fusion-driven cancers, where the anticipation of missense mutations causing resistance lead to the parallel development of a second-generation TRK inhibitor [10]. In doing so, patients that presented with resistance in the early clinical trials were immediately shuffled on to the next clinical trial for the second generation of inhibitor. One of the key challenges in this study was identifying potential missense mutations for targeting with next generation inhibitors. Physical modeling has the potential aid the design of new inhibitors in this

paradigm by predicting *de novo* missense resistance mutations, as well as the impact of rare but already observed mutations that have yet to be studied. Further, reliable predictions of the impact on inhibitor binding could eventually aid in the selection of existing inhibitors to treat a patient with a novel mutation, especially as the pharmacopeia available grows over time.

The work in this chapter represents a step towards addressing the feasibility of using alchemical free energy calculations to predict the impact of clinical kinase mutations on small molecule binding. This field had previously been understudied and *much* work remains to be done before these methodologies can impact the clinic and the way clinicians make decisions directly. Previous work focused on understanding the mechanism of a small number of kinase gatekeeper mutations [74], which does not provide a large enough data set to draw rigorous conclusions about the broad accuracy and applicability of free energy calculations for predicting the impact of mutations on inhibitor binding. After the publication of the work in this chapter, Aldeghi *et al.*, published a broad study assessing non-equilibrium free energy methods across 17 different proteins and 27 ligands [118]. This protocol [119] achieved similar levels of accuracy to the work presented here, but did not include any kinase systems. While the paper looked at a large number of compounds, the number of mutants for each target was relatively limited. In a 2018 publication after this work, Bhati *et al.* apply alchemical free energy calculations to a fibroblast growth factor receptor 3 (FGFR3) gatekeeper mutations [120]. These results are similarly promising for predicting changes in free energy of binding and capturing some conformational changes, but are found to be sensitive to initial structure selection.

Here we present a large set of clinical mutations and binding affinity data for many FDA-approved kinase inhibitors, forming a valuable benchmark set for

future work on new methodologies seeking to improve accuracy or efficiency. This work was a collaborative effort that I was able to contribute to in a number of ways. I identified sources of experimental affinity data and helped do background and quality control searches on the mutations to ensure that they were sourced from the clinic. I contributed to the preparation of the systems used in the calculations, and helped analyze the data, particularly looking at outliers resulting from the free energy calculations. I took part in writing the first draft of the paper as well as subsequent editing efforts.

There are a number of directions that future work can take. While some papers have already followed up on this work and looked at other systems [119] beyond kinases, this data set presents many avenues for future work itself. This paper used a single default protocol and forcefield for all of the calculations. Future work could follow up on optimizing the protocol used for more efficient and accurate results, or study the accuracy of additional publicly available forcefields. Further, there are a number of choices made when setting up the calculations, such as: protein conformation, ligand pose, salt concentration, and pH. Understanding the sensitivity of these calculations to these choices, and the combination of parameters that will yield the most accurate calculations, will undoubtedly be required to move physical modeling for predicting the impact of mutations on ligand binding into the clinic. In particular, the sensitivity to the starting conformation used, both for the protein and the ligand binding mode, are particularly important when considering mutations that may cause conformational changes.

3.2 Abstract

The therapeutic effect of targeted kinase inhibitors can be significantly reduced by intrinsic or acquired resistance mutations that modulate the affinity of the drug for the kinase. In cancer, the majority of missense mutations are rare, making it difficult to predict their impact on inhibitor affinity. Here, we examine the potential for alchemical free-energy calculations to predict how kinase mutations modulate inhibitor affinities to Abl, a major target in chronic myelogenous leukemia (CML). We find these calculations can achieve useful accuracy in predicting resistance for a set of eight FDA-approved kinase inhibitors across 144 clinically-identified point mutations, achieving a root mean square error in binding free energy changes of $1.1_{0.9}^{1.3}$ kcal mol⁻¹ (95% confidence interval) and correctly classifying mutations as resistant or susceptible with 88₈₂⁹³% accuracy. This benchmark establishes the potential for physical modeling to collaboratively support the assessment and anticipation of patient mutations to affect drug potency in clinical applications.

3.3 Introduction

Targeted kinase inhibitors are a major therapeutic class in the treatment of cancer. A total of 38 selective small molecule kinase inhibitors have now been approved by the FDA [3], including 34 approved to treat cancer, and perhaps 50% of all current drugs in development target kinases [84]. Despite the success of selective inhibitors, the emergence of drug resistance remains a challenge in the treatment of cancer [121–128] and has motivated the development of second- and then

third-generation inhibitors aimed at overcoming recurrent resistance mutations [82, 129–132].

While a number of drug resistance mechanisms have been identified in cancer (e.g., induction of splice variants [133], or alleviation of feedback [12]), 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) [11, 128, 134].

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 [135], 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 [136, 137], help prioritize candidate compounds for clinical trials, and aid the development of next-generation inhibitors.

3.3.1 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 3.1A), 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 [138]. 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 [139] 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.

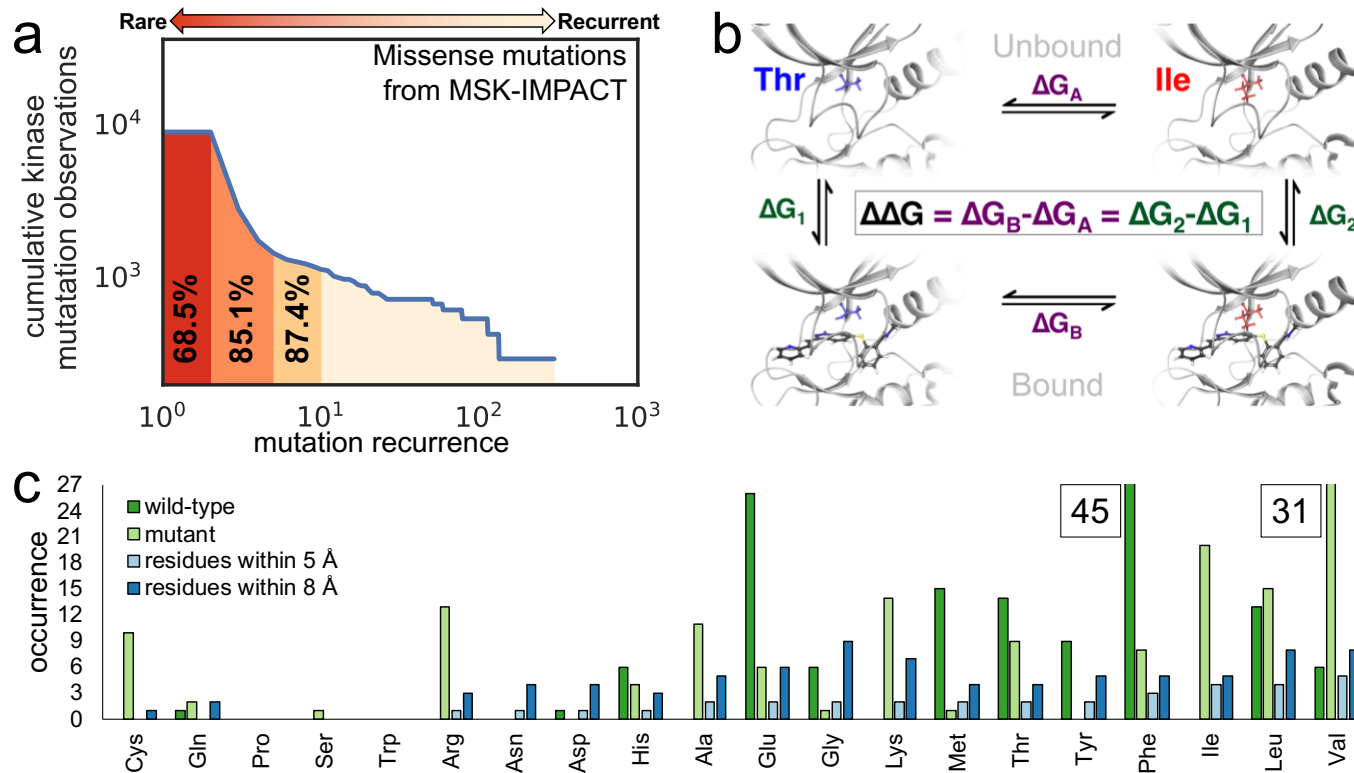


Figure 3.1: **Relative alchemical free-energy calculations can be used to predict affinity changes of FDA-approved selective kinase inhibitors arising from clinically-identified mutations in their targets of therapy.** (a) Missense mutation statistics derived from 10,336 patient samples subjected to Memorial Sloan Kettering-Integrated Mutation Profiling of Actionable Cancer Targets (MSK-IMPACT) deep sequencing panel [135] 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 ?? 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.

3.3.2 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 [54], CHARMM [52], and AMBER FF14SB [53] 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 [140]. Encouragingly, kinase:inhibitor binding affinities have been predicted using alchemical free-energy methods with mean unsigned errors of 1.0 kcal mol⁻¹ for CDK2, JNK1, p38, and Tyk2 [78, 141]. Beyond kinases, alchemical approaches have predicted the binding affinity of BRD4 inhibitors with mean absolute errors of 0.6 kcal mol⁻¹ [142]. Alchemical methods have also been observed to have good accuracy (0.6 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 [143].

3.3.3 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 [144] and protein thermostabilities [145]. 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 [146]. The impact of Gly to D-Ala mutations on protein stability was predicted using an alchemical approach with a similar level of accuracy [147]. 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 [148].

3.3.4 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+ [78], recently demonstrated to achieve sufficiently good accuracy to drive the design of small-molecule inhibitors for a broad range of targets during lead optimization [78, 140, 141, 149], to calculate the effect of point mutation on the binding free energy between the inhibitor and the kinase receptor (Figure 2.1b). Figure 3.1B depicts the thermodynamic cycle that illustrates how we used relative free energy calculations to compute the change in ligand binding free energy in response to the introduction of a point mutation in the kinase (Figure 3.1C). We compare this approach against a fast but approximate physical modeling method implemented in Prime [150] (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 ΔpIC_{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 [3] FDA-approved TKIs are available. While ΔpIC_{50} can approximate a dissociation constant ΔK_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 ΔpIC_{50} and ΔK_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 $88_{82}^{93}\%$ (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 $1.07_{0.89}^{1.26}$ kcal mol⁻¹, and an MUE of $0.79_{0.67}^{0.92}$ kcal mol⁻¹ was achieved.

3.4 Results

3.4.1 A benchmark of ΔpIC_{50} s for predicting mutational resistance

To construct a benchmark evaluation dataset, we compiled a total of 144 ΔpIC_{50} measurements of Abl:TKI affinities, summarized in Table 3.1, taking care to ensure all measurements for an individual TKI were reported in the same study from experiments run under identical conditions. 131 ΔpIC_{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 ΔpIC_{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 (Supplementary Table 3.2). 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 3.1A, 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 3.1: Public ΔpIC_{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_{mut}	R	S	PDB	(kcal mol ⁻¹)	Source	(kcal mol ⁻¹)
					$ \Delta G_{max} - \Delta G_{min} $		ΔG_{WT}
axitinib	26	0	26	4wa9	2.05	[151]	-8.35
bosutinib	21	4	17	3ue4	2.79	[152]	-9.81
dasatinib	21	5	16	4xey	5.08	[152]	-11.94
imatinib	21	5	16	1opj	2.16	[152]	-9.19
nilotinib	21	4	17	3cs9	3.88	[152]	-10.74
ponatinib	21	0	21	3oxz	1.00	[152]	-11.70
subtotal	131	18	113				
erlotinib	7	1	6	<i>Dock to 3ue4</i>	1.73	[21]	-9.77
gefitinib	6	0	6	<i>Dock to 3ue4</i>	1.79	[21]	-8.84
total	144	19	125				

N_{mut} : Total number of mutants for which ΔpIC_{50} data was available.

Number of **R**esistant, **S**usceptible 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.

ΔG_{WT} : Binding free energy of inhibitor to wild-type Abl, as estimated from IC_{50} data.

Experimental ΔpIC_{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 3.1, 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 3.2A,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 $0.32^{0.36}_{0.28}$ 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 ΔpIC_{50} s with those derived from binding affinity measurements (ΔK_d) for which data for a limited set of 27 mutations was available (Figure 3.2D); the larger computed RMSE of $0.81^{1.04}_{0.59}$ 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 ΔK_d s were available, we found a strong correlation between the $\Delta\Delta G$ s derived from those data ($r=0.94$, Figure 3.3); the statistics of that comparison are similar to those of the inter-lab variability comparison.

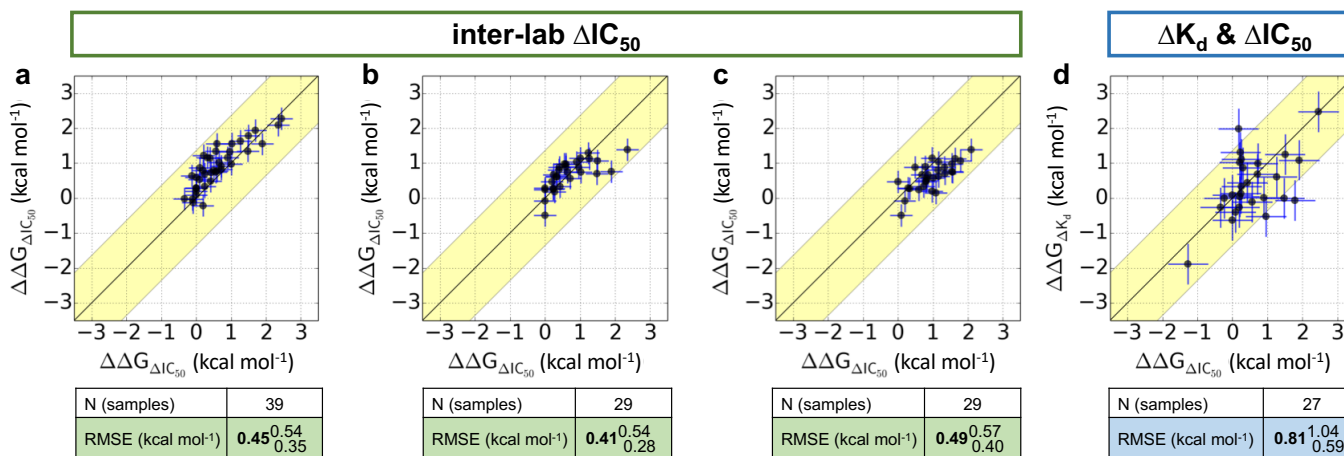


Figure 3.2: **Cross-comparison of the experimentally measured effects that mutations in Abl kinase have on ligand binding, performed by different labs.** $\Delta\Delta G$ was computed from publicly available ΔpIC_{50} or ΔpK_d measurements and these values of $\Delta\Delta G$ were then plotted and the RMSE between them reported. (a) ΔpIC_{50} measurements (X-axis) from [152] compared with ΔpIC_{50} measurements (Y-axis) from [153]. (b) ΔpIC_{50} measurements (X-axis) from [152] compared with ΔpIC_{50} measurements (Y-axis) from [154]. (c) ΔpIC_{50} measurements (X-axis) from [153] compared with ΔpIC_{50} measurements (Y-axis) from [154]. (d) ΔpIC_{50} measurements (X-axis) from [152] compared with ΔpK_d measurements (Y-axis) from [21] 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 $\Delta\Delta G$ s derived from the ΔpIC_{50} measurements, which was $0.32^{0.36}_{0.28}$; the RMSE was $0.45^{0.51}_{0.39}$ kcal mol⁻¹. Scatter plot error bars in d are the \pm standard error (SE) of $\Delta\Delta G$ s derived from ΔpIC_{50} and ΔpK_d from a set of 27 mutations, which is $0.58^{0.74}_{0.42}$ kcal mol⁻¹; the RMSE was $0.81^{1.04}_{0.59}$ kcal mol⁻¹.

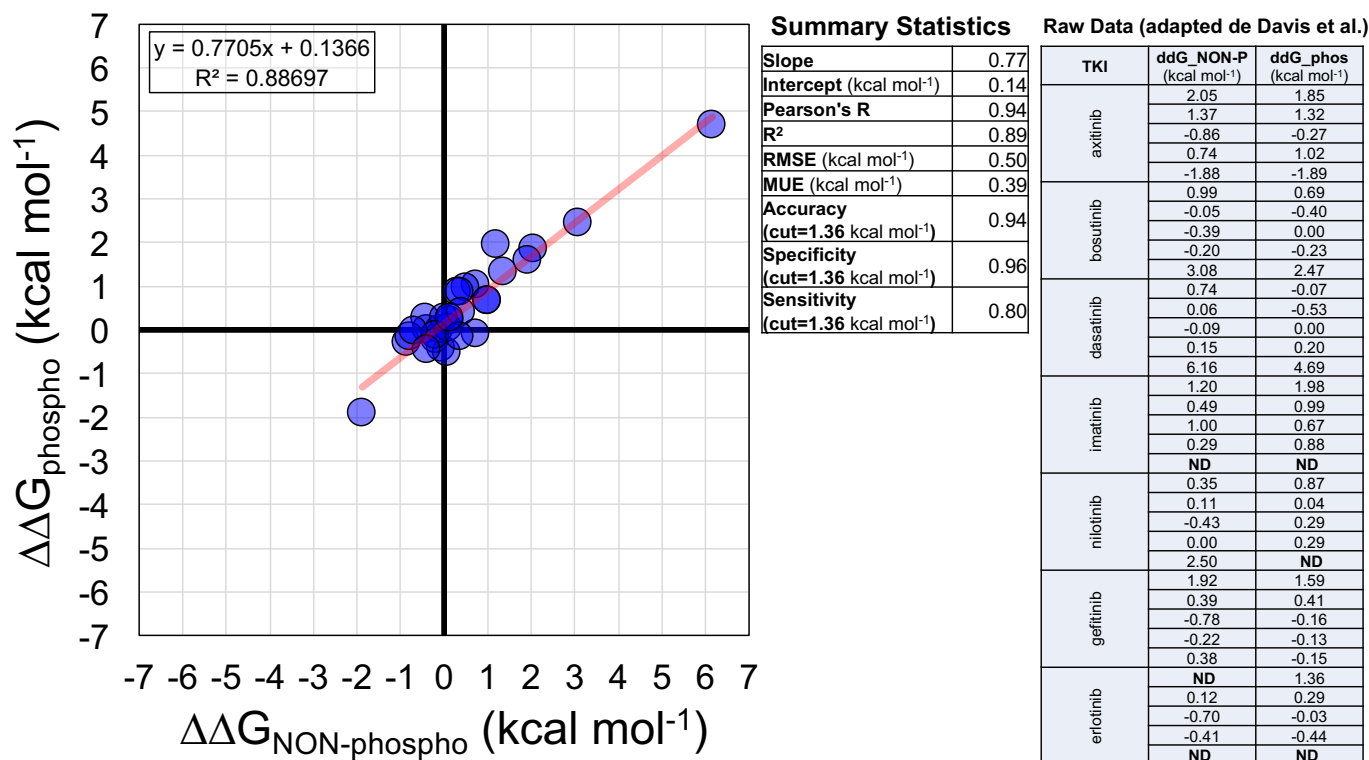


Figure 3.3: Comparison of 31 mutations for which phosphorylated and non-phosphorylated ΔK_d s were available. Scatter plot compares $\Delta\Delta G$ s (derived from the ΔK_d s) and contains the best-fit line with slope 0.77 and intercept 0.14. Summary statistics for this comparison are also shown. The raw $\Delta\Delta G$ s used for this comparison were adapted from [21]; kino-bead data for ponatinib was not available. ND: No data. TKI: Targeted kinase inhibitor.

3.4.2 Most mutations do not significantly reduce TKI potency

The majority of mutations do not lead to resistance by our 10-fold affinity loss threshold: 86.3% of the co-crystal set (n=113) and 86.8% of the total set (n=125). Resistance mutations, which are likely to result in a failure of therapy, constitute 13.7% of the co-crystal set (n=18) and 13.2% of the total set of mutations (n=19). The ΔpIC_{50} s for all 144 mutations are summarized in Table 3.2. Two mutations exceeded the dynamic range of the assays ($IC_{50} > 10,000$ nM); as these two mutations clearly raise resistance, we excluded them from quantitative analysis (RMSE and MUE) but included them in truth table analyses and classification metrics (accuracy, specificity, sensitivity).

Table 3.2: $\Delta\Delta G$ data derived from publicly available ΔpIC_{50} measurements and sources of mutation clinical-observation

Mutation	axitinib $\Delta\Delta G$ (kcal/mol)	bosutinib $\Delta\Delta G$ (kcal/mol)	dasatinib $\Delta\Delta G$ (kcal/mol)	imatinib $\Delta\Delta G$ (kcal/mol)	nilotinib $\Delta\Delta G$ (kcal/mol)	ponatinib $\Delta\Delta G$ (kcal/mol)	gefitinib $\Delta\Delta G$ (kcal/mol)	erlotinib $\Delta\Delta G$ (kcal/mol)	Source of Clinical Observation
M244V	-0.11	0.43	0.00	0.21	-0.13	0.00	nd	nd	A
L248R	0.31	1.50	0.65	2.33	2.15	0.58	nd	nd	B
L248V	0.32	0.56	0.55	0.64	0.33	0.17	nd	nd	A,C
G250E	0.27	0.11	0.41	1.01	0.60	0.30	nd	nd	A,C,D
Q252H	0.20	nd	nd	nd	nd	nd	-0.44	-0.13	A
Y253F	0.26	-0.34	0.24	1.90	1.48	0.30	-0.17	0.00	C
Y253H	0.03	nd	nd	nd	nd	nd	nd	nd	A,C,D
E255K	0.26	0.56	0.90	1.50	1.27	0.41	-0.11	-0.11	A,C,D
E255V	0.30	0.66	1.02	2.22	2.36	1.00	nd	nd	A,C
D276G	0.18	nd	nd	nd	nd	nd	nd	nd	C
E279K	-0.03	nd	nd	nd	nd	nd	nd	nd	C
E292L	0.03	nd	nd	nd	nd	nd	nd	nd	E
V299L	-0.88	1.70	1.24	0.23	0.28	0.17	nd	nd	C
T315A	-0.45	0.32	2.02	0.51	0.72	0.17	nd	nd	C
T315I	-1.27	2.45	5.08	2.32	3.75	0.41	nd	-0.15	C,D
T315V	-1.73	nd	nd	nd	nd	nd	nd	nd	B
F317C	nd	0.50	1.86	0.28	0.04	0.00	nd	nd	A ^g
F317I	nd	0.71	1.79	0.17	0.30	0.51	1.35	1.58	C
F317L	0.23	0.09	0.96	0.72	0.20	0.17	0.29	0.40	C,D
F317R	0.27	nd	nd	nd	nd	nd	nd	nd	B
F317V	0.28	1.72	2.36	0.97	0.33	0.72	nd	nd	C
M343T	0.21	nd	nd	nd	nd	nd	nd	nd	F ^h
M351T	-0.24	0.19	0.00	0.42	0.00	0.17	0.05	-0.08	A,C,D
E355A	nd	0.02	0.24	0.47	0.11	0.51	nd	nd	C
F359C	nd	-0.01	0.00	0.77	0.68	0.41	nd	nd	C
F359I	0.10	0.04	0.24	0.28	0.86	0.77	nd	nd	A
F359V	0.07	-0.11	0.00	0.32	0.60	0.17	nd	nd	A,C
L384M	0.06	nd	nd	nd	nd	nd	nd	nd	F ⁱ

Table 3.2 continued from previous page

Mutation	axitinib $\Delta\Delta G$ (kcal/mol)	bosutinib $\Delta\Delta G$ (kcal/mol)	dasatinib $\Delta\Delta G$ (kcal/mol)	imatinib $\Delta\Delta G$ (kcal/mol)	nilotinib $\Delta\Delta G$ (kcal/mol)	ponatinib $\Delta\Delta G$ (kcal/mol)	gefitinib $\Delta\Delta G$ (kcal/mol)	erlotinib $\Delta\Delta G$ (kcal/mol)	Source of Clinical Observation
H396R	0.25	-0.10	0.00	0.40	0.25	0.17	nd	nd	A ^j
F486S	0.05	nd	nd	nd	nd	nd	nd	nd	A ^k
E459K	nd	0.35	0.41	0.66	0.55	0.30	nd	nd	C

A: Gruber et al. ([155]),

B: Redaelli et al. ([156]),

C: Cortes et al. ([157]),

D: Branford et al. ([158]),

E: Press et al. ([159]),

F: Shah et al. ([121]),

G: F317C observed with $\Delta 27-183$,

^H: M343T observed as compound mutation with H396R,

^I: L384M observed as compound mutation with M343T,

^J: H396R observed as compound mutation with F486S,

^K: F486S observed as compound mutation with H396R

3.4.3 FEP+ predicts affinity changes for clinical Abl mutants

Figure 3.1B depicts the thermodynamic cycle that illustrates how we used relative free energy calculations to compute the change in ligand binding free energy in response to the introduction of a point mutation in the kinase (Figure 3.1C). From prior experience with relative alchemical free-energy calculations for ligand design, good initial receptor-ligand geometry was critical to obtaining accurate and reliable free energy predictions [78], so we first focused on the 131 mutations in Abl kinase across six TKIs for which wild-type Abl:TKI co-crystal structures were available. Figure 3.4 summarizes the performance of predicted binding free-energy changes ($\Delta\Delta G$) for all 131 mutants in this set for both a fast MM-GBSA physics-based method that only captures interaction energies for a single structure (Prime) and rigorous alchemical free-energy calculations (FEP+). Scatter plots compare experimental and predicted free-energy changes ($\Delta\Delta G$) and characterize the ability of these two techniques to predict experimental measurements. Statistical uncertainty in the predictions and experiment-to-experiment variability in the experimental values are shown as ellipse height and widths respectively. The value for experimental variability was $0.32 \text{ kcal mol}^{-1}$, which was the standard error computed from the cross-comparison in Figure 3.2. For FEP+, the uncertainty was taken to be the standard error of the average from three independent runs for a particular mutation, while Prime results are deterministic and are not contaminated by statistical uncertainty (see Methods).

To better assess whether discrepancies between experimental and computed $\Delta\Delta G$ s simply arise for known forcefield limitations or might indicate more significant effects, we incorporated an additional error model in which the forcefield error was taken to be a random error $\sigma_{\text{FF}} \approx 0.9 \text{ kcal mol}^{-1}$, a value

established from previous benchmarks on small molecules absent conformational sampling or protonation state issues [54]. Thin error bars in Figure 3.2 represent the overall estimated error due to both this forcefield error and experimental variability or statistical uncertainty.

To assess overall quantitative accuracy, we computed both root-mean-squared error (RMSE)—which is rather sensitive to outliers, and mean unsigned error (MUE). For Prime, the MUE was $1.16_{0.96}^{1.37}$ kcal mol⁻¹ and the RMSE was $1.72_{1.41}^{2.00}$ kcal mol⁻¹. FEP+, the alchemical free-energy approach, achieved a significantly higher level of quantitative accuracy with an MUE of $0.82_{0.69}^{0.95}$ kcal mol⁻¹ and an RMSE of $1.11_{0.91}^{1.30}$ kcal mol⁻¹. Notably, alchemical free energy calculations come substantially closer than MMGBSA approach to the minimum achievable RMSE of $0.81_{0.59}^{1.04}$ kcal mol⁻¹ (due to experimental error; Figure 3.2) for this dataset.

Abl:TKI Co-Crystal Structures

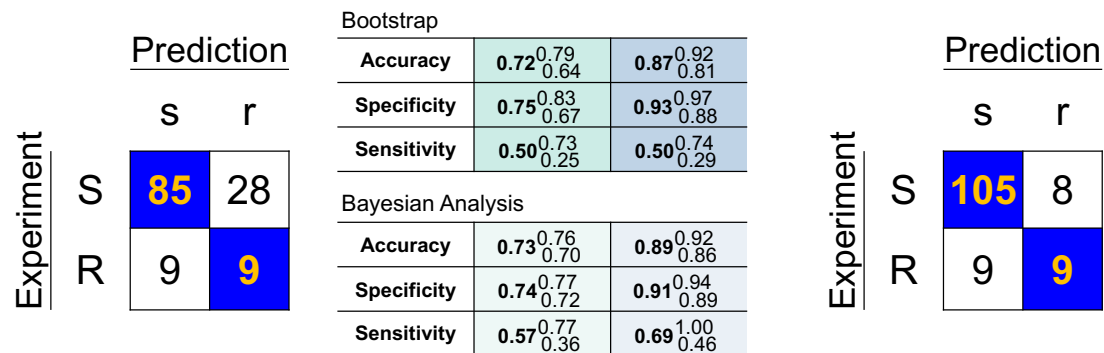
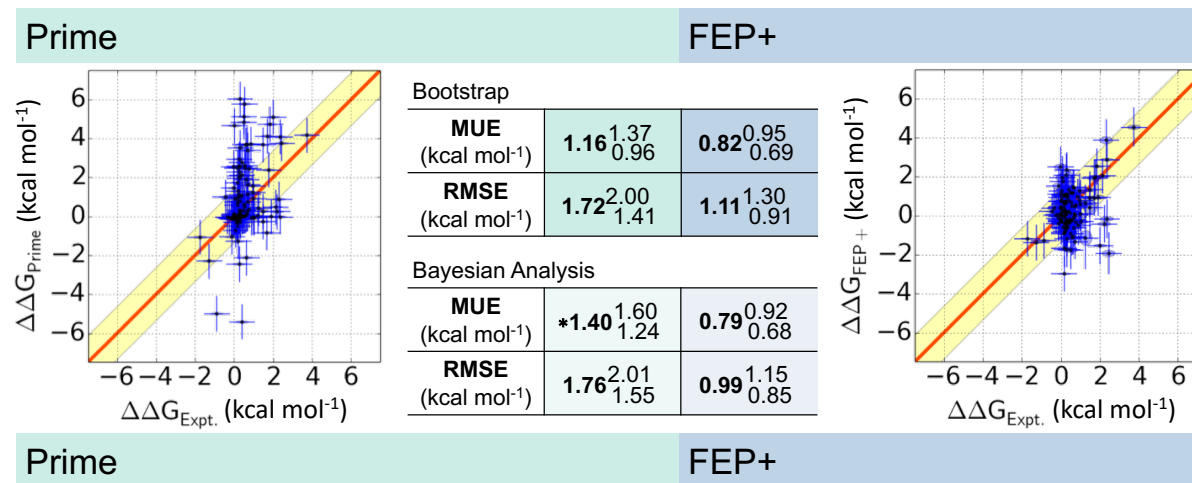
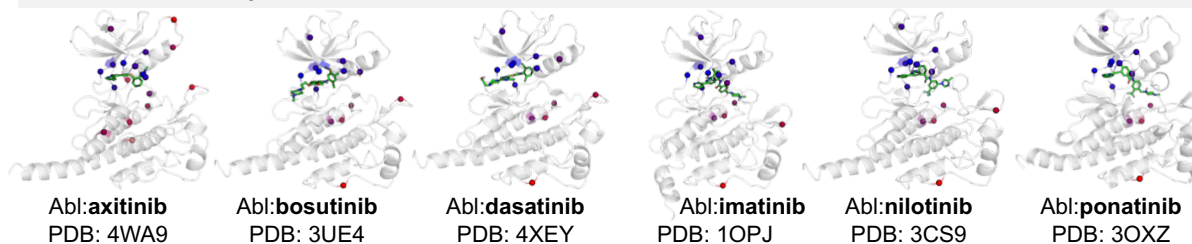


Figure 3.4: Continued on the next page

Figure 3.4: Comparison of experimentally-measured binding free-energy changes ($\Delta\Delta G$) for 131 clinically observed mutations and 6 selective kinase inhibitors for which co-crystal structures of wild-type kinase with inhibitor are available. *Top panel:* Abl:TKI co-crystal structures used in this study with locations of clinical mutants for each inhibitor highlighted (colored from blue to red for residues nearest to farthest from ligand) in relation to TKI (green sticks) on the corresponding Abl:TKI wild-type crystal structure. *Middle panel:* Scatter plots show Prime and FEP+ computed $\Delta\Delta G$ compared to experiment, with ellipse widths and heights ($\pm\sigma$) for experiment and FEP+ respectively. The red diagonal line indicates when prediction equals experiment, while the yellow shaded region indicates area in which predicted $\Delta\Delta G$ is within $1.36 \text{ kcal mol}^{-1}$ of experiment (corresponding to a ten-fold error in predicted affinity change). $\Delta\Delta G < 0 \text{ kcal mol}^{-1}$ denotes the mutation increases the susceptibility of the kinase to the inhibitor, while $\Delta\Delta G > 0 \text{ kcal mol}^{-1}$ denotes the mutation increases the resistance of the kinase to the inhibitor. The two mutations that were beyond the concentration limit of the assay (T315I/dasatinib, L248R/imatinib) were not plotted; 129 points were plotted. Truth tables of classification accuracy, sensitivity and specificity using two-classes. *Bottom panel:* Truth tables and classification results include T315I/dasatinib and L248R/imatinib; 131 points were used. For MUE, RMSE, and truth table performance statistics, sub/superscripts denote 95 % CIs. Variability in the experimental data is shown as ellipse widths and uncertainty in our calculations is shown as ellipse heights. Experimental variability was computed as the standard error between IC_{50} -derived $\Delta\Delta G$ measurements made by different labs, $0.32 \text{ kcal mol}^{-1}$. The statistical uncertainty in the Prime calculations was zero because the method is deterministic ($\sigma_{\text{cal}} = 0$), while the uncertainty in the FEP+ calculations was reported as the standard error, σ_{cal} , of the mean of the predicted $\Delta\Delta G$ s from three independent runs. To better highlight true outliers unlikely to simply result from expected forcefield error, we presume forcefield error ($\sigma_{\text{FF}} \approx 0.9 \text{ kcal mol}^{-1}$ [54]) also behaves as a random error, and represent the total estimated statistical and forcefield error ($\sqrt{\sigma_{\text{FF}}^2 + \sigma_{\text{exp/cal}}^2}$) as vertical error bars. The horizontal error bars for the experiment (σ_{exp}) was computed as the standard error between ΔpIC_{50} and ΔK_d measurements, $0.58 \text{ kcal mol}^{-1}$. For Prime, *MUE highlights that the Bayesian model yields a value for MUE that is noticeably larger than MUE for observed data due to the non-Gaussian error distribution of Prime.

3.4.4 FEP+ accurately classifies affinity changes for Abl mutants

While quantitative accuracy (MUE, RMSE) is a principle metric of model performance, an application of potential interest is the ability to classify mutations as causing resistance to a specific TKI. To characterize the accuracy with which Prime and FEP+ classified mutations in a manner that might be therapeutically relevant, we classified mutations by their experimental impact on the binding affinity as susceptible (affinity for mutant is diminished by no more than 10-fold, $\Delta\Delta G \leq 1.36$ kcal mol⁻¹) or as resistant (affinity for mutant is diminished by least 10-fold, $\Delta\Delta G > 1.36$ kcal mol⁻¹). Summary statistics of experimental and computational predictions of these classes are shown in Figure 3.2 (bottom) as truth tables (also known as confusion matrices).

The simple minimum-energy scoring method Prime correctly classified 9 of the 18 resistance mutations in the dataset while merely 85 of the 113 susceptible mutations were correctly classified (28 false positives). In comparison, the alchemical free-energy method FEP+, which includes entropic and enthalpic contributions as well as explicit representation of solvent, correctly classified 9 of the 18 resistance mutations while a vast majority, 105, of the susceptible mutations were correctly classified (merely 8 false positives). Prime achieved a classification accuracy of $0.72_{0.64}^{0.79}$, while FEP+ achieved an accuracy that is significantly higher (both in a statistical sense and in overall magnitude), achieving an accuracy of $0.87_{0.81}^{0.92}$. Sensitivity (also called true positive rate) and specificity (true negative rate) are also informative statistics in assessing the performance of a binary classification scheme. For Prime, the sensitivity was $0.50_{0.25}^{0.73}$, while the specificity was $0.75_{0.67}^{0.83}$. To put this in perspective, a CML patient bearing a resistance mutation in the kinase domain of Abl has an equal chance of Prime correctly predicting this mutation would be resistant to one of the TKIs considered here,

while if the mutation was susceptible, the chance of correct prediction would be ~75%. By contrast, the classification specificity of FEP+ was substantially better. For FEP+, the sensitivity was $0.50_{0.29}^{0.74}$ while the specificity was $0.93_{0.88}^{0.97}$. There is a very high probability that FEP+ will correctly predict that one of the eight TKIs studied here will remain effective for a patient bearing a susceptible mutation.

3.4.5 How reliant are classification results on choice of cutoff?

Previous work by O'Hare et al. utilized TKI-specific thresholds for dasatinib, imatinib, and nilotinib [160], which were ~ 2 kcal mol⁻¹. Figure 3.5 shows that when our classification threshold was increased to a 20-fold change in binding (1.77 kcal mol⁻¹), FEP+ correctly classified 8 of the 13 resistant mutations and with a threshold of 100-fold change in binding (2.72 kcal mol⁻¹), FEP+ correctly classified the only two resistant mutations (T315I/dasatinib and T315I/nilotinib). With the extant multilayered and multinodal decision-making algorithms used by experienced oncologists to manage their patients' treatment, or by medicinal chemists to propose candidate compounds for clinical trials, the resistant or susceptible cutoffs could be selected with more nuance than the simple 10-fold affinity threshold we consider here. With a larger affinity change cutoff, for example, the accuracy with which physical models predict resistance mutations increases beyond 90% (Figure 3.5). For the alchemical approach, the two-class accuracy was $0.92_{0.87}^{0.96}$ when an affinity change cutoff of 20-fold was used while using an affinity change cutoff of 100-fold further improved the two-class accuracy to $0.98_{0.96}^{1.00}$.

cutoff	Prime			FEP+					
	0.95 kcal/mol	1.77 kcal/mol	2.72 kcal/mol	0.95 kcal/mol	1.77 kcal/mol	2.72 kcal/mol			
axitinib	Prediction s r			Prediction s r			Prediction s r		
	Experiment S 20 6 R 0 0	Experiment S 23 3 R 0 0	Experiment S 24 2 R 0 0	Experiment S 20 6 R 0 0	Experiment S 23 3 R 0 0	Experiment S 26 0 R 0 0			
bosutinib	Prediction s r			Prediction s r			Prediction s r		
	Experiment S 13 4 R 1 3	Experiment S 16 4 R 0 1	Experiment S 16 5 R 0 0	Experiment S 13 4 R 2 2	Experiment S 18 1 R 2 0	Experiment S 21 0 R 0 0			
dasatinib	Prediction s r			Prediction s r			Prediction s r		
	Experiment S 11 2 R 2 6	Experiment S 14 2 R 1 4	Experiment S 17 3 R 1 0	Experiment S 11 2 R 3 5	Experiment S 16 0 R 1 4	Experiment S 19 1 R 0 1			
imatinib	Prediction s r			Prediction s r			Prediction s r		
	Experiment S 7 7 R 5 2	Experiment S 10 7 R 3 1	Experiment S 18 2 R 0 0	Experiment S 13 1 R 4 3	Experiment S 17 0 R 2 2	Experiment S 19 1 R 0 0			
nilotinib	Prediction s r			Prediction s r			Prediction s r		
	Experiment S 8 8 R 4 1	Experiment S 11 7 R 2 1	Experiment S 14 6 R 0 1	Experiment S 10 6 R 3 2	Experiment S 18 0 R 1 2	Experiment S 20 0 R 0 1			
ponatinib	Prediction s r			Prediction s r			Prediction s r		
	Experiment S 14 6 R 1 0	Experiment S 18 3 R 0 0	Experiment S 21 0 R 0 0	Experiment S 17 3 R 1 0	Experiment S 21 0 R 0 0	Experiment S 21 0 R 0 0			

Figure 3.5: TKI-by-TKI truth tables with increasingly large classification cutoffs. Truth tables for the six TKIs (axitinib, bosutinib, dasatinib, imatinib, nilotinib, and ponatinib) using Prime (left, green) and FEP+ (right, blue) with classification cutoff values defining whether mutations are susceptible (S, experiment; s, prediction) or resistant (R, experiment; r, prediction). A mutation is susceptible if $\Delta\Delta G \leq \text{cutoff}$ or resistant if $\Delta\Delta G > \text{cutoff}$.

3.4.6 Bayesian analysis can estimate the true error

The statistical metrics—MUE, RMSE, accuracy, specificity, and sensitivity—discussed above are based on analysis of the apparent performance of the observed modeling results compared with the observed experimental data via sample statistics. However, this analysis considers a limited number of mutants, and both measurements and computed values are contaminated with experimental or statistical error. To obtain an estimate of the *intrinsic performance* of our physical modeling approaches, accounting for known properties of the experimental variability and statistical uncertainties, we used a hierarchical Bayesian model (detailed in the Methods) to infer posterior predictive distributions from which expectations and 95% predictive intervals could be obtained. The results of this analysis are presented in Figure 3.4 (central tables).

FEP+ is significantly better than Prime at predicting the impact of mutations on TKI binding affinities, as the apparent performance (using the original observations) as well as the intrinsic performance (where Bayesian analysis was used to correct for statistical uncertainty or experimental variation) were well-separated outside their 95% confidence intervals in nearly all metrics.

Applying the Bayesian model, the MUE and RMSE for FEP+ was $0.79_{0.68}^{0.92}$ kcal mol⁻¹ and $0.99_{0.85}^{1.15}$ kcal mol⁻¹ respectively (N=129). For the classification metrics accuracy, specificity, and sensitivity, the model yields $0.89_{0.86}^{0.92}$, $0.91_{0.89}^{0.94}$, and $0.69_{0.46}^{1.00}$ respectively (N=131). The intrinsic RMSE and MUE of Prime was $1.76_{1.55}^{2.01}$ kcal mol⁻¹ and $1.40_{1.24}^{1.60}$ kcal mol⁻¹ (N=129) respectively, and the classification accuracy, specificity, and sensitivity was $0.73_{0.70}^{0.76}$, $0.74_{0.72}^{0.77}$, and $0.57_{0.36}^{0.77}$ respectively (N=131). The intrinsic MUE of Prime obtained by this analysis is larger than the observed MUE reflecting the non-Gaussian, fat-tailed error distributions of Prime results.

3.4.7 How transferable is FEP+ across the six TKIs?

The impact of point mutations on drug binding are not equally well predicted for the six TKIs. Figure 3.6 expands the results in Figure 3.4 on a TKI-by-TKI basis to dissect the particular mutations in the presence of a specific TKI. Prime and FEP+ correctly predicted that most mutations in this dataset (N=26) do not raise resistance to axitinib, though FEP+ predicted 4 false positives compared with 3 false positives by Prime. The MUE and RMSE of FEP+ was excellent for this inhibitor, $0.70_{0.50}^{0.93}$ kcal mol⁻¹ and $0.91_{0.64}^{1.14}$ kcal mol⁻¹ respectively. While the classification results for bosutinib (N=21) were equally well predicted by Prime as by FEP+, FEP+ was still able to achieve superior, but not highly significant, predictive performance for the quantitative metrics MUE and RMSE, which were $0.96_{0.55}^{1.42}$ kcal mol⁻¹ and $1.41_{0.77}^{1.97}$ kcal mol⁻¹ respectively (FEP+) and $1.13_{0.60}^{1.83}$ kcal mol⁻¹ and $1.80_{0.92}^{2.62}$ kcal mol⁻¹ respectively (Prime). For dasatinib, FEP+ achieved an MUE and RMSE of $0.76_{0.49}^{1.13}$ kcal mol⁻¹ and $1.07_{0.59}^{1.57}$ kcal mol⁻¹ respectively whereas the results were, as expected, less quantitatively predictive for Prime (N=20). The results for imatinib were similar to those of dasatinib above, where the MUE and RMSE for FEP+ were $0.82_{0.53}^{1.15}$ kcal mol⁻¹ and $1.09_{0.69}^{1.43}$ kcal mol⁻¹ respectively (N=20). Nilotinib, a derivative of imatinib, led to nearly identical quantitative performance results for FEP+ with an MUE and RMSE of $0.82_{0.57}^{1.12}$ kcal mol⁻¹ and $1.06_{0.69}^{1.39}$ kcal mol⁻¹ respectively (N=21). Similar to axitinib, ponatinib presented an interesting case because there were no mutations in this dataset that raised resistance to it. Despite the wide dynamic range in the computed values of $\Delta\Delta G$ for other inhibitors, FEP+ correctly predicted a very narrow range of $\Delta\Delta G$ s for this drug. This is reflected in the MUE and RMSE of $0.87_{0.62}^{1.16}$ kcal mol⁻¹ and $1.09_{0.70}^{1.46}$ kcal mol⁻¹ respectively, which are in-line with the MUEs and RMSEs for the other TKIs.

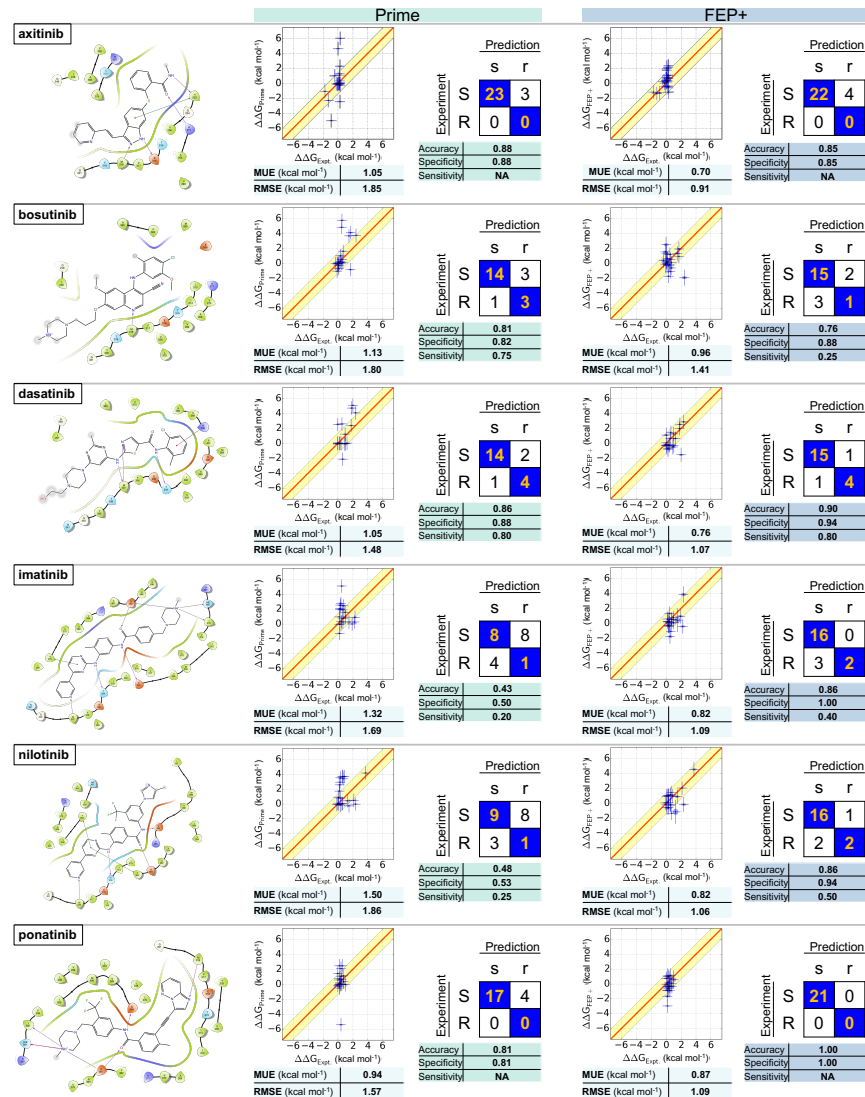


Figure 3.6: Continued on the next page

Figure 3.6: **Physical modeling accuracy in computing the impact of clinical Abl mutations on selective inhibitor binding.** Ligand interaction diagrams for six selective FDA-approved tyrosine kinase inhibitors (TKIs) for which co-crystal structures with Abl were available (left). Comparisons for clinically-observed mutations are shown for FEP+ (right) and Prime (left). For each ligand, computed *vs.* experimental binding free energies ($\Delta\Delta G$) are plotted with MUE and RMSE (units of kcal mol⁻¹) depicted below. Truth tables are shown to the right. Rows denote *true* susceptible (S, $\Delta\Delta G \leq 1.36$ kcal mol⁻¹) or resistant (R, $\Delta\Delta G > 1.36$ kcal mol⁻¹) experimental classes using a 1.36 kcal mol⁻¹ (10-fold change) threshold; columns denote *predicted* susceptible (s, $\Delta\Delta G \leq 1.36$ kcal mol⁻¹) or resistant (r, $\Delta\Delta G > 1.36$ kcal mol⁻¹). Correct predictions populate diagonal elements (orange text), incorrect predictions populate off-diagonals. Accuracy, specificity, and sensitivity for two-class classification are shown below the truth table. Elliptical point sizes and error bars in the scatter plots depict estimated uncertainty/variability and error respectively ($\pm\sigma$) of FEP+ values (vertical size) and experimental values (horizontal size). Note: The sensitivity for axitinib and ponatinib is NA, because there is no resistant mutation for these two drugs.

3.4.8 Understanding the origin of mispredictions

Resistance mutations that are mispredicted as susceptible (false negatives) are particularly critical because they might mislead the clinician or drug designer into believing the inhibitor will remain effective against the target. Which resistance mutations did FEP+ mispredict as susceptible? Nine mutations were classified by FEP+ to be susceptible when experimentally measured ΔpIC_{50} data indicate the mutations should have increased resistance according to our 10-fold affinity cutoff for resistance. Notably, the 95% confidence intervals for five of these mutations spanned the 1.36 kcal mol⁻¹ threshold, indicating these misclassifications are not statistically significant when the experimental error and statistical uncertainty in FEP+ are accounted for: bosutinib/L248R ($\Delta\Delta G_{FEP+}=1.32^{1.94}_{0.70}$ kcal mol⁻¹), imatinib/E255K ($\Delta\Delta G_{FEP+}=0.43^{3.05}_{-2.19}$ kcal mol⁻¹), imatinib/Y253F ($\Delta\Delta G_{FEP+}=0.95^{1.64}_{0.26}$ kcal mol⁻¹), and nilotinib/Y253F ($\Delta\Delta G_{FEP+}=0.89^{1.69}_{0.09}$ kcal mol⁻¹). The bosutinib/V299L mutation was also not significant because the experimental $\Delta\Delta G$, $1.70^{2.33}_{1.08}$ kcal mol⁻¹, included the 1.36 kcal mol⁻¹ cutoff; the value of $\Delta\Delta G$ predicted by FEP+ for this mutation was $0.91^{1.02}_{0.79}$ kcal mol⁻¹, the upper bound of the predicted value was within 0.06 kcal mol⁻¹ of the lower bound of the experimental value.

Four mutations, however, were misclassified to a degree that is statistically significant given their 95% confidence intervals: dasatinib/T315A, bosutinib/T315I, imatinib/E255V, and nilotinib/E255V. For dasatinib/T315A, although the T315A mutations for bosutinib, imatinib, nilotinib, and ponatinib were correctly classified as susceptible, the predicted free energy changes for these four TKIs were consistently much more negative than the corresponding experimental measurements, just as for dasatinib/T315A, indicating there might be a generic driving force contributing to the errors in T315A mutations for these

five TKIs. Abl is known to be able to adopt many different conformations (including DFG-in and DFG-out), and it is very likely that the T315A mutation will induce conformational changes in the apo protein [161], which was not adequately sampled in the relatively short simulations, leading to the errors for T315A mutations for these TKIs. By comparison, the T315I mutations for axitinib, bosutinib, imatinib, nilotinib, and ponatinib were all accurately predicted with the exception of bosutinib/T315I being the only misprediction, suggesting an issue specific to bosutinib. The complex electrostatic interactions between the 2,4-dichloro-5-methoxyphenyl ring in bosutinib and the adjacent positively charged amine of the catalytic Lys271 may not be accurately captured by the fixed-charge OPLS3 force field, leading to the misprediction for bosutinib/T315I mutation.

Insufficient sampling might also belie the imatinib/E255V and nilotinib/E255V mispredictions because they reside in the highly flexible P-loop. Since E255V was a charge change mutation, we utilized a workflow that included a transmutable explicit ion (see Methods). The distribution of these ions in the simulation box around the solute might not have converged to their equilibrium state on the relatively short timescale of our simulations (5 ns), and the insufficient sampling of ion distributions coupled with P-loop motions might lead to misprediction of these two mutations.

3.4.9 How strongly is accuracy affected for docked TKIs?

To assess the potential for utilizing physics-based approaches in the absence of a high-resolution experimental structure, we generated models of Abl bound to two TKIs—erlotinib and gefinitib—for which co-crystal structures with wild-type

kinase are not currently available. In Figure 3.7, we show the Abl:erlotinib and Abl:gefitinib complexes that were generated using a docking approach (Glide-SP, see Methods). These two structures were aligned against the co-crystal structures of EGFR:erlotinib and EGFR:gefitinib to highlight the structural similarities between the binding pockets of Abl and EGFR and the TKI binding mode in Abl versus EGFR. As an additional test of the sensitivity of FEP+ to system preparation, a second set of Abl:erlotinib and Abl:gefitinib complexes was generated in which crystallographic water coordinates were transferred to the docked inhibitor structures (see Methods).

Alchemical free-energy simulations were performed on 13 mutations between the two complexes; 7 mutations for erlotinib and 6 mutations for gefitinib. The quantitative accuracy of FEP+ in predicting the value of $\Delta\Delta G$ was excellent—MUE and RMSE of $0.58_{0.33}^{0.86}$ kcal mol⁻¹ and $0.80_{0.44}^{1.09}$ kcal mol⁻¹ respectively if crystal waters are omitted, and $0.50_{0.26}^{0.78}$ kcal mol⁻¹ and $0.69_{0.35}^{0.97}$ kcal mol⁻¹ if crystal waters were restored after docking. Encouragingly, these results indicate that our initial models of Abl bound to erlotinib and gefitinib were reliable because the accuracy and dependability of our FEP+ calculations were not sensitive to crystallographic waters. Our secondary concern was the accuracy with which the approach classified mutations as resistant or susceptible.

While the results presented in (Figure 3.7) indicate that FEP+ is capable of achieving good quantitative accuracy when a co-crystal structure is unavailable, it is important to understand why a mutation was predicted to be susceptible but was determined experimentally to be resistant. F317I was the one mutation that increased resistance to erlotinib (or gefitinib) because it destabilized binding by more than 1.36 kcal mol⁻¹— $1.35_{1.03}^{1.67}$ kcal mol⁻¹ (gefitinib) and $1.58_{1.26}^{1.90}$ kcal mol⁻¹ (erlotinib), but the magnitude of the experimental uncertainty means we are

unable to confidently discern whether this mutation induces more than 10-fold resistance to either TKI. Therefore, the one misclassification by FEP+ in Figure 3.7 is not statistically significant and the classification metrics presented there underestimate the nominal performance of this alchemical free-energy method.

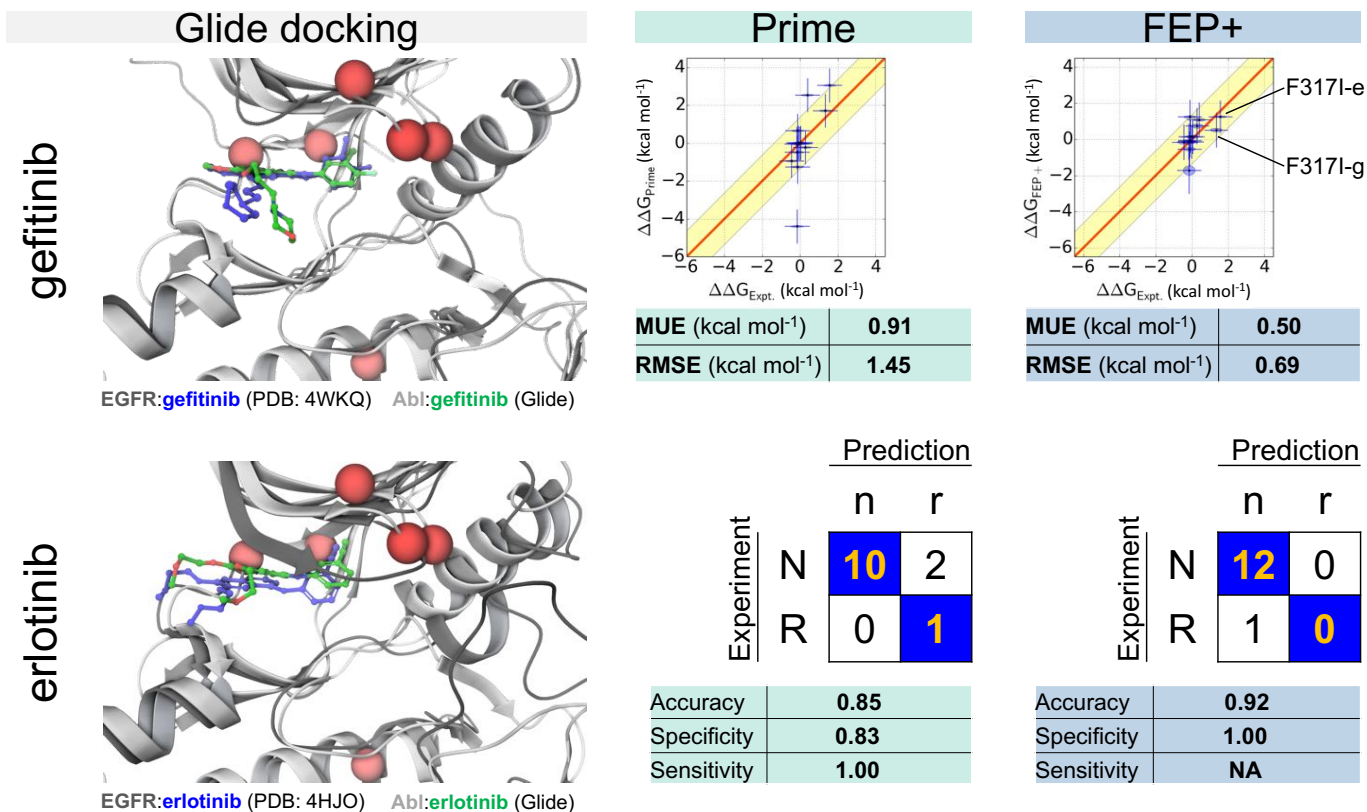


Figure 3.7: **Predicting resistance mutations using FEP+ for inhibitors for which co-crystal structures with wild-type kinase are not available.** The docked pose of Abl:erlotinib is superimposed on the co-crystal structure of EGFR:erlotinib; erlotinib docked to Abl (light gray) is depicted in green and erlotinib bound to EGFR (dark gray) is depicted in blue. The docked pose of Abl:gefitinib is superimposed on the co-crystal structure of EGFR:gefitinib; gefitinib docked to Abl (light gray) is depicted in green and gefitinib bound to EGFR (dark gray) is depicted in blue. The locations of clinical mutants for each inhibitor are highlighted (red spheres). The overall RMSEs and MUEs for Prime (center) and FEP+ (right) and two-class accuracies are also shown in the figure. Computed free energy changes due to the F317I mutation for erlotinib (-e) and gefitinib (-g) are highlighted in the scatter plot. FEP+ results are based on the docked models prepared with crystal waters added back while the Prime (an implicit solvent model) results are based on models without crystallographic water.

3.5 Discussion and Conclusions

Physics-based modeling can reliably predict when a mutation elicits resistance to therapy The results presented in this work are summarized in Table 3.3. The performance metrics summarized in Table 3.3 indicates that the set of 131 mutations for the six TKIs in which co-crystal structures were available is on par with the complete set (144 mutations), which included results based on Abl:TKI complexes generated from docking models. The performance results for the 13 mutations for the two TKIs (erlotinib and gefitinib) in which co-crystal structures were unavailable exhibited good quantitative accuracy (MUE and RMSE) and good classification power.

Overall (N=144), the MM-GBSA approach Prime classified mutations with good accuracy ($0.73_{0.66}^{0.80}$) and specificity ($0.76_{0.69}^{0.84}$) while the alchemical approach FEP+ was a significant improvement in classification accuracy ($0.88_{0.82}^{0.93}$) and specificity ($0.94_{0.89}^{0.98}$). The quantitative accuracy with which Prime was able to predict the experimentally measured change in Abl:TKI binding (N=142) characterized by RMSE and MUE was $1.70_{1.40}^{1.98}$ kcal mol⁻¹ and $1.14_{0.93}^{1.35}$ kcal mol⁻¹ respectively. In stark contrast, the quantitative accuracy of FEP+ was statistically superior to Prime with an RMSE and an MUE of $1.07_{0.89}^{1.26}$ kcal mol⁻¹ and $0.79_{0.67}^{0.92}$ kcal mol⁻¹ respectively.

From the perspective of a clinician, classification rate would be an important metric to measure the predictive power of technologies such as Prime and FEP+. To test the hypothesis that reducing the large spread in Prime predictions could improve its classification rate, we scaled the computed relative free energies and recalculated the performance metrics (Data shown in publication [40]). As expected, the MUE and RMSE were improved but the specificity of Prime was

drastically diminished. Scaling FEP+ eliminated its sensitivity and a naïve model (all $\Delta\Delta G_s = 0.00 \text{ kcal mol}^{-1}$) had zero sensitivity. Lastly, we constructed a consensus model in which free energies were a weighted average of scaled Prime and FEP+. This model also had zero sensitivity.

To address the impact of picking a cutoff to classify predicted free energies as resistant or sensitizing, we computed ROC curves for the various predicted datasets: Prime, FEP+, naïve model, and consensus model (Data shown in publication [40]). ROC curves and ROC-AUCs for scaled and non-scaled Prime were identical, as well as scaled and non-scaled FEP+, because ROC curves are independent of a linear transformation on the data. ROC-AUC for FEP+ was $0.75_{0.61}^{0.90}$ (n=144); ROC-AUC for Prime was $0.66_{0.52}^{0.81}$ (n=144); ROC-AUCs for the naïve model and consensus model were $0.50_{0.50}^{0.50}$ (n=144) and $0.78_{0.67}^{0.90}$ (n=144) respectively. These results show that Prime has poor discriminatory power (ROC-AUC in [0.6,0.7]) while FEP+ has fair discriminatory power (ROC-AUC in [0.7,0.8]).

Table 3.3: Summary of FEP+ and Prime statistics in predicting mutational resistance or sensitivity to FDA-approved TKIs

Dataset	Method	N_{quant}	MUE (kcal mol ⁻¹)	RMSE (kcal mol ⁻¹)	N_{class}	Accuracy	Specificity	Sensitivity
all	FEP+	142	0.79 ^{0.92} _{0.67}	1.07 ^{1.26} _{0.89}	144	0.88 ^{0.93} _{0.82}	0.94 ^{0.98} _{0.89}	0.47 ^{0.69} _{0.25}
all	Prime	142	1.14 ^{1.35} _{0.93}	1.70 ^{1.98} _{1.40}	144	0.73 ^{0.80} _{0.66}	0.76 ^{0.84} _{0.69}	0.53 ^{0.76} _{0.30}
xtals	FEP+	129	0.82 ^{0.95} _{0.69}	1.11 ^{1.30} _{0.91}	131	0.87 ^{0.92} _{0.81}	0.93 ^{0.97} _{0.88}	0.50 ^{0.74} _{0.29}
xtals	Prime	129	1.16 ^{1.37} _{0.96}	1.72 ^{2.00} _{1.41}	131	0.72 ^{0.79} _{0.64}	0.75 ^{0.83} _{0.67}	0.50 ^{0.73} _{0.25}
axitinib	FEP+	26	0.70 ^{0.93} _{0.50}	0.91 ^{1.14} _{0.64}	26	0.85 ^{0.96} _{0.69}	0.85 ^{0.96} _{0.73}	NA
axitinib	Prime	26	1.05 ^{1.71} _{0.53}	1.85 ^{2.61} _{0.96}	26	0.88 ^{1.00} _{0.73}	0.88 ^{1.00} _{0.71}	NA
bosutinib	FEP+	21	0.96 ^{1.42} _{0.55}	1.41 ^{1.97} _{0.77}	21	0.76 ^{0.93} _{0.57}	0.88 ^{1.00} _{0.71}	0.25 ^{1.00} _{0.00}
bosutinib	Prime	21	1.13 ^{1.83} _{0.60}	1.80 ^{2.62} _{0.92}	21	0.81 ^{0.95} _{0.62}	0.82 ^{1.00} _{0.62}	0.75 ^{1.00} _{0.00}
dasatinib	FEP+	20	0.76 ^{1.13} _{0.49}	1.07 ^{1.57} _{0.59}	21	0.90 ^{1.00} _{0.76}	0.94 ^{1.00} _{0.79}	0.80 ^{1.00} _{0.33}
dasatinib	Prime	20	1.05 ^{1.54} _{0.61}	1.48 ^{1.92} _{0.95}	21	0.86 ^{1.00} _{0.71}	0.88 ^{1.00} _{0.69}	0.80 ^{1.00} _{0.33}
imatinib	FEP+	20	0.82 ^{1.15} _{0.53}	1.09 ^{1.43} _{0.69}	21	0.86 ^{1.00} _{0.71}	1.00 ^{1.00} _{1.00}	0.40 ^{0.83} _{0.00}
imatinib	Prime	20	1.32 ^{1.81} _{0.91}	1.69 ^{2.26} _{1.15}	21	0.43 ^{0.67} _{0.24}	0.50 ^{0.75} _{0.25}	0.20 ^{0.67} _{0.00}
nilotinib	FEP+	21	0.82 ^{1.12} _{0.57}	1.06 ^{1.39} _{0.69}	21	0.86 ^{1.00} _{0.67}	0.94 ^{1.00} _{0.80}	0.50 ^{1.00} _{0.00}
nilotinib	Prime	21	1.50 ^{1.97} _{1.06}	1.86 ^{2.25} _{1.43}	21	0.48 ^{0.67} _{0.24}	0.53 ^{0.75} _{0.29}	0.25 ^{1.00} _{0.00}
ponatinib	FEP+	21	0.87 ^{1.16} _{0.62}	1.09 ^{1.46} _{0.70}	21	1.00 ^{1.00} _{1.00}	1.00 ^{1.00} _{1.00}	NA
ponatinib	Prime	21	0.94 ^{1.54} _{0.50}	1.57 ^{2.44} _{0.69}	21	0.81 ^{0.95} _{0.62}	0.81 ^{0.95} _{0.62}	NA
Glide	FEP+	13	0.50 ^{0.78} _{0.26}	0.69 ^{0.97} _{0.35}	13	0.92 ^{1.00} _{0.77}	1.00 ^{1.00} _{1.00}	0.00 ^{0.00} _{0.00}
Glide	Prime	13	0.91 ^{1.56} _{0.39}	1.45 ^{2.22} _{0.54}	13	0.85 ^{1.00} _{0.62}	0.83 ^{1.00} _{0.58}	1.00 ^{1.00} _{0.00}

N_{quant} : Number of mutations for which quantitative metrics were evaluated; N_{class} : Number mutations for which classification metrics were evaluated; All: All mutations; xtals: All mutations for which co-crystal structures were available; Glide: erlotinib and gefitinib
Accuracy, specificity, and sensitivity were computed to assess two-class prediction performance:
resistant ($\Delta\Delta G > 1.36$ kcal mol⁻¹) or *susceptible* ($\Delta\Delta G \leq 1.36$ kcal mol⁻¹).

95% CIs (sub-/superscripts) were estimated from 1000 bootstrap replicates. Note: The sensitivity for axitinib and ponatinib is NA, because there is no resistant mutation for these two drugs.

Hierarchical Bayesian model estimates global performance A hierarchical Bayesian approach was developed to estimate the intrinsic accuracy of the models when the noise in the experimental and predicted values of $\Delta\Delta G$ was accounted for. Utilizing this approach, the MUE and RMSE for Prime was found to be $1.39_{1.23}^{1.58}$ kcal mol⁻¹ and $1.75_{1.55}^{1.98}$ kcal mol⁻¹ (N=142) respectively. The accuracy, specificity, and sensitivity of Prime was found using this method to be $0.74_{0.71}^{0.76}$, $0.75_{0.73}^{0.77}$, and $0.59_{0.40}^{0.78}$ (N=144) respectively. The MUE and RMSE of FEP+ was found to be $0.76_{0.66}^{0.87}$ kcal mol⁻¹ and $0.95_{0.82}^{1.09}$ kcal mol⁻¹ (N=142) respectively, which is significantly better than Prime. Likewise, a clearer picture of the true classification accuracy, specificity, and sensitivity of FEP+ was found— $0.90_{0.86}^{0.93}$, $0.92_{0.90}^{0.95}$, and $0.68_{0.46}^{1.00}$ respectively.

The high accuracy of FEP+ is very encouraging, and the accuracy can be further improved with more accurate modeling of a number of physical chemical effects not currently considered by the method. While highly optimized, the fixed-charged OPLS3 [54] force field can be further improved by explicit consideration of polarizability effects [162], as hinted by some small-scale benchmarks [163]. These features could be especially important for bosutinib, whose 2,4-dichloro-5-methoxyphenyl ring is adjacent to the positively charged amine of the catalytic Lys271. Many simulation programs also utilize a long-range isotropic analytical dispersion correction intended to correct for the truncation of dispersion interactions at finite cutoff, which can induce an error in protein-ligand binding free energies that depends on the number of ligand heavy atoms being modified [164]; recently, efficient Lennard-Jones PME methods [165, 166] and perturbation schemes [164] have been developed that can eliminate the errors associated with this truncation. While the currently employed methodology for alchemical transformations involving a change in system charge (see Methods) reduces artifacts that depend on the simulation box

size and periodic boundary conditions, the explicit ions that were included in these simulations may not have sufficiently converged to their equilibrium distributions in these relatively short simulations. Kinases and their inhibitors are known to possess multiple titratable sites with either intrinsic or effective pK_a s near physiological pH, while the simulations here treat protonation states and proton tautomers fixed throughout the bound and unbound states; the accuracy of the model can be further improved with the protonation states or tautomers shift upon binding or mutation considered [167, 168]. Similarly, some systems display significant salt concentration dependence [169], while the simulations for some systems reported here did not rigorously mimic all aspects of the experimental conditions of the cell viability assays.

Experimentally observed IC_{50} changes can be caused by other physical

mechanisms While we have shown that predicting the direct impact of mutations on the binding affinity of ATP-competitive tyrosine kinase inhibitors for a single kinase conformation has useful predictive capacity, many additional physical effects that can contribute to cell viability are not currently captured by examining only the predicted change in inhibitor binding affinity. For example, kinase missense mutations can also shift the populations of kinase conformations (which may affect ATP and inhibitor affinities differentially), modulate ATP affinity, modulate affinity for protein substrate, or modulate the ability of the kinase to be regulated or bounded by scaffolding proteins. While many of these effects are in principle tractable by physical modeling in general it is valuable to examine our mispredictions and outliers to identify whether any of these cases are likely to induce resistance (as observed by ΔpIC_{50} shifts) by one of these alternative mechanisms.

Other physical mechanisms of resistance are likely similarly computable. A simple threshold of 10-fold TKI affinity change is a crude metric for classifying resistance or susceptibility due to the myriad biological factors that contribute to the efficacy of a drug in a person. Except for affecting the binding affinity of inhibitors, missense mutations can also cause drug resistance through other physical mechanisms including induction of splice variants or alleviation of feedback. While the current study only focused on the effect of mutation on drug binding affinity, resistance from these other physical mechanisms could be similarly computed using physical modeling. For example, some mutations are known to activate the kinase by increasing affinity to ATP, which could be computed using free energy methods like FEP.

Conclusion Revolutionary changes in computing power—especially the arrival of inexpensive graphics processors (GPUs)—and software automation have enabled alchemical free-energy calculations to impact drug discovery and life sciences projects in previously unforeseen ways. In this communication, we tested the hypothesis that FEP+, a fully-automated relative-alchemical free-energy workflow, had reached the point where it can accurately and reliably predict how clinically-observed mutations in Abl kinase alter the binding affinity of eight FDA-approved TKIs. To establish the potential predictive impact of current-generation alchemical free energy calculations—which incorporate entropic and enthalpic effects and the discrete nature of aqueous solvation—compared to a simpler physics-based approach that also uses modern forcefields but scores a single minimized conformation, we employed a second physics-based approach (Prime). This simpler physics-based model, which uses an implicit model of solvation to score the energetic changes in interaction energy that arise from the mutation, was able to capture a useful amount of information

to achieve substantial predictiveness with an MUE of $1.14_{0.93}^{1.35}$ kcal mol⁻¹ (N=142), RMSE of $1.70_{1.40}^{1.98}$ kcal mol⁻¹ respectively (N=142), and classification accuracy of $0.73_{0.66}^{0.80}$ (N=144). Surpassing these good results, we went on to demonstrate that FEP+ is able to achieve superior predictive performance— MUE of $0.79_{0.67}^{0.92}$ kcal mol⁻¹ (N=142), RMSE of $1.07_{0.89}^{1.26}$ kcal mol⁻¹ (N=142), and classification accuracy of $0.88_{0.82}^{0.93}$ (N=144). While future enhancements to the workflows for Prime and FEP+ to account for additional physical and chemical effects are likely to improve predictive performance further, the present results are of sufficient quality and achievable on a sufficiently rapid timescale (with turnaround times ~6 hours/calculation) to impact research projects in drug discovery and the life sciences. This work illustrates how the domain of applicability for alchemical free-energy methods is much larger than previously appreciated, and might further be found to include new areas as research progresses: aiding clinical decision-making in the selection of first- or second-line therapeutics guided by knowledge of likely subclonal resistance; identifying other selective kinase inhibitors (or combination therapies) to which the mutant kinase is susceptible; supporting the selection of candidate molecules to advance to clinical trials based on anticipated activity against likely mutations; facilitating the enrollments of patients in mechanism-based basket trials; and generally augmenting the armamentarium of precision oncology.

3.6 Methods

3.6.1 System preparation

All system preparation utilized the Maestro Suite (Schrödinger) version 2016-4. Comparative modeling to add missing residues using a homologous template made use of the Splicer tool, while missing loops modeled without a template used Prime. All tools employed default settings unless otherwise noted. The Abl wild-type sequence used in building all Abl kinase domain models utilized the ABL1_HUMAN Isoform IA (P00519-1) UniProt gene sequence spanning S229–K512. Models were prepared in non-phosphorylated form. We used a residue indexing convention that places the Thr gatekeeper residue at position 315 to match common usage; an alternate indexing convention utilized in experimental X-ray structures for Abl:imatinib (PDB: 1OPJ) [170] and Abl:dasatinib (PDB: 4XEY) [171] was adjusted to match our convention.

Complexes with co-crystal structures.

Chain B of the experimental structure of Abl:axitinib (PDB: 4WA9) [151] was used, and four missing residues at the N- and C-termini were added using homology modeling with PDB 3IK3 [172] as the template following alignment of the respective termini of the kinase domain. Chain B was selected because chain A was missing an additional 3 and 4 residues at the N- and C-termini, respectively, in addition to 3- and 20-residue loops, both of which were resolved in chain B. All missing side chains were added with Prime. The co-crystal structure of Abl:bosutinib (PDB: 3UE4) [173] was missing 4 and 10 N- and C-terminal residues respectively in chain A that were built using homology

modeling with 3IK3 as the template. All loops were resolved in chain A (chain B was missing two residues in the P-loop, Q252 and Y253). All missing side chains were added with Prime. The co-crystal structure of Abl:dasatinib (PDB: 4XEY) [171] was missing 2 and 9 N- and C-terminal residues, respectively, that were built via homology modeling using 3IK3 as the template. A 3 residue loop was absent in chain B but present in chain A; chain A was chosen. The co-crystal structure of Abl:imatinib (PDB: 1OPJ) [170] had no missing loops. Chain B was used because chain A was missing two C-terminal residues that were resolved in chain B. A serine was present at position 336 (index 355 in the PDB file) and was mutated to asparagine using Prime to match the human wild-type reference sequence (P00519-1). The co-crystal structure of Abl:nilotinib (PDB: 3CS9) [174] contained four chains in the asymmetric unit all of which were missing at least one loop. Chain A was selected because its one missing loop involved the fewest number of residues of the four chains; chain A was missing 4 and 12 N- and C-terminal residues, respectively, that were built using homology modeling with 3IK3 as the template. A 4-residue loop was missing in chain A (chain B and C were missing two loops, chain D was missing a five residue loop) that was built using Prime. The co-crystal structure of Abl:ponatinib (PDB: 3OXZ) [175] contained only one chain in the asymmetric unit. It had two missing loops, one 4 residues (built using Prime) and one 12 residues (built using homology modeling with 3OY3 [175] as the template). Serine was present at position 336 and was mutated to Asn using Prime to match the human wild-type reference sequence (P00519-1). Once the residue composition of the six Abl:TKI complexes were normalized to have the same sequence, the models were prepared using Protein Preparation Wizard. Bond orders were assigned using the Chemical Components Dictionary and hydrogen atoms were added. Missing side chain atoms were built using Prime. Termini were capped with N-acetyl (N-terminus) and N-methyl

amide (C-terminus). If present, crystallographic water molecules were retained. Residue protonation states (e.g. Asp381 and Asp421) were determined using PROPKA [176] with a pH range of 5.0–9.0. Ligand protonation state was assigned using PROPKA with pH equal to the experimental assay. Hydrogen bonds were assigned by sampling the orientation of crystallographic water, Asn and Gln flips, and His protonation state. The positions of hydrogen atoms were minimized while constraining heavy atoms coordinates. Finally, restrained minimization of all atoms was performed in which a harmonic positional restraint ($25.0 \text{ kcal mol}^{-1} \text{ \AA}^{-2}$) was applied only to heavy atoms. Table 3.4 summarizes the composition of the final models used for FEP.

Table 3.4: Summary of the preparation of the 6 Abl:TKI co-crystal structure complexes

PDB	Receptor	Ligand	Chains	Experimental structure				Prepared model used for simulations							
				# Water ^a	# Rec. atoms, (Chain) ^b	# Aminos, (Chain)	Chain used	# Water	# Rec. atoms	# Rec. aminos	# Ash	# Glh	# Hip	# Lig. atoms	Het. atom ^d w/ proton
4wa9	Abl	Axit	A, B	305	2219 (B)	276 (B)	B	131	4580	284	Ash421	0	0	46	neutral
3ue4	Abl	Bosut	A, B	152	2187 (A)	270 (A)	A	89	4581	284	Ash421 Ash381	0	0	66	NBI,4401
4xey	Abl	Dasat	A, B	0	2195 (A)	269 (A)	A	0	4581	284	Ash421 ^c Ash381 ^c	0	0	59	neutral
1opj	Abl	Imat	A, B	231	2336 (B)	288 (B)	B	104	4579	284	0	0	0	69	N51,4767
3cs9	Abl	Nilot	A, B, C, D	266	2142 (A)	264 (A)	A	99	4579	284	0	0	0	61	neutral
3oxz	Abl	Ponat	A	89	2152 (A)	268 (A)	A	89	4580	284	0	0	0	67	N3,2155

^aTotal number of water molecules, ^bCount includes N-Acetyl/N-terminal (6 atoms) and N-methylamide/C-terminal (6 atoms) capping groups, ^cOriginal index in experimental structure was Ash440, Ash400, ^d(PDB atom name) and (PDB atom serial number).
Ash: Neutral form of Asp; **Glh**: Neutral form of Glu; **Hip**: Charged form of His.

Complexes without co-crystal structures.

Co-crystal structures of Abl bound to erlotinib or gefitinib were not publicly available. To generate models of these complexes, Glide-SP [177] was utilized to dock these two compounds into an Abl receptor structure. Co-crystal structures of these two compounds bound to EGFR were publicly available and this information was used to obtain initial ligand geometries and to establish a reference binding mode against which our docking results could be structurally scored. The Abl receptor structure bound to bosutinib was used for docking because its structure was structurally similar to that of EGFR in the erlotinib- (PDB: 4HJO) [178] and gefitinib-bound (PDB: 4WKQ) [179] co-crystal structures. Abl was prepared for docking by using the Protein Preparation Wizard (PPW) with default parameters. Crystallographic waters were removed but their coordinates retained for a subsequent step in which they were optionally reintroduced. Erlotinib and gefitinib protonation states at pH 7.0 ± 2.0 were determined using Epik [180]. Docking was performed using the Glide-SP workflow. The receptor grid was centered on bosutinib. The backbone NH of Met318 was chosen to participate in a hydrogen bonding constraint with any hydrogen bond donor on the ligand. The hydroxyl of T315 was allowed to rotate in an otherwise rigid receptor. Ligand docking was performed with enhanced sampling; otherwise default settings were used. Epik state penalties were included in the scoring. The 16 highest ranked (Glide-SP score) poses were retained for subsequent scoring. To determine the docked pose that would be subsequently used for free energy calculations, the ligand heavy-atom RMSD between the 16 poses and the EGFR co-crystal structures (PDB IDs 4HJO and 4WKQ) was determined. The pose in which erlotinib or gefitinib most structurally resembled the EGFR co-crystal structure (lowest heavy-atom RMSD)

was chosen as the pose for subsequent FEP+. Two sets of complex structures were subjected to free energy calculations to determine the effect of crystal waters: In the first set, without crystallographic waters, the complexes were prepared using Protein Prep Wizard as above. In the second set, the crystallographic waters removed prior to docking were added back, and waters in the binding pocket that clashed with the ligand were removed.

3.6.2 Force field parameter assignment

The OPLS3 forcefield [54] version that shipped with Schrödinger Suite release 2016-4 was used to parameterize the protein and ligand. Torsion parameter coverage was checked for all ligand fragments using Force Field Builder. The two ligands that contained a fragment with a torsion parameter not covered by OPLS3 were axitinib and bosutinib; Force Field Builder was used to obtain these parameters. SPC parameters [181] were used for water. For mutations that change the net charge of the system, counterions were included to neutralize the system with additional Na⁺ and Cl⁻ ions added to achieve 0.15 M excess to mimic the solution conditions of the experimental assay.

3.6.3 Prime (MM-GBSA)

Prime was used to predict the geometry of mutant side chains and to calculate relative changes in free energy using MM-GBSA single-point estimates [150]. VSGB [182] was used as the implicit solvent model to calculate the solvation free energies for the four states (complex/wild-type, complex/mutant, apo protein/wild-type, and apo protein/mutant) and $\Delta\Delta G$ calculated using the

thermodynamic cycle depicted in Figure 3.1B. Unlike FEP (see below), which simulates the horizontal legs of the thermodynamic cycle, MM-GBSA models the vertical legs by computing the interaction energy between the ligand and protein in both wild-type and mutant states, subtracting these to obtain the $\Delta\Delta G$ of mutation on the binding free energy.

3.6.4 Alchemical free energy perturbation calculations using FEP+

Alchemical free energy calculations were performed using the FEP+ tool in the Schrödinger Suite version 2016-4, which offers a fully automated workflow requiring only an input structure (wild-type complex) and specification of the desired mutation. The default protocol was used throughout: It assigns protein and ligand force field parameters (as above), generates a dual-topology [183] alchemical system for transforming wild-type into mutant protein (whose initial structure is modeled using Prime), generates the solvent-leg endpoints (wild-type and mutant apo protein), and constructs intermediate windows spanning wild-type and mutant states. Simulations of the apo protein were setup by removing the ligand from the prepared complex (see System Preparation) followed by an identical simulation protocol as that used for the complex. Charge-conserving mutations utilized 12 λ windows (24 systems) while charge-changing mutations utilized 24 λ windows (48 systems). Each system was solvated in an orthogonal box of explicit solvent (SPC water [181]) with box size determined to ensure that solute atoms were no less than 5 (complex leg) or 10 (solvent leg) from an edge of the box. For mutations that change the net charge of the system, counterions were included to neutralize the charge of the system, and

additional Na⁺ and Cl⁻ ions added to achieve 0.15 M excess NaCl to mimic the solution conditions of the experimental assay. The artifact in electrostatic interactions for charge change perturbations due to periodic boundary conditions in MD simulations are corrected based on the method proposed by Rocklin *et al.* [184], where the difference in solvation free energy of the solute under non-periodic boundary condition and that under periodic boundary condition is approximated by Poisson-Boltzmann method and serves as the correction term for each system.

System equilibration was automated. It followed the default 5-stage Desmond protocol: (i) 100 ps with 1 fs time steps of Brownian dynamics with positional restraints of solute heavy atoms to their initial geometry using a restraint force constant of 50 kcal mol⁻¹ Å⁻²; this Brownian dynamics integrator corresponds to a Langevin integrator in the limit when $\tau \rightarrow 0$, modified to stabilize equilibration of starting configurations with high potential energies; particle and piston velocities were clipped so that particle displacements were limited to 0.1 Å, in any direction. (ii) 12 ps MD simulations with 1 fs time step using Langevin thermostat at 10 K with constant volume, using the same restraints; (iii) 12 ps MD simulations with 1 fs time step using Langevin thermostat and barostat [185] at 10 K and constant pressure of 1 atmosphere, using the same restraints; (iv) 12 ps MD simulations with 1 fs time step using Langevin thermostat and barostat at 300 K and constant pressure of 1 atmosphere, using the same restraints; (v) a final unrestrained equilibration MD simulation of 240 ps with 2 fs time step using Langevin thermostat and barostat at 300 K and constant pressure of 1 atmosphere. Electrostatic interactions were computed with particle-mesh Ewald (PME) [186] and a 9 Å cutoff distance was used for van de Waals interactions. The production MD simulation was performed in the NPT ensemble using the MTK method [187] with integration time steps of 4 fs, 4 fs, and 8 fs respectively for the bonded, near,

and far interactions following the RESPA method [188] through hydrogen mass repartitioning [189]. Production FEP+ calculations utilized Hamiltonian replica exchange with solute tempering (REST) [190], with automated definition of the REST region. Dynamics were performed with constant pressure of 1 atmosphere and constant temperature of 300 K for 5 ns in which exchanges between windows was attempted every 1.2 ps.

Because cycle closure could not be used to reduce statistical errors via path redundancy [190], we instead performed mutational free energy calculations in triplicate by initializing dynamics with different random seeds. The relative free energies for each mutation in each independent run were calculated using BAR [191, 192] The reported $\Delta\Delta G$ was computed as the mean of the computed $\Delta\Delta G$ from three independent simulations. Triplicate simulations were performed in parallel using four NVIDIA Pascal Architecture GPUs per alchemical free-energy simulation (12 GPUs in total), requiring ~6 hours in total to compute $\Delta\Delta G$.

3.6.5 Obtaining $\Delta\Delta G$ from ΔpIC_{50} benchmark set data

Reference relative free energies were obtained from three publicly available sources of ΔpIC_{50} data (Table 3.1). Under the assumption of Michaelis-Menten binding kinetics (pseudo first-order, but relative free energies are likely consistent), the inhibitor is competitive with ATP (Equation 3.1). This assumption has been successfully used to estimate relative free energies [109, 148, 193, 194] using the relationship between IC_{50} and competitive inhibitor affinity K_i ,

$$IC_{50} = \frac{K_i}{1 + \frac{[S_0]}{K_M}} \quad (3.1)$$

If the Michaelis constant for ATP (K_M) is much larger than the initial ATP concentration S_0 , the relation in Equation 3.1 will tend towards the equality $IC_{50} = K_i$. The relative change in binding free energy of Abl:TKI binding due to protein mutation is simply,

$$\Delta\Delta G = -RT \ln \frac{IC_{50,WT}}{IC_{50,mut}} \quad (3.2)$$

where $IC_{50,WT}$ is the IC_{50} value for the TKI binding to the wild-type protein and $IC_{50,mut}$ is the IC_{50} value for the mutant protein. R is the ideal gas constant and T is taken to be room temperature (300 K).

As alluded to above, relating ΔpIC_{50} s to $\Delta\Delta G$ s assumes that the Michaelis constant for ATP is much larger than the initial concentration of ATP, and that the experimentally observed ΔpIC_{50} change is solely from changes in kinase:TKI binding affinity. In practice, not all of these assumptions may hold. For example, the experimentally observed ΔpIC_{50} might depend on the metabolism of drugs, and for drugs with different mechanisms of action than directly binding to the kinase binding pocket (e.g., binding to the transition structures of kinases, target gene amplification, up-/down-regulation of positive-/negative-feedback effectors, diminished synergism of pro-apoptotic machinery, decoupling of the target from cell survival circuits) [195, 196], their inhibition ability might not correlate well with binding affinity. However, the comparison between ΔpIC_{50} and ΔK_D is presented in Figure 3.2D, and this comparison indicates the assumptions we used to relate ΔpIC_{50} to $\Delta\Delta G$ are reasonable for the dataset we studied.

3.6.6 Assessing prediction performance

Quantitative accuracy metrics

Mean unsigned error (MUE) was calculated by taking the average absolute difference between predicted and experimental estimates of $\Delta\Delta G$. Root-mean square error (RMSE) was calculated by taking the square root of the average squared difference between predicted and experimental estimates of $\Delta\Delta G$. MUE depends linearly on errors such that large and small errors contribute equally to the average value, while RMSE depends quadratically on errors, magnifying their effect on the average value.

Truth tables

Two-class truth tables were constructed to characterize the ability of Prime and FEP+ to correctly classify mutations as susceptible ($\Delta\Delta G \leq 1.36 \text{ kcal mol}^{-1}$) or resistant ($\Delta\Delta G > 1.36 \text{ kcal mol}^{-1}$), where the $1.36 \text{ kcal mol}^{-1}$ threshold represents a 10-fold change in affinity. Accuracy was calculated as the fraction of all predictions that were correctly classified as sensitizing, neutral, or resistant. Sensitivity and specificity were calculated using a binary classification of resistant ($\Delta\Delta G > 1.36 \text{ kcal mol}^{-1}$) or susceptible ($\Delta\Delta G \leq 1.36 \text{ kcal mol}^{-1}$). Specificity was calculated as the fraction of correctly predicted non-resistant mutations out of all truly susceptible mutations **S**. Sensitivity was calculated as the fraction of correctly predicted resistant mutations out of all truly resistant mutations, **R**. The number of susceptible mutations was 113 for axitinib, bosutinib, dasatinib, imatinib, nilotinib and ponatinib, and 12 for erlotinib and gefitinib; the number of

resistant mutations **R** was 18 for axitinib, bosutinib, dasatinib, imatinib, nilotinib, and ponatinib, and 1 for erlotinib and gefitinib.

Consensus model

First, Prime and FEP+ (n=142) were scaled by minimizing their RMSE to experiment by optimizing slope using linear regression. The resulting (minimum) RMSE was used in a subsequent step to combine the scaled FEP+ and scaled Prime free energies with inverse-variance weighted averaging.

ROC

A ROC curve was generated by computing the true positive rate (sensitivity) and the true negative rate (specificity) when the classification cutoff differentiating resistant from sensitizing mutations is changed for (only) the predicted values of $\Delta\Delta G$. Cutoffs were chosen by taking the minimum and maximum value of $\Delta\Delta G$ for a data set (Prime or FEP+), and iteratively computing specificity and sensitivity in steps of $0.001 \text{ kcal mol}^{-1}$, which by this definition will be in the range [0,1]. Experimental positives and negatives were classified with the $1.36 \text{ kcal mol}^{-1}$ cutoff. ROC-AUC was computed using the trapezoidal rule.

Estimating uncertainties of physical-modeling results

95% symmetric confidence intervals (CI, 95%) for all performance metrics were calculated using bootstrap by resampling all datasets with replacement, with 1000 resampling events. Confidence intervals were estimated for all performance metrics and reported as $x_{x_{\text{low}}}^{x_{\text{high}}}$ where x is the mean statistic calculated from the

complete dataset (e.g. RMSE), and x_{low} and x_{high} are the values of the statistic at the 2.5th and 97.5th percentiles of the value-sorted list of the bootstrap samples.

Uncertainty for $\Delta\Delta G$ s was computed by the standard deviation between three independent runs (using different random seeds to set initial velocities), where the 95% CI was $[\Delta\Delta G - 1.96 \times \sigma_{\text{FEP+}}, \Delta\Delta G + 1.96 \times \sigma_{\text{FEP+}}]$ kcal mol⁻¹. 1σ used in plots for FEP+ and experiment; 0σ for Prime.

Bayesian hierarchical model to estimate intrinsic error

We used Bayesian inference to estimate the true underlying prediction error of Prime and FEP+ by making use of known properties of the experimental variability (characterized in Figure 3.2) and statistical uncertainty estimates generated by our calculations under weak assumptions about the character of the error.

We presume the true free energy differences of mutation i , $\Delta\Delta G_i^{\text{true}}$, comes from a normal background distribution of unknown mean and variance,

$$\Delta\Delta G_i^{\text{true}} \sim \mathcal{N}(\mu_{\text{mut}}, \sigma_{\text{mut}}^2) \quad i = 1, \dots, M \quad (3.3)$$

where there are M mutations in our dataset. We assign weak priors to the mean and variance

$$\mu_{\text{mut}} \sim U(-6, +6) \quad (3.4)$$

$$\sigma_{\text{mut}} \propto 1 \quad (3.5)$$

where we limit $\sigma > 0$.

We presume the true computational predictions (absent statistical error) differ from the (unknown) true free energy difference of mutation $\Delta\Delta G_i^{\text{true}}$ by

normally-distributed errors with zero bias but standard deviation equal to the RMSE for either Prime or FEP+, the quantity we are focused on estimating:

$$\Delta\Delta G_{i,\text{Prime}}^{\text{true}} \sim \mathcal{N}(\Delta\Delta G_i^{\text{true}}, \text{RMSE}_{\text{Prime}}^2) \quad (3.6)$$

$$\Delta\Delta G_{i,\text{FEP+}}^{\text{true}} \sim \mathcal{N}(\Delta\Delta G_i^{\text{true}}, \text{RMSE}_{\text{FEP+}}^2) \quad (3.7)$$

In the case of Prime, since the computation is deterministic, we actually calculate $\Delta\Delta G_{\text{Prime}}^{\text{true}}$ for each mutant. For FEP+, however, the computed free energy changes are corrupted by statistical error, which we also presume to be normally distributed with standard deviation $\sigma_{\text{calc},i}$,

$$\Delta\Delta G_{i,\text{FEP+}} \sim \mathcal{N}(\Delta\Delta G_{i,\text{FEP+}}, \sigma_{i,\text{FEP+}}^2) \quad (3.8)$$

where $\Delta\Delta G_{i,\text{FEP+}}$ is the free energy computed for mutant i by FEP+, and $\sigma_{i,\text{FEP+}}$ is the corresponding statistical error estimate.

The experimental data we observe is also corrupted by error, which we presume to be normally distributed with standard deviation σ_{exp} :

$$\Delta\Delta G_{i,\text{exp}} \sim \mathcal{N}(\Delta\Delta G_i, \sigma_{\text{exp}}^2) \quad (3.9)$$

Here, we used an estimate of K_d - and IC_{50} -derived $\Delta\Delta G$ variation derived from the empirical RMSE of $0.81 \text{ kcal mol}^{-1}$, where we took $\sigma_{\text{exp}} \approx 0.81/\sqrt{2} = 0.57 \text{ kcal mol}^{-1}$ to ensure the difference between two random measurements of the same mutant would have an empirical RMSE of $0.81 \text{ kcal mol}^{-1}$.

Under the assumption that the true $\Delta\Delta G$ is normally distributed and the calculated value differs from the true value via a normal error model, it can easily be shown that the MUE is related to the RMSE via

$$\text{MUE} = \int dx_{\text{true}} p(x_{\text{true}}) \int dx_{\text{calc}} p(x_{\text{calc}}|x_{\text{true}}) |x_{\text{calc}} - x_{\text{true}}| \quad (3.10)$$

$$\begin{aligned}
&= \int dx_{\text{true}} \frac{1}{\sqrt{2\pi\sigma_{\text{true}}^2}} e^{-\frac{(x_{\text{true}}-\mu_{\text{true}})^2}{2\sigma_{\text{true}}^2}} \int dx_{\text{calc}} \frac{1}{\sqrt{2\pi\sigma_{\text{calc}}^2}} e^{-\frac{(x_{\text{calc}}-\mu_{\text{true}})^2}{2\sigma_{\text{calc}}^2}} |x_{\text{calc}} - x_{\text{true}}| \quad (3.11) \\
&= \sqrt{\frac{2}{\pi}} \text{RMSE} \quad (3.12)
\end{aligned}$$

The model was implemented using PyMC3 [197], observable quantities were set to their computed or experimental values, and 5000 samples drawn from the posterior (after discarding an initial 500 samples to burn-in) using the default NUTS sampler. Expectations and posterior predictive intervals were computed from the marginal distributions obtained from the resulting traces.

3.6.7 Data availability

All relevant data are publicly available: compiled experimental datasets, input files for Prime and FEP+ and computational results that support our findings can be found at GitHub by following the URL: <https://github.com/kehauser/>

[Predicting-resistance-of-clinical-Abl-mutations-to-targeted-kinase-inhibitors-using-FEP](#)

3.6.8 Code availability

Scripts used for statistics analysis (including the Bayesian inference model) can be found at the following URL: <https://github.com/kehauser/>

[Predicting-resistance-of-clinical-Abl-mutations-to-targeted-kinase-inhibitors-using-FEP](#)

3.7 Acknowledgments

We thank Daniel Robinson (Schrödinger), Sonya M. Hanson (MSKCC), and Gregory A. Ross (MSKCC) for helpful discussions. JDC acknowledges support from NIH National Cancer Institute Cancer Center Core Grant P30 CA008748; JDC and SKA acknowledge support from the Sloan Kettering Institute, Cycle for Survival, and NIH grant R01 GM121505. KH acknowledges help from Wei Chen (Schrödinger) and Anthony Clark (Schrödinger) for instructions on running mutations changing the net charge of the system, and Simon Gao (Schrödinger) for assistance in computational resources.

3.8 Author Contributions

KH, JDC, CN, RA, and LW designed the research; KH, SA, TS, and LW identified experimental datasets; KH and LW performed the simulations; KH, CN, SKA, SR, TS, RA, JDC, and LW analyzed the data; KH, JDC, SKA, and LW wrote the paper.

3.9 Competing Interests

JDC is a member of the Scientific Advisory Board for Schrödinger Inc. SR was a former employee of Schrödinger Inc.; and KH, CN, RA, TS, and LW are employees of Schrödinger Inc.

CHAPTER 4

ENABLING HIGH-THROUGHPUT BIOPHYSICAL EXPERIMENTS ON CLINICALLY-OBSERVED MUTATIONS

4.1 Gloss

The work in this chapter was published in *Biochemistry* and is reprinted with permission. Copyright 2018 American Chemical Society.

An open library of human kinase domain constructs for automated bacterial expression

Steven K. Albanese^{†,1,2}, Daniel L. Parton^{†,2}, Mehtap Işık^{‡,2,3}, Lucelenie Rodríguez-Laureano^{‡,2}, Sonya M. Hanson², Julie M. Behr^{2,7}, Scott Gradia^{&,4}, Chris Jeans⁴, Nicholas M. Levinson⁶, Markus A. Seeliger⁵, John D. Chodera^{2,*}

¹ *Louis V. Gerstner, Jr. Graduate School of Biomedical Sciences, Memorial Sloan Kettering Cancer Center, New York, NY 10065*

² *Computational and Systems Biology Program, Sloan Kettering Institute, Memorial Sloan Kettering Cancer Center, New York, NY 10065*

³ *Tri-Institutional PhD Program in Chemical Biology, Weill Cornell Graduate School of Medical Sciences, Cornell University, New York, NY 10065*

⁴ *QB3 MacroLab, University of California, Berkeley, CA 94720*

⁵ *Department of Pharmacological Sciences, Stony Brook University Medical School, Stony Brook, NY 11794*

⁶ *Department of Pharmacology, University of Minnesota, Minneapolis, MN 55455*

⁷ *Tri-Institutional Program in Computation Biology and Medicine, Weill Cornell Graduate School of Medical Sciences, Cornell University, New York, NY 10065*

[†] or [‡] *These authors contributed equally to this work*

^{*} *Corresponding Author*

[&] *Present Address: Caribou Biosciences, Berkeley, CA 94720*

A common roadblock in testing computational methodologies to predict resistance and selectivity is the relative scarcity of high-quality, publicly available, and quantitative experimental data for the same compounds measured against multiples kinases or multiple forms of the kinase. Further, what data is available is often reported without reliable uncertainty estimates. Studies of internal and public databases have provided estimates of uncertainties for IC₅₀ [56, 103] and K_i [198], as well as the pitfalls that come from combining heterogeneous datasets.

Even for kinases, many of which are extensively studied and are the subjects of many drug development programs, quantitative data is scarce and difficult to access publicly. The previous chapter uniquely highlighted this problem: the RMSE of FEP+ for predicting the impact of mutations was roughly equal to the RMSE of K_d and IC50 measurements for the same ligand and mutant kinase pairs.

Rather than manually intensive gathering of unreliable public data, an idealized pipeline for experiments benchmarking computational methods for their ability to predict selectivity or the impact of mutations would be: express the wildtype or clinical mutant kinases of interest, measure the affinities using a minimally perturbative method that allows for the estimation of a $\Delta\Delta G_{\text{binding}}$ with well-characterized uncertainty and then benchmark the computational calculations or methodologies. In the future, one can envision a scenario in which a novel mutation is mapped onto all relevant structures and computationally tested for the mutations impact on all available FDA-approved treatments. Any promising calculations can be confirmed by rapidly expressing the mutant and performing the biophysical experiments in a high-throughput manner, with a consistent protocol to enable comparisons between different mutations and ligands. Engineering this type of pipeline in either direction requires a number of components. First, we need to be able to identify clinical mutations, through an assay such as the MSK-IMPACT panel [38]. Then, we need ways to automate the setup and execution of alchemical free energy calculations for all combinations of structural conformations and inhibitors of interest. Finally, we need ways to rapidly engineer and express the mutants to validate the computational experiments with biophysical measurements. In this chapter, we present work moving towards enabling the high-throughput expression of kinases and their clinical mutants using a one-size fits all, automated protocol. We also present two different high-throughput biophysical measurements that are used to validate the

quality of the expressed protein, which can be extended to test computational predictions in a high-throughput, quantitative manner in future work. Finally, the resulting wild-type kinase plasmids from this work have been made available on Addgene, which we hope will provide a useful resource to the community for kinase oriented biophysical studies.

This work was the result of a long collaboration between the Chodera Lab and the QB3 MacroLab at UC Berkeley. Daniel Parton and John Chodera designed the structural informatics pipeline and initial expression experiments, which were carried out at the MacroLab by Scott Gradua and Chris Jeans. Mehtap Işık, Lucelenie Rodríguez-Laureano and I planned and troubleshot the biophysical experiments, which built on Sonya Hanson's work in the lab. Julie Behr and I planned the mutation expression experiments, which were carried out at the MacroLab. I analyzed and curated the data, as well as wrote and edit the manuscript, with a great deal of input from all of the authors, especially John Chodera, Mehtap Işık, Nick Levinson and Markus Seeliger.

The most immediate future work in this area involves further optimizing the fluorescence-based assay presented in this chapter so that quantitative measurements of binding affinity can be made. Much of this work is presently under way already. Exploring novel kinases that have yet to be expressed in *E. coli* using this protocol could enable drug discovery and biophysical work on kinases that have yet to be extensively characterized. While we see promising yields co-expressing the kinases with phosphatases, many kinases and clinical mutants still did not express well using this protocol. While future work could focus on optimizing the protocol used to get better yields from the TEV cleavage, co-expressing with chaperone protein such as Hsp90 and Cdc37 may rescue insoluble protein from the cell pellet. Additionally, future work may focus on

expanding the number of clinical mutants considered for Src and Abl, as well as additional kinases of interest.

4.2 Abstract

Kinases play a critical role in cellular signaling and are dysregulated in a number of diseases, such as cancer, diabetes, and neurodegeneration. Therapeutics targeting kinases currently account for roughly 50% of cancer drug discovery efforts. The ability to explore human kinase biochemistry and biophysics in the laboratory is essential to designing selective inhibitors and studying drug resistance. Bacterial expression systems are superior in terms of simplicity and cost-effectiveness compared to insect or mammalian cells, but have historically struggled with human kinase expression. Following the discovery that phosphatase coexpression produced high yields of Src and Abl kinase domains in bacteria, we have generated a library of 52 His-tagged human kinase domain constructs that express above 2 $\mu\text{g}/\text{mL}$ culture in an automated bacterial expression system utilizing phosphatase coexpression (YopH for Tyr kinases, Lambda for Ser/Thr kinases). Here, we report a structural bioinformatics approach to identify kinase domain constructs previously expressed in bacteria and likely to express well in our protocol, experiments demonstrating our simple construct selection strategy selects constructs with good expression yields in a test of 84 potential kinase domain boundaries for Abl, and yields from a high-throughput expression screen of 96 human kinase constructs. Using a fluorescence-based thermostability assay and a fluorescent ATP-competitive inhibitor, we show that the highest-expressing kinases are folded and have well-formed ATP binding sites. We also demonstrate that these constructs can

enable characterization of clinical mutations by expressing a panel of 48 Src and 46 Abl mutations. The wild-type kinase construct library is available publicly via Addgene.

4.3 Introduction

Kinases play a critical role in cellular signaling pathways, controlling a number of key biological processes that include growth and proliferation. There are over 500 kinases in the human genome [199, 200], many of which are of therapeutic interest. Perturbations due to mutation, translocation, or upregulation can cause one or more kinases to become dysregulated, often with disastrous consequences [201]. Kinase dysregulation has been linked to a number of diseases, such as cancer, diabetes, and inflammation. Cancer alone is the second leading cause of death in the United States, accounting for nearly 25% of all deaths; in 2015, over 1.7 million new cases were diagnosed, with over 580,000 deaths [202]. Nearly 50% of cancer drug development is targeted at kinases, accounting for perhaps 30% of *all* drug development effort globally [84, 203].

The discovery of imatinib, an inhibitor that targets the Abelson tyrosine kinase (Abl) dysregulated in chronic myelogenous leukemia (CML) patients, was transformative in revealing the enormous therapeutic potential of selective kinase inhibitors, kindling hope that this remarkable success could be recapitulated for other cancers and diseases [204]. While there are now 39 FDA-approved selective kinase small molecule inhibitors (as of 16 Jan 2018) [3, 205], these molecules were approved for targeting only 22 out of ~500 human kinases¹, with the vast majority

¹These targets are, currently: Abl, DDR1, EGFR, HER2, VGFR1/2/3, Alk, Met, BRAF, JAK1/2/3, Btk, Pi3K, CDK4, CDK6, MEK, ROS1, FLT3, IGF1R, Ret, Kit, Axl, TrkB, and mTOR [3].

developed to target just a handful of kinases [206]. The discovery of therapeutically effective inhibitors for other kinases has proven remarkably challenging.

While these inhibitors have found success in the clinic, many patients cease to respond to treatment due to resistance caused by mutations in the targeted kinase [9], activation of downstream kinases [201], or relief of feedback inhibition in signaling pathways [207]. These challenges have spurred the development of a new generation of inhibitors aimed at overcoming resistance [34, 35], as well as mutant-specific inhibitors that target kinases bearing a missense mutation that confers resistance to an earlier generation inhibitor [36]. The ability to easily engineer and express mutant kinase domains of interest would be of enormous aid in the development of mutant-selective inhibitors, offering an advantage over current high-throughput assays [208–210], which typically include few clinically-observed mutant kinases.

Probing human kinase biochemistry, biophysics, and structural biology in the laboratory is essential to making rapid progress in understanding kinase regulation, developing selective inhibitors, and studying the biophysical driving forces underlying mutational mechanisms of drug resistance. While human kinase expression in baculovirus-infected insect cells can achieve high success rates [211, 212], it cannot compete in cost, convenience, or speed with bacterial expression. *E. coli* expression enables production of kinases without unwanted post-translational modifications, allowing for greater control of the system. A survey of 62 full-length non-receptor human kinases found that over 50% express well in *E. coli* [211], but often expressing only the soluble kinase domains are sufficient, since these are the molecular targets of therapy for targeted kinase inhibitors and could be studied even for receptor-type kinases. While removal of

regulatory domains can negatively impact expression and solubility, coexpression with phosphatase was shown to greatly enhance bacterial kinase expression in Src and Abl tyrosine kinases, presumably by ensuring that kinases remain in an unphosphorylated inactive form where they can cause minimal damage to cellular machinery [213].

The protein databank (PDB) now contains over 100 human kinases that were expressed in bacteria, according to PDB header data. Many of these kinases were expressed and crystallized as part of the highly successful Structural Genomics Consortium (SGC) effort to increase structural coverage of the human kinome [214]. Since bacterial expression is often complicated by the need to tailor construct boundaries, solubility-promoting tags, and expression and purification protocols individually for each protein expressed, we wondered whether a simple, uniform, automatable expression and purification protocol could be used to identify tractable kinases, select construct boundaries, express a large number of human kinases and their mutant forms, and produce a convenient bacterial expression library to facilitate kinase research and selective inhibitor development. As a first step toward this goal, we developed a structural informatics pipeline to use available kinase structural data and associated metadata to select constructs from available human kinase libraries to clone into a standard set of vectors intended for phosphatase coexpression under a simple automatable expression and purification protocol. Using an expression screen for multiple construct domain boundaries of Abl, we found that transferring construct boundaries from available structural data can produce constructs with useful expression levels, enabling simple identification of construct domain boundaries. We then completed an automated expression screen in Rosetta2 cells of 96 different kinases and found that 52 human kinase domains express with yields greater than 2 $\mu\text{g}/\text{mL}$ culture. To investigate whether these kinases are

properly folded and useful for biophysical experiments, we performed a fluorescence-based thermostability assay on the 14 highest expressing kinases in our panel and a single-well high-throughput fluorescence-based binding affinity measurement on 39 kinases. These experiments demonstrated that many of the expressed kinases were folded, with well formed ATP binding sites capable of binding a small molecule kinase inhibitor. To demonstrate the utility of these constructs for probing the effect of clinical mutations on kinase structure and ligand binding, we subsequently screened 48 Src and 46 Abl mutations, finding that many clinically-derived mutant kinase domains can be expressed with useful yields in this uniform automated expression and purification protocol.

All source code, data, and wild-type kinase plasmids associated with this project are freely available online:

- **Source code and data:**

- <https://github.com/choderalab/kinase-ecoli-expression-panel>

- **Interactive table of expression data:**

- <http://choderalab.org/kinome-expression>

- **Plasmids:** <https://www.addgene.org/kits/chodera-kinase-domains>

4.4 Results

4.4.1 Construct boundary choice impacts Abl kinase domain expression

To understand how alternative choices of expression construct boundaries can modulate bacterial expression of a human kinase domain, we carried out an expression screen of 84 unique construct boundaries encompassing the kinase domain of the tyrosine protein kinase ABL1.

Three constructs known to express in bacteria were chosen from the literature and used as controls, spanning Uniprot residues 229–500 (PDBID: 3CS9) [215], 229–512 (PDBID: 2G2H) [216] and 229–515 (PDBID: 2E2B) [217]. 81 constructs were generated combinatorially by selecting nine different N-terminal boundaries spanning residues 228–243 and nine different C-terminal boundaries spanning residues 490–515, chosen to be near the start and end points for the control constructs (Figure 4.1A). Each of the three control constructs included six replicates to provide an estimate of the typical standard error in expression readout for the experimental constructs, which was found to be between 0.42–1.5 $\mu\text{g}/\text{mL}$ (Figure 4.1A, green constructs).

Briefly, the impact of construct boundary choice on Abl kinase domain expression was tested as follows (see Methods for full details). His₁₀-TEV N-terminally tagged wild-type Abl constructs² were coexpressed with YopH phosphatase in a 96-well format with control replicates distributed randomly throughout the plate. His-tagged protein constructs were recovered via a single nickel affinity

²Parent plasmid is a pET His₁₀ TEV LIC cloning vector and is available on Addgene (Plasmid #78173).

chromatography step, and construct yields were quantified using microfluidic capillary electrophoresis following thermal denaturation. Expression yields are summarized in Figure 4.1A, and a synthetic gel image from the constructs with detectable expression is shown in Figure 4.2. Abl construct bands are present at sizes between 29 and 35 kDa (due to the variation in construct boundaries), and YopH phosphatase (which is not His-tagged but has substantial affinity for the nickel beads) is present in all samples at its expected size of 50 kDa. Strikingly, despite the fact that N-terminal and C-terminal construct boundaries only varied over 15–25 residues, only a small number of constructs produced detectable expression (Figure 4.1B). As highlighted in Figure 4.1C (left), the best N-terminal boundaries (residues 228, 229, 230) are located on a disordered strand distant from any secondary structure; N-terminal boundaries closer to the beta sheet of the N-lobe gave poor or no detectable expression (Figure 4.1B).

The best C-terminal construct boundaries (residues 511 and 512) occur in an α -helix (Figure 4.1C, right). Of note, this α -helix is not resolved in PDBID:2E2B [217], suggesting this structural element may only be weakly thermodynamically stable in the absence of additional domains. In previous work, this α -helix was shown to undergo a dramatic conformational change which introduces a kink at residue 516, splitting the α -helix into two [20]. This suggests a high potential for flexibility in this region.

Two of the control constructs (which differ in construct boundary by only one or two residues) were in the top six expressing constructs (Figure 4.1A), and were in fact within 60% of the maximum observed expression yield. From this, we concluded that transferring construct boundaries from existing kinase domain structural data would be sufficient to bias our constructs toward useful expression levels for a large-scale screen of multiple kinases.

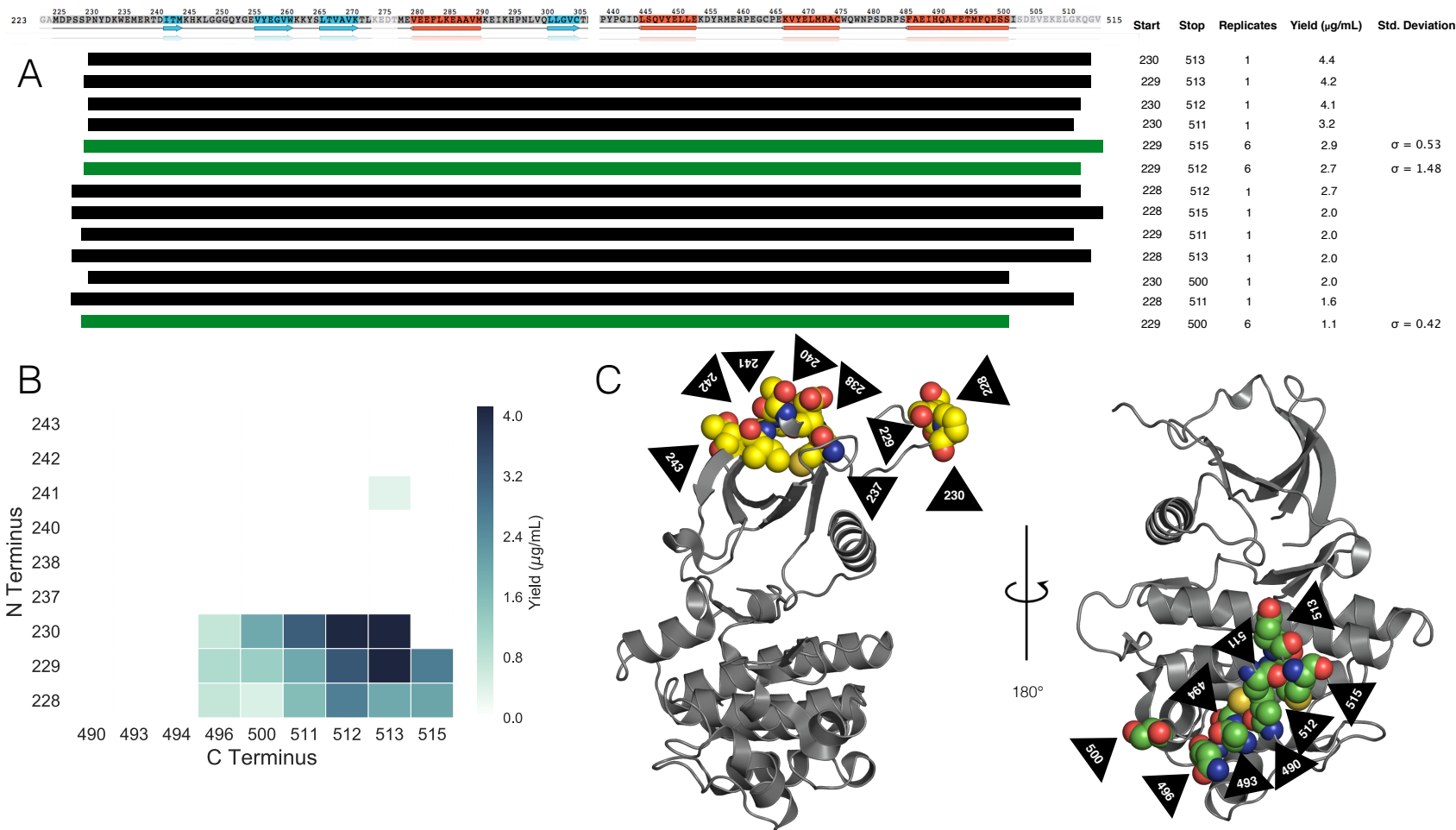


Figure 4.1: **Abl kinase domain construct expression screen illustrates high sensitivity to construct boundaries.** (A) Abl kinase domain construct boundaries with highest expression yields. Standard deviations of the yield are listed for control constructs for which six replicates were performed to give an indication of the uncertainty in experimental constructs. Secondary structure is indicated on the sequence. Beta sheets are colored blue and alpha helices are colored orange. (B) Heatmap showing average yields for constructs (in $\mu\text{g/mL}$ culture) with detectable expression as a function of N- and C-terminal construct boundaries. (C) *left*: PDBID: 2E2B with the nine N-terminal construct boundary amino acids shown as yellow spheres. *right*: PDBID: 4XEY with the nine C-terminal construct boundary amino acids shown as green spheres. Black arrows indicate residue numbers.

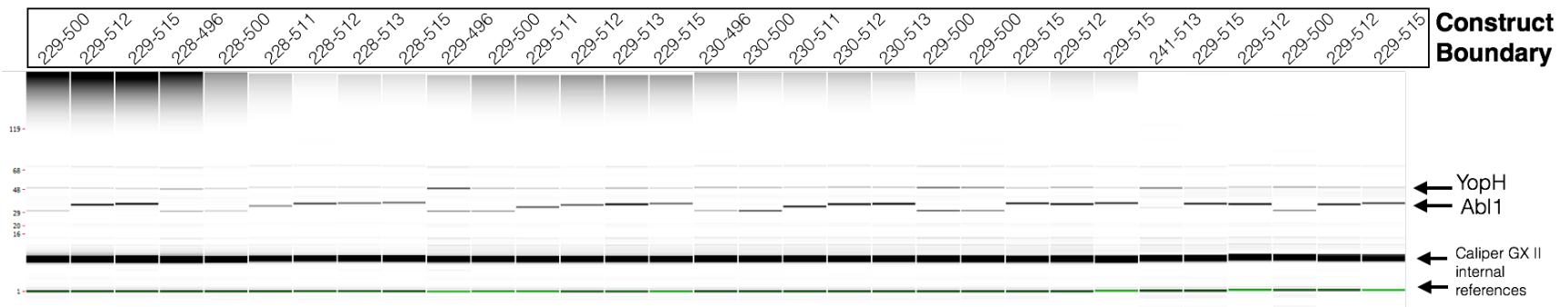


Figure 4.2: **Expression yields of Abl kinase domain constructs for all constructs with detectable expression.** A synthetic gel image rendering generated from Caliper GX II microfluidic gel electrophoresis data following Ni-affinity purification and thermal denaturation for all Abl constructs with detectable expression. Each well is marked with the Abl kinase domain construct residue boundaries (Uniprot canonical isoform numbering). Bands for YopH164 phosphatase (50 kDA) and Abl kinase domain constructs (28–35 kDA) are labeled.

4.4.2 Screen of 96 kinases finds 52 with useful levels of automated *E. coli* expression

To begin exploring which human kinase domains can achieve useful expression in *E. coli* using a simple automatable expression and purification protocol, a panel of kinase domain constructs for 96 kinases, for which bacterial expression has been previously demonstrated, was assembled using a semi-automated bioinformatics pipeline. Briefly, a database was built by querying Uniprot [218] for human protein kinase domains that were both active and not truncated. This query returned a set of target sequences that were then matched to their relevant PDB constructs and filtered for expression system (as determined from PDB header EXPRESSION_SYSTEM records), discarding kinases that did not have any PDB entries with bacterial expression. As a final filtering step, the kinases were compared to three purchased kinase plasmid libraries (described in Methods), discarding kinases without a match. Construct boundaries were selected from PDB constructs and the SGC plasmid library, both of which have experimental evidence for *E. coli* expression, and subcloned from a plasmid in a purchased library (see Methods). Selecting the kinases and their constructs for this expression trial in this method rested on the basis of expected success: these specific kinase constructs were bacterially expressed and purified to a degree that a crystal structure could be solved. While expression protocols used to produce protein for crystallographic studies are often individually tailored, we considered these kinases to have a high likelihood of expressing in our semi-automated pipeline where the *same* protocol is utilized for all kinases. Statistics of the number of kinases obtained from the PDB mining procedure are shown in Figure 4.3A. Surprisingly, the most highly sampled family was the CAMK family, suggesting researchers may have found this family particularly amenable to

bacterial expression. Based on the results of the previous experiment scanning Abl constructs for expression, we decided to use construct boundaries that were reported in the literature for each kinase. This process resulted in a set of 96 plasmid constructs distributed across kinase families (Figure 4.3B).

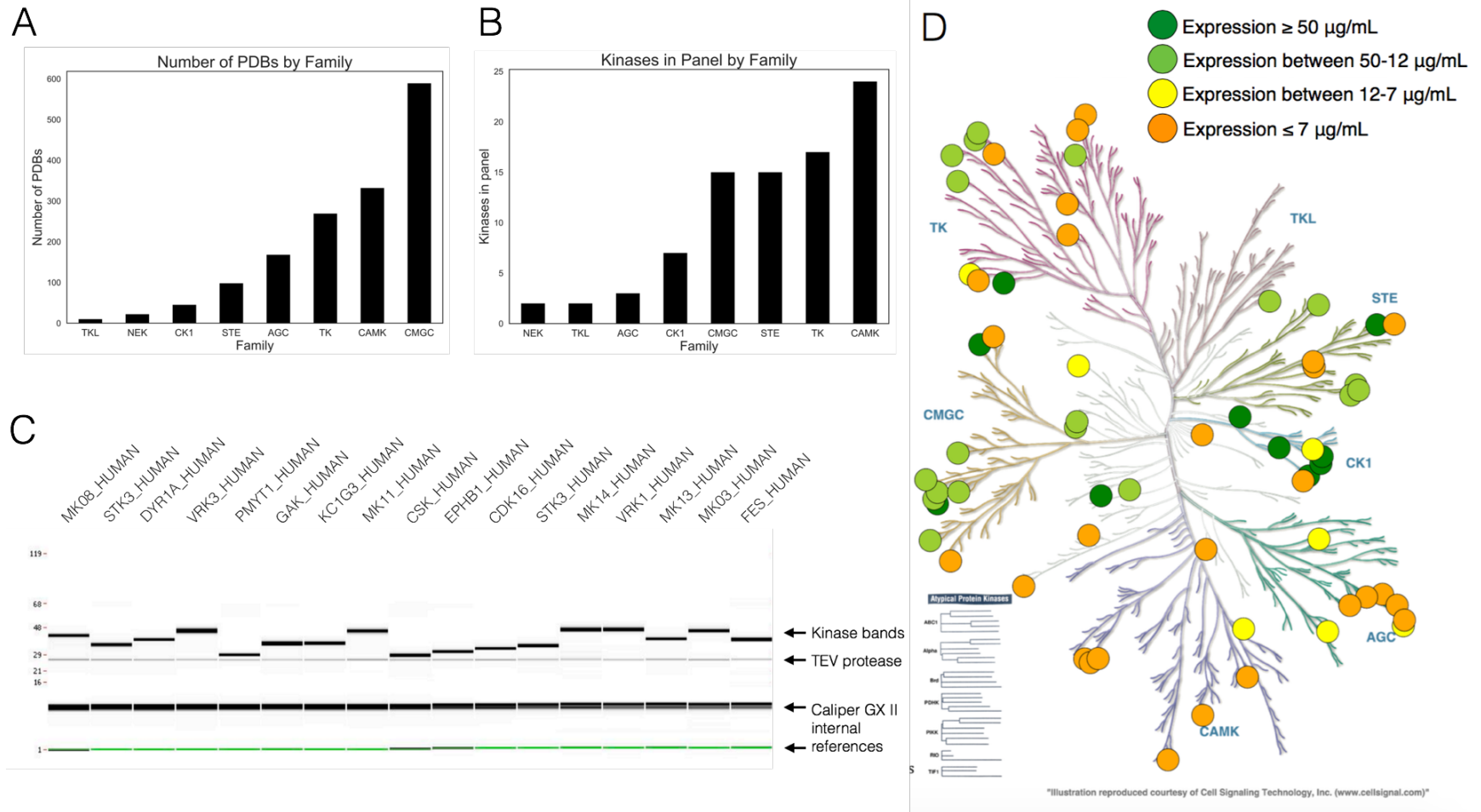


Figure 4.3: Kinome wide search for expressible kinases. (A) The number of PDB structures per kinase family, from the database built to select kinases for expression. (B) The distribution among families of candidate kinases in our expression screen. (C) Caliper GX II synthetic gel image rendering of the highest expressing kinases, quantified using microfluidic capillary electrophoresis. (D) Kinome distribution of expression based on our 96 kinase screen. Dark green circles represent kinases with expression above 50 $\mu\text{g}/\text{mL}$ culture yield. Light green circles represent kinases with expression between 50 and 12 $\mu\text{g}/\text{mL}$ yield. Yellow circles represent kinases with expression between 12 and 7 $\mu\text{g}/\text{mL}$ yield. Orange circles represent kinases with any expression (even below 2 $\mu\text{g}/\text{mL}$) up to 7 $\mu\text{g}/\text{mL}$ yield. Image made with KinMap: <http://www.kinhub.org/kinmap>.

Table 4.1: **Kinase domain constructs with yields >2 $\mu\text{g}/\text{mL}$ culture for 96-kinase expression screen.** Kinases are listed by Uniprot designation and whether they were co-expressed with Lambda or truncated YopH164 phosphatase. Yield (determined by Caliper GX II quantitation of the expected size band) reported in $\mu\text{g}/\text{mL}$ culture, where total eluate volume was 120 μL from 900 μL bacterial culture. Yields are shaded green (yield > 12 $\mu\text{g}/\text{mL}$), yellow (12 > yield > 7 $\mu\text{g}/\text{mL}$) and orange (yield < 7 $\mu\text{g}/\text{mL}$); kinase domain constructs with yields that were undetectable or < 2 $\mu\text{g}/\text{mL}$ are not listed. ‡ denotes that the second kinase domain of KS6A1_HUMAN was expressed; all other kinases were the first or only kinase domain occurring in the ORF. Construct boundaries are listed in UniProt residue numbering for the UniProt canonical isoform. An interactive table of expression yields and corresponding constructs is available at <http://choderalab.org/kinome-expression>

Kinase	Construct Boundary	Plasmid Source and ID	Phosphatase	Yield ($\mu\text{g}/\text{mL}$)
MK14_HUMAN	1–360	Addgene 23865	Lambda	70.7
VRK3_HUMAN	24–352	SGC Oxford VRK3A-c016	Lambda	67.5
GAK_HUMAN	24–359	SGC Oxford GAKA-c006	Lambda	64.7
CSK_HUMAN	186–450	Addgene 23941	YopH	62.5
VRK1_HUMAN	3–364	Addgene 23496	Lambda	62.3
KC1G3_HUMAN	24–351	SGC Oxford CSNK1G3A-c002	Lambda	56.3
FES_HUMAN	448–822	Addgene 23876	YopH	44.0
PMYT1_HUMAN	24–311	SGC Oxford PKMYT1A-c004	Lambda	38.0
MK03_HUMAN	1–379	Addgene 23509	Lambda	36.4
STK3_HUMAN	16–313	Addgene 23818	Lambda	34.3
DYR1A_HUMAN	24–382	SGC Oxford DYRK1AA-c004	Lambda	34.1
KC1G1_HUMAN	24–331	SGC Oxford CSNK1G1A-c013	Lambda	34.1
MK11_HUMAN	24–369	SGC Oxford MAPK11A-c007	Lambda	31.7
MK13_HUMAN	1–352	Addgene 23739	Lambda	31.7
EPHB1_HUMAN	602–896	Addgene 23930	YopH	28.9
MK08_HUMAN	1–363	HIP pJP1520 HsCD00038084	Lambda	28.5
CDK16_HUMAN	163–478	Addgene 23754	Lambda	26.9
EPHB2_HUMAN	604–898	HIP pJP1520 HsCD00038588	YopH	25.1
PAK4_HUMAN	291–591	Addgene 23713	Lambda	23.9
CDKL1_HUMAN	2–304	SGC Oxford CDKL1A-c024	Lambda	23.2
SRC_HUMAN	254–536	Addgene 23934	YopH	22.0
STK16_HUMAN	24–316	SGC Oxford STK16A-c002	Lambda	20.7
MAPK3_HUMAN	33–349	Addgene 23790	Lambda	18.8
PAK6_HUMAN	383–681	Addgene 23833	Lambda	18.0

Table 4.1 continued from previous page

Kinase	Construct Boundary	Plasmid Source and ID	Phosphatase	Yield ($\mu\text{g/mL}$)
CSK22_HUMAN	1–334	HIP pJP1520 HsCD00037966	Lambda	17.9
MERTK_HUMAN	570–864	Addgene 23900	YopH	16.8
PAK7_HUMAN	24–318	SGC Oxford PAK5A-c011	Lambda	14.7
CSK21_HUMAN	1–335	Addgene 23678	Lambda	14.5
EPHA3_HUMAN	606–947	Addgene 23911	YopH	14.1
BMPR2_HUMAN	1–329	SGC Oxford BMPR2A-c019	Lambda	14.1
M3K5_HUMAN	659–951	HIP pJP1520 HsCD00038752	Lambda	14.0
KCC2G_HUMAN	24–334	SGC Oxford CAMK2GA-c006	Lambda	13.3
E2AK2_HUMAN	254–551	HIP pJP1520 HsCD00038350	Lambda	11.6
MK01_HUMAN	1–360	HIP pJP1520 HsCD00038281	Lambda	11.2
CSKP_HUMAN	1–340	HIP pJP1520 HsCD00038384	Lambda	10.1
CHK2_HUMAN	210–531	Addgene 23843	Lambda	8.1
KC1G2_HUMAN	4–312	SGC Oxford CSNK1G2A-c002	Lambda	7.6
DMPK_HUMAN	24–433	SGC Oxford DMPK1A-c026	Lambda	7.6
KCC2B_HUMAN	11–303	Addgene 23820	Lambda	7.1
FGFR1_HUMAN	456–763	Addgene 23922	YopH	6.1
KS6A1_HUMAN [‡]	413–735	SGC Oxford RPS6KA1A-c036	Lambda	5.7
DAPK3_HUMAN	9–289	Addgene 23436	Lambda	4.0
STK10_HUMAN	18–317	HIP pJP1520 HsCD00038077	Lambda	3.7
KC1D_HUMAN	1–294	Addgene 23796	Lambda	3.7
KC1E_HUMAN	1–294	Addgene 23797	Lambda	3.5
NEK1_HUMAN	23–350	SGC Oxford NEK1A-c011	Lambda	3.3
CDK2_HUMAN	1–297	Addgene 23777	Lambda	3.1
ABL1_HUMAN	229–512	HIP pJP1520 HsCD00038619	YopH	2.5
DAPK1_HUMAN	2–285	HIP pJP1520 HsCD00038376	Lambda	2.4
DYRK2_HUMAN	23–417	SGC Oxford DYRK2A-c023	Lambda	2.4
HASP_HUMAN	24–357	SGC Oxford GSG2A-c009	Lambda	2.3
FGFR3_HUMAN	449–759	Addgene 23933	YopH	2.3

From these constructs, a set of 96 His10-TEV N-terminally tagged kinase domain constructs were generated, coexpressed with a phosphatase in *E. coli*, purified via nickel bead pulldown, and quantified using microfluidic gel electrophoresis. The 96 kinases were coexpressed with either Lambda phosphatase (for Ser/Thr kinases) or a truncated form of YopH phosphatase³ (for Tyr kinases).

Instead of eluting with imidazole, purified kinase was cleaved off nickel beads by the addition of 10% TEV protease to minimize phosphatase contamination in the resulting eluate, allowing us to assess whether resulting yields would be sufficient (and sufficiently free of phosphatase) to permit activity assays. While the initial panel of 96 kinases was well-distributed among kinase families (Figure 4.3B), the most highly expressing kinases (yield of more than 12 μg kinase/mL culture) were not evenly distributed (Figure 4.3D). While many of the kinases chosen from the CMGC and CK1 families expressed well in our panel, nearly all of the kinases from the CAMK and AGC family express below 12 μg kinase/mL (Figure 4.3D). 52 kinases demonstrated a useful level of soluble protein expression, here defined as greater than 2 μg /mL, naïvely expected to scale up to better than 2 mg/L culture (Table 4.1). Some kinases (shaded green in Table 4.1) demonstrated very high levels of expression, while others (shaded orange in Table 4.1) would likely benefit from further rounds of construct boundary optimization or solubility tags to boost soluble expression. The 17 most highly expressing kinases showed relatively high purity after elution, though we note that eluting via TEV site cleavage results in a quantity of TEV protease in the eluate (Figure 4.3C), but does not cause the elution of the His-tagged phosphatases which would hinder the ability to perform kinase activity assays.

³Yoph164 phosphatase, engineered to minimize intrinsic affinity for nickel purification resin by the QB3 MacroLab based on parent plasmid pCDFDuet1-YOPH, a gift from the Kuriyan Lab.

Further optimization of elution conditions may be required for optimizing kinase recovery via TEV cleavage [219–221].

Constructs with expression yields above 2 $\mu\text{g}/\text{mL}$ have been made available via Addgene: <https://www.addgene.org/kits/chodera-kinase-domains>

4.4.3 High-expressing kinases are folded with a well-formed ATP binding site

To determine whether the expressed kinases were properly folded, we performed both a fluorescence-based thermostability assay (Figure 4.4) as well as a fluorescent ATP-competitive ligand binding measurement to quantify whether the ATP binding site was well-formed (Figure 4.5).

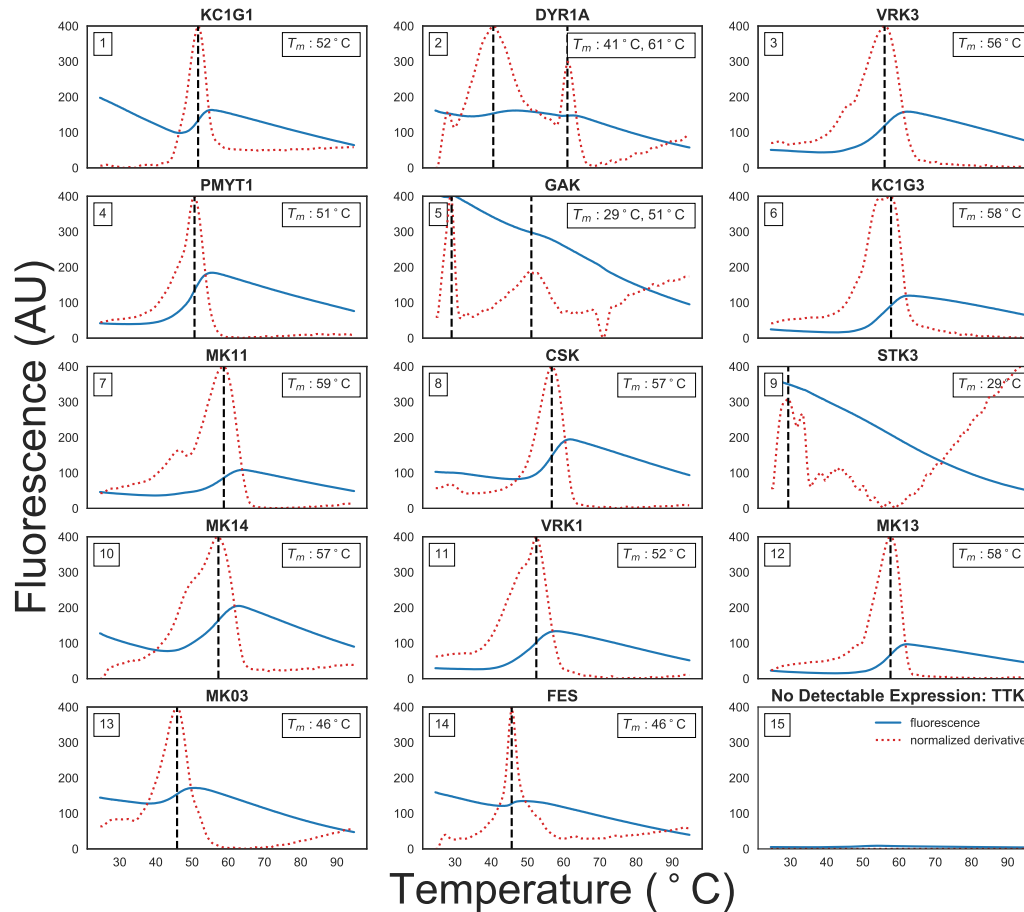


Figure 4.4: **Fluorescence-based thermostability assay demonstrates many high-expressing kinases are well-folded.** A fluorescence-based thermostability assay was performed on the 14 kinases shown to express above a minimum 0.24 mg/mL concentration after elution. SYPRO Orange fluorescence (solid blue line) was measured at 580 nm (half bandwidth 20 nm) after excitation at 465 nm (half bandwidth 25 nm) as the temperature was ramped from (x-axis) in Nickel Buffer A (25 mM HEPES pH 7.5, 5% glycerol, 400 mM NaCl, 20 mM imidazole, 1 mM BME). The temperature was held at 25°C for 15 sec before ramping up to 95°C with a ramp rate of 0.06°C/s. The unfolding temperature T_m (black dashed line and insert) was determined from the maxima of the normalized first derivative of fluorescence (red dashed line). Fluorescence emission at 580 nm is shown on the left y-axis. To control for signals resulting from TEV protease contamination present at 0.01–0.03 mg/mL, TTK, a kinase with no detectable expression in our panel as determined via Caliper GX II quantitation was included (panel 15).

4.4.4 Fluorescence-based thermostability assay

A fluorescence-based thermostability assay was performed with the hydrophobic dye SYPRO Orange to determine whether a strong two-state unfolding signal could be observed (see Methods). Also referred to as *thermofluor* or *differential scanning fluorimetry (DSF)*, as the temperature is slowly increased, unfolded proteins will expose hydrophobic patches that SYPRO orange will bind to, causing an increase in fluorescence [222–224]. While the fluorescence of solvated SYPRO Orange is temperature-dependent, clear unfolding temperatures (T_m) can often be identified from peaks in the first derivative of the observed fluorescence signal. Figure 4.4 shows the fluorescence (blue line), the absolute value of its derivative (red dashed line), and the unfolding temperature determined from the maximum absolute derivative (T_m) for the 14 kinases that were eluted to concentrations above 0.24 mg/mL eluate, which was determined to be the minimum concentration required for optimal resolution of melting curves upon dilution to 10 μ L. Because TEV-eluted kinase was used directly in this assay, TEV protease contaminant varies from 0.01–0.03 mg/mL in the resulting assay mix. The selected minimum concentration ensured that the kinase was roughly an order of magnitude higher concentration than the contaminating TEV.

Most of the kinases assayed had strong peaks above room temperature, suggesting that they are well-folded in the elution buffer (25 mM HEPES pH 7.5, 5% glycerol, 400 mM NaCl, 20 mM imidazole, 1 mM BME) at room temperature. Some kinases, such as a DYR1A and GAK (Figure 4.4, panels 6 and 9), had two shallow inflection points in SYPRO fluorescence as a function of temperature. While STK3 does not have a strong peak above room temperature, titration with an ATP-competitive inhibitor suggests this kinase either has a well-formed ATP binding site or folding can be induced by ligand binding (Figure 4.5, panel 10). As

a control, a sample with no detectable kinase expression (TTK from our expression panel) was assayed (Figure 4.4, panel 9), which showed nearly no fluorescence signal.

4.4.5 ATP-competitive inhibitor binding fluorescence assay

To determine whether expressed kinases had well-folded ATP binding sites, we probed their ability to bind an ATP-competitive inhibitor. While a pan-kinase inhibitor such as staurosporine could be used as a fluorescent probe [225], the ATP-competitive inhibitor bosutinib shows a much stronger increase in fluorescence around 450–480 nm when bound to kinases with well-folded ATP binding sites [226, 227]. While excitation at 350 nm can be used, excitation at 280 nm results in lower background, potentially due to fluorescent energy transfer between kinase and ligand. Despite the weak affinity of bosutinib for many kinases, its aqueous solubility is sufficient to provide a quantitative assessment of ATP-competitive binding to many kinases at sufficiently high concentrations to function as a useful probe [226, 227].

Here, we utilized this approach as a *qualitative* probe for ATP-competitive ligand binding, due to uncertainty in the ligand concentration caused by significant evaporation over the course of the sequential titration experiment (see Methods section for a more in depth discussion). 33 of the kinases in our expression panel had sufficient yields to prepare 100 μL of 0.5 μM kinase assay solutions, and were assessed for binding to bosutinib (Figure 4.5, panels 1-33), with a concentration-dependent increase in fluorescence signal (colored spectra) over the baseline ligand fluorescence titrated into buffer (gray spectra) providing evidence of a well-formed ATP binding site. Six of the lowest expression kinase constructs

(Figure 4.5, panels 39-44) were prepared by diluted 20 μL to a reaction volume of 100 μL and assessed for bosutinib binding. Unexpectedly, these kinases also showed evidence of binding, suggesting this assay is able to detect a well-formed ATP binding site even for protein concentrations less than 0.5 μM . To demonstrate that unfolded kinases do not demonstrate this increase in fluorescence over ligand-only baseline, thermally denatured MK14 was included as a control next to folded MK14 from a large-scale expression prep (Figure 4.5, panels 37–38), with thermally denatured MK14 exhibiting little difference from titrating ligand into buffer alone.

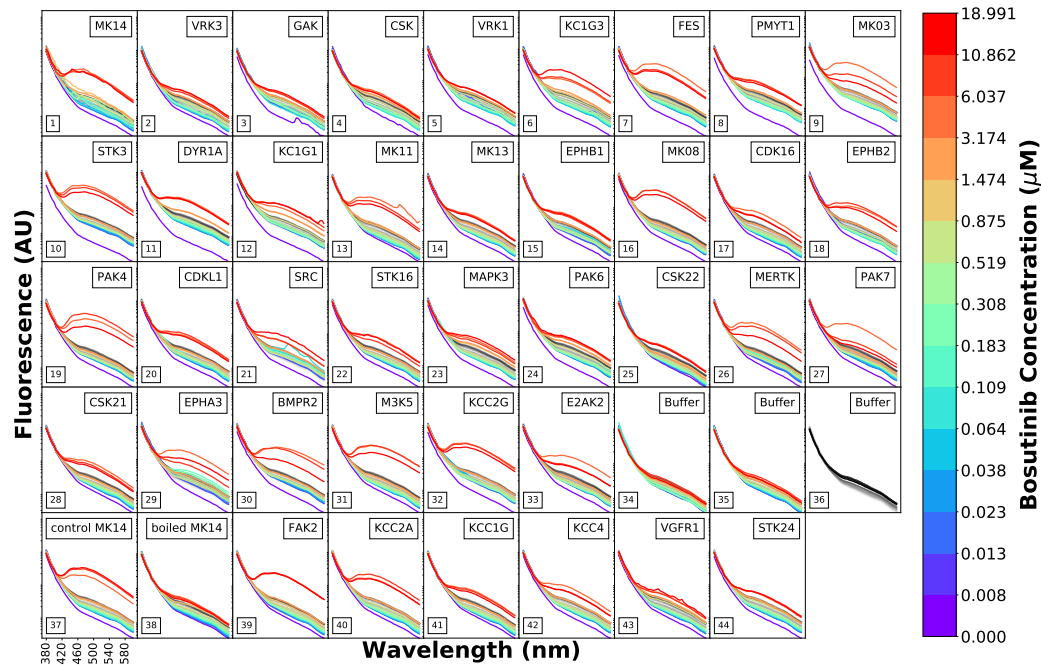


Figure 4.5: Fluorescence emission spectra as a function of the fluorescent ATP-competitive kinase inhibitor bosutinib demonstrates the presence of a well-formed ATP binding pocket. The ATP-competitive inhibitor bosutinib shows a strong increase in fluorescence centered around 450 nm when bound to kinases with well-folded ATP binding sites upon excitation at 280 nm [226]. To assess whether the kinases from the high-throughput expression screen were well-folded, bosutinib was titrated in a 15-concentration series geometrically spanning 0.008 μM to 18.99 μM (colored lines, higher concentrations are shown in warmer colors) in 15 increments for 39 expressing kinases with protein concentration adjusted to $\sim 0.5 \mu\text{M}$ in 100 μL assay volume. Eluted TEV protease contaminant varies from 0.01–0.03 mg/mL in the assay volumes. The control MK14 and boiled MK14 (boiled for 10 min at 95°C) were produced in a large scale expression from the same plasmid as used in the high-throughput expression protocol and they were included as positive and negative controls for bosutinib binding to ATP binding pocket. Fluorescence emission spectra (y-axis, bandwidth 20 nm) were measured from 370 nm to 600 nm (x-axis) for excitation at 280 nm (bandwidth 10 nm). For reference, the fluorescence of bosutinib titrated into buffer titration (panel 36) is shown in grayscale in each panel. Significant increases in fluorescence signal above baseline qualitatively indicate the presence of a well-formed ATP binding site.

4.4.6 Expressing clinically-derived Src and Abl mutants

Next-generation sequencing has enabled generation of massive datasets rich with missense alterations in kinases observed directly in the clinic [37–39], and has been particularly transformative in the field of oncology. To determine how well our human kinase domain panel supports the automated expression of clinically-identified missense mutants for biophysical, biochemical, and structural characterization, we attempted to express 96 missense mutations mined from sequencing studies of cancer patients. The mutations were gathered using cBioPortal [228] from publicly available sources and a large clinical tumor sequencing dataset from the Memorial Sloan Kettering Cancer Center [38] sequenced in the MSK-IMPACT panel [229].

Table 4.2: **Expression yields for engineered clinical missense mutants of Abl kinase domains with yields > 2 $\mu\text{g}/\text{mL}$ culture.** Abl kinase domain constructs with engineered clinical mutations with expression yields >2 $\mu\text{g}/\text{mL}$ culture are listed, sorted by yield. Yield was determined by Caliper GX II quantitation of the expected size band and reported in $\mu\text{g}/\text{mL}$ culture, where total eluate volume was 80 μL purified from 900 μL bacterial culture. Wild-type (WT) controls for both Src and Abl (here, a single well for each) are shown as the first entry for each gene.

Abl1 (229–512)	Mutation^a	Functional Impact Score^b	yield ($\mu\text{g}/\text{mL}$)	% of WT expression
	WT	–	5.1	–
	I403T	Low	17.8	350
	I293M	Low	9.8	193
	P309S	Neutral	7.8	153
	E453K	Low	7.3	144
	Y440H	Medium	7.1	140
	E292D	Low	6.9	135
	G251C	High	5.2	102
	E282Q	Neutral	5.1	102
	G250R	Neutral	5.1	100
	G254R	High	5.0	98
	Y312C	Neutral	4.7	93
	E453Q	Low	3.7	73
	R328K	Low	3.5	69
	D482E	Neutral	2.5	49
	F382L	Medium	2.1	41
	G390W	Medium	2.1	41

^a Uniprot amino acid sequence numbering of primary isoform

^b MutationAssesor Score [230, 231], which predicts functional impact via conservation

Table 4.3: **Expression yields for engineered clinical missense mutants of Src kinase domains with yields > 2 $\mu\text{g}/\text{mL}$ culture.** Src kinase domain constructs with engineered clinical mutations with expression yields >2 $\mu\text{g}/\text{mL}$ culture are listed, sorted by yield. Yield was determined by Caliper GX II quantitation of the expected size band and reported in $\mu\text{g}/\text{mL}$ culture, where total eluate volume was 80 μL purified from 900 μL bacterial culture. Wild-type (WT) controls for both Src and Abl (here, a single well for each) are shown as the first entry for each gene.

Src (254–536)	Mutation ^a	Functional Impact Score ^b	yield ($\mu\text{g}/\text{mL}$)	% of WT expression
	WT	–	35.7	–
	T456S	Neutral	80.9	227
	R388G	Medium	61.5	172
	K298E	High	54.5	153
	V380M	Neutral	51.7	145
	D368N	Neutral	49.9	140
	D521N	Low	42.8	120
	R463Q	Neutral	38.4	108
	R391C	Neutral	37.5	105
	E323D	Low	37.2	104
	A309V	Low	35.9	98
	G303D	Neutral	34.1	96
	R362Q	Neutral	33.6	94
	L361M	Medium	31.7	89
	A421V	Neutral	30.7	86
	V402L	Neutral	30.6	86
	V397M	Medium	29.8	84
	Q278E	Neutral	29.6	83
	Q312H	Low	29.5	83
	L353V	Medium	29.0	81
	L454V	Neutral	29.0	81
	P307R	Neutral	28.6	80
	V340I	Low	28.0	78
	P307S	Neutral	24.2	68
	D476N	Neutral	23.3	65
	D351N	Neutral	22.9	64
	T293A	Neutral	22.2	62

Table 4.3 continued from previous page

Src (254–536)	Mutation ^a	Functional Impact Score ^b	yield ($\mu\text{g/mL}$)	% of WT expression
	S345C	Low	22.2	62
	P428S	Medium	22.2	62
	E507D	Neutral	20.7	58
	D389E	High	20.0	56
	R503Q	Neutral	17.3	49
	D407H	High	15.9	45
	R463L	Neutral	14.9	42
	G291C	Medium	11.9	33
	G347E	Medium	10.2	29
	R483W	High	9.8	27
	P487L	Medium	6.0	17
	R463W	Medium	5.2	15
	R362W	Low	3.9	11
	S493F	Low	3.0	8
	P491S	Low	2.2	6

^a Uniprot amino acid sequence numbering of primary isoform

^b MutationAssesor Score [230, 231], which predicts functional impact via conservation

Using our structural informatics pipeline, a database was built focusing on the kinases we found to be expressible in *E. coli*. To add the mutation data, we retrieved public datasets from cBioPortal [232, 233] along with annotations from Oncotator [234] through their respective web service APIs. We then added mutations and annotations from the MSKCC dataset [38] by extracting the mutations from a local copy of the dataset and retrieving annotations from Oncotator. The annotated mutations were filtered for mutations that occurred within the construct boundaries of our kinase domains. We found 63 unique clinical mutations appearing within our kinase domain construct boundaries for Abl and 61 for Src. We subsequently selected 48 mutants for Abl and 46 for Src to express, aiming for a panel of mutants distributed throughout the kinase domain (Figure 4.6A), with wild-type sequences included as controls. Mutations were introduced using site-directed mutagenesis and assayed for expression yields (Figure 4.6B). Those with yields above 2 μg kinase/mL culture are listed in Tables 4.3 and 4.2.

High-expressing mutants appear to be distributed relatively uniformly throughout the kinase domain (Figure 4.6A). While the vast majority of the Src mutants expressed at a usable level, many of the Abl mutants expressed below the 2 μg /mL threshold. This can primarily be attributed to the low level of expression for wild-type Abl construct (Table 4.1). In instances where kinase activity is not required, yield could be increased via the introduction of inactivating mutations [213] or further tailoring of expression and purification protocols.

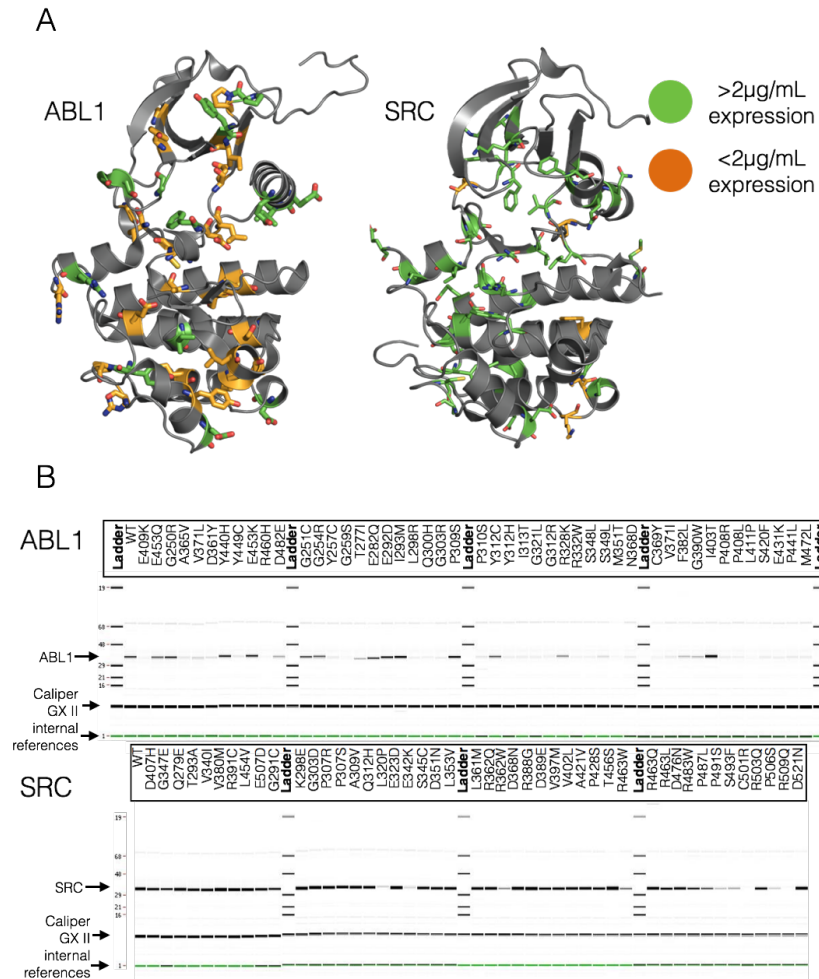


Figure 4.6: Expression yields for engineered clinically-derived Src and Abl missense mutants. (A) All Abl and Src clinically-identified mutants assessed in the expression screen are displayed as sticks. Mutants with expression yields $>2 \mu\text{g/mL}$ are colored green, while those with yields $<2 \mu\text{g/mL}$ are colored orange. Rendered structures are Abl (PDBID: 2E2B) and Src (PDBID: 4MXO) [227]. (B) Synthetic gel images showing Abl (*top*) or Src (*bottom*) expression, with wells labeled by missense mutation. Yield was determined by Caliper GX II quantitation of the expected size band and reported in $\mu\text{g/mL}$ culture, where total eluate volume was $120 \mu\text{L}$ following nickel bead pulldown purification from $900 \mu\text{L}$ bacterial culture. Residue mutations use numbering for the Uniprot canonical isoform.

4.5 Discussion

We have demonstrated that a simple, uniform, automatable protocol is able to achieve useful bacterial expression yields for a variety of kinase domain constructs. While yields could likely be further improved by a variety of methods—such as the addition of solubility-promoting tags, construct domain boundary and codon optimization, or mutations to improve the solubility or ablate catalytic activity—the simplicity of this approach suggests widespread utility of automated bacterial expression for biophysical, biochemical, and structural biology work for the further study of human kinase domains.

Our expression test of 81 different construct boundaries of the Abl kinase domain demonstrated a surprising sensitivity of expression yields to the precise choice of boundary. This sensitivity may be related to where the construct is truncated with respect to the secondary structure of the protein, as disrupting secondary structure could cause the protein to improperly fold, leading to low soluble protein yield even when total expression is high. Of note, the highest expressing C-terminal boundaries for Abl were residues 511 and 512. These residues fall in the regulatory alpha helix αI [20]. This helix has been shown to undergo a dramatic conformational change upon binding to the myristoylated N-terminal cap, which introduces a sharp “kink” in residues 516–519. These residues may lead to higher levels of soluble expression by truncating an secondary structural element that is unusually flexible. Indeed, this helix is not resolved in some X-ray structures (PDBID:2E2B) [217], further suggesting that this helix is less thermodynamically stable than expected. Control replicates of three constructs indicate good repeatability of expression yields in the high-throughput format. This screen suggests that optimization of construct boundaries could potentially further greatly increase yields of poorly expressing kinase domains. Codon

optimization for bacterial expression could also increase expression for kinase domains with low yield due to codon bias [235], as could coexpression with chaperones [236].

For those kinases that did express, a fluorescence-based thermostability assay indicated that many of the highest-expressing kinases are well folded. An ATP-competitive inhibitor binding fluorescent assay provides qualitative evidence that the 39 kinases that had sufficiently high expression levels to be assayed have a well-formed ATP-binding site capable of binding bosutinib, a small molecule ATP-competitive kinase inhibitor. Taken together, these two experiments demonstrate that our expression protocol produces folded kinases with utility for biophysical experiments and drug design.

The tolerance of these bacterial constructs to many engineered clinical missense mutations suggests a promising route to the high-throughput biophysical characterization of the effect of clinical mutants on anticancer therapeutics. Mutations that did not express well may destabilize the protein, or may increase the specific activity of the kinase. A higher specific activity would require more phosphatase activity, wasting ATP to prevent high levels of phosphorylation that have been hypothesized to cause difficulty expressing kinases without a coexpressed phosphatase in bacteria [213]. Mutations that are destabilizing may show improved expression if coexpressed with more elaborate chaperones such as GroEL and Trigger factor [236]. Mutations that increase the specific activity of the kinase might also express better when combined with an inactivating mutation.

High-throughput automated kinase expression could be combined with enzymatic or biophysical techniques for characterizing the potency of a variety of clinical kinase inhibitors to assess which mutations confer resistance or sensitivity. While the process of engineering, expressing, purifying, and assaying

mutants currently takes approximately two weeks, it is possible that new techniques for cell-free bacterial expression [237, 238] may reduce this time to a matter of days or hours in a manner that might be compatible with clinical time frames to impact therapeutic decision-making.

We hope that other laboratories find these resources useful in their own work.

4.6 Methods

4.6.1 Semi-automated selection of kinase construct sequences for *E. coli* expression

Selection of human protein kinase domain targets

Human protein kinases were selected by querying the UniProt API (query date 30 May 2014) for any human protein with a domain containing the string "protein kinase", and which was manually annotated and reviewed (i.e. a Swiss-Prot entry). The query string used was:

```
taxonomy:"Homo sapiens (Human) [9606]" AND domain:"protein kinase" AND reviewed:yes
```

Data was returned by the UniProt API in XML format and contained protein sequences and relevant PDB structures, along with many other types of genomic and functional information. To select active protein kinase domains, the UniProt domain annotations were searched using the regular expression `^Protein kinase(?!; truncated)(?!; inactive)`, which excludes certain domains annotated "Protein kinase; truncated" and "Protein kinase; inactive". Sequences

for the selected domains, derived from the canonical isoform as determined by UniProt, were then stored.

Matching target sequences with relevant PDB constructs

Each target kinase gene was matched with the homologous in any other species, if present, and all UniProt data was downloaded. This data included a list of PDB structures which contain the protein, and their sequence spans in the coordinates of the UniProt canonical isoform. PDB structures which did not include the protein kinase domain or truncated more than 30 residues at each end were filtered out. PDB coordinate files were then downloaded for each remaining PDB entry. The coordinate files contain various metadata, including the `EXPRESSION_SYSTEM` annotation, which was used to filter PDB entries for those which include the phrase "ESCHERICHIA COLI". The majority of PDB entries returned had an `EXPRESSION_SYSTEM` tag of "ESCHERICHIA COLI", while a small number had "ESCHERICHIA COLI BL21" or "ESCHERICHIA COLI BL21(DE3)".

The PDB coordinate files also contain SEQRES records, which should contain the protein sequence used in the crystallography or NMR experiment. According to the PDB-101 (<http://pdb101.rcsb.org/learn/guide-to-understanding-pdb-data/primary-sequences-and-the-pdb-format>), the SEQRES should include the "sequence of each chain of linear, covalently-linked standard or modified amino acids or nucleotides. It may also include other residues that are linked to the standard backbone in the polymer." However, we found that these records are very often misannotated, instead representing only the crystallographically resolved residues. Since expression levels can be greatly affected by insertions or deletions of only one or a few residues at either terminus [239], it is important to

know the full experimental sequence. To measure the authenticity of a given SEQRES record, we developed a simple metric by hypothesizing that most crystal structures would likely have at least one or more unresolved residues at one or both termini and that the presence of an expression tag, which is typically not crystallographically resolved, would indicate an authentic SEQRES record. To achieve this, unresolved residues were first defined by comparing the SEQRES sequence to the resolved sequence, using the SIFTS service to determine which residues were not present in the canonical isoform sequence [240]. Regular expression pattern matching was used to detect common expression tags at the N- or C-termini. Sequences with a detected expression tag were given a score of 2, while those with any unresolved sequence at the termini were given a score of 1, and the remainder were given a score of 0. This data was stored to allow for subsequent selection of PDB constructs based on likely authenticity in later steps. The number of residues extraneous to the target kinase domain, and the number of residue conflicts with the UniProt canonical isoform within that domain span were also stored for each PDB sequence.

Plasmid libraries

As a source of kinase DNA sequences for subcloning, we purchased three kinase plasmid libraries: the [Addgene Human Kinase ORF kit](#), a kinase library from the Structural Genomics Consortium (SGC), Oxford (<http://www.thesgc.org>), and a kinase library from the [PlasmID Repository](#) maintained by the Dana-Farber/Harvard Cancer Center. Annotated data for the kinases in each library was used to match them to the human protein kinases selected for this project. The plasmid open reading frames (ORFs) were translated into protein sequences and aligned against the target kinase domain sequences from UniProt.

Also calculated were the number of extraneous protein residues in the ORF, relative to the target kinase domain sequence, and the number of residue conflicts with the UniProt sequence. Our aim was to subclone the chosen sequence constructs from these library plasmids into our expression plasmids.

Selection of sequence constructs for expression

Of the kinase domain targets selected from UniProt, we filtered out those with no matching plasmids in our available plasmid libraries or no suitable PDB construct sequences. For this purpose, a suitable PDB construct sequence was defined as any with an authenticity score greater than zero (see above). Library plasmid sequences and PDB constructs were aligned against each UniProt target domain sequence, and various approaches were considered for selecting the construct boundaries to use for each target, and the library plasmid to subclone it from. Candidate construct boundaries were drawn from two sources: PDB constructs and the SGC plasmid library, has been successfully tested for *E. coli* expression.

For most of the kinase domain targets, multiple candidate construct boundaries were available. To select the most appropriate construct boundaries, we sorted them first by authenticity score, then by the number of conflicts relative to the UniProt domain sequence, then by the number of residues extraneous to the UniProt domain sequence span. The top-ranked construct was then chosen. In cases where multiple library plasmids were available, these were sorted first by the number of conflicts relative to the UniProt domain sequence, then by the number of residues extraneous to the UniProt domain sequence span, and the top-ranked plasmid was chosen. This process resulted in a set of 96 kinase domain constructs, which (by serendipity) matched the 96-well plate format we

planned to use for parallel expression testing. We selected these constructs for expression testing.

An interactive table of the selected plasmids, constructs, and aligned PDB files can be viewed at <http://choderalab.org/kinome-expression>.

Automation of the construct selection process

While much of this process was performed programmatically, many steps required manual supervision and intervention to correct for exceptional cases. While these exceptions were encoded programmatically as overrides to ensure the scheme could be reproduced from existing data, we hope to eventually develop a fully automated software package for the selection of expression construct sequences for a given protein family, but this was not possible within the scope of this work.

4.6.2 Mutagenesis protocol

Point mutations were introduced with a single-primer QuikChange reaction. Primers were designed to anneal at 55°C both upstream and downstream of the point mutation, and with a total length of approximately 40 bases. At the codon to be modified, the fewest possible number of bases was changed. Plasmid template (160 ng) was mixed with 1 μ M primer in 1x PfuUltra reaction buffer, with 0.8 mM dNTPs (0.2 mM each) and 1 U PfuUltra High-Fidelity DNA polymerase (Agilent), in a total volume of 20 μ L. Thermocycler settings were 2 min at 95°C, followed by 18 cycles of 20s at 95°C, 1 min at 53°C, 12 min at 68°C (2min/kb), then 1 minute at 68°C. After cooling to room temperature, 4 μ L of the

PCR reaction was added to 16 μ L CutSmart Buffer (NEB) containing 10 U DpnI (NEB). After incubation for 2.5 hours at 37°C, 6 μ L of this mixture was used to directly transform XL1-Blue chemically competent cells (Agilent) according to the manufacturer's protocol. Transformants were picked for plasmid mini-preps and the presence of the point mutations was confirmed by sequencing.

4.6.3 Expression testing

For each target, the selected construct sequence was subcloned from the selected DNA plasmid. Expression testing was performed at the QB3 MacroLab (QB3 MacroLab, University of California, Berkeley, CA 94720)

[<http://qb3.berkeley.edu/macrolab/>], a core facility offering automated gene cloning and recombinant protein expression and purification services.

Each kinase domain was tagged with a N-terminal His10-TEV and coexpressed with either the truncated YopH164 for Tyr kinases or lambda phosphatase for Ser/Thr kinases. All construct sequences were cloned into the 2BT10 plasmid, an AMP resistant ColE1 plasmid with a T7 promoter, using ligation-independent cloning (LIC). The inserts were generated by PCR using the LICv1 forward (TACTTCCAATCCAATGCA) and reverse (TTATCCACTTCCAATGTTATTA) tags on the primers. Gel purified PCR products were LIC treated with dCTP. Plasmid was linearized, gel purified, and LIC-treated with dGTP. LIC-treated plasmid and insert were mixed together and transformed into XL1-Blues for plasmid preps.

Expression was performed in Rosetta2 cells (Novagen) grown with Magic Media (Invitrogen autoinducing medium), 100 μ g/mL of carbenicillin and 100 μ g/mL of spectinomycin. Single colonies of transformants were cultivated with 900 μ L of MagicMedia into a gas permeable sealed 96-well block. The cultures were

incubated at 37°C for 4 hours and then at 16°C for 40 hours while shaking. Next, cells were centrifuged and the pellets were frozen at -80°C overnight. Cells were lysed on a rotating platform at room temperature for an hour using 700 μ L of SoluLyse (Genlantis) supplemented with 400 mM NaCl, 20 mM imidazole, 1 μ g/mL pepstatin, 1 μ g/mL leupeptin and 0.5 mM PMSF.

For protein purification, 500 μ L of the soluble lysate was added to a 25 μ L Protino Ni-NTA (Machery-Nagel) agarose resin in a 96-well filter plate. Nickel Buffer A (25 mM HEPES pH 7.5, 5% glycerol, 400 mM NaCl, 20 mM imidazole, 1 mM BME) was added and the plate was shaken for 30 min at room temperature. The resin was washed with 2 mL of Nickel Buffer A. For the 96-kinase expression experiment, target proteins were eluted by a 2 hour incubation at room temperature with 10 μ g of TEV protease in 80 μ L of Nickel Buffer A per well and a subsequent wash with 40 μ L of Nickel Buffer A to maximize protein release. Nickel Buffer B (25 mM HEPES pH 7.5, 5% glycerol, 400 mM NaCl, 400 mM imidazole, 1 mM BME) was used to elute TEV resistant material remaining on the resin. Untagged protein eluted with TEV protease was run on a LabChip GX II Microfluidic system to analyze the major protein species present.

For the clinical mutant and Abl1 construct boundaries expression experiments, target proteins were washed three times with Nickel Buffer A prior to elution in 80 μ L Nickel Buffer B. The eluted protein was run on a LabChip GX II Microfluidic system to analyze with major protein species were present.

4.6.4 Fluorescence-based thermostability assay

To assess whether the highly-expressed wild-type kinase constructs are folded, a thermofluor thermostability assay [222–224] was performed for kinase constructs

that have a minimum of 0.24 mg/mL protein concentration in the eluate. After diluting 9 μ L of eluate by 1 μ L dye, the effective assay concentration is 0.216 mg/mL minimum in 10 μ L assay volume. Previous optimization efforts in the lab determined that 0.20 mg/mL was the lower limit of well-defined T_m detection. This minimum concentration also ensured that the kinase was present at roughly an order of magnitude concentration higher than contaminating TEV protease.

Kinase expression panel eluates, which were kept in 96-well deep well plate frozen at -80°C for 2 years prior to the thermal stability assay, were thawed in an ice-water bath for 30 min. 9 μ L of each kinase eluate was added to a 384 well PCR plate (4titude-0381). 100X SYPRO Orange dye solution was prepared from a 5000X DMSO solution of SYPRO Orange Protein Gel Stain (Life Technologies, Ref S6650, LOT 1790705) by dilution in distilled water. In initial experiments, SYPRO Orange dye solution was diluted in kinase binding assay buffer (20 mM Tris 0.5 mM TCEP pH 8), which caused the dye to precipitate out of solution. Particulates in the dye solution were pelleted by tabletop centrifugation (2 min, 5000 RCF) and the solution was kept covered with aluminum foil in the dark to prevent photodamage. 1 μ L of 100X dye solution was added to each kinase eluate sample in 384-well PCR plate. The plate was sealed with Axygen UC-500 Ultra Clear Pressure Sensitive sealing film. To remove any air bubbles, the sample plate was centrifuged for 30 sec with 250 g using Bionex HiG4 centrifuge. Sample mixing was performed by orbital shaking with Inheco shakers for 2 min at 1000 RPM.

A thermofluor melt was performed using a LightCycler 480 (Roche) qPCR instrument using an excitation filter of 465 nm (half bandwidth 25 nm) and emission filter at 580 nm (half bandwidth 20 nm). LightCycler 480 Software

Version 1.5.1 was used to operate the instrument and analyze the results. The temperature was held at 25°C for 15 s before ramping up to 95°C with a ramp rate of 0.06°C/s. During temperature ramp 10 fluorescence acquisitions/°C were recorded with dynamic integration time mode, melt factor of 1, quant factor of 10, and maximum integration time of 2 sec. Thermal protein denaturation causes hydrophobic patches of protein to be exposed, which SYPRO Orange dye can bind. Binding of SYPRO Orange dye is detected as an increase in fluorescence at 580 nm. Presence of a clear thermal denaturation peak in the absolute value of the derivative of the fluorescence as a function of temperature serves as an indication that the proteins were well-folded. Observed fluorescence was plotted as a function of temperature, and a melting temperature T_m was determined as the maximum of the absolute value of its first derivative.

4.6.5 ATP-competitive inhibitor binding fluorescence assay

To determine whether the expressed kinases had a well-folded ATP-binding site, we assessed whether the eluted kinase was capable of binding the ATP-competitive small molecule kinase inhibitor bosutinib. We designed fluorescence-based binding assays following earlier work reporting that this quinoline-scaffold inhibitor undergoes a strong increase in fluorescence upon binding (even weakly) to kinase ATP-binding sites [226]. By titrating in the ligand to close to the solubility limit, even weak binding to the ATP-binding site can be detected by observing emission increases around 450 nm during excitation at 280 nm.

For 33 of the kinases in our expression panel, 0.5 μ M kinase solutions from kinase expression panel eluates were prepared in kinase binding assay buffer (20 mM

Tris 0.5 mM TCEP pH 8) for a final volume of 100 μ L in a black 96-well vision plate (4titude-0223). Six low-expressing kinases (Figure 4.5, panels 39-44) were prepared by diluting 20 μ L of eluate in kinase binding assay buffer (20 mM Tris 0.5 mM TCEP pH 8) to a final volume of 100 μ L, for final concentrations below 0.5 μ M. The plate was shaken for 2 min clockwise and 2 min counter-clockwise by orbital shaking with Inheco shakers at 2000 RPM and centrifuged for 30 sec with 1000 g using Bionex HiG4 centrifuge. Fluorescence emission spectra were measured from 370 nm to 600 nm (20 nm bandwidth) in 5 nm steps using 280 nm excitation (10 nm bandwidth) from both the top and bottom of the well using a Tecan Infinite M1000 PRO.

Bosutinib free base (LC Labs, cat no. B-1788, lot no. BSB-103, M.W. 530.45 Da) was dispensed directly from a roughly 10 mM DMSO stock solution to the assay solution using a Tecan HP D300 Digital Dispenser. The 10 mM DMSO stock solution was prepared gravimetrically using an automated balance (Mettler Toledo Balance XPE205 with LabX Laboratory Software) by dispensing 39.02 mg solid Bosutinib powder stored under nitrogen gas at 25°C into 8.0499 g DMSO (Alfa Aesar, cat no. 42780, log no. Y25B604, density 1.1004 g/mL at ambient temperature) which is kept dry under argon gas at 25°C. To minimize atmospheric water absorption due to the hygroscopic nature of DMSO, the 10 mM stock solution was pipetted into wells of a 96-well stock plate by an automated liquid handling device (Tecan EVO 200 with air LiHa) and sealed with foil seal (PlateLoc). Ligand was dispensed to the assay plate with HP D300 (using aliquots of stock solution pipetted from a freshly pierced stock plate well) targeting a roughly geometrically-increasing series of ligand concentrations in each well to achieve the following total ligand concentrations after each dispense: 0.008 μ M, 0.013 μ M, 0.023 μ M, 0.038 μ M, 0.064 μ M, 0.109 μ M, 0.183 μ M, 0.308 μ M, 0.519 μ M, 0.875 μ M, 1.474 μ M, 3.174 μ M, 6.037 μ M, 10.862 μ M, 18.991 μ M. The

plate was shaken by HP D300 for 10 sec after usage of each dispensehead. After each titration, the plate was shaken with Inheco shakers (2 min clockwise and counter-clockwise, 2000 RPM, orbital shaking) and centrifuged (30 sec, 1000 g) using a Bionex HiG4 centrifuge. Fluorescence spectra from 370 nm to 600 nm (bandwidth 20 nm) in 5 nm steps using 280 nm excitation (bandwidth 10 nm) were read from both the top and bottom of the well using a Tecan Infinite M1000 PRO. In total, the experiment took 17.5 hours to complete due to the time-consuming spectral read after each dispense, likely resulting in significant evaporation from some wells during the experiment.

ATP-competitive binding was analyzed qualitatively for each kinase by plotting the fluorescence spectra as a function of concentration to detect concentration-dependent increases in fluorescence. As a control for background ligand fluorescence independent of protein binding, fluorescence spectra of three replicates of ligand into buffer titrations were plotted. As a positive control, MK14 produced by a validated large scale expression protocol (see Supplementary Methods) from the same plasmid used in the high-throughput protocol was included. To control for non-specific binding to unfolded protein, we included boiled MK14 (prepared from the large scale expression of MK14 by boiling at 95°C for 10 min). A concentration-dependent increase in fluorescence was interpreted as evidence that the ATP-binding site of the kinase was well folded and allowed for bosutinib binding. Due to the length of the experiment, it is possible that evaporation reduced the well volume below 100 μ L and potentially caused bosutinib to reach higher concentration levels than expected. This creates uncertainty for data points, as bosutinib may either be a higher concentration (due to evaporation) or a lower concentration (due to potential precipitation caused by lower well volumes) than expected. For this reason, we have interpreted the experiment as qualitative evidence of binding, instead of

quantitatively. Bosutinib binding is an indication of proper folding of the ATP binding pocket of these recombinantly expressed kinase constructs.

4.6.6 Large Scale expression and purification protocol for MK14

Large scale expression of MK14 was performed at the QB3 MacroLab (QB3 MacroLab, University of California, Berkeley, CA 94720

[<http://qb3.berkeley.edu/macrolab/>], a core facility offering automated gene cloning and recombinant protein expression and purification services.

Rosetta2(DE3)pLysS cells (Novagen) were used to co-express MK14 (same plasmid as from the high-throughput kinase expression panel) and 13SA Lamda phosphatase. The cells were grown in 2YT Medium (16 g/L Tryptone, 10 g/L Yeast Extract, 5 g/L NaCl) to OD₆₀₀ of 0.5 at 37°C. The culture was cooled to 16°C and induced with 0.5 mM IPTG overnight. The cultures were pelleted at 5000 rpm for 30 min and resuspended in 20 mL Nickel buffer A (25 mM HEPES pH 7.5, 10% glycerol, 400mM NaCl, 20 mM imidazole, 5 mM BME) with the following protease inhibitors: 1 µg/mL leupeptin, 1 µg/mL pepstatin, and 0.5 mM PMSF). The resuspended cells were frozen at -80°C.

When ready for purification, the cells were thawed and ruptured using a homogenizer (Avestin C3, 15000psi, 3 passes). The broken cells were pelleted at 15000 rpm for 30 min (SS34 rotor). Clarified lysate was loaded onto a 5 mL HisTrap FF Crude column (GE Healthcare) and washed with Nickel buffer A to remove any unbound material. The protein was eluted with Nickel buffer B (25 mM HEPES pH 7.5, 10% glycerol, 400mM NaCl, 400 mM imidazole, 5 mM BME) and pooled for buffer exchange into Nickel buffer A on a HiPrep 26/10 Desalting Column (GE Healthcare). Rough protein yields were quantified using

theoretical extinction coefficients calculated using ProtParam (<http://ca.expasy.org/tools/protparam.html>). The His tag was cleaved off of MK14 by incubation with TEV protease (25°C, 2 hours, 1:20 mass ratio).

After tag cleavage, the sample was run over a 5 mL HisTrap FF Crude column (GE Healthcare) with Nickel buffer A. The flow-through was collected, concentrated to roughly 5mL using centrifugal concentrators (10 kDA MWCO, Millipore) and loaded onto a HiPrep 16/60 Sephacryl S-200 HR column (GE Healthcare). The sample was equilibrated into Gel Filtration buffer (20 mM Tris-HCl pH 8.0, 150 mM NaCl, 5% glycerol, 1 mM DTT) and fractions containing MK14 were pooled and concentrated (10 kDA MWCO centrifugal concentrators, Millipore). 500 μ L aliquots of MK14 were snap frozen in liquid nitrogen and stored at -80°C. Quantification by theoretical extinction coefficient suggests the final MK14 concentration was roughly 4.0 mg/mL (97 μ M), roughly 22.4 mg/L of culture yield.

4.7 Author Contributions

Conceptualization, JDC, DLP, SKA, MI, LRL, SMH, NML, MAS; Methodology, DLP, MI, LRL, SMH, SKA, JDC, NML, MAS; Software, DLP, JDC, SMH; Formal Analysis, SKA, JDC, MI, SMH; Investigation, MI, LRL, SG, CJ, SKA, SMH; Resources, CJ, SG; Data Curation, SKA, MI, LRL, DLP, JMB; Writing-Original Draft, SKA, LRL, DLP, JDC, SG, SMH, MI; Writing - Review and Editing, SKA, JDC, MI, LRL, SHM, SG, CJ, NML, MAS; Visualization, SKA, JDC, MI, SMH; Supervision, JDC, NML, MAS; Project Administration, SKA, JDC, MI, SMH; Funding Acquisition, JDC, SMH

4.8 Acknowledgments

DLP, SMH, LRL, SKA, MI, and JDC acknowledge support from the Sloan Kettering Institute. This work was funded in part by the Marie-Josée and Henry R. Kravis Center for Molecular Oncology, the National Institutes of Health (NIH grant R01 GM121505 and National Cancer Institute Cancer Center Core grant P30 CA008748), the Functional Genomics Institute (FGI) at MSKCC, and a Louis V. Gerstner Young Investigator Award. MAS acknowledges funding support by NIH grant R35 GM119437. The authors are grateful to Gregory Ross (MSKCC) for assistance in preparing the computational infrastructure for selecting clinical point mutants, and to Sarah E. Boyce (current address: Schrödinger, New York, NY) for assistance with multiple stages of this project. We gratefully acknowledge the members of the MSKCC Molecular Diagnostics Service in the Department of Pathology for their efforts in collecting and compiling mutations for Abl and Src kinases used here. We thank the Kuriyan lab for the gift of pCDFDuet1-YOPH plasmid. The authors are grateful to [Addgene](#) for their help in making the plasmids generated by this work available to the research community at minimal cost.

CHAPTER 5

CONCLUSION

The work contained in this dissertation comprises contributions to the field of physical modeling to predict selectivity and resistance, as well as steps towards enabling the generation of the data sets needed to better evaluate free energy methodologies. Chapter 2 provides insight into the speedup in selectivity optimization we can expect from free energy calculations based on correlation in the forcefield error, using kinases as model system relevant to drug discovery efforts. Chapter 3 evaluates the ability of free energy calculations to predict the impact of missense mutations on kinase inhibitor binding, an important first step towards using physical modeling to characterize rare mutations, or develop next-generation inhibitors. Finally, Chapter 4 presents the first stage of the idealized pipeline to automate the process of identifying novel mutations, predicting their impact on inhibitor binding, and rapidly testing those predictions using biophysical experiments. The high-throughput expression protocol and biophysical assays discussed in Chapter 4 are key to generating that types of data needed to further validate the promising results discussed in Chapters 2 and 3.

Chapter 2 shows that correlation in the forcefield error can accelerate optimization of compounds for selectivity in a congeneric series of ligands. Maintaining a running estimate of the correlation coefficient ρ over the course of a discovery project will allow computational chemists to better understand the uncertainty in their selectivity predictions. Further, the work presented in Chapter 2 suggests that when the correlation of forcefield errors is high, expending additional computational effort to reduce statistical error will yield further speed ups. In Chapter 3, we present evidence that alchemical free energy calculations can accurately predict the impact of clinical kinase mutations on Abl

binding affinity for a number of different ligands. This work suggests that working from a single crystal structure for each ligand, or even docking a ligand into the crystal structure of a somewhat related ligand, is sufficient to predict the impact of mutations with good classification power. Chapter 4 presents an automated, one size fits all protocol for expressing a wide array of human protein kinases in *E. coli*. This protocol was shown to be useful for expressing clinical mutations in Src and Abl kinase, enabling its use for the development of selective and mutant-specific inhibitors, as well as for testing the methodologies presented in Chapters 2 and 3.

Each Chapter discusses the work that remains before free energy calculations can be used to their fullest potential for developing selective inhibitors, or next-generation inhibitors to overcome resistance. Larger data sets covering a wider range of proteins, even within the kinase family, will allow us to draw conclusions about the correlation of forcefield error, as well develop heuristics to predict expected correlation before running any calculations. Further work on additional forcefields, and the parameters that free energy calculations are sensitive to when predicting the impact of mutations, is important for moving free energy tools closer to being used to prioritize mutants for biophysical characterization, predicting potential resistance mutations, or understanding a novel clinical mutation. Importantly, much progress remains to be made on understanding the importance of including multiple conformations in predicting ligand binding affinity, especially for conformationally-plastic proteins like kinases.

While this dissertation primarily focuses using on alchemical free energy calculations, we expect that these approaches could be fruitfully combined with statistical learning methods. With the rapid accumulation of high-throughput

binding assays, clinical sequencing, and structural information, exciting future work will combine physical modeling with machine learning to utilize the wealth of data available for the kinome to improve predictions about selectivity and resistance.

Appendices

APPENDIX A
SUPPLEMENTAL FIGURES FROM CHAPTER 2

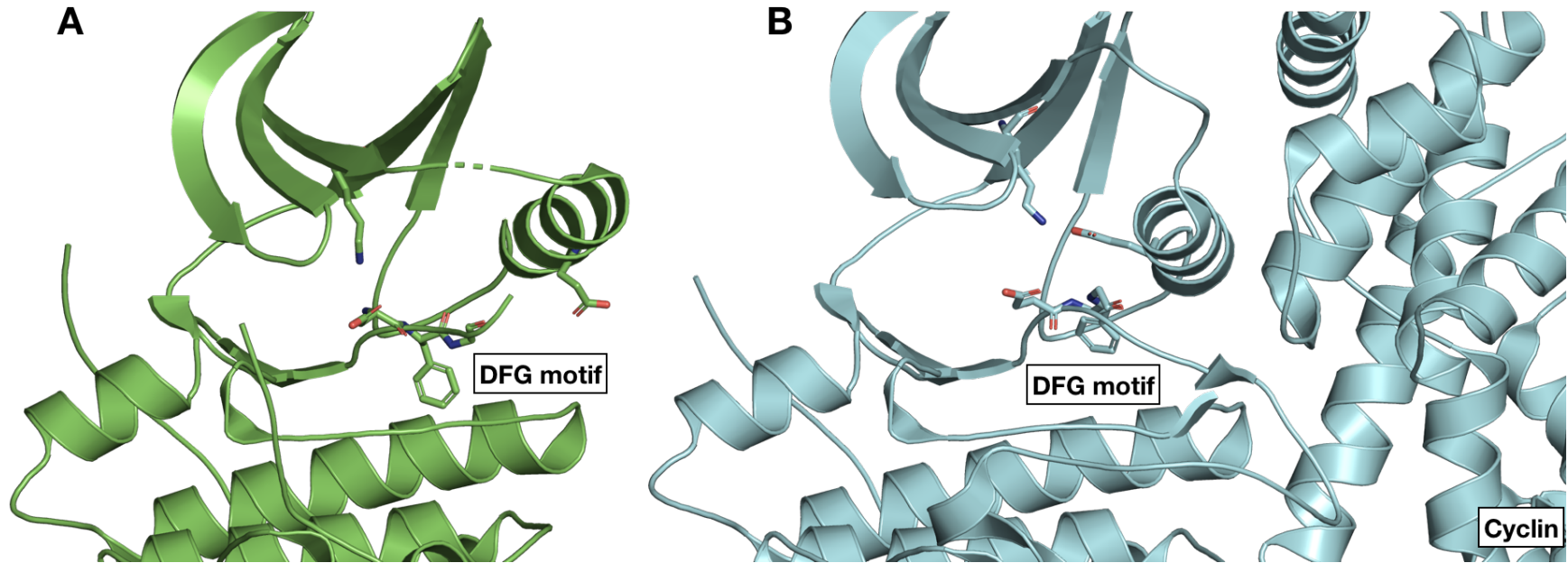


Figure A.1: **CDK2 adopts an inactive conformation in the crystal structure used for the CDK2/ERK2 calculations** (A) CDK2 (5K4J) adopts an inactive conformation in the absence of its cyclin. The DFG motif is in a DFG-in conformation, with the α C helix rotated outwards, breaking the salt bridge between K33 and E51 (Uniprot numbering) that is typically a marker of an active conformation. Notably, the Phe in the DFG motif does not completely form the hydrophobic spine due to the rotation of the α C helix [241] (B) The CDK2 structure used for the CDK2/CDK9 calculations (4BCK) contains cyclin A and adopts a DFG-in/ α C helix-in conformation that forms the salt bridge between K33 and E51. This is typically indicative of a fully active kinase [18, 104].

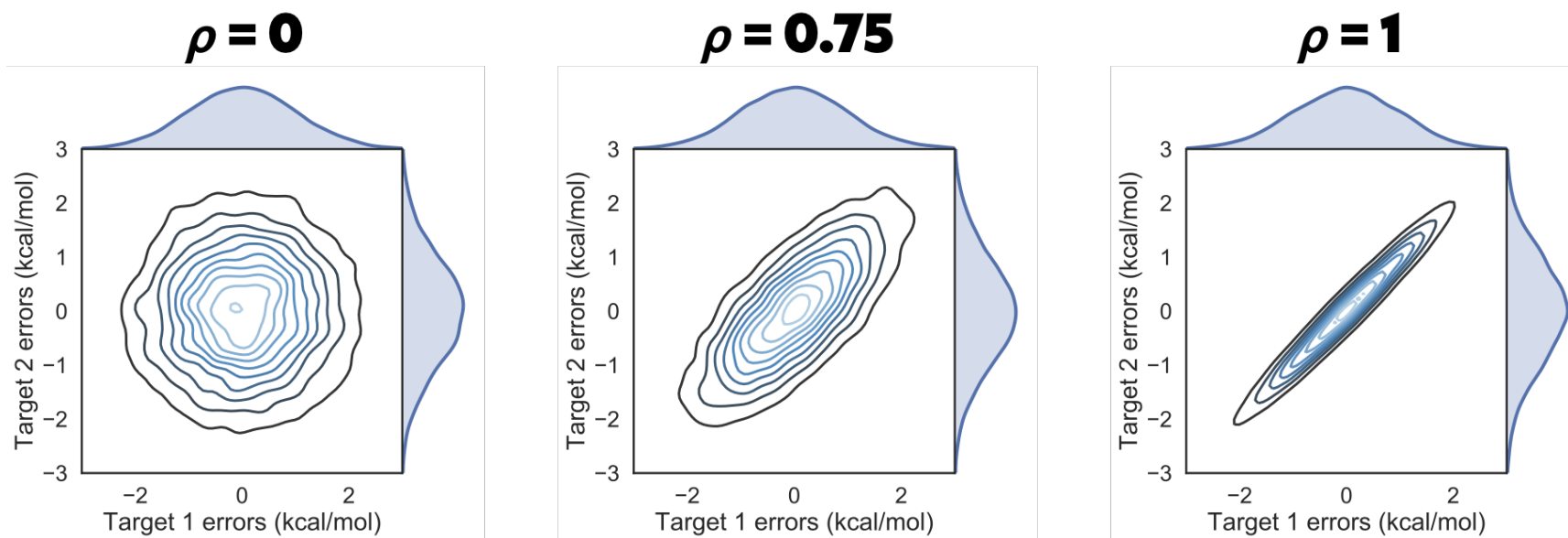


Figure A.2: **Correlation coefficient ρ controls the shape of the joint marginal distribution of errors** As ρ increases, the joint marginal distribution of errors become more diagonal. Each panel shows 10000 samples drawn from a multivariate normal distribution centered around 0 kcal/mol, where the per target error was set to 1 kcal/mol and ρ to the value indicated in bold over the plot.

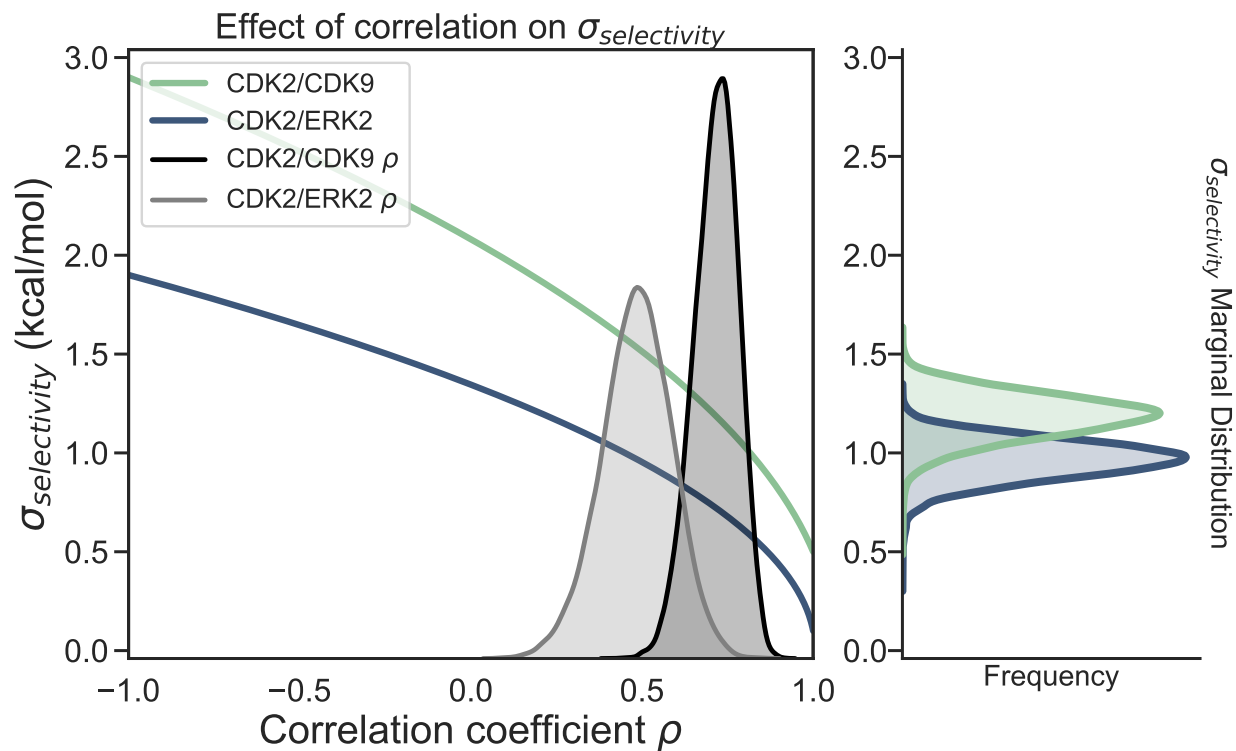


Figure A.3: **Correlation reduces the expected error for selectivity predictions** As correlation coefficient ρ increases, $\sigma_{selectivity}$ decreases. The intersection between CDK2/CDK9 $\sigma_{selectivity}$ (green curve) and ρ (black distribution) indicates the range of expected $\sigma_{selectivity}$ values. The intersection for CDK2/ERK $\sigma_{selectivity}$ (blue curve) and ρ (gray distribution) suggests the expected $\sigma_{selectivity}$ range for that set of calculations.

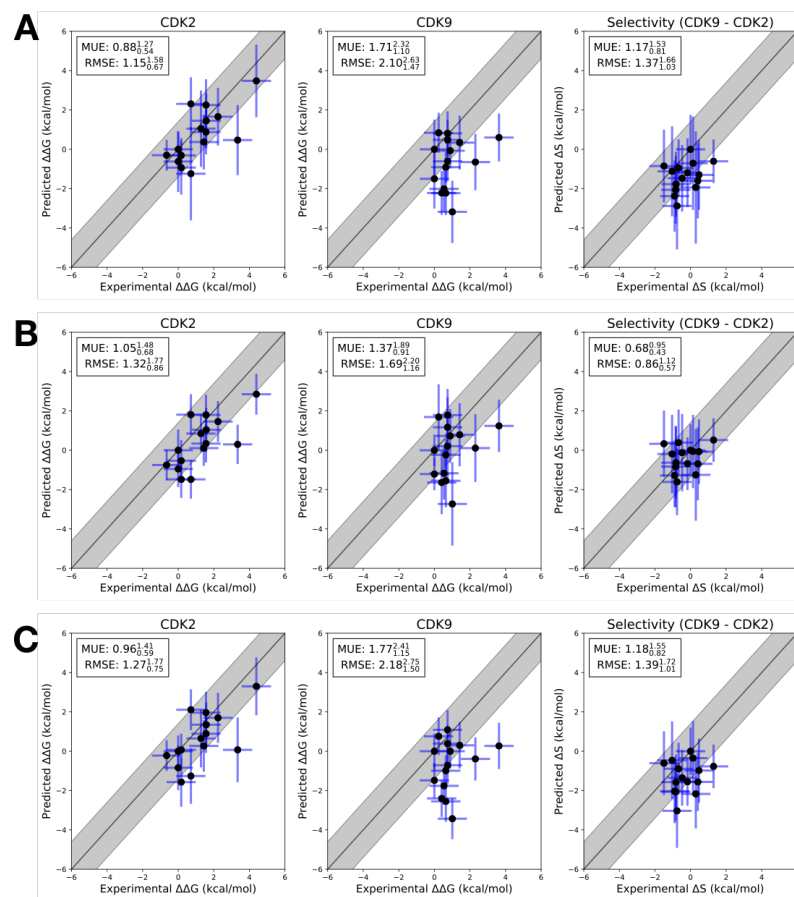


Figure A.4: Each replicate of the CDK2/CDK9 calculations yields a consistent RMSE and MUE

Three replicates of the CDK2/CDK9 calculations with different random seeds, but otherwise the same input structures, files, and parameters. The experimental values are shown on the X-axis and calculated values on the Y-axis. Each data point corresponds to a transformation between a ligand i to a set reference ligand j (Compound 1a) for a given target. All values are shown in units of kcal/mol. The horizontal error bars show the the 95% CI based on an assumed experimental uncertainty of 0.3 kcal/mol[56] expanded assuming no correlation between each measurement. For selectivity, the errors were propagated under the assumption that they were completely uncorrelated. 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. (A) Replicate 1 (B) Replicate 2 (C) Replicate 3

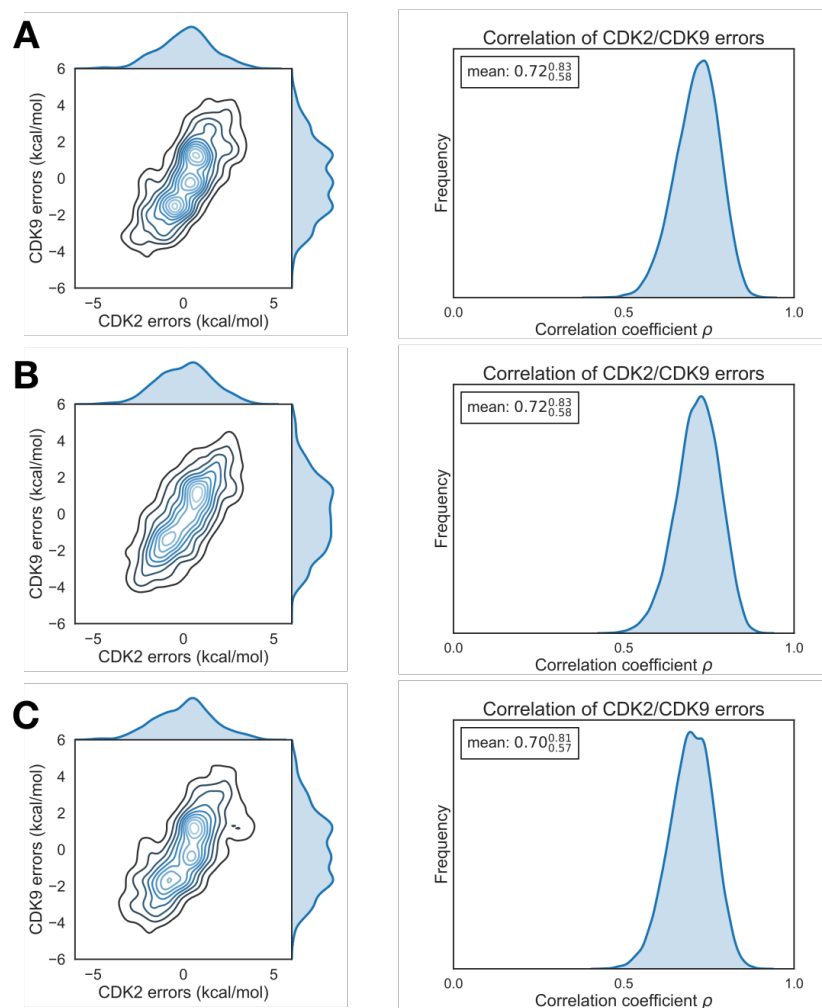


Figure A.5: Each replicate of the CDK2/CDK9 calculations yields consistent errors and correlation coefficient

(A) (left) The joint posterior distribution of the prediction errors for CDK2 (X-axis) and CDK9 (Y-axis) from the Bayesian graphical model for replicate 1. (right) The posterior marginal distribution of the correlation coefficient (ρ) is shown in gray for replicate 1. The inserted box shows the mean and 95% confidence interval for the correlation coefficient. (B) and (C) The same as above, but for replicates 2 and 3, respectively

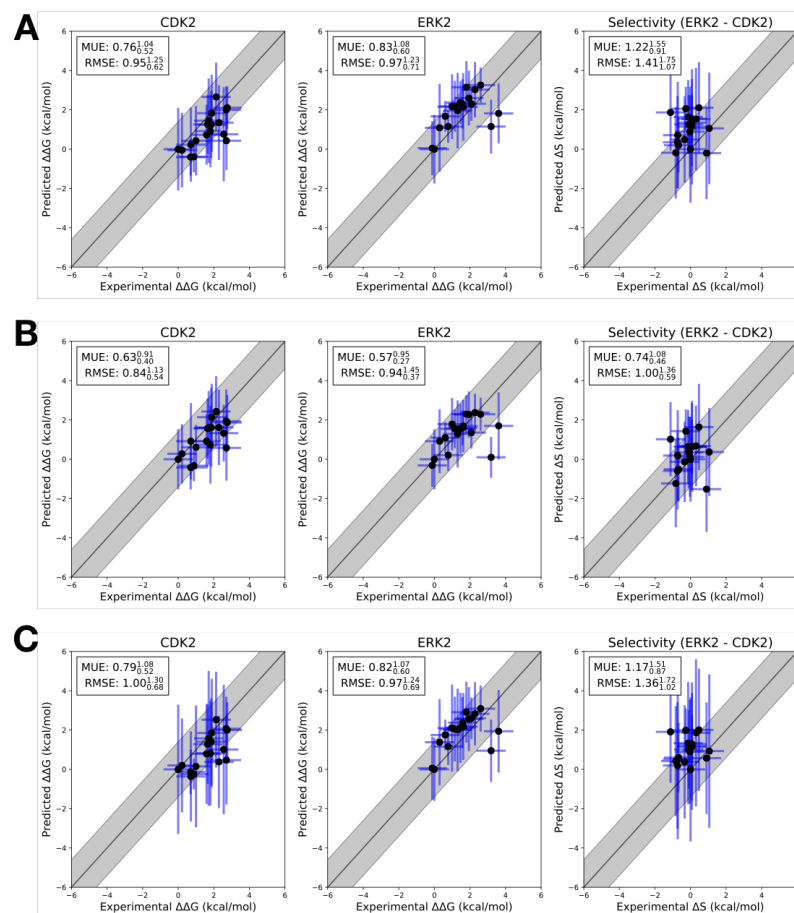


Figure A.6: Each replicate of the CDK2/ERK2 calculations yields a consistent RMSE and MUE

Three replicates of the CDK2/ERK2 calculations with different random seeds, but otherwise the same input structures, files, and parameters. The experimental values are shown on the X-axis and calculated values on the Y-axis. Each data point corresponds to a transformation between a ligand i to reference ligand j (Compound 6) for a given target. All values are shown in units of kcal/mol. The horizontal error bars show the the 95% CI based on an assumed experimental uncertainty of 0.3 kcal/mol[56] expanded assuming no correlation between each measurement. We show the 95% CI based on the estimated statistical as vertical blue error bars. For selectivity, the errors were propagated under the assumption that they were completely uncorrelated. 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. (A) Replicate 1 (B) Replicate 2 (C) Replicate 3

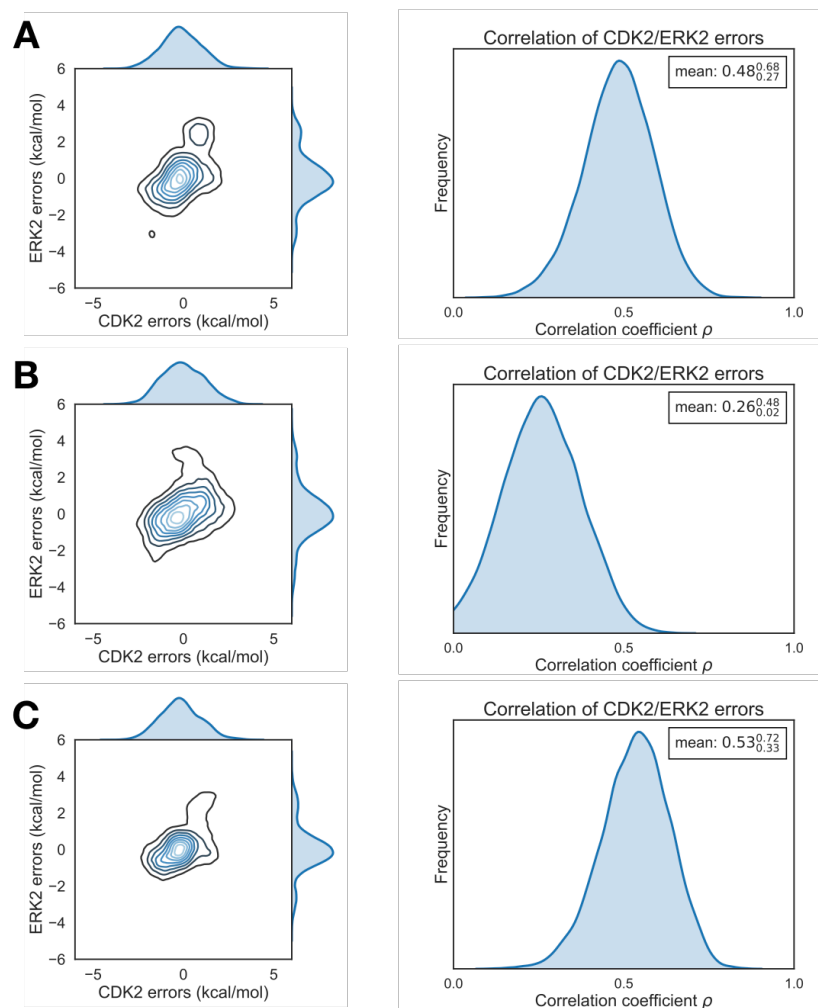


Figure A.7: Each replicate of the CDK2/ERK2 calculations yields consistent errors and correlation coefficient

(A) (left) The joint posterior distribution of the prediction errors for CDK2 (X-axis) and ERK2 (Y-axis) from the Bayesian graphical model for replicate 1. (right) The posterior marginal distribution of the correlation coefficient (ρ) is shown in gray for replicate 1. The inserted box shows the mean and 95% confidence interval for the correlation coefficient. (B) and (C) The same as above, but for replicates 2 and 3, respectively

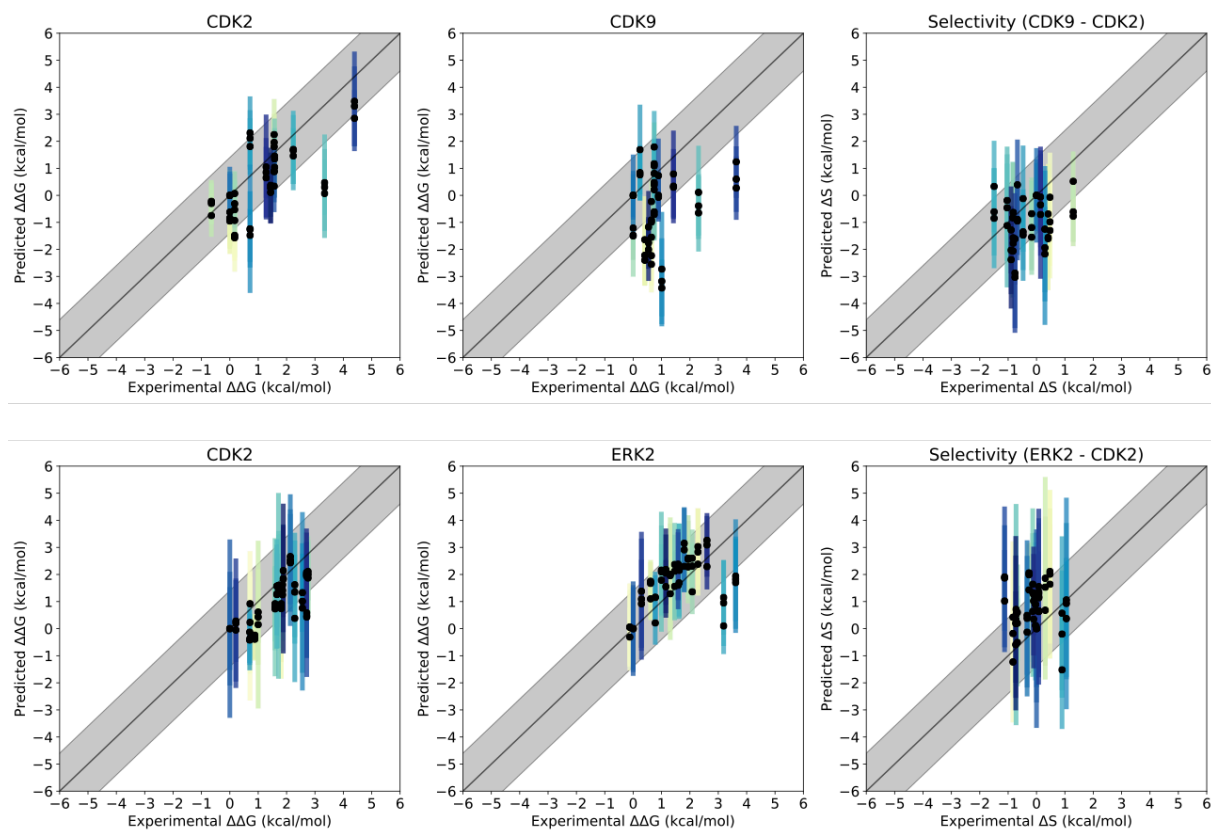


Figure A.8: The pooled replicates show good agreement in predictions for individual ligands

The experimental values are shown on the X-axis and calculated values on the Y-axis. Each data point corresponds to a transformation between a ligand i to reference ligand j (Compound 6 for CDK2/ERK2, Compound 1a for CDK2/CDK9) 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^[56] for each individual measurement, expanded assuming the error is uncorrelated. We show the 95% CI based on the estimated statistical as vertical blue error bars. For selectivity, the errors were propagated under the assumption that they were completely uncorrelated. 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. **(Top)** CDK2/CDK9 replicates **(Bottom)** CDK2/ERK2 replicates

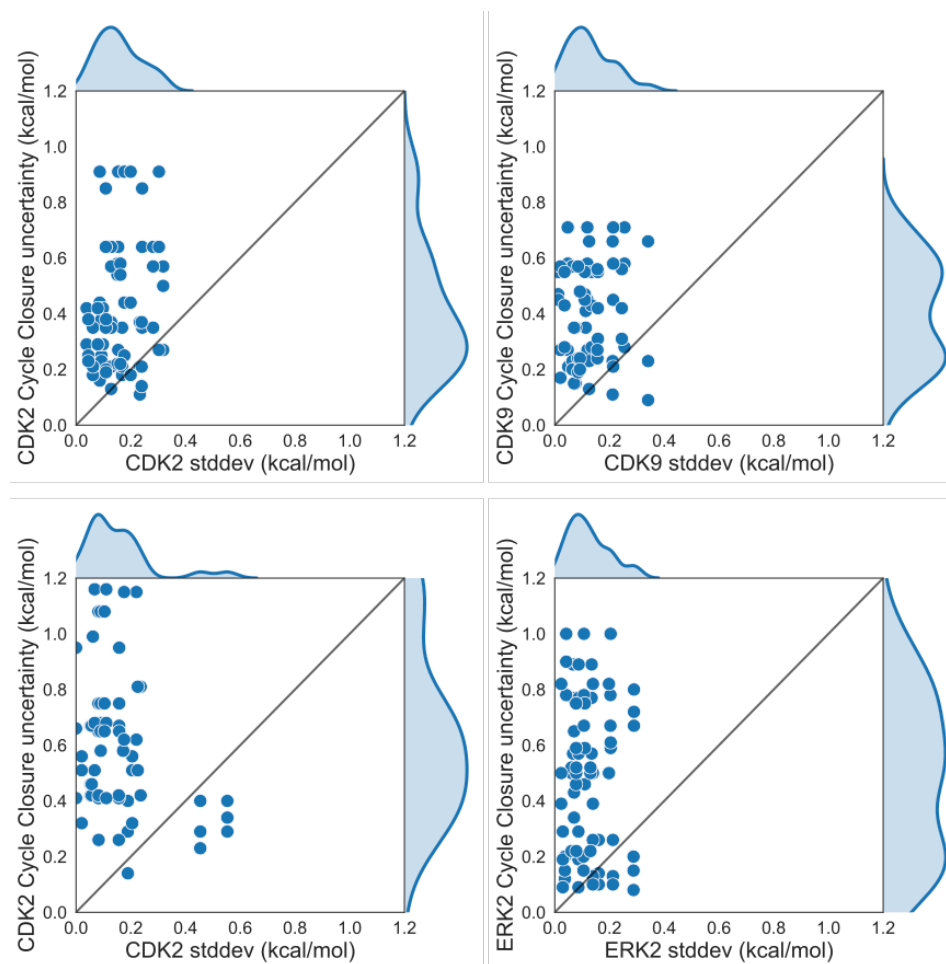


Figure A.9: **The standard deviation for each edge is smaller than the estimated cycle closure uncertainties**

The cycle closure uncertainty for each edge of the map is shown on the Y-axis and the standard deviation for that edge in all three replicate calculations is shown on the X-axis, in kcal/mol. Each point corresponds to an edge of the FEP map. The edges for all three replicates are pooled and shown together. **(Top)** CDK2/CDK9 calculations **(Bottom)** CDK2/ERK2 calculations.

Table A.1: **The CDK2 and CDK9 binding sites are more similar than the CDK2 and ERK2 binding sites**

Sequence based similarity of the binding sites based on multiple sequence alignments of the 85 residues annotated by the KLIFS database [100, 101]

Kinase	CDK2	CDK9	ERK2
CDK2	1.0	0.57	0.52
CDK9	0.57	1.0	0.52
ERK2	0.52	0.52	1.0

BIBLIOGRAPHY

- [1] **Siegel RL**, Miller KD, Jemal A. Cancer statistics, 2018. *CA: a cancer journal for clinicians*. 2018 Jan; 68(1):7–30.
- [2] **Cohen P**, Tcherpakov M. Will the ubiquitin system furnish as many drug targets as protein kinases? *Cell*. 2010 Nov; 143(5):686–693.
- [3] **Robert Roskoski Jr**. USFDA Approved Protein Kinase Inhibitors. . 2017; <http://www.brimr.org/PKI/PKIs.htm>, updated 3 May 2017.
- [4] **Paul, Steven M**, Mytelka, Daniel S, Dunwiddie, Christopher T, Persinger, Charles C, Munos, Bernard H, Lindborg, Stacy R, Schacht, Aaron L. How to improve R&D productivity: the pharmaceutical industry's grand challenge. *Nature reviews Drug discovery*. 2010 Mar; 9(3):203–214.
- [5] **Sanchez-Vega F**, Hechtman JF, Castel P, Ku GY, Tuvy Y, Won H, Fong CJ, Bouvier N, Nanjangud GJ, Soong J, Vakiani E, Schattner M, Kelsen DP, Lefkowitz RA, Brown K, Lacouture ME, Capanu M, Mattar M, Qeriqi B, Cecchi F, et al. EGFR and MET Amplifications Determine Response to HER2 Inhibition in ERBB2-Amplified Esophagogastric Cancer. *Cancer discovery*. 2018 Nov; .
- [6] **Bose R**, Kavuri SM, Searleman AC, Shen W, Shen D, Koboldt DC, Monsey J, Goel N, Aronson AB, Li S, Ma CX, Ding L, Mardis ER, Ellis MJ. Activating HER2 mutations in HER2 gene amplification negative breast cancer. *Cancer discovery*. 2013 Feb; 3(2):224–237.
- [7] **Prahallad A**, Sun C, Huang S, Di Nicolantonio F, Salazar R, Zecchin D, Beijersbergen RL, Bardelli A, Bernards R. Unresponsiveness of colon cancer to BRAF(V600E) inhibition through feedback activation of EGFR. *Nature*. 2012 Mar; 483(7387):100–103.
- [8] **Engelman JA**, Zejnullahu K, Mitsudomi T, Song Y, Hyland C, Park JO, Lindeman N, Gale CM, Zhao X, Christensen J, Kosaka T, Holmes AJ, Rogers AM, Cappuzzo F, Mok T, Lee C, Johnson BE, Cantley LC, Jänne PA. MET amplification leads to gefitinib resistance in lung cancer by activating ERBB3 signaling. *Science*. 2007 May; 316(5827):1039–1043.
- [9] **Pao W**, Miller VA, Politi KA, Riely GJ, Somwar R, Zakowski MG, Maureen F and Kris, Varmus H. Acquired resistance of lung adenocarcinomas to gefitinib or erlotinib is associated with a second mutation in the EGFR kinase domain. *PLoS medicine*. 2005 Mar; 2(3):e73.
- [10] **Drilon A**, Nagasubramanian R, Blake JF, Ku N, Tuch BB, Ebata K, Smith S, Lauriault V, Kolakowski GR, Brandhuber BJ, Larsen PD, Bouhana KS, Winski SL, Hamor R, Wu WI, Parker A, Morales TH, Sullivan FX, DeWolf WE, Wollenberg LA, et al. A Next-Generation TRK Kinase Inhibitor Overcomes Acquired Resistance to Prior TRK Kinase Inhibition in Patients with TRK Fusion-Positive Solid Tumors. *Cancer discovery*. 2017 Sep; 7(9):963–972.
- [11] **Knight ZA**, Lin H, Shokat KM. Targeting the Cancer Kinome through Polypharmacology. *Nat Rev Cancer*. 2010; 10(2):130.
- [12] **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.

- [13] **Rudmann DG**. On-target and off-target-based toxicologic effects. *Toxicol Pathol*. 2013 Feb; 41(2):310–314.
- [14] **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.
- [15] **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.
- [16] **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.
- [17] **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.
- [18] **Huse M**, Kuriyan J. The conformational plasticity of protein kinases. *Cell*. 2002 Jan; 109(3):275–282.
- [19] **Harrison SC**. Variation on an Src-like theme. *Cell*. 2003 Mar; 112(6):737–740.
- [20] **Nagar B**, Hantschel O, Young MA, Scheffzek K, Veach D, Bornmann W, Clarkson B, Superti-Furga G, Kuriyan J. Structural basis for the autoinhibition of c-Abl tyrosine kinase. *Cell*. 2003 Mar; 112(6):859–871.
- [21] **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.
- [22] **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.
- [23] **Hopkins AL**, Mason JS, Overington JP. Can we rationally design promiscuous drugs? *Curr Opin Struct Biol*. 2006 Feb; 16(1):127–136.
- [24] **Hopkins AL**. Network pharmacology: the next paradigm in drug discovery. *Nat Chem Biol*. 2008 Nov; 4(11):682–690.
- [25] **Zhang J**, Yang PL, Gray NS. Targeting cancer with small molecule kinase inhibitors. *Nat Rev Cancer*. 2009 Jan; 9(1):28–39.
- [26] **Huggins DJ**, Sherman W, Tidor B. Rational approaches to improving selectivity in drug design. *J Med Chem*. 2012 Feb; 55(4):1424–1444.
- [27] **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.
- [28] **Tricker EM**, Xu C, Uddin S, Capelletti M, Ercan D, Ogino A, Pratilas CA, Rosen N, Gray NS, Wong KK, Janne PA. Combined EGFR/MEK Inhibition Prevents the Emergence of Resistance in EGFR-Mutant Lung Cancer. *Cancer Discov*. 2015 Sep; 5(9):960–971.

- [29] **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.
- [30] **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.
- [31] **Littlefield P**, Moasser MM, Jura N. RETRACTED: An ATP-Competitive Inhibitor Modulates the Allosteric Function of the HER3 Pseudokinase. *Chem Biol*. 2014 Apr; 21(4):453–458. doi: [10.1016/j.chembiol.2014.02.011](https://doi.org/10.1016/j.chembiol.2014.02.011).
- [32] **Yun, Cai-Hong**, Boggon, Titus J, Li, Yiqun, Woo, Michele S, Greulich, Heidi, Meyerson, Matthew, Eck, Michael J. Structures of lung cancer-derived EGFR mutants and inhibitor complexes: mechanism of activation and insights into differential inhibitor sensitivity. *Cancer Cell*. 2007 Mar; 11(3):217–227.
- [33] **Gajiwala KS**, Feng J, Ferre R, Ryan K, Brodsky O, Weinrich S, Kath JC, Stewart A. Insights into the Aberrant Activity of Mutant EGFR Kinase Domain and Drug Recognition. *Structure*. 2013; 21(2):209–219.
- [34] **Jia Y**, Yun CH, Park E, Ercan D, Manuia M, Juarez J, Xu C, Rhee K, Chen T, Zhang H, Palakurthi S, Jang J, Lelais G, DiDonato M, Bursulaya B, Michellys PY, Epple R, Marsilje TH, McNeill M, Lu W, et al. Overcoming EGFR(T790M) and EGFR(C797S) resistance with mutant-selective allosteric inhibitors. *Nature*. 2016 Jun; 534(7605):129–132.
- [35] **Politi K**, Ayeni D, Lynch T. The Next Wave of EGFR Tyrosine Kinase Inhibitors Enter the Clinic. *Cancer Cell*. 2015 Jun; 27(6):751–753.
- [36] **Song Z**, Wang M, Zhang A. Alectinib: a novel second generation anaplastic lymphoma kinase (ALK) inhibitor for overcoming clinically-acquired resistance. *Acta Pharmaceutica Sinica B*. 2015 Jan; 5(1):34–37.
- [37] **Varghese AM**, Berger MF. Advancing clinical oncology through genome biology and technology. *Genome biology*. 2014 Aug; 15(8):427.
- [38] **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. *Nature medicine*. 2017 May; 159:676.
- [39] **Garraway LA**. Genomics-driven oncology: framework for an emerging paradigm. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology*. 2013 May; 31(15):1806–1814.
- [40] **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.
- [41] **Durrant JD**, McCammon JA. Molecular dynamics simulations and drug discovery. *BMC biology*. 2011; 9(1):71.

- [42] Lin JH, Perryman AL, Schames JR, McCammon JA. The relaxed complex method: Accommodating receptor flexibility for drug design with an improved scoring scheme. *Biopolymers*. 2003 Jan; 68(1):47–62.
- [43] Kitchen DB, Decornez H, Furr JR, Bajorath J. Docking and scoring in virtual screening for drug discovery: methods and applications. *Nature reviews Drug discovery*. 2004 Nov; 3(11):935–949.
- [44] Amaro RE, Baron R, McCammon JA. An improved relaxed complex scheme for receptor flexibility in computer-aided drug design. *Journal of computer-aided molecular design*. 2008 Jan; 22(9):693–705.
- [45] Hazuda DJ, Anthony NJ, Gomez RP, Jolly SM, Wai JS, Zhuang L, Fisher TE, Embrey M, Guare JP, Egbertson MS, Vacca JP, Huff JR, Felock PJ, Witmer MV, Stillmock KA, Danovich R, Grobler J, Miller MD, Espeseth AS, Jin L, et al. A naphthyridine carboxamide provides evidence for discordant resistance between mechanistically identical inhibitors of HIV-1 integrase. *Proceedings of the National Academy of Sciences*. 2004 Aug; 101(31):11233–11238.
- [46] Norman RA, Toader D, Ferguson AD. Structural approaches to obtain kinase selectivity. *Trends in pharmacological sciences*. 2012 May; 33(5):273–278.
- [47] Shoichet BK. Virtual screening of chemical libraries. *Nature*. 2004 Dec; 432(7019):862–865.
- [48] Warren GL, Andrews CW, Capelli AM, Clarke B, LaLonde J, Lambert MH, Lindvall M, Nevins N, Semus SF, Senger S, Tedesco G, Wall ID, Woolven JM, Peishoff CE, Head MS. A Critical Assessment of Docking Programs and Scoring Functions. *Journal of medicinal chemistry*. 2006 Oct; 49(20):5912–5931.
- [49] 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.
- [50] Aldeghi M, Heifetz A, Bodkin MJ, Knapp S, Biggin PC. Accurate calculation of the absolute free energy of binding for drug molecules. *Chemical science*. 2016 Jan; 7(1):207–218.
- [51] Mikulskis P, Genheden S, Ryde U. A large-scale test of free-energy simulation estimates of protein-ligand binding affinities. *Journal of chemical information and modeling*. 2014 Oct; 54(10):2794–2806.
- [52] 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).
- [53] 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).
- [54] 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).

- [55] **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.
- [56] **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>.
- [57] **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.
- [58] **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.
- [59] **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.
- [60] **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.
- [61] **Jorgensen WL**. Computer-aided discovery of anti-HIV agents. *Bioorganic & medicinal chemistry*. 2016 Oct; 24(20):4768–4778.
- [62] **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.
- [63] **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.
- [64] **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.
- [65] **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.
- [66] **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.
- [67] **Park J**, McDonald JJ, Petter RC, Houk KN. Molecular Dynamics Analysis of Binding of Kinase Inhibitors to WT EGFR and the T790M Mutant. *Journal of Chemical Theory and Computation*. 2016 Apr; 12(4):2066–2078.
- [68] **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.

- [69] **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.
- [70] **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.
- [71] **Seeliger D**, de Groot BL. Protein thermostability calculations using alchemical free energy simulations. *Biophysical journal*. 2010 May; 98(10):2309–2316.
- [72] **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 Apr; 429(7):948–963.
- [73] **Ford MC**, Babaoglu K. Examining the Feasibility of Using Free Energy Perturbation (FEP+) in Predicting Protein Stability. *Journal of chemical information and modeling*. 2017 Jun; 57(6):1276–1285.
- [74] **Mondal J**, Tiwary P, Berne BJ. How a Kinase Inhibitor Withstands Gatekeeper Residue Mutations. *Journal of the American Chemical Society*. 2016 Apr; 138(13):4608–4615.
- [75] **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.
- [76] **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.
- [77] **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.
- [78] **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.
- [79] **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.
- [80] **Pao W**, Miller V, Zakowski M, Doherty J, Politi K, Sarkaria I, Singh B, Heelan R, Rusch V, Fulton L, Mardis E, Kupfer D, Wilson R, Kris M, Varmus H. EGF receptor gene mutations are common in lung cancers from “never smokers” and are associated with sensitivity of tumors to gefitinib and erlotinib. *Proceedings of the National Academy of Sciences*. 2004 Sep; 101(36):13306–13311.

- [81] 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.
- [82] Juchum M, Günther M, Laufer SA. Fighting Cancer Drug Resistance: Opportunities and Challenges for Mutation-Specific EGFR Inhibitors. *Drug Resist Updat*. 2015 May; 20:12–28. doi: [10.1016/j.drug.2015.05.002](https://doi.org/10.1016/j.drug.2015.05.002).
- [83] Din OS, Woll PJ. Treatment of gastrointestinal stromal tumor: focus on imatinib mesylate. *Ther Clin Risk Manag*. 2008 Feb; 4(1):149–162.
- [84] 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](https://doi.org/10.1038/nrd.2016.230).
- [85] 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.
- [86] Wu P, Nielsen TE, Clausen MH. FDA-approved small-molecule kinase inhibitors. *Trends Pharmacol Sci*. 2015 Jul; 36(7):422–439.
- [87] 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.
- [88] 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.
- [89] 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.
- [90] Klaeger S, Heinzlmeir 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).
- [91] Sun C, Hobor S, Bertotti A, Zecchin D, Huang S, Galimi F, Cottino F, Prahallad A, Grenrum W, Tzani A, Schlicker A, Wessels LFA, Smit EF, Thunnissen E, Halonen P, Liefstink C, Beijersbergen RL, Di Nicolantonio F, Bardelli A, Trusolino L, et al. Intrinsic resistance to MEK inhibition in KRAS mutant lung and colon cancer through transcriptional induction of ERBB3. *CellReports*. 2014 Apr; 7(1):86–93.
- [92] Machado 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.
- [93] 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.

- [94] **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.
- [95] **Bosc N**, Meyer C, Bonnet P. The use of novel selectivity metrics in kinase research. *BMC bioinformatics.* 2017 Jan; 18(1):17.
- [96] **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.
- [97] **Shirts MR**, Mobley DL, Brown SP. Free energy calculations in structure-based drug design. In: *Structure Based Drug Design* Cambridge University Press; 2009.
- [98] **Abel R**, Manas ES, Friesner RA, Farid RS, Wang L. Modeling the value of predictive affinity scoring in preclinical drug discovery. *Current opinion in structural biology.* 2018 Oct; 52:103–110.
- [99] **Abel R**, Wang L, Mobley DL, Friesner RA. A Critical Review of Validation, Blind Testing, and Real- World Use of Alchemical Protein-Ligand Binding Free Energy Calculations. *Current topics in medicinal chemistry.* 2017; 17(23):2577–2585.
- [100] **Kooistra AJ**, Kanev GK, van Linden OPJ, Leurs R, de Esch IJP, de Graaf C. KLIFS: a structural kinase-ligand interaction database. *Nucleic acids research.* 2016 Jan; 44(D1):D365–D371.
- [101] **van Linden OPJ**, Kooistra AJ, Leurs R, de Esch IJP, de Graaf C. KLIFS: a knowledge-based structural database to navigate kinase-ligand interaction space. *Journal of medicinal chemistry.* 2014 Jan; 57(2):249–277.
- [102] **Hole AJ**, Baumli S, Shao H, Shi S, Huang S, Pepper C, Fischer PM, Wang S, Endicott JA, Noble ME. Comparative Structural and Functional Studies of 4-(Thiazol-5-yl)-2-(phenylamino)pyrimidine-5-carbonitrile CDK9 Inhibitors Suggest the Basis for Isotype Selectivity. *J Med Chem.* 2013 Feb; 56(3):660–670.
- [103] **Kalliokoski T**, Kramer C, Vulpetti A, Gedeck P. Comparability of Mixed IC50 Data—a Statistical Analysis. *PloS One.* 2013; 8(4):e61007.
- [104] **Hari SB**, Merritt EA, Maly DJ. Sequence determinants of a specific inactive protein kinase conformation. *Chemistry & biology.* 2013 Jun; 20(6):806–815.
- [105] **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.
- [106] **Graczyk PP**. Gini coefficient: a new way to express selectivity of kinase inhibitors against a family of kinases. *Journal of medicinal chemistry.* 2007 Nov; 50(23):5773–5779.
- [107] **Duong-Ly KC**, Devarajan K, Liang S, Horiuchi KY, Wang Y, Ma H, Peterson JR. Kinase Inhibitor Profiling Reveals Unexpected Opportunities to Inhibit Disease-Associated Mutant Kinases. *CellReports.* 2016 Feb; 14(4):772–781.

- [108] **Uitdehaag JCM**, Zaman GJR. A theoretical entropy score as a single value to express inhibitor selectivity. *BMC bioinformatics*. 2011 Apr; 12:94.
- [109] **Michel J**, Verdonk ML, Essex JW. Protein-Ligand Binding Affinity Predictions by Implicit Solvent Simulations: A Tool for Lead Optimization? *J Med Chem*. 2006 Dec; 49(25):7427–7439. doi: 10.1021/jm061021s.
- [110] **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.
- [111] **Sastry GM**, Adzhigirey M, Day T, Annabhimoju R, Sherman W. Protein and ligand preparation: parameters, protocols, and influence on virtual screening enrichments. *J Comput Aided Mol Des*. 2013 Mar; 27(3):221–234.
- [112] **Salvatier J**, Wiecki TV, Fonnesbeck C. Probabilistic programming in Python using PyMC3. *PeerJ Computer Science*. 2016; 2:e55.
- [113] **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, Blecher 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>.
- [114] **Knight ZA**, Shokat KM. Features of selective kinase inhibitors. *Chem Biol*. 2005; 12:621–637.
- [115] **Chuang JC**, Salahudeen AA, Wakelee HA. Rociletinib, a third generation EGFR tyrosine kinase inhibitor: current data and future directions. *Expert opinion on pharmacotherapy*. 2016; 17(7):989–993.
- [116] **Tan CS**, Kumarakulasinghe NB, Huang YQ, Ang YLE, Choo JRE, Goh BC, Soo RA. Third generation EGFR TKIs: current data and future directions. *Molecular cancer*. 2018 Feb; 17(1):29.
- [117] **Gainor JF**, Dardaei L, Yoda S, Friboulet L, Leshchiner I, Katayama R, Dagogo-Jack I, Gadgeel S, Schultz K, Singh M, Chin E, Parks M, Lee D, DiCecca RH, Lockerman E, Huynh T, Logan J, Ritterhouse LL, Le LP, Muniappan A, et al. Molecular Mechanisms of Resistance to First- and Second-Generation ALK Inhibitors in ALK-Rearranged Lung Cancer. *Cancer discovery*. 2016 Oct; 6(10):1118–1133.
- [118] **Aldeghi M**, Gapsys V, de Groot BL. Accurate Estimation of Ligand Binding Affinity Changes upon Protein Mutation. *ACS Central Science*. 2018 Dec; 4(12):1708–1718.
- [119] **Aldeghi M**, de Groot BL, Gapsys V. Accurate Calculation of Free Energy Changes upon Amino Acid Mutation. In: *Computational Methods in Protein Evolution* New York, NY: Springer New York; 2018.p. 19–47.
- [120] **Bhati AP**, Wan S, Coveney PV. Ensemble-based replica exchange alchemical free energy methods: the effect of protein mutations on inhibitor binding. *Journal of Chemical Theory and Computation*. 2018 Dec; p. acs.jctc.8b01118.
- [121] **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 (STI571) in Chronic Phase and Blast Crisis Chronic Myeloid Leukemia. *Cancer Cell*. 2002 Aug; 2(2):117–125.

- [122] 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>.
- [123] Huang L, Fu L. Mechanisms of Resistance to EGFR Tyrosine Kinase Inhibitors. *Acta Pharm Sin B*. 2015; 5(5):390–401.
- [124] 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).
- [125] 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).
- [126] 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).
- [127] Rani S, Corcoran C, Shiels L, Germano S, Breslin S, Madden S, McDermott MS, Browne BC, O'Connell N, Donovan N, Crown J, Gogarty M, Byrne AT, Driscoll 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).
- [128] 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).
- [129] 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).
- [130] 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).
- [131] 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).
- [132] 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](https://doi.org/10.1038/s41698-017-0007-0).
- [133] 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.

- [134] **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.
- [135] **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.
- [136] **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.
- [137] **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.
- [138] **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.
- [139] **Melnikov A**, Rogov P, Wang L, Gnirke 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.
- [140] **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.
- [141] **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.
- [142] **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.
- [143] **Cappel D**, Hall ML, Lenseink 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.
- [144] **Clark AJ**, Gindin T, Zhang B, Wang L, Abel R, Murrett 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.
- [145] **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.
- [146] **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.

- [147] 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](https://doi.org/10.1021/jacs.6b09511).
- [148] 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](https://doi.org/10.1021/jacs.6b01232).
- [149] 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).
- [150] 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).
- [151] 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).
- [152] 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.
- [153] Soverini S, Colarossi S, Gnani A, Rosti G, Castagnetti F, Poerio A, Iacobucci I, Amabile M, Abruzzese E, Orlandi E, Radaelli F, Ciccone F, Tiribelli 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).
- [154] 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.
- [155] Gruber FX, Lundán T, Goll R, Silye A, Mikkola I, Rekvig OP, Knuutila S, Remes K, Gedde-Dahl T, Porkka K, Hjorth-Hansen H. BCR-ABL Isoforms Associated with Intrinsic or Acquired Resistance to Imatinib: More Heterogeneous than Just ABL Kinase Domain Point Mutations? *Med Oncol.* 2012 Mar; 29(1):219–226. doi: [10.1007/s12032-010-9781-z](https://doi.org/10.1007/s12032-010-9781-z).
- [156] Redaelli S, Mologni L, Rostagno R, Piazza R, Magistroni V, Ceccon M, Viltadi M, Flynn D, Gambacorti-Passerini C. Three novel patient-derived BCR/ABL mutants show different sensitivity to second and third generation tyrosine kinase inhibitors. *American journal of hematology.* 2012 Nov; 87(11):E125–8.
- [157] Cortes JE, Kantarjian H, Shah NP, Bixby D, Mauro MJ, Flinn I, O’Hare T, Hu S, Narasimhan NI, Rivera VM, Clackson T, Turner CD, Haluska FG, Druker BJ, Deininger MWN, Talpaz M. Ponatinib in Refractory Philadelphia

- Chromosome-Positive Leukemias. *N Engl J Med.* 2012 Nov; 367(22):2075–2088. doi: 10.1056/NEJMoa1205127.
- [158] **Branford S.** High Frequency of Point Mutations Clustered within the Adenosine Triphosphate-Binding Region of BCR/ABL in Patients with Chronic Myeloid Leukemia or Ph-Positive Acute Lymphoblastic Leukemia Who Develop Imatinib (STI571) Resistance. *Blood.* 2002 May; 99(9):3472–3475. doi: [10.1182/blood.V99.9.3472](https://doi.org/10.1182/blood.V99.9.3472).
- [159] **Press RD, Willis SG, Laudadio J, Mauro MJ, Deininger MWN.** Determining the rise in BCR-ABL RNA that optimally predicts a kinase domain mutation in patients with chronic myeloid leukemia on imatinib. *Blood.* 2009 Sep; 114(13):2598–2605.
- [160] **O’Hare T.** Combined Abl Inhibitor Therapy for Minimizing Drug Resistance in Chronic Myeloid Leukemia: Src/Abl Inhibitors Are Compatible with Imatinib. *Clin Cancer Res.* 2005 Oct; 11(19):6987–6993. doi: [10.1158/1078-0432.CCR-05-0622](https://doi.org/10.1158/1078-0432.CCR-05-0622).
- [161] **Shan Y, Seeliger MA, Eastwood MP, Frank F, Xu H, Jensen MØ, Dror RO, Kuriyan J, Shaw DE.** A Conserved Protonation-Dependent Switch Controls Drug Binding in the Abl Kinase. *Proc Natl Acad Sci.* 2009; 106(1):139–144.
- [162] **Demerdash O, Yap EH, Head-Gordon T.** Advanced Potential Energy Surfaces for Condensed Phase Simulation. *Annu Rev Phys Chem.* 2014 Apr; 65(1):149–174. doi: 10.1146/annurev-physchem-040412-110040.
- [163] **Jiao D, Golubkov PA, Darden TA, Ren P.** Calculation of Protein–ligand Binding Free Energy by Using a Polarizable Potential. *Proc Natl Acad Sci.* 2008; 105(17):6290–6295.
- [164] **Shirts MR, Mobley DL, Chodera JD, Pande VS.** Accurate and Efficient Corrections for Missing Dispersion Interactions in Molecular Simulations. *J Phys Chem B.* 2007 Nov; 111(45):13052–13063. doi: 10.1021/jp0735987.
- [165] **Essmann U, Perera L, Berkowitz ML, Darden T, Lee H, Pedersen LG.** A Smooth Particle Mesh Ewald Method. *J Chem Phys.* 1995 Nov; 103(19):8577–8593. doi: [10.1063/1.470117](https://doi.org/10.1063/1.470117).
- [166] **Wennberg CL, Murtola T, Hess B, Lindahl E.** Lennard-Jones Lattice Summation in Bilayer Simulations Has Critical Effects on Surface Tension and Lipid Properties. *J Chem Theory Comput.* 2013 Aug; 9(8):3527–3537. doi: 10.1021/ct400140n.
- [167] **Onufriev AV, Alexov E.** Protonation and pK Changes in Protein–ligand Binding. *Q Rev Biophys.* 2013 May; 46(02):181–209. doi: 10.1017/S0033583513000024.
- [168] **Martin YC.** Let’s Not Forget Tautomers. *J Comput Aided Mol Des.* 2009 Oct; 23(10):693–704. doi: 10.1007/s10822-009-9303-2.
- [169] **Jensen J.** Calculating pH and Salt Dependence of Protein-Protein Binding. *Curr Pharm Biotechnol.* 2008 Apr; 9(2):96–102. doi: 10.2174/138920108783955146.
- [170] **Nagar B, Hantschel O, Young MA, Scheffzek K, Veach D, Bornmann W, Clarkson B, Superti-Furga G, Kuriyan J.** Structural Basis for the Autoinhibition of C-Abl Tyrosine Kinase. *Cell.* 2003; 112(6):859–871.
- [171] **Lorenz S, Deng P, Hantschel O, Superti-Furga G, Kuriyan J.** Crystal Structure of an SH2-kinase Construct of C-Abl and Effect of the SH2 Domain on Kinase Activity. *Biochem J.* 2015 Jun; 468(2):283–291. doi: 10.1042/BJ20141492.

- [172] O'Hare T, Shakespeare WC, Zhu X, Eide CA, Rivera VM, Wang F, Adrian LT, Zhou T, Huang WS, Xu Q, Metcalf CA, Tyner JW, Loriaux MM, Corbin AS, Wardwell S, Ning Y, Keats JA, Wang Y, Sundaramoorthi R, Thomas M, et al. AP24534, a Pan-BCR-ABL Inhibitor for Chronic Myeloid Leukemia, Potently Inhibits the T315I Mutant and Overcomes Mutation-Based Resistance. *Cancer Cell*. 2009 Nov; 16(5):401–412. doi: [10.1016/j.ccr.2009.09.028](https://doi.org/10.1016/j.ccr.2009.09.028).
- [173] Levinson NM, Boxer SG. Structural and Spectroscopic Analysis of the Kinase Inhibitor Bosutinib and an Isomer of Bosutinib Binding to the Abl Tyrosine Kinase Domain. *PLoS ONE*. 2012 Apr; 7(4):e29828. doi: [10.1371/journal.pone.0029828](https://doi.org/10.1371/journal.pone.0029828).
- [174] Weisberg E, Manley PW, Breitenstein W, Brügger J, Cowan-Jacob SW, Ray A, Huntly B, Fabbro D, Fendrich G, Hall-Meyers E, Kung AL, Mestan J, Daley GQ, Callahan L, Catley L, Cavazza C, Mohammed A, Neuberg D, Wright RD, Gilliland DG, et al. Characterization of AMN107, a Selective Inhibitor of Native and Mutant Bcr-Abl. *Cancer Cell*. 2005 Feb; 7(2):129–141. doi: [10.1016/j.ccr.2005.01.007](https://doi.org/10.1016/j.ccr.2005.01.007).
- [175] Zhou T, Commodore L, Huang WS, Wang Y, Thomas M, Keats J, Xu Q, Rivera VM, Shakespeare WC, Clackson T, Dalgarno DC, Zhu X. Structural Mechanism of the Pan-BCR-ABL Inhibitor Ponatinib (AP24534): Lessons for Overcoming Kinase Inhibitor Resistance: Structural Mechanism of Ponatinib. *Chem Biol Drug Des*. 2011 Jan; 77(1):1–11. doi: [10.1111/j.1747-0285.2010.01054.x](https://doi.org/10.1111/j.1747-0285.2010.01054.x).
- [176] Li H, Robertson AD, Jensen JH. Very Fast Empirical Prediction and Rationalization of Protein pKa Values. *Proteins Struct Funct Bioinforma*. 2005 Oct; 61(4):704–721. doi: [10.1002/prot.20660](https://doi.org/10.1002/prot.20660).
- [177] 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. *J Med Chem*. 2004 Mar; 47(7):1739–1749. doi: [10.1021/jm0306430](https://doi.org/10.1021/jm0306430).
- [178] Park JH, Liu Y, Lemmon MA, Radhakrishnan R. Erlotinib Binds Both Inactive and Active Conformations of the EGFR Tyrosine Kinase Domain. *Biochem J*. 2012 Dec; 448(3):417–423. doi: [10.1042/BJ20121513](https://doi.org/10.1042/BJ20121513).
- [179] Yosaatmadja Y, Squire CJ. 1.85 Angstrom Structure of EGFR Kinase Domain with Gefitinib. . 2014 Nov; doi: [10.2210/pdb4wkq/pdb](https://doi.org/10.2210/pdb4wkq/pdb).
- [180] Shelley JC, Cholleti A, Frye LL, Greenwood JR, Timlin MR, Uchimaya M. Epik: A Software Program for pK a Prediction and Protonation State Generation for Drug-like Molecules. *J Comput Aided Mol Des*. 2007 Dec; 21(12):681–691. doi: [10.1007/s10822-007-9133-z](https://doi.org/10.1007/s10822-007-9133-z).
- [181] Berendsen HJC, Postma JPM, van Gunsteren WF, Hermans J. Interaction Models for Water in Relation to Protein Hydration. In: Pullman B, editor. *Intermolecular Forces*, vol. 14 Dordrecht: Springer Netherlands; 1981.p. 331–342. doi: [10.1007/978-94-015-7658-1_21](https://doi.org/10.1007/978-94-015-7658-1_21).
- [182] Shivakumar D, Williams J, Wu Y, Damm W, Shelley J, Sherman W. Prediction of Absolute Solvation Free Energies Using Molecular Dynamics Free Energy Perturbation and the OPLS Force Field. *J Chem Theory Comput*. 2010 May; 6(5):1509–1519. doi: [10.1021/ct900587b](https://doi.org/10.1021/ct900587b).
- [183] Pearlman DA. A Comparison of Alternative Approaches to Free Energy Calculations. *J Phys Chem*. 1994 Feb; 98(5):1487–1493. doi: [10.1021/j100056a020](https://doi.org/10.1021/j100056a020).

- [184] **Rocklin GJ**, Mobley DL, Dill KA, Hünenberger PH. Calculating the Binding Free Energies of Charged Species Based on Explicit-Solvent Simulations Employing Lattice-Sum Methods: An Accurate Correction Scheme for Electrostatic Finite-Size Effects. *J Chem Phys*. 2013 Nov; 139(18):184103. doi: [10.1063/1.4826261](https://doi.org/10.1063/1.4826261).
- [185] **Feller SE**, Zhang Y, Pastor RW, Brooks BR. Constant pressure molecular dynamics simulation: The Langevin piston method. *The Journal of Chemical Physics*. 1995; 103(11):4613–4621. <https://doi.org/10.1063/1.470648>, doi: [10.1063/1.470648](https://doi.org/10.1063/1.470648).
- [186] **Essmann U**, Perera L, Berkowitz ML, Darden T, Lee H, Pedersen LG. A smooth particle mesh Ewald method. *The Journal of Chemical Physics*. 1995; 103(19):8577–8593. <https://doi.org/10.1063/1.470117>, doi: [10.1063/1.470117](https://doi.org/10.1063/1.470117).
- [187] **Martyna GJ**, Tobias DJ, Klein ML. Constant pressure molecular dynamics algorithms. *The Journal of Chemical Physics*. 1994; 101(5):4177–4189. <https://doi.org/10.1063/1.467468>, doi: [10.1063/1.467468](https://doi.org/10.1063/1.467468).
- [188] **Tuckerman M**, Berne BJ, Martyna GJ. Reversible multiple time scale molecular dynamics. *The Journal of Chemical Physics*. 1992; 97(3):1990–2001. <https://doi.org/10.1063/1.463137>, doi: [10.1063/1.463137](https://doi.org/10.1063/1.463137).
- [189] **Hopkins CW**, Le Grand S, Walker RC, Roitberg AE. Long-Time-Step Molecular Dynamics through Hydrogen Mass Repartitioning. *Journal of Chemical Theory and Computation*. 2015; 11(4):1864–1874. <http://dx.doi.org/10.1021/ct5010406>, doi: [10.1021/ct5010406](https://doi.org/10.1021/ct5010406), pMID: 26574392.
- [190] **Wang L**, Berne BJ, Friesner RA. On Achieving High Accuracy and Reliability in the Calculation of Relative Protein-Ligand Binding Affinities. *Proc Natl Acad Sci*. 2012 Feb; 109(6):1937–1942. doi: [10.1073/pnas.1114017109](https://doi.org/10.1073/pnas.1114017109).
- [191] **Bennett CH**. Efficient Estimation of Free Energy Differences from Monte Carlo Data. *J Comput Phys*. 1976; 22:245–268.
- [192] **Shirts MR**, Bair E, Hooker G, Pande VS. Equilibrium Free Energies from Nonequilibrium Measurements Using Maximum-Likelihood Methods. *Phys Rev Lett*. 2003 Oct; 91(14). doi: [10.1103/PhysRevLett.91.140601](https://doi.org/10.1103/PhysRevLett.91.140601).
- [193] **Price DJ**, Jorgensen WL. Computational Binding Studies of Human Pp60c-Src SH2 Domain with a Series of Nonpeptide, Phosphophenyl-Containing Ligands. *Bioorg Med Chem Lett*. 2000 Sep; 10(18):2067–2070. doi: [10.1016/S0960-894X\(00\)00401-7](https://doi.org/10.1016/S0960-894X(00)00401-7).
- [194] **Luccarelli J**, Michel J, Tirado-Rives J, Jorgensen WL. Effects of Water Placement on Predictions of Binding Affinities for P38 α MAP Kinase Inhibitors. *J Chem Theory Comput*. 2010 Dec; 6(12):3850–3856. doi: [10.1021/ct100504h](https://doi.org/10.1021/ct100504h).
- [195] **Barouch-Bentov R**, Sauer K. Mechanisms of drug resistance in kinases. *Expert Opin Invest Drugs*. 2011 Feb; 20(2):153–208.
- [196] **McDermott U**, Sharma SV, Dowell L, Greninger P, Montagut C, Lamb J, Archibald H, Raudales R, Tam A, Lee D, Rothenberg SM, Supko JG, Sordella R, Ulkus LE, Iafrate AJ, Maheswaran S, Njauw CN, Tsao H, Drew L, Hanke JH, et al. Identification of genotype-correlated sensitivity to selective kinase inhibitors by using high-throughput tumor cell line profiling. *Proc Natl Acad Sci*. 2007 Dec; 104(50):19936–19941.

- [197] **Salvatier J**, Wiecki TV, Fonnesbeck C. Probabilistic programming in Python using PyMC3. *PeerJ Computer Science*. 2016 Apr; 2:e55. <https://doi.org/10.7717/peerj-cs.55>, doi: 10.7717/peerj-cs.55.
- [198] **Kramer C**, Kalliokoski T, Gedeck P, Vulpetti A. The Experimental Uncertainty of Heterogeneous Public K_i Data. *J Med Chem*. 2012 Jun; 55(11):5165–5173. doi: 10.1021/jm300131x.
- [199] **Manning G**, Whyte DB, Martinez R, Hunter T, Sudarsanam S. The protein kinase complement of the human genome. *Science*. 2002; 298:1912–1934.
- [200] **Hunter T**, Manning G. The Eukaryotic Protein Kinase Superfamily and the Emergence of Receptor Tyrosine Kinases. In: Wheeler DL, Yarden Y, editors. *Receptor Tyrosine Kinases: Structure, Functions and Role in Human Disease* New York, NY: Springer New York; 2015.p. 1–15.
- [201] **Knight ZA**, Lin H, Shokat KM. Targeting the cancer kinome through polypharmacology. *Nature Reviews Cancer*. 2010 Feb; 10(2):130–137. <http://www.nature.com.proxy.library.cornell.edu/nrc/journal/v10/n2/full/nrc2787.html>, doi: 10.1038/nrc2787.
- [202] **American Cancer Society**, Cancer Facts & Figures 2015. American Cancer Society; 2015.
- [203] **Cohen P**, Tcherpakov M. Will the Ubiquitin System Furnish as Many Drug Targets as Protein Kinases? *Cell*. 2010 Nov; 143(5):686–693. <http://linkinghub.elsevier.com/retrieve/pii/S0092867410012961>, doi: 10.1016/j.cell.2010.11.016.
- [204] **Stegmeier F**, Warmuth M, Sellers WR, Dorsch M. Targeted cancer therapies in the twenty-first century: Lessons from imatinib. *Clin Pharm & Therap*. 2010; 87:543–552. doi: 10.1038/clpt.2009.297.
- [205] **Wu P**, Nielsen TE, Clausen MH. FDA-approved small-molecule kinase inhibitors. *Trends in Pharmacological Sciences*. 2015 Jul; 36(7):422–439. <http://linkinghub.elsevier.com/retrieve/pii/S0165614715000772>, doi: 10.1016/j.tips.2015.04.005.
- [206] **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. *Journal of Chemical Information and Modeling*. 2015; 55(3):538–549. <http://dx.doi.org/10.1021/ci500624s>, doi: 10.1021/ci500624s, pMID: 25557645.
- [207] **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. <http://linkinghub.elsevier.com/retrieve/pii/S1535610810004332>, doi: 10.1016/j.ccr.2010.10.031.
- [208] **Karaman MW**, Herrgard S, Treiber DK, Gallant P, Atteridge CE, Campbell BT, Chan KW, Ciceri P, Davis MI, Edeen PT, Faraoni R, Floyd M, Hunt JP, Lockhart DJ, Milanov ZV, Morrison MJ, Pallares G, Patel HK, Pritchard S, Wodicka LM, et al. A quantitative analysis of kinase inhibitor selectivity. *Nature Biotech*. 2008; 26:127–132.

- [209] **Davis MI**, Hunt JP, Herrgard S, Ciceri P, Wodicka LM, Pallares G, Hocker M, Treiber DK, Zarrinkar PP. Comprehensive analysis of kinase inhibitor selectivity. *Nature biotechnology*. 2011 Nov; 29(11):1046–1051.
- [210] **Médard G**, Pachel F, Ruprecht B, Klaeger S, Heinzlmeir S, Helm D, Qiao H, Ku X, Wilhelm M, Kuehne T, Wu Z, Dittmann A, Hopf C, Kramer K, Kuster B. Optimized chemical proteomics assay for kinase inhibitor profiling. *Journal of proteome research*. 2015 Mar; 14(3):1574–1586.
- [211] **Chambers SP**, Austen DA, Fulghum JR, Kim WM. High-throughput screening for soluble recombinant expressed kinases in *Escherichia coli* and insect cells. *Protein Expression and Purification*. 2004; 36:40–47. doi: 10.1016/j.pep.2004.03.003.
- [212] **Wang L**, Foster M, Zhang Y, Tschantz WR, Yang L, Worrall J, Loh C, Xu X. High yield expression of non-phosphorylated protein tyrosine kinases in insect cells. *Protein Express Pur*. 2008; 61:204–211. doi: 10.1016/j.pep.2008.05.017.
- [213] **Seeliger MA**, Young M, Henderson MN, Pellicena P, King DS, Falick AM, Kuriyan J. High yield bacterial expression of active c-Abl and c-Src tyrosine kinases. *Protein Sci*. 2005; 14:3135–3139.
- [214] **The Structural Genomics Consortium**, SGC: Progress in protein kinase structural biology; 2012. <http://www.thesgc.org/scientists/resources/kinases>, updated 18 Jul 2012.
- [215] **Weisberg E**, Manley PW, Breitenstein W, Brügger J, Cowan-Jacob SW, Ray A, Huntly B, Fabbro D, Fendrich G, Hall-Meyers E, Kung AL, Mestan J, Daley GQ, Callahan L, Catley L, Cavazza C, Azam M, Mohammed A, Neuberg D, Wright RD, et al. Characterization of AMN107, a selective inhibitor of native and mutant Bcr-Abl. *Cancer Cell*. 2005 Feb; 7(2):129–141.
- [216] **Levinson NM**, Kuchment O, Shen K, Young MA, Koldobskly M, Karplus M, Cole PA, Kuriyan J. A Src-like inactive conformation in the Abl tyrosine kinase domain. *Plos Biology*. 2006; 4:e144.
- [217] **Horio T**, Hamasaki T, Inoue T, Wakayama T. Structural factors contributing to the Abl/Lyn dual inhibitory activity of 3-substituted benzamide derivatives. *Bioorganic & medicinal . . .* 2007; .
- [218] **UniProt Consortium**. UniProt: The universal protein knowledgebase. *Nucl Acids Res*. 2017; 45:D158–D169. doi: 10.1093/nar/gkh131.
- [219] **Puhl AC**, Giacomini C, Irazoqui G, Batista Viera F, Villarino A, Terenzi H. Covalent immobilization of tobacco-etch-virus NIa protease: a useful tool for cleavage of the histidine tag of recombinant proteins. *Biotechnology and Applied Biochemistry*. 2009 May; 53(3):165–174.
- [220] **Nallamsetty S**, Kapust RB, Tözsér J, Cherry S, Tropea JE, Copeland TD, Waugh DS. Efficient site-specific processing of fusion proteins by tobacco vein mottling virus protease in vivo and in vitro. *Protein expression and purification*. 2004 Nov; 38(1):108–115.
- [221] **Sun C**, Liang J, Shi R, Gao X, Zhang R, Hong F, Yuan Q, Wang S. Tobacco etch virus protease retains its activity in various buffers and in the presence of diverse additives. *Protein expression and purification*. 2012 Mar; 82(1):226–231.

- [222] **Lo MC**, Aulabaugh A, Jin G, Cowling R, Bard J, Malamas M, Ellestad G. Evaluation of fluorescence-based thermal shift assays for hit identification in drug discovery. *Analytical Biochemistry*. 2004 Sep; 332(1):153–159.
- [223] **Ericsson UB**, Hallberg BM, Detitta GT, Dekker N, Nordlund P. Thermofluor-based high-throughput stability optimization of proteins for structural studies. *Analytical Biochemistry*. 2006 Oct; 357(2):289–298.
- [224] **Matulis D**, Kranz JK, Salemme FR, Todd MJ. Thermodynamic stability of carbonic anhydrase: measurements of binding affinity and stoichiometry using ThermoFluor. *Biochemistry*. 2005 Apr; 44(13):5258–5266.
- [225] **Iyer GH**, Taslimi P, Pazhanisamy S. Staurosporine-based binding assay for testing the affinity of compounds to protein kinases. *Analytical Biochemistry*. 2008 Feb; 373(2):197–206.
- [226] **Levinson NM**, Boxer SG. Structural and Spectroscopic Analysis of the Kinase Inhibitor Bosutinib and an Isomer of Bosutinib Binding to the Abl Tyrosine Kinase Domain. *PLoS One*. 2012; 7:e29828.
- [227] **Levinson NM**, Boxer SG. A conserved water-mediated hydrogen bond network defines bosutinib's kinase selectivity. *Nature chemical biology*. 2014 Feb; 10(2):127–132.
- [228] **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: Figure 1. *Cancer Discovery*. 2012 May; 2(5):401–404. doi: [10.1158/2159-8290.CD-12-0095](https://doi.org/10.1158/2159-8290.CD-12-0095).
- [229] **Cheng DT**, Mitchell TN, Zehir A, Shah RH, Benayed R, Syed A, Chandramohan R, Liu ZY, Won HH, Scott SN, Brannon AR, O'Reilly C, Sadowska J, Casanova J, Yannes A, Hechtman JF, Yao J, Song W, Ross DS, Oultache A, et al. Memorial Sloan Kettering-Integrated Mutation Profiling of Actionable Cancer Targets (MSK-IMPACT). *The Journal of Molecular Diagnostics*. 2015 May; 17(3):251–264. doi: [10.1016/j.jmoldx.2014.12.006](https://doi.org/10.1016/j.jmoldx.2014.12.006).
- [230] **Reva B**, Antipin Y, Sander C. Determinants of protein function revealed by combinatorial entropy optimization. . 2007; 8(11):R232. <http://genomebiology.biomedcentral.com/articles/10.1186/gb-2007-8-11-r232>, doi: 10.1186/gb-2007-8-11-r232.
- [231] **Reva B**, Antipin Y, Sander C. Predicting the functional impact of protein mutations: application to cancer genomics. *Nucleic Acids Research*. 2011; 39(17):e118. <http://dx.doi.org/10.1093/nar/gkr407>, doi: 10.1093/nar/gkr407.
- [232] **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.
- [233] **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):p11–p11.

- [234] Ramos AH, Lichtenstein L, Gupta M, Lawrence MS, Pugh TJ, Saksena G, Meyerson M, Getz G. Oncotator: cancer variant annotation tool. *Human mutation*. 2015 Apr; 36(4):E2423–9.
- [235] Sørensen HP, Mortensen KK. Advanced genetic strategies for recombinant protein expression in *Escherichia coli*. *Journal of Biotechnology*. 2005; 115(2):113 – 128. <http://www.sciencedirect.com/science/article/pii/S0168165604004560>, doi: <http://dx.doi.org/10.1016/j.jbiotec.2004.08.004>.
- [236] Haacke A, Fendrich G, Ramage P, Geiser M. Chaperone Over-Expression in *Escherichia Coli*: Apparent Increased Yields of Soluble Recombinant Protein Kinases Are Due Mainly to Soluble Aggregates. *Protein Expr Purif*. 2009 Apr; 64(2):185–193. doi: [10.1016/j.pep.2008.10.022](https://doi.org/10.1016/j.pep.2008.10.022).
- [237] Kim DM, Swartz JR. Prolonging Cell-Free Protein Synthesis with a Novel ATP Regeneration System. *Biotechnol Bioeng*. 1999; 66(3):180–188. doi: [10.1002/\(SICI\)1097-0290\(1999\)66:3<180::AID-BIT6>3.0.CO;2-S](https://doi.org/10.1002/(SICI)1097-0290(1999)66:3<180::AID-BIT6>3.0.CO;2-S).
- [238] Sawasaki T, Ogasawara T, Morishita R, Endo Y. A Cell-Free Protein Synthesis System for High-Throughput Proteomics. *Proc Natl Acad Sci*. 2002 Nov; 99(23):14652–14657. doi: [10.1073/pnas.232580399](https://doi.org/10.1073/pnas.232580399).
- [239] Klock HE, Koesema EJ, Knuth MW, Lesley SA. Combining the polymerase incomplete primer extension method for cloning and mutagenesis with microscreening to accelerate structural genomics efforts. *Proteins: Structure, Function, and Bioinformatics*. 2008 May; 71(2):982–994. doi: [10.1002/prot.21786](https://doi.org/10.1002/prot.21786).
- [240] Velankar S, Dana JM, Jacobsen J, van Ginkel G, Gane PJ, Luo J, Oldfield TJ, O'Donovan C, Martin MJ, Kleywegt GJ. SIFTS: Structure Integration with Function, Taxonomy and Sequences resource. *Nucleic Acids Research*. 2013; 41(D1):D483. <http://dx.doi.org/10.1093/nar/gks1258>, doi: [10.1093/nar/gks1258](https://doi.org/10.1093/nar/gks1258).
- [241] 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.