

THE UNIVERSITY OF CHICAGO

A KINETIC FRAMEWORK TO UNRAVEL THE PRINCIPLES OF COVALENT
INHIBITION

A DISSERTATION SUBMITTED TO
THE FACULTY OF THE DIVISION OF THE BIOLOGICAL SCIENCES
AND THE PRITZKER SCHOOL OF MEDICINE
IN CANDIDACY FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY
GRADUATE PROGRAM IN BIOCHEMISTRY AND MOLECULAR BIOPHYSICS

BY
KYLE SALIM GHABY

CHICAGO, ILLINOIS
AUGUST 2026

© 2026 by Kyle Salim Ghaby

ORCID iD: <https://orcid.org/0000-0002-6071-9274>

For Tati

Corpora non agunt nisi fixata.

—PAUL EHRLICH, 1913

TABLE OF CONTENTS

LIST OF FIGURES	viii
LIST OF TABLES	x
ACKNOWLEDGMENTS	xi
ABSTRACT	xiv
1 INTRODUCTION	1
1.1 Covalent inhibitors in pharmacology: a brief history	3
1.2 Enablers of modern covalent inhibitor design	6
1.3 Fundamental challenges require fundamental solutions	7
1.3.1 The limits of established metrics	7
1.3.2 A problem of many scales	8
1.3.3 A framework built on mass action	10
1.3.4 Microscopic input from molecular dynamics	12
1.4 Overview of dissertation	14
2 KINETIC MODELING OF COVALENT INHIBITION: EFFECTS OF RAPIDLY FLUCTUATING INTERMEDIATE STATES	17
2.1 Introduction	18
2.2 Theoretical framework	21
2.2.1 Kinetic schemes	21
2.2.2 Simulation methodology	23
2.3 Results	26
2.3.1 Noncovalent vs covalent inhibition	26
2.3.2 Effects on observable outcomes	28
2.4 Conclusion	37
3 COMPUTATIONAL INVESTIGATION OF THE COVALENT INHIBITION MECH- ANISM OF BRUTON’S TYROSINE KINASE BY IBRUTINIB	39
3.1 Introduction	40
3.2 Methods	43
3.2.1 Simulation system	43
3.2.2 Hybrid QM/MM simulations	44
3.2.3 PMF calculation of hydronium	49
3.3 Results	50
3.3.1 Dynamics of the unlinked inhibitor associated with the binding pocket	50
3.3.2 Initial reaction pathways from SMD simulations	54
3.3.3 Optimized reaction pathway from the string method	56
3.3.4 Kinetic implications	62
3.4 Conclusion	68

4	MOLECULAR DYNAMICS-GUIDED DESIGN AND CHEMOPROTEOMIC PROFILING OF COVALENT KINASE ACTIVITY PROBES	70
4.1	Introduction	71
4.2	Results	73
4.2.1	Computational design of covalent kinase activity probes identifies K60P as candidate	73
4.2.2	Proteome-wide profiling reveals kinase and non-kinase targets of K60P and KW-2449	76
4.2.3	Covalent probes display specific modes of active site labeling despite disparate kinetics for ABL1, a model target kinase	80
4.2.4	A BMMC kinetic modeling framework captures key determinants of covalent probe efficiency	86
4.2.5	Molecular dynamics modeling rationalizes “active” kinase conformation preference and inherent differences in covalent probe kinetic efficiency	89
4.3	Discussion	92
5	DETERMINANTS OF COVALENT INHIBITOR RESIDENCE TIME REVEALED BY MOLECULAR AND KINETIC SIMULATIONS	96
5.1	Introduction	97
5.2	Methods	100
5.2.1	Atomic models	100
5.2.2	Free energy landscape of C $_{\alpha}$ -H bond cleavage	103
5.2.3	Noncovalent binding free energy	104
5.2.4	Kinetic modeling	106
5.3	Results and discussion	108
5.3.1	Dynamics of associated inhibitor	108
5.3.2	Noncovalent association	114
5.3.3	Kinetic modeling	116
5.4	Conclusion	121
6	CONCLUSION AND OUTLOOK	123
6.1	Summary of dissertation	123
6.2	Implications for covalent inhibitor design	125
6.3	Refining the kinetic model	128
6.3.1	Model fidelity and its limits	128
6.3.2	Resolving the covalent step	129
6.3.3	State representation	132
6.4	Extending beyond <i>in vitro</i> kinetics	134
6.4.1	Pharmacokinetics	134
6.4.2	Systems biology	137
6.5	Concluding Remarks	138

APPENDIX A SUPPLEMENTARY INFORMATION FOR CHAPTER 3	141
A.1 Structural dynamics	141
A.2 Reaction pathways	144
A.3 Expanded kinetic scheme	147
APPENDIX B SUPPLEMENTARY INFORMATION FOR CHAPTER 4	149
B.1 Supplementary figures and tables	149
B.2 Methods for biology	166
B.2.1 Cell culture	166
B.2.2 <i>In vitro</i> recombinant Abl labeling assays	166
B.2.3 BSA labeling assays	168
B.2.4 K50P and K60P probe dosing on live cells and lysates for immunoblot analysis	169
B.2.5 Viability assay	170
B.2.6 ABL1 site of labeling assays	170
B.2.7 SILAC cell culture methods and proteomic sample preparation	172
B.2.8 Sample preparation and streptavidin enrichment of biotin-labelled pro- teins in live-cell experiments	172
B.2.9 Sample preparation and streptavidin enrichment of biotin-labelled pro- teins in lysate experiments	173
B.2.10 LC-MS/MS acquisition and analysis	175
B.3 Supplementary molecular modeling methodology	176
B.4 Supplementary molecular modeling discussion	180
B.5 Supplementary kinetic modeling methodology	181
B.6 Supplementary kinetic modeling discussion	184
APPENDIX C SUPPLEMENTARY INFORMATION FOR CHAPTER 5	187
REFERENCES	193

LIST OF FIGURES

1.1	Noncovalent vs. covalent inhibition of enzyme activity over time	4
1.2	A discretization of timescales governing covalent inhibitor binding kinetics . . .	9
2.1	Occupancy for noncovalent vs. covalent inhibition for different values of p_r . . .	27
2.2	Noncovalent residence time of a bound ligand as a function of k_{off} and p_r	27
2.3	Simulated dose-rate data for different values of p_r	30
2.4	Analytical vs. numerical dose-responses	33
2.5	Simulated dose-response data for different parameter and time values	34
3.1	Chemical structure of ibrutinib	41
3.2	Schematic of the BTK–ibrutinib reaction mechanism	43
3.3	Structure of BTK bound to ibrutinib	46
3.4	Conformational energy landscape for ibrutinib in BTK	51
3.5	Two states of distance between linking atoms of ibrutinib and BTK	52
3.6	Conformational distribution of BTK-ibrutinib is nonreactive and fast	53
3.7	Evolution of reaction coordinates during SMD simulations	55
3.8	Optimized string of the BTK-ibrutinib chemical reaction	57
3.9	Energy landscape of a hydronium leaving/entering the binding pocket	61
3.10	Numerical kinetic simulations of the detailed kinetic scheme	64
4.1	Computation-guided design for covalent, conserved lysine-targeting kinase probes	76
4.2	Proteome-wide labeling by K60P-Bio in lysates and live cells	81
4.3	K60P specificity for kinases and known novel targets of KW-2449	82
4.4	K60P, K50P, and X4K covalently label recombinant ABL1 with distinct kinetics	83
4.5	K60P and X4K covalently label distinct sites in ABL1	85
4.6	Kinetic modeling reveals complex ABL1-binding kinetics of K60P, K50P, and X4K	88
4.7	Molecular dynamics reveals probe preference for DFG-in conformation of kinases	92
5.1	Chemical structures of the investigated reversible inhibitors	101
5.2	Linked ensemble over a key dihedral that governs retro-reactivity	108
5.3	Energy landscape and water accessibility over deprotonation	110
5.4	Reactive and nonreactive conformations in the unlinked ensemble	113
5.5	Noncovalent dissociation constant vs. overall residence time	115
5.6	Experimental and simulated dissociation curves of total occupancy over time . .	117
A.1	RMSD of the inhibitor and protein pocket over time	141
A.2	Conformational fluctuations of the unlinked inhibitor over time	142
A.3	Conformational energy landscape with respect to COM deviation and RMSD . .	142
A.4	Visualizations of the two distinct unlinked conformations	143
A.5	All 10 initial pathways generated from SMD trajectories	144
A.6	Completing the reaction for the indirect pathway with SMD	145
A.7	Indirect and direct pathway observed from SMD trajectories	145
A.8	Optimized string of indirect reaction path	146

A.9	Time course of the kinetic scheme including the unbound thiolate	148
A.10	Free energy profile of the thiol and thiolate reaction pathways	148
B.1	Modeled and final poses of screened compounds	149
B.2	Alternative alkene rotation for K5* compounds	150
B.3	Kinome scope of X4K	150
B.4	K60P-Bio labeling across cell lines	151
B.5	PC-2-52 and XO44 competition vs K60P-Bio	152
B.6	PC-2-52 competition with X4K	153
B.7	Gene Ontology enrichment of KW-2449 competition	153
B.8	KW-2449 competition with X4K in live cells	154
B.9	KW-2449 viability curve	154
B.10	Uncropped blots for ABL1 dose/time responses	155
B.11	Uncropped blots for PC-2-52 competition with X4K and K60P	156
B.12	MS/MS of K271-labeled peptides	157
B.13	MS/MS of Y393- and Y226-labeled peptides	157
B.14	apo-ABL1 conformations with Y393 closest to K60P	158
B.15	Solvent-exposed tyrosine residues of ABL1	158
B.16	Time-evolution of BMMC parameters and energy	161
B.17	Minimized fit and BMMC variation for each K50P scheme	162
B.18	Species evolution for K60P minimized fit	162
B.19	Experimental $k_{\text{inact}}/K_{\text{I}}$ determination	163
B.20	Blots of K50P and K60P time responses	164
B.21	Intrinsic reactivity of X4K, K60P, K50P	164
B.22	Empirical Cumulative Distribution Function of lysine–warhead distance	165
C.1	The two dihedrals used to resolve the conformational states in the unlinked ensemble	188
C.2	The two dihedrals restrained during the steered proton-abstraction simulations .	188
C.3	Conformational energy landscapes for a set of diverse kinase–inhibitor systems .	189
C.4	Conformational reactivity vs. noncovalent dissociation constant	189
C.5	Time courses of all enzyme-containing species during incubation and post-incubation	190
C.6	Time evolution of the BMMC parameters and energy	191

LIST OF TABLES

3.1	Estimated rates from transition state theory	60
3.2	Experimental estimates of kinetic parameters of BTK inhibition by ibrutinib . .	65
5.1	Estimated rate constants from transition state theory	110
5.2	Computed absolute binding and solvation free energies	114
5.3	Predicted timescales of coarse-grained rate constants	119
B.1	Solvent-exposed tyrosine residues of ABL1	159
B.2	Medians and ranges of BMMC parameters	160
B.3	Experimental vs BMMC k_{inact} , K_{I} , $k_{\text{inact}}/K_{\text{I}}$	163
C.1	Kinase-inhibitor systems investigated for unlinked conformational reactivity . .	187
C.2	Medians and ranges of the BMMC parameters	192

ACKNOWLEDGMENTS

Well, here we are. July 9th, 2026. This dissertation is basically finished and all that really remains to be written are the following sentiments, and all I really feel is immense gratitude. Gratitude and motivation to not forget why.

Carving a message so personal into stone, one as semi-sempiternal as anyone can hope for a dissertation (or any piece of academic work for that matter), is made easier by the supportive, intelligent, and kind people who have given me something to say. We ourselves are stones, smoothened and faceted and engraved until it is our turn to practice our own iteration of lapidary.

Those who come to be subjected to my mentorship should know the same thing I will come back here to remember: what I am is only a fraction me. I can only be proud of myself because of the vast amount of impressive people that left their impression on me, and for that I am so thankful.

First, my doctoral advisor, Benoît Roux. Benoît taught me to be pedagogical in how I present my work, and to know who I am presenting to. When solving statistical mechanical problems, he taught me how to pick and choose the details that matter. His mastery over the fundamentals is obvious in his practical advice, and it is a standard I follow. Benoît has a contagious enthusiasm for learning. This, together with his somehow unending support and patience, made him a perfect mentor for me. Benoît reminds me of my Lebanese family. Slightly related is the fact that he is French-Canadian. More importantly, he has this stentorian voice that could make him appear intimidating, but in reality, Benoît is only good-hearted. So, Benoît, thank you for making me feel at home here.

Second is my thesis committee: Anthony Kossiakoff, Raymond Moellering, and Tobin Sosnick. Given how closely this thesis is tied to experiment, both in its applications and in the field it primarily serves, Benoît and I wanted perspectives beyond those focused mainly on molecular dynamics. That is exactly what we got. I did not always make it easy for them,

but Tony, Ray, and Tobin continually provided invaluable advice throughout the years. They emphasized the importance of taking insight from molecular dynamics and turning it into an actionable message for other fields. A special thank you goes to Ray for (i) his perspective in the medicinal chemistry and industry realities of covalent inhibitor design, and (ii) the fruitful collaboration that gifted me several eureka moments when it came to modeling biochemical data.

I would next like to thank my peers at UChicago. It was an honor to be surrounded by such intelligent and passionate people throughout my entire PhD. Every week in Roux lab group meetings, when someone would present an update on their research, I left humbled and excited. Everyone had such a positive impact on me, but some people deserve a special mention for their help. I worked closely with Trayder Thomas throughout my entire PhD. Thank you, Trayder, for being such a fantastic brain-storming partner, for throwing yourself so deeply into problems I need help with, for sending me papers, for teaching me so much about molecular modeling, and for being a great friend. Thank you to Angela Barragán for mentoring me from my rotation to about halfway through my PhD. Thank you to Pratyasha Chakraborty and Anthony Carlos for being great collaborators and for hearing me out while I yapped about kinetics. Thank you to Phoebe Rice, who is not so much a peer as she is a mentor, for supporting me throughout my PhD and for the brief but memorable experience in her lab. Phoebe, I hope my scientific explanations can be half as whimsical, and my critiques be half as trenchant, as yours. Thank you, Alessia Ghidini, for helping me navigate free energy calculations, for proof-reading my writing, and for challenging my opinions.

Thank you to my undergraduate mentors, Matthew Salanga and Gerrick Lindberg, who taught me the fundamentals of what it means to be scientist in molecular biology and computational chemistry, respectively. I am so proud to have been a part of both of your labs. Thank you for believing in me.

Thank you to all my friends, especially Joseph Guimond, Phillip Kalaniopio, Sumit

Bandekar, Peter Wang, Stephen Gee, and Ramon Mendoza Uriarte. You all inspire me.

To my family, thank you so, so much. My whole life, I have been surrounded by a large and loving family. They provided me with a safety net that allowed me to take risks and pursue my passion, and being with them — my parents, sister, grandparents, cousins, aunts/uncles, everyone — is the quickest way for me to recharge. In reaching the age that my parents were when they had me, I have often found myself in awe at what they accomplished then and still accomplish now (yes, I am familiar with the (apocryphal) Mark Twain quote, thanks to my dad). Thank you, Mom and Dad, for this life you both have given me. Thank you for your guidance. I look up to you both. Karley, you are such an amazing sister and an even more amazing person. Thank you for the general wisdom, thank you for the fashion advice, thank you for carrying the burden of remembering every minute detail from our youth, and thank you for always being down to watch a random two-hour-long YouTube video.

And thank you, dear reader, for your attention.

ABSTRACT

Covalent inhibitors, *i.e.*, small-molecule drugs that form covalent bonds with their targets, have made for some of the most successful stories in pharmacology. In principle, their tunable reactivity provides another dial for drug design on top of the noncovalent scaffold, offering selectivity and potency that scaffold alone cannot reach. Only after decades of eschewal during the late 1900s have their rational design and systematic application become possible. While the last few decades have seen a rapid acceleration to this end, the initiative is far from over, and many theoretical challenges stand in the way of a satisfactory understanding.

Understanding covalent inhibitors requires fundamentally different concepts from those that describe noncovalent inhibitors. An experimental evaluation of inhibition strength is forced to measure a timescale leading to equilibrium rather than a population distribution at equilibrium. However, covalent inhibition is controlled by a hierarchy of steps including diffusional encounter, noncovalent association, conformational rearrangements of the enzyme-inhibitor complex, covalent bond formation, and possibly bond cleavage. These processes can occur on vastly disparate timescales, spanning nanoseconds for fast conformational fluctuations to days for slow covalent bond turnover. The challenge is that an experimental assay yields a nonlinear combination of the underlying rates, from which the individual contributions are inseparable without additional insight.

To unravel the principles between the macroscopic mechanisms we observe and the microscopic mechanisms we predict, we introduce a kinetic framework comprising numerical simulations of the time evolution of arbitrary kinetic schemes, lossless experimental fitting, and Bayesian parameter exploration. The goal is to relate the data collected by atomistic simulations to the data collected by biochemical assays. In doing so, two questions become answerable. Which microscopic step must change to produce a desired effect, and which structural modification changes it?

The first part of this dissertation demonstrates the importance of this approach with a

formal description of the surprisingly large effect that rapid conformational intermediates can have on experimental measurements of irreversible inhibition. The second part applies the same kinetic modeling machinery to the binding mechanism between Bruton’s tyrosine kinase and the irreversible covalent inhibitor ibrutinib, where free energy calculations of the chemical steps and of the conformational preferences of the unlinked ligand together recover the measured covalent efficiency. The third part combines a general workflow of kinetic modeling with bespoke analysis to interpret unconventional experimental data on the binding of novel covalent kinase probes to ABL1 kinase, resolving kinetic parameters with less data and more accuracy. The fourth part applies the framework developed over the previous three parts to understand how small structural differences among a triad of ibrutinib-derived reversible inhibitors can cause drastic changes in residence time. Taken together, these four parts show that productive drug design requires resolving the microscopic steps underlying an observable and identifying their molecular determinants.

CHAPTER 1

INTRODUCTION

“A substance will not work unless it is bound” — Paul Ehrlich’s century-old precept remains a core tenet of molecular medicine.¹ If a drug acts only insofar as it is associated with its target, then covalent inhibitors are the absolutist’s instantiation. After diagnosing the state of the field, this dissertation provides a theoretical and computational prescription for predicting and evaluating the efficacy of covalent inhibitors.

Every second, a biological cell manages anywhere from millions of chemical reactions in the simplest of cases to billions in a metabolically-active human cell.^a Electrons are taken, given, and shared in a chaotic economy in which essentially every atom that finds itself within the bounds of a biological system participates. Two atoms — usually those of a core group of nonmetal elements: carbon, hydrogen, oxygen, nitrogen, phosphorus, and sulfur — can sometimes share their most vulnerable (*i.e.*, valence) electrons, forming an aptly-named covalent bond. Molecules are defined by the particular arrangement of their covalent bonds, and the existing set of molecules changes with every reaction.

The formation and cleavage of biochemically-relevant covalent bonds are rare events on their own, so biological systems employ a class of biomolecules called enzymes to catalyze them. Enzymes vary in size, from $\sim 10^3$ to $\sim 10^5$ atoms.^{6;7} They are typically polymeric chains of either amino acids or ribonucleic acids that fold into a particular three-dimensional shape that confers the ability to bind specific substrates and catalyze specific reactions.

Almost every vital chemical process in biology is catalyzed by an enzyme. At any given

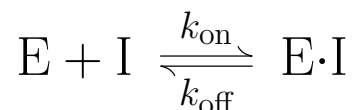
a. These estimates used adenosine triphosphate (ATP) turnover as a proxy for metabolic flux, since metabolism is the bulk of covalent biochemistry. An *E. coli* cell turns over $\sim 0.2\text{--}6.4 \times 10^6$ ATP/s depending on the growth phase.² A human body turns over ~ 60 kg ATP/day (BNID 105606³). With $M_{\text{ATP}} = 507$ g/mol and $\sim 3.7 \times 10^{13}$ cells in the human body⁴, $\frac{60,000 \text{ g/day}}{507 \text{ g/mol}} \times \frac{N_A \text{ ATP/mol}}{86,400 \text{ s/day}} \times \frac{1}{3.7 \times 10^{13} \text{ cells}} \approx 2 \times 10^7$ ATP/s/cell. This is a body-wide mean biased low by quiescent cells. Metabolically-active cells run much hotter, more consistent with its $\sim 10^3$ -fold greater volume and protein content relative to a prokaryote.⁵ ATP is made and spent through multi-step pathways, and there exist many other metabolic pathways not connected to ATP. Therefore, estimates of covalent reactions per second exceed ATP turnover by over an order of magnitude.

moment, each of the trillions of cells in the human body operates with several thousand distinct types of enzymes, present in anywhere from hundreds to tens of millions of copies apiece.⁸ They are distributed through the cytoplasm, sequestered in specialized compartments, or secreted into the extracellular space. Many of these are essential for keeping the cell alive.

Kinases are particularly important. This class of enzymes acts as a management layer that switches functional pathways, such as those governing gene expression or cell-cycle progression, on or off.

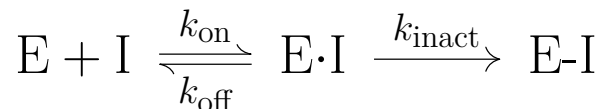
Many diseases can be blamed on problematic enzymes. These might be enzymes that are malfunctioning, leading to dangerous chemical imbalances. They could be foreign, hijacking the host cell on behalf of a pathogen. Maybe the enzyme is doing its job exactly as planned, but it is working for a greedy cancer cell. Sometimes the body cannot return to a healthy state without external aid. For these cases, efforts in molecular medicine aim to develop drugs that stop enzymes from doing their jobs.⁹ These are ligands that bind enzymes to inhibit their function. These are enzyme inhibitors.

Enzyme inhibitors can be partitioned into two mechanistic classes. A noncovalent inhibitor (I) binds its target enzyme (E) through a reversible association (Scheme 1.1). The



Scheme 1.1: Generic mechanism of noncovalent inhibition.

fraction of target bound is fixed by a reversible equilibrium that is typically quick to relax. A covalent inhibitor adds an irreversible (or slowly reversible) chemical step that converts the noncovalent complex into a linked adduct E-I (Scheme 1.2).



Scheme 1.2: Generic mechanism of covalent inhibition.

The qualitative consequence is shown in Figure 1.1. An excess of noncovalent inhibitor depresses enzyme activity to a plateau based on the amount of inhibitor and the stability of the bimolecular complex. This inhibitor may comprise only a few dozen heavy atoms arranged to be complementary to the primary binding pocket of an enzyme, classifying it as a small-molecule and competitive inhibitor. The inhibitor is competing for occupancy of the binding site with the enzyme’s small-molecule endogenous substrate.

These molecules often provide a starting template (*i.e.*, a scaffold) for covalent inhibitors. The modernized concept of a covalent inhibitor has two main structural components: the scaffold and a reactive chemical group called the “warhead”. The inhibitor is designed such that once it binds in the pocket, the electrophilic warhead would be situated nearby a specific nucleophilic “target residue”, a reactive side chain found somewhere near the primary binding site. With any luck, a covalent bond will form between the warhead and target residue, trapping the inhibitor in the pocket until the covalent bond breaks or, if the reaction is effectively irreversible, the protein is degraded.

As shown in Figure 1.1, covalent inhibitors drive enzyme activity toward zero given enough time (and given that the concentration of inhibitor is greater than or equal to the concentration of enzyme), because every encounter that proceeds to the adduct is removed permanently from the noncovalent equilibrium. This is what covalent inhibitors do. They act on enzymes with conditional reactivity for their target residue and with greater potency than their noncovalent counterparts.¹⁰ What they have achieved for pharmacology, and why designing them has nevertheless remained so difficult, motivates this dissertation.

1.1 Covalent inhibitors in pharmacology: a brief history

A natural place to begin is aspirin. Patented at the turn of the twentieth century and still among the most widely used medications in the world, aspirin is one of the earliest covalent drugs, acetylating an active-site serine of cyclooxygenase.¹¹ Penicillin followed a few decades

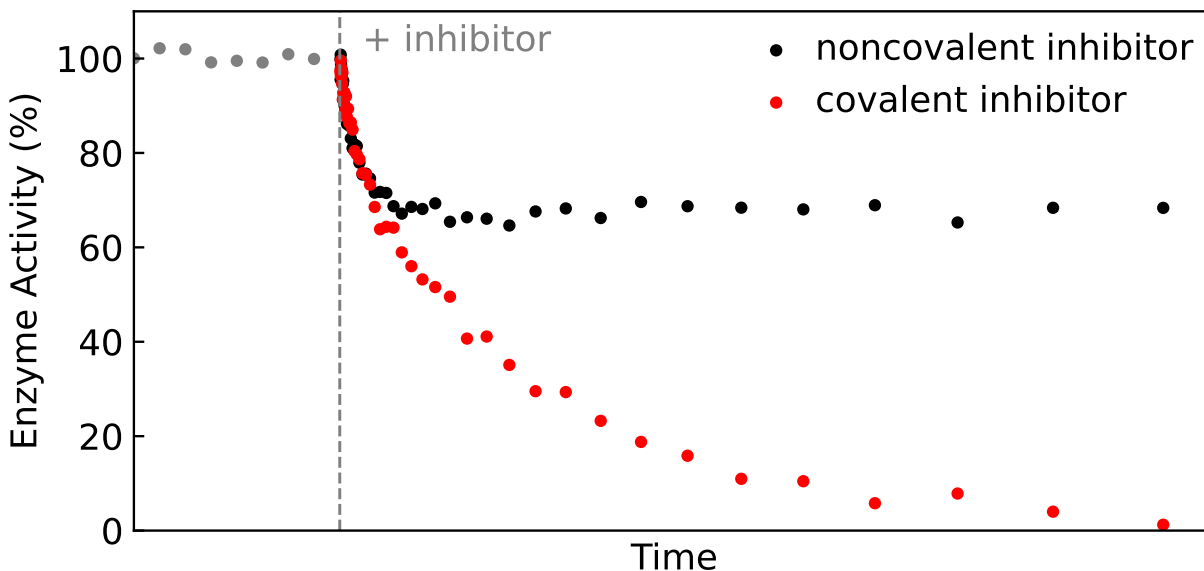


Figure 1.1: Simulated biochemical data observing enzyme activity over time. The gray dashed line marks the point at which either a noncovalent (black) or covalent (red) inhibitor is added.

later, discovered in the late 1920s and pressed into clinical use during World War II.¹² In neither case did anyone know the drug acted irreversibly via covalent bond formation.

It was not until the 1960s that irreversible inhibition attracted serious mechanistic attention. During this time, Kitz and Wilson popularized the canonical kinetic treatment of irreversible inactivation, giving the field a quantitative, assay-feasible protocol for evaluating these compounds and a surface-level grasp of their mechanism.¹³ A few years later, penicillin was recognized to acylate a transpeptidase serine.¹⁴

The liabilities became apparent too. Highly reactive molecules implicated off-target labeling and reactive-metabolite toxicity, and their reactivity made them difficult to study by assay and to deliver. A stigma formed: covalent inhibitors were regarded as reckless, too risky to design deliberately.¹⁰

The mechanism-based (“suicide”) inhibitors developed by Rando, Abeles, and Walsh in the 1970s and 80s constituted the first genuine framework for rational covalent design.¹⁵ These inhibitors are nonreactive until the target enzyme’s own catalytic action converts them into a highly reactive species, which then forms a covalent bond and inactivates the

enzyme. This strategy is therefore limited to targets bearing the catalytic machinery needed to unmask the latent warhead.

It is worth emphasizing that the modern “targeted covalent inhibitor” operates in the opposite direction. The warhead (typically an acrylamide targeting a cysteine as of 2026) is intrinsically reactive but kinetically slow, and selectivity is supplied by noncovalent positioning rather than by catalytic activation at the site.

Targeted covalent inhibitors continued to be found and used in the interim — a proton-pump inhibitor reached the market in the 80s and hit “blockbuster” status (*i.e.* it made a billion bucks) — but such compounds were discovered, not designed, emerging from phenotypic and biochemical screens with their covalent mechanism elucidated only afterward.

This situation began to change in the late 1990s with the first rationally designed, targeted covalent inhibition of a protein kinase, characterized in detail *in vitro* and *in vivo* and shown to work as intended.¹⁶ Because kinases are implicated across cancers and inflammatory and immune disease, as well as their own controversial history for being difficult to target with drug-like specificity, this garnered attention. Interest in irreversible inhibition revived, and covalent compounds were applied to indications ranging from oncology to vascular and central-nervous-system disorders, though most still arrived by serendipity. A decade later the rationally-designed, irreversible Bruton’s tyrosine kinase (BTK) inhibitor ibrutinib was approved for B-cell malignancies, and its clinical success effectively reversed the field’s bias.^{17;18} Since then a full-fledged rational design initiative has been ongoing, with intentionally-designed covalent inhibitors entering the clinic at a steady rate for especially cancers and viral infections.

The initiative is far from finished. There is no shortage of disease, and some “solved” targets are unsolved again by resistance that emerges in a fraction of the time it took to develop the inhibitor. For example, it took almost a decade and hundreds of analogs to inhibit KRAS(G12C), while acquired resistance was documented after only a couple years.¹⁹

1.2 Enablers of modern covalent inhibitor design

To see where one can contribute, it is necessary to understand the modern flavor of rational design and what enabled its acceleration over the past three decades. The acceleration was permitted and in large part defined by methodological gains across several fields.

Structural biology. One must see what they are working with. Foundational structures of pharmaceutically-relevant covalent targets from the mid-1990s and methodological advances in this field made it possible to observe where in a pocket a candidate nucleophile sits and how a warhead would be presented to it.²⁰

Kinome mapping. Sequence and structural surveys of the human kinome established that noncatalytic cysteines could be exploited as a systematic targeting strategy.^{21;22} Cysteine is a fortunate choice for three reasons. (i) The thiol is the most nucleophilic standard protein side chain at physiological pH. (ii) the acrylamide chemistry that targets it is mature, tunable, and clinically de-risked. (iii) cysteine is rare on protein surfaces ($\sim 2\%$ of residues),²³ conferring inherent selectivity. Roughly 200 of the ~ 518 human kinases carry a targetable cysteine near the active site.²²

Chemoproteomics. Activity-based protein profiling and its quantitative descendants extended target identification to proteome scale, allowing selectivity to be de-risked empirically across thousands of reactive residues rather than inferred one target at a time.^{24–26} These methods also generalized the notion of a targetable residue beyond cysteine.

Warhead chemistry. Maturation of electrophilic chemistry from the late 2000s onward provided chemists with a menu of warheads calibrated for distinct nucleophiles and tunable in intrinsic reactivity.²⁷

Molecular modeling. Molecular modeling is broad and immensely deep. In principle, it gives life to static structures, so one might expect it to be as central to covalent design as it has become for noncovalent design, where computed affinities are integral to lead optimization. In practice, its covalent contributions have primarily been intrinsic warhead-reactivity

calculations and on ranking docked poses.^{28;29} Many of the most elegant results have been post hoc dissections of a mechanism for a single ligand.³⁰ Today, the bottleneck of computation has shifted from the hardware’s ability to generate enough data to the quality with which we interpret and use atomistic simulations. The fact that theoretical or computational findings often postdate the design suggests that there are improvements to be made in this area.

1.3 Fundamental challenges require fundamental solutions

1.3.1 *The limits of established metrics*

In a test tube containing some concentration of target enzyme equilibrated with a larger concentration of noncovalent inhibitor, the probability that each enzyme is bound to an inhibitor (or the fraction of all enzymes bound) has a simple closed form,

$$p_{\text{bound}}^{\text{eq}} = \left(1 + \frac{K_{\text{d}}}{[\text{I}]_0}\right)^{-1} \quad (1.1)$$

where the K_{d} is the equilibrium dissociation constant, a thermodynamic measure of binding affinity that marks the initial inhibitor concentration ($[\text{I}]_0$) required for a bound fraction of 50%. Thermodynamic and equilibrium measurements here are informative because the system rests in a stationary distribution around this probability. For a covalent inhibitor, the long-time outcome is $\sim 100\%$ occupancy. The covalent state becomes truly stable only once the entire target population has been consumed, so the equilibrium distribution is uninformative and the value of any equilibrium-like potency (*e.g.*, a time-dependent EC_{50}) is ambiguous. What remains useful is not where the system ends up but how it relaxes there. The kinetics of irreversible binding must therefore be characterized with respect to both time and dose.

The canonical way of doing so descends from Michaelis and Menten’s famous 1913 work

in the early days of understanding enzyme kinetics.³¹ It was adapted and popularized for covalent inactivation by Kitz and Wilson half a century later.¹³ One fits analytical expressions derived from a simplified binding scheme to a biochemical observable measured during the approach to equilibrium, summarizing time- and dose-dependence with a few apparent parameters: the inactivation rate constant k_{inact} and inactivation steady-state constant K_{I} . The protocol’s end result is a curve of observed rate with respect to dose.

The approach is feasible *in vitro* and has served the field well, but it is quite limited. Collecting enough data can be experimentally expensive and cumbersome relative to the information it aims to extract. Yet, it discards most of a time-course by compressing it into a single rate. It relies on many assumptions (*e.g.*, pseudo-first-order, rapid equilibrium, steady state) that are often violated. It demands a rigid protocol that is sometimes impractical, since fitting the parameters of particularly fast ligands could require timepoints too short or initial concentrations too high for standard assays. It is inflexible to abnormalities in the experimental outcome. It yields little mechanistic insight. It was designed this way, like other pragmatic compressions of a biochemical binding process, driven by observability and parameter identifiability constraints.

1.3.2 *A problem of many scales*

The apparent parameters provide a macroscopic, bird’s-eye description, but covalent inhibition is controlled by a hierarchy of microscopic processes. It is the scope of these processes to reveal kinetic inefficiencies. Because each apparent rate absorbs contributions from several microscopic steps, the resulting kinetic degeneracy makes it difficult to diagnose which step limits, differentiates, or contributes to the observed behavior. A medicinal chemist modifying a molecule has, in effect, no clear readout which lever is being pulled, and there are several levers. As shown in Figure 1.2, the relevant processes span an enormous range of timescales, both on the individual level and as a whole. It is worth noting that this is all competing

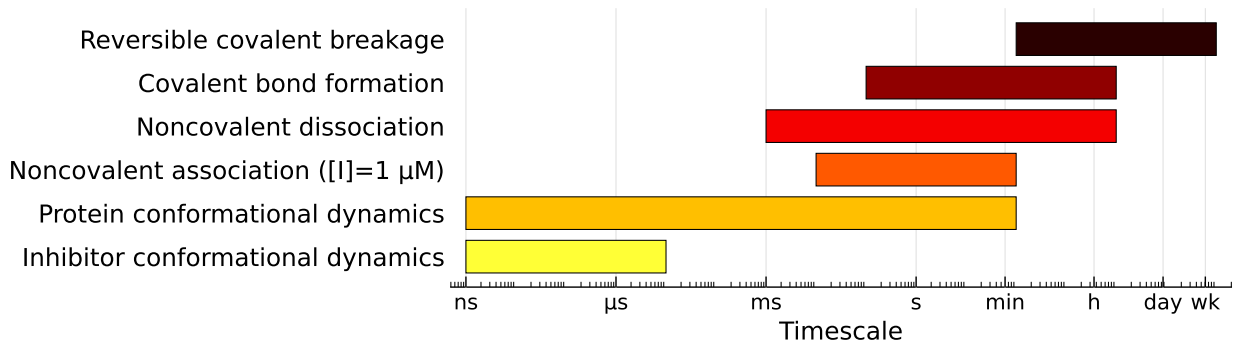


Figure 1.2: A discretization of timescales governing covalent inhibitor binding kinetics.

with the timescale of target resynthesis, which sets the dosing interval and determines the relevance of residence time.³²

Current efforts to understand covalent inhibition partition broadly into two camps to meet this scale problem from opposite sides. Experimentalists interrogate the macroscopic behavior. They mix trillions of protein with even more trillions of ligand, measure the binding over time, and describe the resulting curve with one or two parameters. Computational chemists interrogate the microscopic behavior with physics-based simulation. They follow, for example, the warhead of a single inhibitor as it fluctuates in a kinase active site toward a geometry competent to react with the target residue, resolving the action of dozens of atoms in arbitrary detail. It is very difficult to extract deep mechanistic insight from a biochemical assay, and equally difficult to see the macroscopic picture from a single fragment of the molecular mechanism.

When the disparity in scale is this large, several questions become unavoidable. How can we bridge these two sources of data? How do we correctly interpret the experimental observable? How do we understand the broader relevance of each molecular feature? How can apparent slow timescales emerge from much faster microscopic timescales? How can we be more precise from an experimental perspective and more pragmatic from an atomistic perspective? How do we understand the microscopic steps underlying macroscopic data, and how do we understand the macroscopic consequences of microscopic steps?

These questions matter because there is, at present, only a limited understanding of how chemical modifications affect the outcome. We know how to make molecules, and we know what we want in the end (faster covalent bond formation, more specific noncovalent binding, etc.). What we do not know are the microscopic changes needed to achieve it. Even when strategizing in the broader sense, there is no consensus.³³ Many follow the paradigm that the best designs come from retrofitting a noncovalent scaffold for covalent inhibition, because they inherit the tried-and-true specificity of noncovalent recognition for some particular binding site. Others start by screening a library of reactive fragments instead, aiming to discover ligandable sites incompatible with existing scaffolds. Then there are those who claim that neither route is reliable and that success is largely luck. Sometimes, it can indeed appear that way. It is shown in Chapter 5 that a minor change to the molecular structure, as minor as the addition of a single methyl group, can perturb several microscopic steps in the binding process. When observing the process from a birds-eye-view, like with what is illustrated in Figure 1.1, these steps are completely convoluted.

1.3.3 A framework built on mass action

The central aim of this dissertation is to answer these questions by establishing a multi-scale framework for rationalizing covalent inhibition, a self-consistent account that connects atomistic detail to macroscopic observables. The connective tissue, in parts found in adjacent fields, resurrected from history, or developed here, is a strategy centered around kinetic modeling that can fit, simulate, and combine the timescales of arbitrary steps in the binding process, allowing one to walk between levels of detail and ask at which level a given understanding generalizes into a design principle.

First, the binding process is written out as a network of states, where each state is connected to at least one other through a defined transition. Each state comprises some type(s) of molecular species, like how “E” and “I” together compose the unbound state in

the simple kinetic Scheme 1.1. The system moves from one state to another with known (or parametric) stoichiometry and rate constants. A rate constant, k , is the average frequency with which particles in the source state transition to the target state. The inverse of a rate constant is the average waiting time between transitions, *i.e.*, the characteristic timescale of that process.

To understand how a bulk of particles flows through some kinetic scheme, we apply the law of mass-action. This describes how the concentration, or probability, of each species smoothly changes over time using a set of ordinary differential equations (ODEs). The set of ODEs for Scheme 1.1 is,

$$\frac{d[E]}{dt} = k_{\text{off}}[E \cdot I] - k_{\text{on}}[E][I] \quad (1.2)$$

$$\frac{d[I]}{dt} = k_{\text{off}}[E \cdot I] - k_{\text{on}}[E][I] \quad (1.3)$$

$$\frac{d[E \cdot I]}{dt} = k_{\text{on}}[E][I] - k_{\text{off}}[E \cdot I] \quad (1.4)$$

where $[\cdot]$ denotes the concentration of the contained species. Deterministic simulations of the concentrations are obtained by integrating these ODEs numerically. With rapid-equilibrium approximations, series combinations of timescales, and other tricks for compressing Markov models, or with simple empirical fits to simulated experimental data, we can move from resolved schemes informed by atomistic simulations or experimental indicators to the generic model (Scheme 1.2). We fit the resolved schemes to experiment, then coarse-grain the resolved parameters into more universal ones (*e.g.*, k_{inact} and K_I) that are used to benchmark binding quality and draw broad conclusions. By now it should be clear that such coarse-grained parameters hide many microscopic steps and, in that sense, are “apparent.” These kinetic simulations indicate what to analyze next in the atomistic dynamics and what kinetic regime a molecule occupies, and hence how it might be improved.

The reason we represent the binding kinetics with mass-action models rather than em-

pirical models like Kitz-Wilson is that they are flexible in both construction and application. One can add any amount of species and transitions to the network, as long as there is a scientific reason. When starting from a generic model, the existing species and transitions, coarse-grained as they are, can be split as reasonably desired into more specific partitions. This methodology also supports second-order reactions or any other forms of nonlinear kinetics. After simulating a time-course or several, it is almost always possible to process the output to match what is being detected in some binding experiment. In fact, some of the innovations within this dissertation were realized through stubborn attempts to relate idealized time-courses and experimental measurements.

These benefits alone solve most of the limitations associated with traditional analyses, but mass-action carries limitations of its own. Writing the master equation of a kinetic system as a set of ODEs assumes that there is a large, well-mixed copy number of particles and that the transitions are memoryless (*i.e.*, Markovian). The former is guaranteed by the trillions of molecules in a test tube, and the latter is justified because the underlying microscopic dynamics decorrelate far faster than the processes made explicit in the scheme. The genuine downsides are risk of overparameterization and loss of identifiability, both due to the fact that every transition introduces at least one parameter (the rate constant). This concern is mitigated by evaluating the proportionality of parameters (with, *e.g.*, $K_d = k_{\text{off}}/k_{\text{on}}$) or by Bayesian Metropolis Monte Carlo (BMMC) sampling of parameter space.

1.3.4 *Microscopic input from molecular dynamics*

This framework is incomplete without information of the atomistic environment. The computational method we use to obtain this information is called molecular dynamics (MD).³⁴ It is sometimes referred to as a “computational microscope,” since it provides a microscopic view with atomistic resolution across a handful of nanometers and nanoseconds. The atoms of a single protein and ligand typically begin at coordinates provided by a crystallographic

structure. The surrounding space is filled with water and salt to mimic basic cytosolic condition. The initial velocity of each atom is sampled from the Maxwell-Boltzmann distribution of the desired temperature. For each cycle of the MD algorithm, the net force on each atom is calculated, which dictates the acceleration of each atom, which dictates the movement of each atom over a billionth of a millionth of a second, which dictates the coordinates and velocities to use in the next cycle.

Many technical choices are made based on the size of the system to approximate the most realistic level of chemical theory with something cheaper to compute. These choices are the difference between a simulation taking days instead of decades. For covalent inhibitors, it is necessary to examine the dynamics under multiple different levels of theory. MD always propagates the nuclei classically via Newton’s second law, but the level of theory determines how the potential energy and forces are evaluated. Molecular mechanics (MM), an empirical energy function that does not permit bonds to break, is used to sample the unlinked and linked complexes extensively while bypassing the expensive quantum step. A semiempirical quantum-mechanical/molecular-mechanical (QM/MM) treatment is used to interrogate the covalent binding mechanism where bonds form and break. After billions of MD cycles, we collect the accumulated statistical information over an arbitrary force or geometrical feature, relate this to the free energy landscape, identify interesting metastable states, and calculate from the free energy landscape the average time it takes to move from one state to another. These states and rates are the details that turn a generic binding scheme into one tailored for some enzyme–inhibitor pair.

It may be interesting to note that the pragmatic model of irreversible binding kinetics found in Michaelis-Menten and Kitz-Wilson analyses was actually derived from mass-action-based rate laws³⁵ just over a century prior to this dissertation. A century full of statistical mechanical theory, the inventions behind today’s powerful computer hardware, and millions of lines of molecular simulation code have brought us here, leveraging mass-action far more

fully than before.

1.4 Overview of dissertation

The first study, Chapter 2, establishes the principle of rapid intermediates and the mass-action machinery used throughout. By expanding the conventional two-step scheme to distinguish reactive and nonreactive configurations of the noncovalent complex, it shows how a rapidly equilibrating intermediate can dominate the apparent inactivation rate even though it has nothing to do with the warhead chemistry itself. This simple abstraction of conformational gating demonstrates that rapid equilibria should be exploited for (i) drug optimization, where the apparent inactivation rate must be deconvolved into its chemical and conformational contributions to take full advantage of a safer warhead with low/moderate reactivity and to aim the optimization in a productive direction and (ii) parameter identification, since k_{inact} , k_{off} , and the conformational probabilities have distinct effects on dose-dependency. The expanded scheme is only one instance of a procedure applicable to any kinetic mechanism.

Chapter 3 demonstrates how to use MD simulations together with mass-action kinetic simulations to calculate macroscopic parameters from a microscopic binding pathway. We use the binding of the famous drug ibrutinib to BTK as an example. A decade after FDA approval, the mechanism has not been well-characterized. Our investigation results in a kinetic scheme that contains ten states in its most complex form. Integrating these microscopic steps into a kinetic simulation makes it possible to benchmark the collective kinetic parameters against experiment, to partition the contributions of competing pathways, and to quantify the effect of rapid conformational gating on the overall binding kinetics. Furthermore, we show how the kinetics of solvent-mediated reactions can be dependent on the $\text{p}K_{\text{a}}$ of the target residue and the pH of the solvent. Optimizing for these dependencies may prove useful when targeting biological environments with unusual pH values, such as tumors.

Chapter 4 advances the treatment of experimental data. My contribution lies in the kinetic modeling section, where we evaluate the binding of irreversible kinase probes to the kinase ABL1. A mass-action-based fitting protocol made it possible to model assay abnormalities, such as nonspecific binding and apparent sub-stoichiometry, and to fit the data losslessly, making for a much more efficient and insightful extraction of kinetic behavior. The result is a workflow that is more accurate, requires substantially less experimental data, and able to recover reversible affinity (K_d). The underlying program can be used to predict the experimental conditions needed to break any parameter degeneracy and determine potencies and rate constants poorly estimated in a previous round.

Chapter 5 turns to the microscopic origins of residence time and principles for reversible covalent inhibitors. Residence time is an attractive parameter for estimating *in vivo* efficacy.³² Apparent *in vivo* k_{on} is of little pharmacological importance, largely because target engagement is gated by slow absorption and distribution rather than by free diffusion in a closed volume, and a poor k_{on} can be offset by dose (although covalent inhibitor dose regimens should be less cavalier). Improving k_{on} with structure-activity-relationships is largely empirical and very complex. By contrast, k_{off} is independent of absorption, distribution, and local concentration, and is governed by the same interactions that determine thermodynamic affinity. Where irreversible covalent inhibitors bind their target until the target is recycled, a liability even for some on-targets, reversible covalent inhibitors offer a tunable compromise between prolonged residence time and safety. What sets residence time for reversible covalent inhibitors apart from noncovalent inhibitors is that it can comprise multiple kinetically relevant timescales that aren't necessarily the ones targeted in drug optimization, as is shown in this section. This chapter examines three ibrutinib-derived reversible BTK inhibitors from Bradshaw et al.,³⁶ whose residence times differ markedly (~ 0.5 , 91, and >200 hours) despite minor structural differences. We used a series of molecular modeling methods to understand the underlying workings of these drugs, which resulted in a partial

answer to why the residence times differ. Still, without the full molecular basis, we were able to use the kinetic fitting protocol established in previous chapters with the experimental data digitized from Bradshaw et al.³⁶ and deduce a broader conclusion: prolonged residence time is not the property of any single apparent rate constant. For the two well-performing inhibitors, it is an emergent consequence of repeated covalent reengagement, in which the reformation of the broken bond outpaces escape from the pocket. The methodology thereby identifies which kinetic steps are worth modulating and predicts how modulating them will affect the overall kinetic behavior.

These four studies assemble a framework in which atomistic simulations and macroscopic observables are coupled through mass-action kinetic modeling, allowing covalent inhibition to be understood, and ultimately designed, across the full range of timescales on which it operates. The concluding chapter, Chapter 6, lays out the implications of these studies and suggests where the framework goes next. The outlook includes a more detailed treatment of the covalent step, the use of continuous state variables, and the extension of the same reasoning beyond *in vitro* binding kinetics.

CHAPTER 2

KINETIC MODELING OF COVALENT INHIBITION: EFFECTS OF RAPIDLY FLUCTUATING INTERMEDIATE STATES

Reproduced with permission from Ghaby, K. & Roux, B. Kinetic Modeling of Covalent Inhibition: Effects of Rapidly Fluctuating Intermediate States. *Journal of Chemical Information and Modeling* **65**, 8730–8739 (2025).³⁷ Copyright 2025 American Chemical Society. The published version is available at <https://doi.org/10.1021/acs.jcim.5c00952>.

Abstract

There is increasing interest in the discovery of small-molecule inhibitors that form covalent bonds with their targets for therapeutic applications. Nevertheless, identifying clear rational design principles remains challenging because the action of these molecules cannot be understood as common noncovalent inhibition. Conventional kinetic models often reduce the binding of covalent inhibitors to a two-step irreversible process, overlooking rapid complex dynamics of the associated unlinked inhibitor before the formation of the covalent bond with its target. In the present analysis, we expand the intermediate state into two conformations: reactive (E·I) and nonreactive (E··I). To illustrate the consequences of such simplification, the expanded kinetic model can be reduced to an effective two-step scheme expressed in terms of the equilibrium probability of the unlinked inhibitor to form either conformation. A mass-action-based numerical workflow is implemented to simulate time-dependent kinetics, overcoming the common limitations of empirical models. The numerical workflow helps relate microscopic states observed in molecular dynamics simulations to macroscopic observables like EC₅₀ and the apparent rate of covalent inhibition, showing the impact of transient intermediates on dissociation rates and potency. The proposed framework refines the inter-

pretation of dose-response data, aiding medicinal chemists in optimizing covalent inhibitors and providing a quantitative platform that relates molecular conformational distributions to empirical parameters.

2.1 Introduction

For a long time, inhibitory small-molecules that form covalent bonds with their targets, *i.e.*, covalent inhibitors, were largely eschewed in drug development due to concerns over off-target effects, assay and delivery interference, and metabolic toxicity.^{10;38–44} With the improvements in drug design and optimization over the past couple decades, targeted covalent inhibition is becoming a promising and widely recognized strategy for exceeding the potency and selectivity of noncovalent inhibition.^{33;45–61} These enhancements should lead to improvements in the translation of *in vitro* to *in vivo* efficacy and safety profile.^{62;63}

Nevertheless, the rational design of covalent inhibitors remains extremely challenging. Today’s dominant strategy to design a covalent inhibitor is to start from a molecular scaffold that displays some reversible binding affinity to a protein of interest and attach to it a reactive functional group or “warhead” that targets a specific residue in the binding site.^{33;64–72} Conventional docking methods developed to deal with noncovalent protein-ligand interactions have been adapted to model the binding of a covalent ligand with the protein.^{28;73–78} It is commonly assumed that poses with the warhead in close proximity to the target residue can be used to predict a covalently-bound adduct, and computational workflows based on these ideas have been proposed.^{68–70;79–83} However, there is a wide range of issues. In particular, conventional docking often falls short of an accurate estimation of potency when either the ligand or protein target is flexible and can adopt multiple distinct conformations. It is also worth emphasizing that the mere presence of a reactive residue near a ligand binding site is not sufficient to guarantee a successful covalent strategy; the local chemical environment of the target residue must be carefully considered to determine the suitability of this ap-

proach.²⁹ Residues near the warhead may act as catalysts or steric obstacles at various steps along the reaction pathway. The situation is very different than for a reaction occurring in bulk solution. Dynamics before the formation of the covalent bond is also thought to play a key role.^{84;85} Traditional schematic descriptions of covalent inhibition typically ignore the impact of rapidly fluctuating transient intermediates on the overall potency of a covalent inhibitor. Despite their short lifetime, these may have a profound and complex impact on the apparent reaction rate and potency. Such issues may be overcome by incorporating the probabilities of these states in kinetic analyses.

The purpose of kinetic models, in the context of a drug development project, is to help draw accurate lessons from experimental observations in order to make judicious decisions to orient the design strategy in a productive direction. Due to the limited resolution, kinetic measurements of covalent inhibition have commonly been interpreted in terms of a simple two-step mechanism.^{13;86;87} The simple picture is that the inhibitor initially diffuses in the bulk solution to encounter the kinase enzyme and associate with the active site, followed by the irreversible inactivation of the enzyme due to the formation of the covalent adduct. A critical issue with such a minimal two-step model is that it may obscure important microscopic factors that are at play, which in turn may misdirect the design strategy. For example, kinetic data interpreted in light of the two-step model tend to ascribe the apparent rate of inactivation to the final chemical step that leads to the formation of the covalent linkage. Thus, in the case where a slow inactivation rate constant is caused by a low probability of a reaction-ready pose by the unlinked ligand occupying the binding site, the observation of an apparent slow rate might erroneously be attributed to poor reactivity of the warhead, thereby leading the design strategy in a wrong direction.

While the mathematical simplicity of a model is an important advantage to analyze experimental data, it should not be a determining factor. Computational methods for numerically predicting molecular behavior and solving complex kinetic schemes have grown

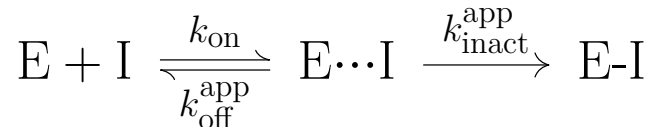
much more accessible and reliable. The precision of analysis depends on the precision of the kinetic scheme. Also of interest, the fluctuations predicted by classical molecular dynamics (MD) simulations based on atomic models can provide quantitative information that can be incorporated into a kinetic scheme to affect the observable covalent occupancy. A recent computational study of the irreversible covalent inhibitor of Bruton’s tyrosine kinase (BTK) ibrutinib illustrates these ideas.⁸⁴ Classical MD simulations of the unlinked compound associated with the kinase binding pocket showed that only a small fraction (about 3-5%) of the configurations were reaction-ready despite the fact that the inhibitor remains very close to the crystallographic binding pose. Quantitative agreement with experimental data was achieved only once this information was included in the overall kinetics with the result of hybrid quantum mechanics/molecular mechanics (QM/MM) calculation of the covalent reaction. The kinetic effects of rapidly fluctuating intermediate conformational states have been characterized and discussed in other fields⁸⁸⁻⁹⁰, but they have not been extensively considered in covalent inhibition.

Owing to their growing significance, it is important to develop more accurate and more realistic mathematical models of covalent inhibitors.⁹¹⁻⁹⁹ It is the central goal of the present work to help clarify the impact of rapid conformational fluctuations on the apparent potency of covalent inhibitors, which should hopefully lead to more effective design strategies. In the following, we first derive the kinetic framework that allows one to simulate covalent inhibition with and without rapid intermediates. Then, we compare the kinetic simulations of covalent and noncovalent schemes as a function of intermediate probabilities. Lastly, we investigate how dose-rate and dose-response data are affected by the intermediate probability distribution.

2.2 Theoretical framework

2.2.1 Kinetic schemes

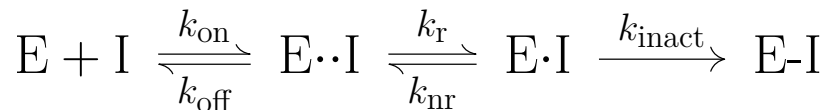
At the simplest level, the kinetics of irreversible covalent inhibition are often described as a two-step mechanism shown in Scheme 2.1,^{13;86;87}



Scheme 2.1: Generic two-step mechanism of irreversible kinetics.

where k_{on} is the bimolecular association rate, $k_{\text{off}}^{\text{app}}$ is the apparent dissociate rate, and $k_{\text{inact}}^{\text{app}}$ is the apparent inactivation rate constant. Initially, the inhibitor (I) diffuses in the bulk solution to encounter the kinase enzyme (E) and associate with the active site, resulting in the reversible noncovalent complex (E $\cdots$ I). The final step describes the inactivation of the enzyme as an irreversible process that results in the covalent adduct (E-I). Intermediates such as noncovalent fluctuations and the QM mechanisms of bond formation are combined in E $\cdots$ I and accounted for in $k_{\text{off}}^{\text{app}}$ and $k_{\text{inact}}^{\text{app}}$. We explicitly attach the label ‘app’ to these rates because they implicitly subsume additional processes, as will be clarified in the following.

Scheme 2.2 splits the state E $\cdots$ I into nonreactive (E $\cdot\cdot$ I) and reactive (E $\cdot$ I) intermediate states, where the reactive intermediate state meets some distance and orientation requirements necessary for the progress of the covalent reaction between the warhead and the target residue. For the sake of simplicity, it is assumed that only the nonreactive state can dissociate.



Scheme 2.2: Mechanism where the unlinked state is partitioned into nonreactive and reactive substates.

The equilibrium probability of the unlinked inhibitor to form the either reactive (p_{r}) or

nonreactive complex (p_{nr}) can be defined from the species concentrations as,

$$p_{\text{r}} \equiv \frac{[\text{E}\cdot\text{I}]}{[\text{E}\cdot\text{I}] + [\text{E}\cdots\text{I}]} \quad (2.1)$$

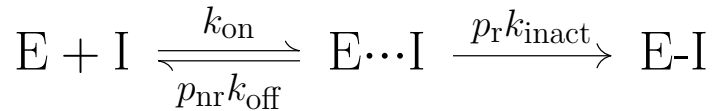
$$p_{\text{nr}} \equiv \frac{[\text{E}\cdots\text{I}]}{[\text{E}\cdot\text{I}] + [\text{E}\cdots\text{I}]} = 1 - p_{\text{r}} \quad (2.2)$$

Classical MD simulations indicate that the fluctuations of the unlinked inhibitor between the reactive and nonreactive pose may take place on a timescale of microseconds.⁸⁴ This is significantly faster than the other transitions, whose rate constants are typically on the order of minutes or even hours. As a rule of thumb, the dynamics of the intermediate states can be considered fast if the transition rate constants k_{r} and k_{nr} are larger than the slowest rate constant (commonly k_{inact}) by at least 1-2 order of magnitude. Since these states are rapidly equilibrated relative to the rest of the scheme,

$$p_{\text{r}} \approx \frac{k_{\text{r}}}{k_{\text{r}} + k_{\text{nr}}} \quad (2.3)$$

$$p_{\text{nr}} \approx \frac{k_{\text{nr}}}{k_{\text{r}} + k_{\text{nr}}} \quad (2.4)$$

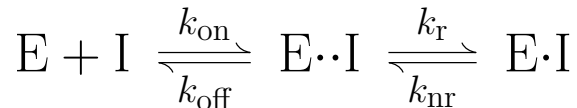
and Scheme 2.2 can be reduced and made equivalent to a two-step system, as shown in Scheme 2.3,



Scheme 2.3: Split-intermediate mechanism reduced to two steps.

where the apparent dissociation rate constant ($k_{\text{off}}^{\text{app}}$) equals $p_{\text{nr}}k_{\text{off}}$, and the apparent inactivation rate constant ($k_{\text{inact}}^{\text{app}}$) equals $p_{\text{r}}k_{\text{inact}}$. This equivalence was verified with numerical simulations, *i.e.*, for the cases examined in the present study, the results obtained by solving the ordinary differential equations (ODEs) numerically for Scheme 2.2 or Scheme 2.3 were identical.

To highlight the effects of p_r on the residence time (τ) and occupancy of a bound ligand in the absence of the irreversible covalent step from Scheme 2.2, it is of interest to consider a non-covalent version as Scheme 2.4.



Scheme 2.4: Noncovalent mechanism that includes a rapid equilibrium between two bound states.

The residence time derived from Scheme 2.4 can be expressed by substituting $k_{\text{off}}^{\text{app}}$ for $k_{\text{off}}(1 - p_r)$,

$$\tau = \frac{1}{k_{\text{off}}(1 - p_r)} \quad (2.5)$$

The validity of this substitution depends on the rapid-equilibrium of the bound conformations to keep p_r constant over time. Note that dissociation is defined only to occur from the nonreactive pose. The sensitivity of τ with respect to p_r is,

$$\frac{\partial \tau}{\partial p_r} = \frac{1}{k_{\text{off}}(1 - p_r)^2} \quad (2.6)$$

When p_r is small, the sensitivity is approximately $1/k_{\text{off}}$. Therefore, engineering warheads to nudge p_r from near-zero will only have small linear effects on τ proportional to k_{off} .

2.2.2 Simulation methodology

The time course of species concentrations in a well-defined (the rate constants and stoichiometry are known or parametric for each transition), Markovian kinetic scheme can be solved deterministically by integrating a set of ODEs for concentration with respect to time.³² These ODEs are derived from the law of mass action and formulated based on the stoichiometric coefficients and rate constants of each transition in the scheme. This fast, flexible, and straightforward framework adequately models the *in vitro* kinetics¹⁰⁰ of covalent inhibition.

Let $C_S(t)$ represent the concentration of species S at time t . The time derivative dC_S/dt

is given by the sum of contributions from all transitions involving species S ,

$$\frac{dC_S}{dt} = \sum_T \left(\Delta\nu_S k_T \prod_{S' \in \text{Src}(T)} C_{S'}^{\nu_{S', \text{Src}(T)}} \right) \quad (2.7)$$

where k_T is the rate constant of transition T , and $\Delta\nu_S$ is the difference between the stoichiometric coefficients of species S as a target and as a source in transition T (*i.e.*, $\nu_{S, \text{Tgt}(T)} - \nu_{S, \text{Src}(T)}$).

A system of these rate laws with any degree of (non)linearity can be expressed and solved in matrix notation as,

$$\frac{d\vec{C}}{dt} = \vec{C}^{\text{N}_{\text{Src}}} \cdot \text{diag}(\vec{k}) \cdot \mathbf{N} \quad (2.8)$$

The complete stoichiometry matrix, $\mathbf{N}$, is the difference between the target and source stoichiometry matrices (*i.e.*, $\mathbf{N} = \mathbf{N}_{\text{Tgt}} - \mathbf{N}_{\text{Src}}$). The stoichiometric matrices have m rows corresponding to the species and n columns corresponding to the transitions. $\text{diag}(\vec{k})$ is the diagonal matrix constructed from the vector of transition rate constants. The adjusted concentration vector, $\vec{C}^{\text{N}_{\text{Src}}}$, is the column-wise product of the element-wise exponentiation of the concentration vector with the source stoichiometry matrix, *i.e.*,

$$\vec{C}^{\text{N}_{\text{Src}}} \equiv \prod_n \left[\vec{C}^{\text{N}_{\text{Src}}} \right]_{m,n} \quad (2.9)$$

eq 2.8 is numerically integrated over time with the `scipy` implementation of LSODA.^{101–103} This vectorized rate function together with the rest of the infrastructure for modeling the schemes exists in GeKiM, an in-house and open-source Python package available at <https://github.com/GeKiM/GeKiM>:

//github.com/kghaby/GeKiM. For example, the system of ODEs for Scheme 2.1 is,

$$\begin{aligned}
d[E]/dt &= -[E][I]k_{\text{on}} + [E\cdots I]k_{\text{off}}^{\text{app}} \\
d[I]/dt &= -[E][I]k_{\text{on}} + [E\cdots I]k_{\text{off}}^{\text{app}} \\
d[E\cdots I]/dt &= [E][I]k_{\text{on}} - [E\cdots I]k_{\text{off}}^{\text{app}} - [E\cdots I]k_{\text{inact}}^{\text{app}} \\
d[E-I]/dt &= [E\cdots I]k_{\text{inact}}^{\text{app}}
\end{aligned} \tag{2.10}$$

The combined stoichiometry matrix ($\mathbf{N} = \mathbf{N}_{\text{Tgt}} - \mathbf{N}_{\text{Src}}$) is,

$$\mathbf{N} = \begin{array}{c|ccc} & k_{\text{on}} & k_{\text{off}}^{\text{app}} & k_{\text{inact}}^{\text{app}} \\ \hline [E] & -1 & 1 & 0 \\ [I] & -1 & 1 & 0 \\ [E\cdots I] & 1 & -1 & -1 \\ [E-I] & 0 & 0 & 1 \end{array} \tag{2.11}$$

The diagonal matrix of the rate constants is,

$$\text{diag}(\vec{k}) = \begin{bmatrix} k_{\text{on}} & 0 & 0 \\ 0 & k_{\text{off}}^{\text{app}} & 0 \\ 0 & 0 & k_{\text{inact}}^{\text{app}} \end{bmatrix} \tag{2.12}$$

The adjusted concentration vector is,

$$\vec{C}_{\mathbf{N}_{\text{src}}} = \begin{bmatrix} [E]^1 \times [I]^1 \times [E\cdots I]^0 \times [E-I]^0 \\ [E]^0 \times [I]^0 \times [E\cdots I]^1 \times [E-I]^0 \\ [E]^0 \times [I]^0 \times [E\cdots I]^1 \times [E-I]^0 \end{bmatrix} = \begin{bmatrix} [E] \times [I] \\ [E\cdots I] \\ [E\cdots I] \end{bmatrix} \tag{2.13}$$

The dot product of the matrices produces a vectorized version of the system of ODEs,

$$\mathbf{N} \cdot \text{diag}(\vec{k}) \cdot \vec{C}_{\mathbf{N}_{\text{src}}} = \begin{bmatrix} d[\text{E}]/dt \\ d[\text{I}]/dt \\ d[\text{E}\cdots\text{I}]/dt \\ d[\text{E-I}]/dt \end{bmatrix} = \begin{bmatrix} -[\text{E}][\text{I}]k_{\text{on}} + [\text{E}\cdots\text{I}]k_{\text{off}}^{\text{app}} \\ -[\text{E}][\text{I}]k_{\text{on}} + [\text{E}\cdots\text{I}]k_{\text{off}}^{\text{app}} \\ [\text{E}][\text{I}]k_{\text{on}} - [\text{E}\cdots\text{I}]k_{\text{off}}^{\text{app}} - [\text{E}\cdots\text{I}]k_{\text{inact}}^{\text{app}} \\ [\text{E}\cdots\text{I}]k_{\text{inact}}^{\text{app}} \end{bmatrix} \quad (2.14)$$

A Jupyter notebook to reproduce the simulations and plots herein is available at <https://github.com/RouxLab/covalent-kinetics-with-rapid-intermediates-2025>. All data is deterministically generated therein. The numerical values of the rate constants used in the simulations were chosen to reflect the regime commonly observed for covalent inhibitors, where $k_{\text{inact}} \ll k_{\text{off}} < k_{\text{on}}[\text{I}]_0$.¹⁰⁴

2.3 Results

2.3.1 Noncovalent vs covalent inhibition

Time courses of the total occupied populations of Scheme 2.2 and its noncovalent version Scheme 2.4 are compared in Figure 2.1 for various probabilities of a reactive conformation, p_{r} . The purpose of Scheme 2.4 is to highlight the effects of p_{r} in the absence of the irreversible covalent step from Scheme 2.2. p_{r} has a similar impact on the occupancy of both of the schemes. The noncovalent time courses demonstrate how p_{r} can compensate for poor affinity and limit maximal occupancy. The irreversible covalent time courses demonstrate the dependence of covalent action rate on p_{r} . Reversible covalent inhibition falls somewhere in between, depending on the ratio of covalent forward and backward rate constants.

As shown in Figure 2.2, the propensity to adopt a conformation in which the ligand cannot dissociate significantly increases the lifetime of binding despite the short lifespan of the conformations themselves.

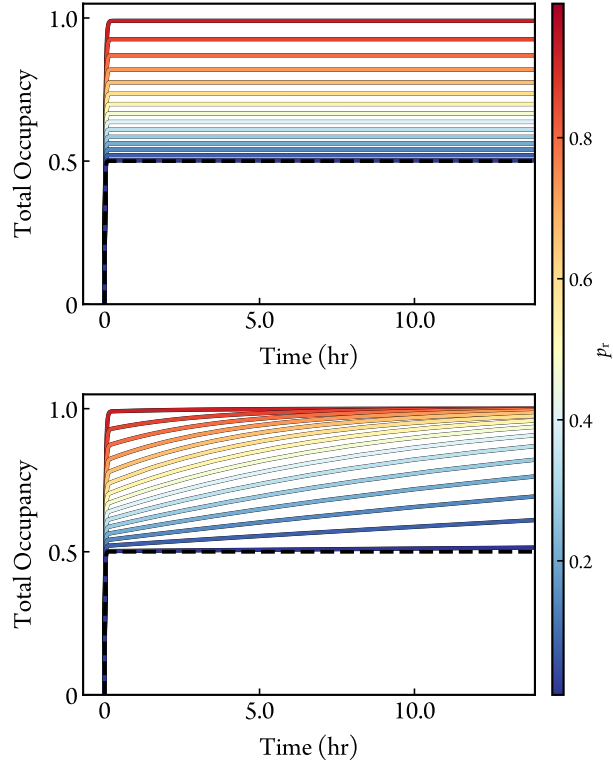


Figure 2.1: Simulated time courses of the total occupancy from noncovalent Scheme 2.4 (top) and covalent Scheme 2.2 (bottom) for different values of p_r . $[E]_0 = 1$ nM, $[I]_0 = 1$ μ M, $k_{\text{on}} = 1 \times 10^{-2}$ $\mu\text{M}^{-1}\text{s}^{-1}$, $k_{\text{off}} = 1 \times 10^{-2}$ s^{-1} , and $k_{\text{inact}} = 1 \times 10^{-4}$ s^{-1} . The occupancy of a one-step scheme is also included (black-dashed) to illustrate the contribution of binding affinity, $k_{\text{off}}/k_{\text{on}}$.

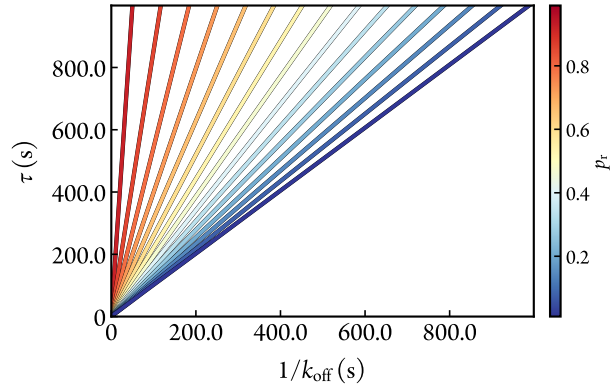


Figure 2.2: Noncovalent residence time (τ) of a bound ligand as a function of k_{off} and p_r . $[E]_0 = 1$ nM, $[I]_0 = 1$ μ M, and $k_{\text{on}} = 1 \times 10^{-2}$ $\mu\text{M}^{-1}\text{s}^{-1}$. The residence times calculated analytically from eq 2.5 or numerically from simulations of Scheme 2.4 are the same.

2.3.2 Effects on observable outcomes

Dose-rate analysis

The kinetics of covalent inhibitors are often studied in a pseudo-first order regime in the spirit of Michaelis-Menten kinetics,^{31;35} with an excess of either inhibitor or enzyme. For the sake of convention, we will treat the inhibitor as the saturating species, so the bimolecular step can be linearly approximated as shown in Scheme 2.5.



Scheme 2.5: Generic two-step mechanism with a pseudo-first order approximation.

In conventional experimental procedures of scoring an irreversible inhibitor’s on-target behavior,⁸⁷ the observed rate constant of inhibition, k_{obs} , is first obtained by fitting the fractional population of the covalently-bound state (*i.e.*, covalent occupancy, CO) to the pseudo-first order growth factor,

$$\begin{aligned} \text{Occ}_{\text{cov}}(t) &\equiv \frac{[\text{E-I}]}{[\text{E}]_0} \\ &\approx 1 - e^{-k_{\text{obs}}t} \end{aligned} \tag{2.15}$$

at various inhibitor concentrations. The variation of k_{obs} as a function of initial inhibitor concentration, $[\text{I}]_0$, is then fitted to the Michaelis-Menten analog,

$$k_{\text{obs}} = \frac{k_{\text{inact}}^{\text{app}} [\text{I}]_0}{K_{\text{I}}^{\text{app}} + [\text{I}]_0} \tag{2.16}$$

The apparent rate constant of inactivation, $k_{\text{inact}}^{\text{app}}$, is the maximal observed rate constant that occurs when the entire free enzyme population rapidly transitions into the reversible complex state. This limit is approached when $k_{\text{on}}[\text{I}]_0 \gg k_{\text{off}}^{\text{app}}$. The apparent inactivation

constant, K_I^{app} , is the concentration of the inhibitor that results in a half-maximal observed rate constant, and it can be expressed as,

$$K_I^{\text{app}} = \frac{k_{\text{off}}^{\text{app}} + k_{\text{inact}}^{\text{app}}}{k_{\text{on}}} \quad (2.17)$$

A reaction may appear one-step in both non-saturating ($[I]_0 < K_I$) and saturating ($[I]_0 \gg K_I$) conditions. These one-step approximations are distinguishable by the y -intercept, where an intercept of 0 corresponds to a slope of k_{obs} over $[I]_0$ equivalent to the inactivation efficiency (*i.e.*, the potency of the irreversible inhibitor).¹⁰⁵ For the three-step paradigm (Scheme 2.2),

$$\begin{aligned} \text{Eff}_{\text{inact}} &= \frac{k_{\text{inact}}^{\text{app}}}{K_I^{\text{app}}} \\ &= \frac{p_r k_{\text{inact}}}{(p_{\text{nr}} k_{\text{off}} + p_r k_{\text{inact}}) / k_{\text{on}}} \end{aligned} \quad (2.18)$$

The probability distribution of the intermediates has a significant role in the potency of inhibition despite the non-limiting magnitude of the arming and disarming rate constants. From an empirical perspective, the probability distribution between conformations limits inhibition by linearly scaling the maximal observed rate constant. K_I^{app} also depends linearly on p_r , with,

$$\lim_{p_r \rightarrow 0} K_I^{\text{app}} = \frac{k_{\text{off}}}{k_{\text{on}}} = K_d \quad (2.19)$$

and

$$\lim_{p_r \rightarrow 1} K_I^{\text{app}} = \frac{k_{\text{inact}}}{k_{\text{on}}} \quad (2.20)$$

Rearranging eq 2.20 for k_{on} shows that $\text{Eff}_{\text{inact}}$ approaches k_{on} as p_r approaches 1; a reminder that simplifying eq 2.18 at $k_{\text{off}}^{\text{app}} = 0$ yields k_{on} . The dramatic relationship between simulated empirical data from the numerical solution to the ODEs of Scheme 2.3 and p_r is illustrated

in Figure 2.3.

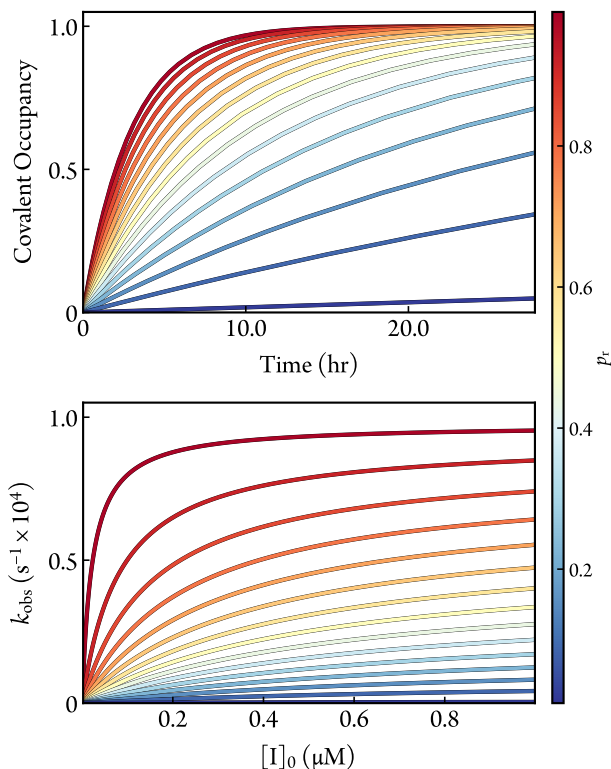


Figure 2.3: Simulated time-responses of covalent occupancy ($[I]_0 = 1 \mu\text{M}$) (top) and dose-rate (bottom) data for different values of p_r with $[E]_0 = 1 \text{ nM}$, $k_{\text{on}} = 1 \times 10^{-2} \mu\text{M}^{-1}\text{s}^{-1}$, $k_{\text{off}} = 1 \times 10^{-2} \text{s}^{-1}$, and $k_{\text{inact}} = 1 \times 10^{-4} \text{s}^{-1}$. The ODEs of Scheme 2.3 were solved numerically for the time-responses (top), and the fraction of the covalent complex was fit for k_{obs} with eq 2.15 (bottom).

Rapid saturation of the reversibly-bound state, when $k_{\text{on}}[I]_0 \gg k_{\text{off}}^{\text{app}}$, occurs (as a rule of thumb) when $[I]_0 > k_{\text{off}}^{\text{app}}/k_{\text{on}} \times 10$. With the parameters used in Figure 2.3, this threshold is 100 nM when $p_r = 0.99$ and 9.9 μM when $p_r = 0.01$.

Dose-response analysis

The dose-dependence of irreversible kinetics can alternatively be examined with a dose-response curve. Unlike reversible kinetics, the dose-response must be examined under non-equilibrium conditions where the reaction is stopped before completion. For simplicity, we will only consider noncompetitive experimental procedures where some proportional marker of covalent occupancy (*i.e.*, response), would be measured at the same timepoint for an array

of saturating ligand concentrations. While the numerical simulations of the ODEs provide the most faithful representation of a given kinetic scheme, the familiar Hill equation,¹⁰⁶

$$\text{Response} ([I]_0) = \lambda_{\text{DR}} \frac{1}{1 + (\text{EC}_{50}/[I]_0)^n} \quad (2.21)$$

can serve as a useful shorthand analytical framework for describing the dose-dependence data because it produces a simple sigmoidal curve with clearly identified parameters. In eq 2.21, λ_{DR} represents a scaling factor to match the arbitrary amplitude of the data, EC_{50} represents the dose that yields half the maximal scaled response, and n represents the Hill coefficient.

The Hill equation works exceedingly well for reversible one-step systems measured at equilibrium, where the EC_{50} approximately equals the apparent binding affinity, $K_{\text{d}}^{\text{app}} \approx k_{\text{off}}^{\text{app}}/k_{\text{on}}$ (*i.e.*, the apparent inhibition constant, $K_{\text{i}}^{\text{app}}$). However, special attention is required in the case of a covalent inhibitor because the covalent occupancy becomes truly stable only when the entire enzyme population has been irreversibly occupied. Thus, the value of EC_{50} is time-dependent and its kinetic significance is ambiguous.^{32;92} In the following, we derive a Hill-like model in terms of $k_{\text{inact}}^{\text{app}}$ and $K_{\text{I}}^{\text{app}}$ for the purpose of describing dose-dependence data for a covalent inhibitor.

First, the expression must be properly normalized to covalent occupancy. This can be done with the exponential growth factor for approximating a time-response curve of covalent occupancy (eq 2.15). Recall the empirical definition of $k_{\text{inact}}^{\text{app}}$: the maximal observed rate constant that occurs when $k_{\text{on}}[I]$ is large enough to instantly and steadily force the entire (remaining) protein population into the reversibly-bound state. The Hill equation is then scaled by the maximal covalent occupancy achieved within the dose-response time limit (t_{DR}) to get,

$$\text{Occ}_{\text{cov}} ([I]_0) = \frac{1 - e^{-k_{\text{inact}}^{\text{app}} t_{\text{DR}}}}{1 + (\text{EC}_{50}/[I]_0)^n} \quad (2.22)$$

Thus, the amplitude of eq 2.22, *i.e.*, the limit of covalent occupancy given a set time and an infinitely high dose, is solely dependent on $k_{\text{inact}}^{\text{app}}$ (or by extension k_{inact} and p_r).

To characterize the impact of the kinetic parameters on the location of the inflection point of eq 2.22, *i.e.*, the dose required for half-maximal covalent occupancy within the time limit, we solve for the isosurface of EC₅₀ in terms of $k_{\text{inact}}^{\text{app}}$ and $K_{\text{I}}^{\text{app}}$. We start by solving for k_{obs} from the concentration at half maximum inhibition, defining $[I]_0 = \text{EC}_{50}$. Using the definition of the EC₅₀, a relationship of familiar first-order growth factors can be expressed at this condition,

$$1 - e^{-k_{\text{obs}} t_{\text{DR}}} = \frac{1}{2} \left(1 - e^{-k_{\text{inact}}^{\text{app}} t_{\text{DR}}} \right) \quad (2.23)$$

We then solve for k_{obs} ,

$$k_{\text{obs}} = \frac{-\eta \text{EC}_{50}}{t_{\text{DR}}} \quad (2.24)$$

where

$$\eta \text{EC}_{50} = \ln \left(\frac{1 + e^{-k_{\text{inact}}^{\text{app}} t_{\text{DR}}}}{2} \right) \quad (2.25)$$

Then, we substitute k_{obs} with the Michaelis-Menten analog (eq 2.16) when $[I]_0 = \text{EC}_{50}$ to get the approximation,

$$\frac{k_{\text{inact}}^{\text{app}} \text{EC}_{50}}{K_{\text{I}}^{\text{app}} + \text{EC}_{50}} \approx \frac{-\eta \text{EC}_{50}}{t_{\text{DR}}} \quad (2.26)$$

Finally, we solve for the EC₅₀,

$$\text{EC}_{50} \approx \frac{-K_{\text{I}}^{\text{app}} \eta \text{EC}_{50}}{k_{\text{inact}}^{\text{app}} t_{\text{DR}} + \eta \text{EC}_{50}} \quad (2.27)$$

As is expected from reversible one-step kinetics, the EC_{50} approaches the reversible binding affinity, K_d , as k_{inact}^{app} approaches 0. This expression of the EC_{50} is substituted into eq 2.22 to guide the understanding of dose-dependence with respect to the microscopic kinetic parameters of irreversible binding.

Still, this expression is an approximation of the kinetics. Covalent occupancy calculated from eq 2.22 and ODE-based simulations of Scheme 2.3 is compared in Figure 2.4 for three demonstrative cases. The same rate constants are used in both solutions for each case, and

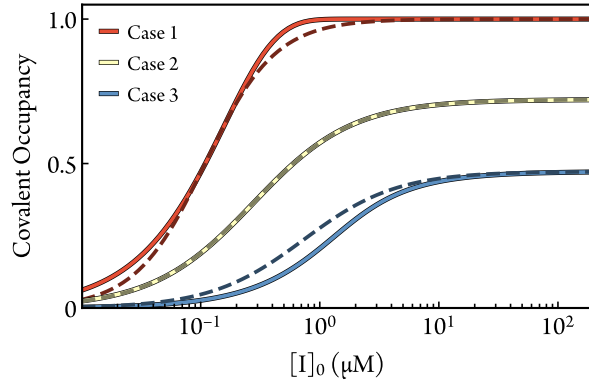


Figure 2.4: Comparisons between eq 2.22 (darkened dashed line) and ODE-based numerical simulations of Scheme 2.3 (solid line) for three representative cases. In Case 2, the parameters are $t_{DR} = 7.1$ h, $[E]_0 = 1$ nM, $k_{on} = 1 \times 10^{-2} \mu M^{-1} s^{-1}$, $k_{off}^{app} = 5 \times 10^{-3} s^{-1}$, and $k_{inact}^{app} = 5 \times 10^{-5} s^{-1}$. The altered parameters for Case 1 are $k_{on} = 5 \times 10^{-4} \mu M^{-1} s^{-1}$ and $k_{inact}^{app} = 5 \times 10^{-3} s^{-1}$. The altered parameters for Case 3 are $[E]_0 = 1 \mu M$, $k_{on} = 5 \times 10^{-4} \mu M^{-1} s^{-1}$, and $k_{inact}^{app} = 5 \times 10^{-3} s^{-1}$. The fitted values of n for cases 1, 2, and 3 are 1.47, 1.03, and 1.11, respectively.

the value of n in eq 2.22 is fit to the simulated covalent occupancy. Eq 2.22 deviates from the numerical covalent occupancy in Cases 1 and 3. Case 1 represents the regime where $k_{on}[I]_0 < k_{inact}^{app}$. In this regime, a single dose-response is likely not enough to resolve these rate constants anyway. Case 3 includes nonlinear effects from the bimolecular association step. The nonlinearity violates the first-order approximation used to derive eq 2.27, and therefore leads to a poor analytical prediction of the EC_{50} .

According to the present analysis, the EC_{50} depends both on k_{inact}^{app} and K_I^{app} . A smaller EC_{50} can either be attributed to a smaller K_d (or faster k_{on} for systems lacking reversible equilibrium), larger p_r , or faster k_{inact} . Therefore, the EC_{50} by itself is a vague and possibly

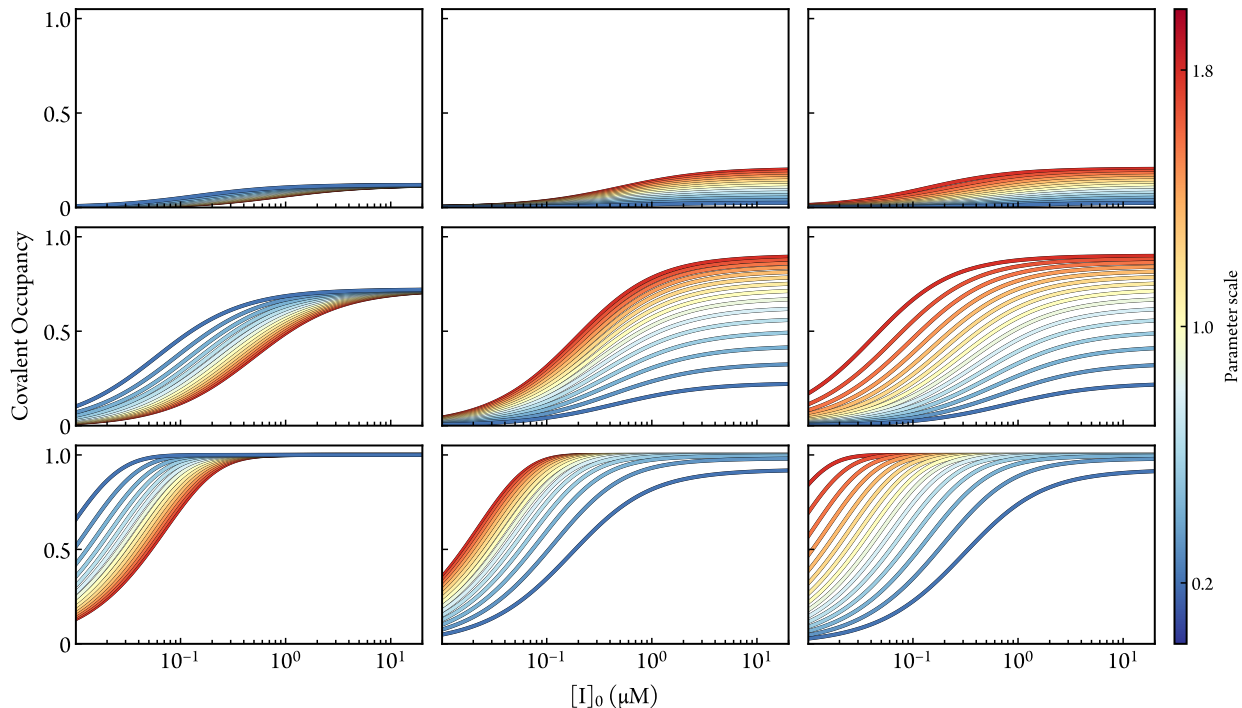


Figure 2.5: The variation of dose-response data (simulated by numerically solving the ODEs of Scheme 2.3) as a function of k_{off} (left column), k_{inact} (middle column), and p_r (right column) when t_{DR} is 0.7 (top row), 7.1 (middle row), and 71.1 (bottom row) hours. The default parameters are $[E]_0 = 1 \text{ nM}$, $k_{\text{on}} = 1 \times 10^{-2} \mu\text{M}^{-1}\text{s}^{-1}$, $k_{\text{off}} = 1 \times 10^{-2} \text{ s}^{-1}$, $k_{\text{inact}} = 1 \times 10^{-4} \text{ s}^{-1}$, and $p_r = 0.5$.

misleading parameter for evaluating irreversible systems. This mathematical analysis is illustrated with numerical kinetic simulations in Figure 2.5.

The amplitude and inflection point of eq 2.22 taken together reveal the unique roles of K_d , p_r , and k_{inact} . As shown in Figure 2.5, proportional variations of these simulation parameters distinctly affect the dose-response curves. The lack of complete covariance for these parameters suggests that fitting dose-dependent data with simulations of the models is a viable strategy for empirical approximation of K_d , p_r , and k_{inact} . If a set of dose-response curves (for different, *e.g.*, experimental conditions or inhibitor modifications) is used in the fitting process, then one may discern the conditional impact on each parameter.

The dose-response simulations also demonstrate the importance of a well-chosen time-point. If t_{DR} is too short, as is shown in Figure 2.5 (top row), then the analysis of covalent occupancy is more obscured by absolute sources of error. If t_{DR} is too long and the reaction

reaches 100% covalent occupancy, as is shown in Figure 2.5 (bottom row), then $k_{\text{inact}}^{\text{app}}$ and k_{off} cannot be resolved in the fitting process.

By separating the contribution of rapid intermediates (*e.g.*, conformations of the warhead or target residue, translation of the scaffold, protonation states) from the dissociation and inactivation steps, the kinetic comparisons between drug candidates become more precise, which should make it more straightforward to associate molecular characteristics with more potent and selective inhibition. This is especially beneficial when observed rates from reactivity or non-specific binding assays do not correlate with observed rates of inhibition or when the covalent action displays nonlinear complexities (as is the case for lysine-binders).^{84;107}

While eq 2.22 is a useful approximation expressing the relationships between covalent occupancy and kinetic parameters, *in vitro* kinetics are more accurately represented with ODE-based simulations. This avoids the vague nature of the Hill coefficient n for fitting irreversible kinetics since noncooperative causes of a fitted $n > 1$ (*e.g.*, ultrasensitivity and significant depletion of [I]) are unlikely to symmetrically affect the dose-response around the midpoint in the manner n is designed to modulate.^{108;109} ODE-based simulations also sidestep the assumptions and limitations of a pseudo-first order model such as the exponential growth factor,

$$\text{Occ}_{\text{cov}} ([I]_0) \approx 1 - e^{-k_{\text{obs}}([I]_0)t_{\text{DR}}} \quad (2.28)$$

where $k_{\text{obs}}([I]_0)$ uses eq 2.16. This extends to the dose-rate analysis using a time-dependent growth factor (eq 2.15) and eq 2.16.

If one globally fits parameterized simulations to multiple time-responses across varying doses—data required for dose-rate analysis anyway—then the fitting quality is no longer dependent on a single, well-chosen timepoint. The dose-dependence is included without the loss of information that occurs when an entire time course is compressed into a single value of k_{obs} . Since ODE-based simulations accommodate low doses without concern for the validity

of a first-order approximation, this strategy may enable a more accurate and practical (especially if k_{obs} would only plateau at unreasonably high doses) estimation of $k_{\text{inact}}^{\text{app}}$ and $K_{\text{I}}^{\text{app}}$ compared to the dose-rate workflow. The estimation of inactivation efficiency is also likely to be more accurate, because the initial slope of $k_{\text{obs}}([I]_0)$ is susceptible to the nonlinearity of the bimolecular association step that emerges at low doses.

Admittedly, redundancy in these parameters and the danger of over-fitting are important challenges. Within experimental error, some combination of k_{inact} and K_{d} may have a p_{r} -like effect on dose-response data. Importantly, $k_{\text{inact}}^{\text{app}}$ and $K_{\text{d}}^{\text{app}}$ can be deconvoluted from this type of data with high reliability. Using values of p_{r} predicted from MD simulations, the k_{inact} and K_{d} can then be derived from their apparent counterparts. Alternatively, one may exploit information from various sources, including classical MD simulations, to establish reasonable bounds on various parameters which could then be used as priors in a Bayesian Metropolis Monte Carlo protocol.^{110;111}

In ending this discussion, it is important to recall that the present analysis is meant to serve mainly as a conceptual illustration of how microscopic factors, that are not immediately related to the warhead’s chemistry, can impact the apparent observables such as the inactivation rate and residence time. Clearly, even with the expanded associated-unlinked intermediate state, Scheme 2.2 remains fairly simple and is unable to account for all the complex microscopic aspects of covalent inhibition encountered with different warheads. Expanding this type of analysis to realistic situations requires detailed atomic studies of the inhibition mechanism for specific cases. As an example, we previously represented the overall time-course of inhibition of BTK by ibrutinib, which targets a conserved cysteine residue, using an elaborate multi-state kinetic scheme based on the wealth of information extracted from QM/MM simulations.⁸⁴ The mechanisms and reaction pathways of other cysteine-targeting covalent inhibitors could be examined using a similar strategy.^{36;112} Beyond cysteines, an understanding of covalent inhibition in the case of lysine-targeting warheads¹¹³ or other

chemistries^{58;59} will require detailed atomic studies.

2.4 Conclusion

A critical issue with the minimal two-step kinetic model, which represents the process of covalent inhibition as an association followed by formation of a chemical bond, is that it may obscure important microscopic factors that affect the apparent efficiency of lead compounds. The propensity of the warhead and the target residue to adopt a reaction-ready configuration affects the apparent rate of inactivation but is not, in fact, specifically related to the warhead itself. The lack of clarity on this issue and misinterpretation of experimental kinetic data can, in turn, misdirect the design strategy and pull a project away from a productive direction. To make judicious design decisions, it is critical to draw accurate lessons from experimental data. To this end, we explored a more realistic kinetic scheme, which can be used as a framework to incorporate information about the fluctuations of the associated-unlinked inhibitor generated from classical MD simulations. Analysis based on numerical simulations showed that rapidly-equilibrating intermediates taking place on a short timescale can dramatically affect the observed inhibition rate via the relative populations of a reactive vs. nonreactive configurations. Explicitly including the probability distribution of these states also refines the values of dissociation and inactivation rate constants, informing for example the amount of inactivation efficiency that can be attributed to only the warhead. Optimizing drug candidates with knowledge of the factors associated to the rapid fluctuations of the associated-unlinked inhibitor may offer a promising route for improving the apparent affinity or apparent inactivation rate, while ignoring those could lead one to erroneously attribute a poor apparent inactivation efficiency to the warhead chemistry itself. Refining the evaluation with computational prediction of noncovalent behavior can provide a more precise evaluation of a drug’s potential and streamline drug optimization. Further work to develop kinetic frameworks incorporating microscopic states gleaned from classical MD simulations may be

able to provide a more precise understanding of the molecular characteristics underlying apparent kinetic parameters.

CHAPTER 3

COMPUTATIONAL INVESTIGATION OF THE COVALENT INHIBITION MECHANISM OF BRUTON’S TYROSINE KINASE BY IBRUTINIB

Reproduced with permission from Barragán, A. M.,[†] Ghaby, K.,[†] Pond, M. P. & Roux, B. Computational Investigation of the Covalent Inhibition Mechanism of Bruton’s Tyrosine Kinase by Ibrutinib. *Journal of Chemical Information and Modeling* **64**, 3488–3502 (2024).⁸⁴ Copyright 2024 American Chemical Society. The published version is available at <https://doi.org/10.1021/acs.jcim.4c00023>. [†]These authors contributed equally. My main contribution was the kinetic modeling.

Abstract

Covalent inhibitors represent a promising class of therapeutic compounds. Nonetheless, rationally designing covalent inhibitors remains extremely challenging. To better understand the covalent binding mechanism, these issues are examined with computational methods using the irreversible covalent inhibitor of Bruton tyrosine kinase (BTK) ibrutinib as an example. A multi- μ s classical molecular dynamics trajectory of the unlinked inhibitor is generated to explore the fluctuations of the compound associated with the kinase binding pocket. Then, the reaction pathway leading to the formation of the covalent bond with the cysteine residue at position 481 via a Michael addition is determined using the string method in collective variables on the basis of hybrid quantum mechanical - molecular mechanical (QM/MM) simulations. The reaction pathway shows a strong correlation between the covalent bond formation and the protonation/deprotonation events taking place sequentially in the covalent inhibition reaction, consistent with a 3-step reaction with transient thiolate and enolates intermediate states. Two possible atomistic mechanisms affecting de-

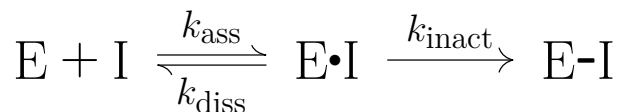
protonation/deprotonation events from the thiolate to the enolate intermediate were observed: a highly correlated direct pathway involving proton transfer to the C $_{\alpha}$ of the acrylamide warhead from the cysteine involving one or a few water molecules, and a more indirect pathway involving a long-lived enolate intermediate state following the escape of the proton to the bulk solution. The results are compared with experiments by simulating the long-time kinetics of the reaction using kinetic modeling.

3.1 Introduction

Covalent drugs represent a class of small therapeutic compounds that can covalently bind to specific target proteins, thereby inhibiting their biological functions.^{10;33;41} The potential benefits of covalency have inspired medicinal chemists to explore the covalent drug space despite concerns about reactivity. In many cases, compromises between reactivity, selectivity, and potency have produced safe and effective drugs. However, despite the increasing number of FDA-approved covalent inhibitors,⁴¹ rationally designing covalent drugs remains extremely challenging. The formation of a chemical bond between the compound and its target can lead to prolonged engagement and exceptional potency, which is advantageous. However, the imperfect selectivity of a highly reactive molecule binding irreversibly to off-target proteins can also lead to severe side effects. To alleviate these issues, the dominant design strategy has been to add a mildly reactive functional group (warhead) targeting non-conserved amino acids to an existing non-covalently binding framework that is already selective for a desired protein.⁶⁵ Conventional docking methods developed to describe non-covalent protein-ligand interactions have been expanded to model the covalent docking of a ligand with the protein.^{28;73–78} The general philosophy is that poses, where the warhead is in close proximity to a reactive amino acid, are automatically identified and used to construct a covalently bound adduct. Despite the progress, going beyond simple considerations based on static binding poses is difficult and many mechanistic aspects of covalent binding remain

poorly understood.^{70;81} Owing to their growing significance, it is important to advance our fundamental understanding of covalent inhibitors, and develop accurate computational treatments to support the rational design of these compounds.

The two fundamental steps of irreversible covalent inhibition are represented in Scheme 3.1.¹³ Initially, the inhibitor (I) diffuses in the bulk solution to encounter the kinase en-



Scheme 3.1: Mechanism of covalent inhibition of the enzyme E by the inhibitor I (E·I and E-I denotes the equilibrium noncovalent complex and the covalent adduct, respectively).

zyme (E), resulting in the non-covalent association with the kinase active site. Then, the non-covalently associated ligand fluctuates dynamically in the binding pocket, until the electrophilic warhead is favorably positioned to undergo a reaction with the nucleophilic amino acid residue of the target enzyme. At this point, the chemistry may proceed through a series of transitions leading to the covalent inhibition of the kinase. All the events preceding the chemical changes occur for non-covalent as well as covalent inhibitors.

An interesting prototypical example of a covalent inhibitor is offered by the anti-cancer covalent drug ibrutinib (PCI-32765).¹¹⁴ Ibrutinib (Figure 3.1) is an irreversible covalent

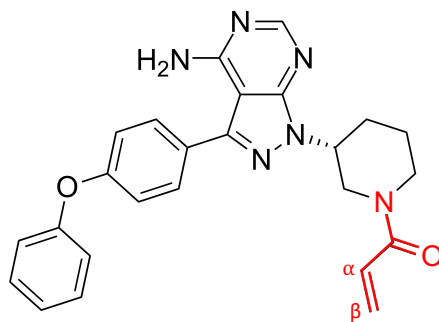


Figure 3.1: The ibrutinib molecule with its acrylamide warhead (red) which covalently binds to cysteine 481 in BTK.

inhibitor of Bruton tyrosine kinase (BTK), which plays a key role in the proliferation and survival of leukemic cells in many B-cell malignancies and is a critically important therapeutic

target in the treatment of B-cell malignancies, such as certain leukemia and lymphomas, as well as inflammatory disorders and autoimmunity.^{17;115–117} Protein kinases represent a critically important class of signaling enzymes that are implicated in numerous diseases including cancer. Because the active site of kinases has evolved to bind ATP, designing highly specific inhibitors is generally difficult. The purposeful addition of a reactive functional group to an existing ligand in order to covalently target non-conserved amino acids in the kinase active site offers a promising solution to the problem.¹¹⁸ In the early design of ibrutinib, sequence alignments analysis pointed toward a cysteine residue at position 481 in the ATP binding pocket of BTK as a potential nucleophilic site to form a covalent complex with an electrophilic inhibitor.¹¹⁴ It was also noted that this residue was in the same position as Cys773 in EGFR family kinases, which was known to be targeted by several irreversible kinase inhibitors in clinical trials.¹¹⁹ More generally, only a few kinases have a non-catalytic cysteine at this position, which suggests that it could be exploited in the design of selective covalent inhibitors.¹¹⁴

To bind irreversibly, ibrutinib utilizes a thiol-Michael addition to assist in the formation of the covalent bond between the electrophilic acrylamide “warhead” on the inhibitor and the nucleophilic Cys481 in the active site of the BTK (Figure 3.1). This last step occurs via a Michael addition with thiolate and enolate intermediate states, as depicted in Figure 3.2.

Computations can shed light on the dynamical factors affecting the Michael addition reaction in the confined environment of the kinase active site, leading to the covalently inhibited BTK. The nature of the problem of covalent inhibition requires a hybrid quantum mechanical/molecular mechanical (QM/MM) representation of the system.^{120–122} In a pioneering study, Voice et al¹²³ explored possible reaction pathways of BTK inactivation by ibrutinib using umbrella sampling QM/MM simulations. However, many key mechanistic details about the solvent-mediated deprotonation and protonation events have not been elucidated. Of particular interest is the existence of the transient anionic thiolate and enolate

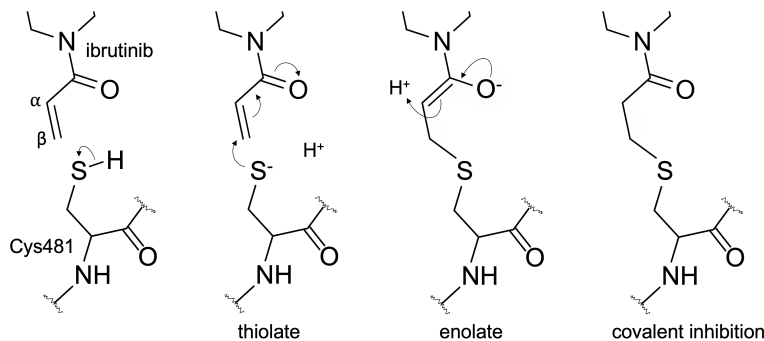


Figure 3.2: Schematic representation of the reaction steps of the Michael addition leading to the covalent inhibition of BTK by ibrutinib. First, the ligand in the active site of BTK is positioned to start the chemical reaction. The deprotonation of Cys481 yields a thiolate, which then attacks the carbon C_β of the acrylamide warhead to form a covalent bond, resulting in the anionic enolate intermediate. In the final step, the carbon C_α gets protonated, yielding the covalently inhibited BTK.

intermediate states, which are accompanied by subtle electronic changes.^{124;125}

Our overarching goal is to comprehensively describe the key steps leading to the covalent inhibition of BTK by ibrutinib. To this end, we first generated a multi- μs classical molecular dynamics (MD) trajectory of the unlinked inhibitor associated with the kinase and then determined the reaction pathway leading to the formation of the covalent inhibition using QM/MM simulations using the string method in collective variables.^{126–131} The results are finally put into a general context by simulating a simple kinetic model representing the microscopic transitions involved in covalent inhibition of BTK by ibrutinib.

3.2 Methods

3.2.1 Simulation system

A complete atomistic model of BTK kinase with ibrutinib was constructed. The protein and ligand coordinates were taken from the 1.08 Å resolution crystallographic structure with PDB ID: 5P9J.¹³² The binding pose of ibrutinib is similar to that of dasatinib,¹³³ with the additional covalent bond with residue Cys481. The system was solvated with 18397 water molecules, 59 K^+ and 53 Cl^- ions, corresponding to a physiological salt concentration of 140

mM. The model was built using the CHARMM-GUI web server,^{134–136} and subsequently equilibrated with molecular dynamics (MD). The CHARMM36 force field^{137;138} was used for the protein, and the TIP3P model was used for the water molecules.¹³⁹ The unlinked ibrutinib molecule was parameterized with CGenFF v2.5.1^{140–142} through CHARMM-GUI.^{135;136} The topology and parameter files generated in CHARMM format by CGENFF are given at <https://github.com/RouxLab/ibrutinib>. A classical trajectory of 6.6 μ s was generated using the GPU-resident version of the AMBER program.^{143–145} Periodic conditions were applied. The dimension of the cubic simulation system is approximately 80 Å. Long-range electrostatic interactions were handled by PME.¹⁴⁶ Lennard-Jones non-bonded interactions were truncated at a 14 Å pairlist cutoff, with a smooth forced-based switching from 10–12 Å consistent with the CHARMM force field.¹⁴⁷ The equations of motion were integrated using a time step of 4 fs with mass repartitioning.¹⁴⁸ The bonds involving hydrogen atoms were constrained using the rattle algorithm.¹⁴⁹ The temperature was set to 303.15 K via a Langevin thermostat with a friction of 0.5 ps⁻¹. The pressure was set to 1 atm via Monte Carlo volume fluctuations.^{143–145} The trajectory was visualized using VMD for analysis.¹⁵⁰

3.2.2 Hybrid QM/MM simulations

A hybrid QM/MM approach is necessary to accurately represent the chemical rearrangements taking place during the binding of ibrutinib with BTK.^{120;121} We used the implementation in NAMD 2.14,¹²² as it supports a number of additional tools that are required for the calculations.^{151;152} The QM/MM simulation system was constructed directly from the periodic box used for the classical MM simulation described above. The Michael addition leading to covalent inhibition of BTK corresponds to the deprotonation of the S_γ atom of Cys481, the formation of a covalent bond between the C_β atom of ibrutinib and the S_γ atom of Cys481, and the protonation of the C_α atom of ibrutinib. Accordingly, the QM region was chosen to include all relevant atoms to accurately model the chemical rearrangement

involved in the reactions of interest. Regarding the ibrutinib ligand, the QM region includes the acrylamide warhead and its neighboring pyrrolopyrimidine group. The rest of the inhibitor is represented with classical MM. On the protein side, the QM region includes the entire Cys481 residue together with the backbone atoms of Gly480 and Leu482 which are the nearest neighboring residues along the sequence. The bonds crossing the QM/MM boundary in the protein and inhibitor were capped by hydrogen atoms to satisfy the valences of the QM region.¹²² In addition, 16 water molecules, selected based on a proximity criterion of 3.6 Å from the C $_{\alpha}$, C $_{\beta}$, and S $_{\gamma}$ atoms, were included in the QM region to account for solvent effects on the reactions, corresponding to a first shell of solvation around the heavy atoms of interest in the QM region. In total, the QM region comprises 106 atoms. The overall QM/MM ibrutinib-BTK simulation system is illustrated in Figure 3.3.

The semi-empirical NDDO method PM7,¹⁵³ supported by the MOPAC 2016 program,^{154;155} was used to represent the QM region. Again, the CHARMM36 force field^{137;138} was used with TIP3P water model¹³⁹ to represent the MM region of the system. The QM/MM trajectories were integrated with a time step of 0.4 fs to increase numerical stability. The standard MD protocol in NAMD 2.14 was followed:¹⁵² the temperature was set to 303.15 K with the Langevin thermostat and the pressure was set to 1 atm with the Langevin piston. Long-range electrostatic interactions were handled by PME. Short-range non-bonded interactions were handled with a 14 Å pairlist cutoff, and a smooth 10-12 Å force switching for van der Waals interactions. The equations of motion were integrated with a time step of 0.4 fs. As the simulations progressed, the configurations were wrapped with PBCTools as needed and visualized using VMD.¹⁵⁰ After the final setup, the QM/MM system was relaxed briefly for 1,000 MD steps prior to initiating the pathway calculations.

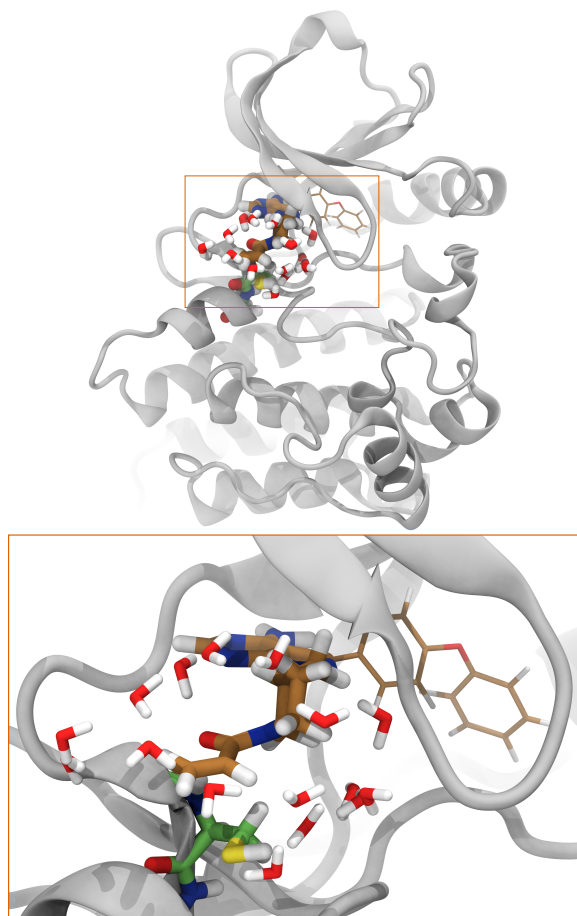


Figure 3.3: QM/MM simulation of ibrutinib bound to BTK. (top) The QM/MM region in relation to the complete kinase structure. (bottom) Close-up view of the QM/MM region and binding pocket. The atoms of the QM region, shown in color with thick bonds, include the acrylamide warhead and the neighboring pyrrolopyrimidine group (the rest of the inhibitor simulated with a classical force field is represented with thin bonds), Cys481 and the backbone atoms of Gly480 and Leu482, and 16 water molecules. In total, the QM region comprises 106 atoms. The entire system was solvated in a box of aqueous solution at 140 mM KCl and simulated with periodic boundary conditions.

String and collective variables

The string method represents a reaction pathway connecting the reactant and product as a chain of M states, or “images”, which is an ordered sequence $\{\mathbf{z}_1, \mathbf{z}_2, \dots, \mathbf{z}_M\}$ parametrized within the subspace of “collective variables” (CVs) $\mathbf{z}$.^{126–131} First, a set of meaningful CVs able to monitor the progress of the reaction must be chosen. As depicted in Figure 3.2, the thiol-Michael addition leading to covalent inhibition requires the deprotonation of the S_γ atom of Cys481, the formation of a covalent bond between the C_β atom of the ibrutinib

acrylamide warhead and the S_γ atom of the cysteine residue at position 481, and the protonation of the C_α atom of ibrutinib. One obvious CV is directly associated with the formation of the chemical bond between ibrutinib and BTK. Neglecting orientational effects, we chose the interatomic distance between the C_β atom of ibrutinib and the S_γ atom of Cys481, r_{SC} for the sake of simplicity. Two additional CVs are needed to monitor the deprotonation of S_γ and protonation of C_α along the Michael addition reaction. Importantly, such CVs must be constructed to reflect the fact that the protons in the QM region are indistinguishable particles. To this end, a generalized mean distance between one atom A and a group of N protons was defined as,

$$r_{A-[H]} = \left(\frac{1}{N} \sum_{i=1}^N \frac{1}{\|\mathbf{r}_A - \mathbf{r}_i\|^n} \right)^{-1/n} \quad (3.1)$$

This expression treats all the protons as identical particles but it weights more heavily the one at the shortest distance from atom A. In the present work, a value of 6 was used for n . The CVs monitoring the deprotonation of the S_γ atom and the protonation of the C_α atom are $r_{S-[H]}$ and $r_{C-[H]}$, respectively. The three CVs were implemented using the colvars module¹⁵⁶ of NAMD.¹⁵²

The algorithm consists of improving an initial string pathway starting from a reasonable initial guess by making small adjustments to the values of the CVs $\mathbf{z}_i$ of each image ($i = 1, \dots, M$). Here, initial pathways were generated via steered MD (SMD) simulations by imposing a time dependence on the two first CVs, r_{SC} and $r_{S-[H]}$, while allowing the third CV $r_{C-[H]}$ to evolve freely. This initial SMD pathway was then discretized into an ordered sequence of M images to produce an initial string in the subspace of CVs that was initially relaxed for 1.2 ps of MD with harmonic restraints on the protein backbone prior to the iterative optimization procedure. In the string method with swarms-of-trajectories (SMwST), the small adjustments to the string are calculated as the average dynamical drift from an ensemble of unbiased short MD trajectories of length τ , initiated from each of the images.¹³⁰

At iteration k of the optimization cycle, each image i is first thermalized with a harmonic biasing restraint with force constant K_i to keep its 3 CVs near $\mathbf{z}_i^{(k)}$. A progressive force constant K_i of 1,000 kcal/mol·Å⁻² was used to reset the CVs at their updated position. Those restraints are then released and the mean drift is determined from an ensemble (swarms) of unbiased trajectories of length τ initiated from each image as, $\langle \Delta \mathbf{z}_i(\tau) \rangle^{(k)} = \langle \mathbf{z}_i(\tau) - \mathbf{z}_i(0) \rangle^{(k)}$. The position of each image i is then updated via $\mathbf{z}_i^{(k+1)} = \mathbf{z}_i^{(k)} + \langle \Delta \mathbf{z}_i(\tau) \rangle^{(k)}$. During the string relaxation, images are kept at an equal Euclidean distance between adjacent images i and $i + 1$ in the space of the CVs.¹²⁶ Reparametrization is necessary to prevent the images from pooling at the free energy minima and keep the path well-resolved, especially in high-energy transition regions. The SMwST iteration cycle is repeated until the mean drift of all the images becomes parallel to the local tangent of the string and displays no orthogonal component, $\langle \Delta \mathbf{z}_i^\perp(\tau) \rangle^{(k)} = 0$.¹³⁰ Each cycle is made of 1) thermalization of the atomic conformations via restrained simulations at the collective variables values defining the images for 20 fs, and 2) launching from each image a swarm of 20, independent, 2 fs long, unbiased trajectories. In the present calculation, the string reached convergence in about 174 iterations, after which the path was further refined for 7 iterations using the mean force algorithm¹³¹ with trajectories lasting 4 ps. To monitor convergence at each iteration, we calculate the average of the RMSD of each image with respect to the same image in the initial path (path-RMSD) with respect to the initial path (see Supporting Figure A.1). A plateau of this quantity means that the path is not changing anymore after updates. An alternative approach to optimize a string pathway, which is consistent with the mean force acting on each image in the space of the CVs, is to update the position of the images directly from the deviations of the CVs relative to the reference position during the restrained trajectory of iteration k , via $\mathbf{z}_i^{(k+1)} = \mathbf{z}_i^{(k)} + \langle \Delta \mathbf{z}_i \rangle_{(i)}^{(k)}$.¹³¹ Because the deviation of the CVs relative to the reference position is related to the mean force exerted on each image in the CV subspace at iteration k , $\langle \mathbf{F} \rangle_i = K_i \langle \Delta \mathbf{z}_i \rangle_{(i)}$, convergence is reached when $\langle \mathbf{F} \rangle_i$ is parallel to the local

tangent of the string for each image.¹³¹

Once the optimization has converged, the free energy profile along the resulting string, $\{\mathbf{z}_1, \mathbf{z}_2, \dots, \mathbf{z}_M\}$, can be calculated by integrating the mean force for image 1 to n using the trapezoidal rule,

$$W(\mathbf{z}_n) = W(\mathbf{z}_0) - \sum_{i=1}^{n-1} \frac{1}{2} \left(\langle \mathbf{F} \rangle_i + \langle \mathbf{F} \rangle_{i+1} \right) \cdot (\mathbf{z}_{i+1} - \mathbf{z}_i) \quad (3.2)$$

where $\langle \mathbf{F} \rangle_i$ corresponds to the mean force calculated on image i , and $\mathbf{z}_i$ corresponds to the coordinates of image i in the CV space. The mean force of each image, $\langle \mathbf{F} \rangle_i$, was calculated from equilibrium QM/MM simulations of about 200 ps in the presence of harmonic restraints applied to the 3 CVs. A force constant K_i of 500 kcal/mol·Å⁻² was used. The mean force was determined from the deviation relative to the reference values of the optimized string, $\langle \mathbf{F} \rangle_i = K_i \langle \Delta \mathbf{z}_i \rangle_{(i)}$. The statistical uncertainty with respect to the free energy profile $W(\mathbf{z}_i)$ along the string stems from the estimated value of the mean force $\langle \mathbf{F} \rangle_i$ at each image. Using a bootstrapping block-averaging procedure, the statistical uncertainty on the mean force at each image was determined to be on the order of 0.2 to 2.2 kcal/mol/Å. The cumulated statistical error integrated from the bottom of a well to the top of the nearest barrier best represents the uncertainty on the free energy profile. Those values are reported in Table 3.1.

3.2.3 PMF calculation of hydronium

To further characterize the reaction, the potential of mean force (PMF) of a hydronium cation (OH_3^+) leaving the binding pocket was determined from MD simulations based on a classical MM force field. To compute the PMF, we used the enhanced sampling algorithm WTM-eABF^{157;158} built upon a hybrid of the extended adaptive biasing force (eABF) method^{159–162} with well-tempered metadynamics (WTM),^{163–165} implemented within the colvars module¹⁵⁶ of NAMD 2.14.¹⁵² The PMF is calculated as a function of the distance between the center-of-mass (COM) of the atoms S_γ , C_β , and C_α and that of the OH_3^+ cation.

The simulation system comprises the BTK kinase with bound ibrutinib in the anionic enolate intermediate state, one hydronium cation, 18,396 water molecules, and 59 K^+ and 52 Cl^- ions corresponding to a concentration of 0.176 and 0.156 M, respectively. The parameters of the classical OH_3^+ model were taken from Sagnella and Voth.¹⁶⁶ The partial charges of the negatively charged ibrutinib in the enolate intermediate state were approximated as Mulliken charges from the QM/MM simulation (given in <https://github.com/RouxLab/ibrutinib>). The initial configurations for each WTM-eABF window were generated by steered MD (SMD) with the colvars module.^{152;156} For all of the simulations, the WTM bias temperature was set to 3000 K. For the eABF algorithm, the extended fluctuation, or standard deviation between the extended CV and the actual CV was set to 0.05 Å. The width of the bins for accumulating forces in the one-dimensional PMF calculation was set to 0.2 Å. For efficiency, the reaction pathway was separated into 3 windows, with the center window overlapping its neighbors by 1 Å. The value of the distance was recorded from 3.0 to 15.0 Å. The free-energy calculations were determined to have converged by the histogram of the sampling in each bin along the distance and by differences in the gradient of the PMF between the first and second halves of the simulation. The PMF and the associated gradient with the corrected z -averaged restraint (CZAR) estimator¹⁶⁰ were output by the colvars module from the eABF data. The PMF was obtained from 2.6 μs of sampling in each window.

3.3 Results

3.3.1 *Dynamics of the unlinked inhibitor associated with the binding pocket*

The first step leading to BTK inhibition is the reversible bimolecular association of the ibrutinib molecule to the kinase active site without the formation of a covalent bond with Cys481. The initial diffusive association of ibrutinib with the kinase may occur according to two scenarios, in which the side chain of Cys481 is either protonated (thiol) or deprotonated

(thiolate). With an elevated pK_a of about 7.7,¹⁶⁷ the encounter is more likely to involve the thiol form, which is the situation that we consider here. This part of the problem essentially mirrors the familiar thermodynamic and kinetic behavior of a standard non-covalent kinase inhibitor. But more information on the state of the system following this initial step, before the chemical reactions begin, is needed to better understand how the covalent inhibition is likely to proceed. To characterize the structural dynamics of the unlinked inhibitor associated with the kinase active site, a 6.6 μ s classical MD simulation was generated.

To provide a global view of the fluctuations of the unlinked ligand associated with the BTK binding pocket from the 6.6 μ s classical MD simulation, we calculated the 2D-PMF as a function of the distance between the C_β atom of the acrylamide warhead and the S_γ atom of Cys481, together with the RMSD of the inhibitor from its crystallographic binding pose (PDB ID: 5P9J¹³²). The result is shown in Figure 3.4. The 2D-PMF shows

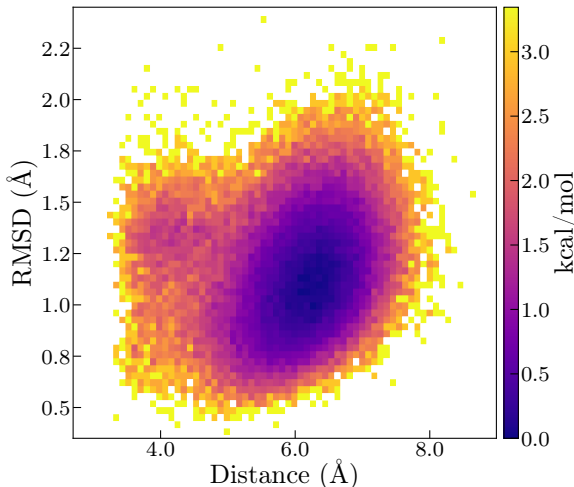


Figure 3.4: 2D-PMF calculated from the 6.6 μ s classical MD simulation as a function of the distance between the C_β atom of the acrylamide warhead and the S_γ atom of Cys481 along the x -axis, and the symmetry-corrected RMSD of the inhibitor’s heavy atoms from its crystallographic binding pose along the y -axis (PDB ID: 5P9J¹³²). The deviation was calculated post-alignment of the binding pocket backbone C_α ’s (within 10 Å of the inhibitor) to the crystallographic structure.

that the ibrutinib molecule in the binding pocket fluctuates significantly while retaining its binding pose. The average root-mean-square deviation (RMSD) of the inhibitor relative to its crystallographic binding pose is 1.15 Å. The hydrogen bond between the carbonyl

oxygen of the acrylamide warhead and the amide hydrogen of Cys481, which is present in the crystallographic structure,¹³²) fluctuates but is nonetheless maintained 93% of the time according to a distance cutoff of 3.5 Å and angle cutoff of 135°. (Figure 3.5).

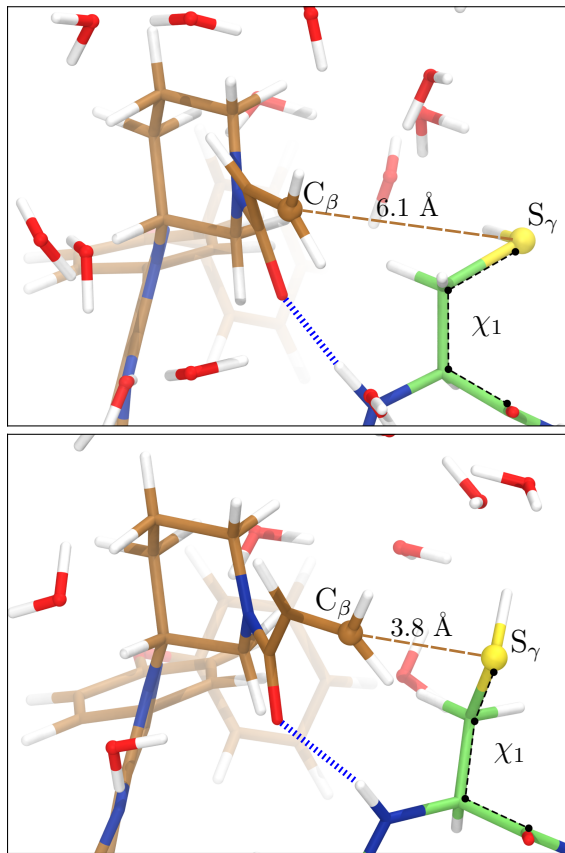


Figure 3.5: Two distinct orientations of the Cys481 side chain occurring during the MD trajectory of unlinked ibrutinib associated with the BTK binding site have a direct impact on the interatomic distance, r_{CS} (sienna dashed), between the C_β atom (sienna) of the warhead and the S_γ atom (yellow) of Cys481. The larger interatomic distance of 6.1 Å shown in (top) occurs the most frequently ($\sim 95\%$ of the time) when the χ_1 dihedral of Cys481 (black dashed) is $+50^\circ$. The smaller interatomic distance of 3.8 Å shown in (bottom) occurs infrequently ($\sim 5\%$ of the time) when the χ_1 dihedral of Cys481 is -50° . The hydrogen bond (blue dashed) between the carbonyl oxygen (red) of the acrylamide warhead and the amide hydrogen (white) of Cys481 observed in the crystallographic structure¹³²) is present in both configurations.

A key feature of interest is the interatomic distance r_{CS} between the C_β atom of the warhead and the S_γ atom of Cys481, the two atoms most directly involved in covalent inhibition. The main results are shown in Figure 3.6. Two distinctive values of r_{CS} occur, corresponding to two different atomic configurations of the system (Supporting Figures A.2 and A.4). Figure 3.6 displays the histogram of this distance during the 6.6 μs classical MD

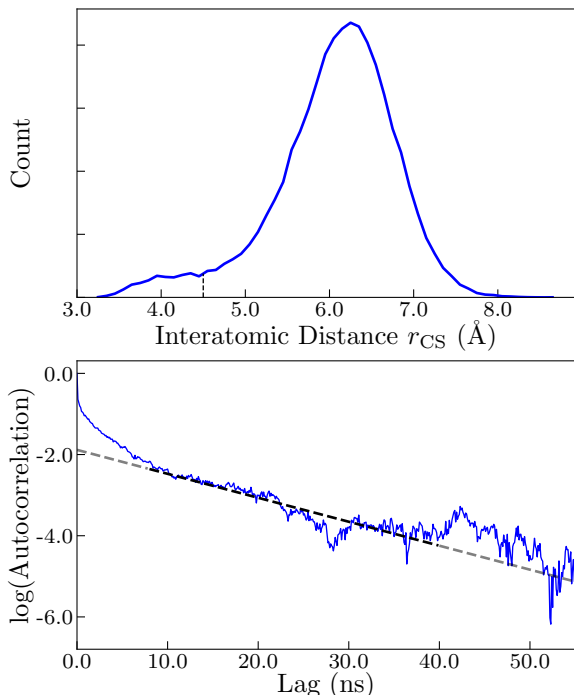


Figure 3.6: Dynamics of the associated but not covalently bonded inhibitor from the 6.6 μ s classical MD simulation. (top) Histogram of the interatomic distance, r_{CS} , between the C_β atom of the warhead and the S_γ atom of Cys481. Two states distinguished by the vertical dash line corresponding to a short and long distance are populated by about 5% and 95%, respectively. They correspond to two different orientations of the Cys481 side chain (Figure 3.5). (bottom) Log plot of the normalized time correlation function of the two states displaying an initial rapid relaxation followed by a near exponential decay (slope of the fitted dashed line is -0.059 ns^{-1} corresponding to a relaxation time of 17 ns).

simulation. The most prevalent state corresponds to a distance r_{CS} of about 6.1 Å. Figure 3.5(top) displays this configuration, in which the side chain of the Cys residue is rotated, keeping the S_γ atom away from the drug’s warhead. Figure 3.5(bottom) displays a second state, occurring infrequently, with an average distance of 3.8 Å. This corresponds to an “activated complex” where the chemical reactions leading to covalent inhibition may start. This configuration is achieved by rotation of the cysteine side chain, bringing the S_γ atom closer to the C_β atom of the warhead of ibrutinib. Transitions between the two metastable states are controlled by rotation of the rotation of the χ_1 dihedral angle $S_\gamma-C_\beta-C_\alpha-C$ of Cys481 switching between -50° and 50° (see Supporting Figure A.2). About 95% of the time, the inhibitor and the side chain of Cys481 are separated by a distance on the order of 6.1 Å, too far from one another to allow the initiation of the chemical reactions leading to

covalent inhibition. According to the classical trajectory, the system is configured favorably about 5% of the time to allow the formation of a covalent bond between the C_β atom of the warhead and the S_γ atom of Cys481. This configuration is shown in Figure 3.5(bottom).

3.3.2 *Initial reaction pathways from SMD simulations*

The string method aims at discovering the minimum free energy path and the free energy along the path in the subspace corresponding to a large but finite set of selected CVs.^{126–131} Whereas transition path sampling (TPS) offers a sophisticated strategy for determining the reaction mechanism of enzymes,^{168–170} it can be computationally prohibitive when multiple metastable states are encountered along a reaction pathway. The string method offers a computationally attractive alternative by aiming to construct an optimal reaction pathway directly over the free energy surface in the subspace of CVs.

To obtain an optimized string mapping of the covalent inhibition process, it is necessary to provide an initial reaction pathway that includes all three chemical events of interest. For this purpose, a series of 10 steered MD (SMD) simulations were generated with the QM/MM model starting from 10 independent equilibrium configurations of the reactant state (Supporting Figure A.5). During the SMD simulations, the dynamical evolution of the CVs was enforced via time-dependent harmonic restraints to induce two of the chemical steps involved in the covalent inhibition. The change in the interatomic distance between the warhead C_β atom and the Cys481- S_γ atom (r_{CS}) enforces the formation of the covalent bond, and the change in the generalized mean distance ($r_{S-[H]}$) controls the deprotonation of the S_γ atom to its surrounding solvent molecules. Including only these two distances in the SMD while allowing the generalized mean distance ($r_{C-[H]}$) controlling the protonation of the C_α atom to evolve freely resulted spontaneously in a complete reaction pathway in 6 out of 10 SMD simulations. A typical SMD trajectory of this type of event is shown in Figure 3.7a. In the remaining 4 cases, the protonation of the C_α atom did not occur

spontaneously upon the formation of the S_γ - C_β covalent bond and the deprotonation of the S_γ atom. In this case, the proton is transferred from S_γ to the nearby water molecules and takes the form of a solvated hydronium cation in the neighborhood of the ligand. To complete the reaction in these 4 cases, the generalized mean distance ($r_{C-[H]}$) controlling the protonation of the C_α atom was controlled via a time-dependent harmonic restraint in a subsequent SMD simulation (Supporting Figure A.6). A typical SMD trajectory of this type of event is shown in Figure 3.7b. Two distinct mechanistic pathways could be

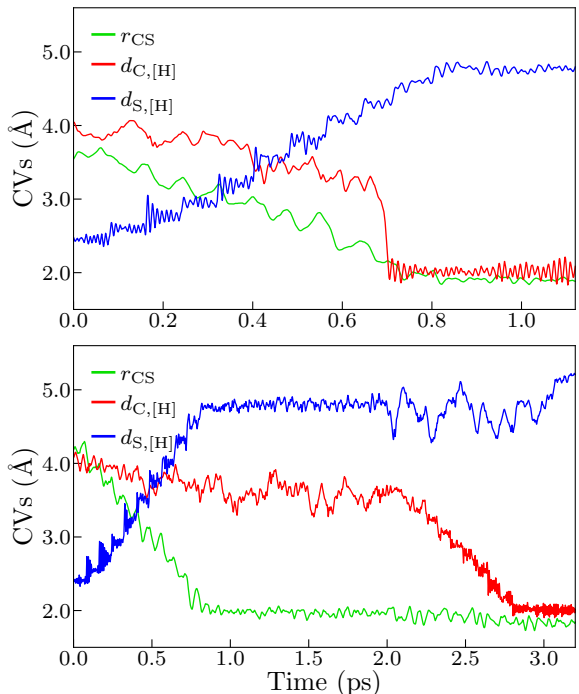


Figure 3.7: Dynamical evolution of the CVs during SMD simulations. The evolution of the first two CVs, r_{CS} (green line) and $r_{S-[H]}$ (blue line), was enforced through time-dependent harmonic constraints. The third CV, $r_{C-[H]}$ (red line), was allowed to evolve freely during the SMD simulations. (top) The entire covalent inhibition reaction including the C_α protonation reaction takes place spontaneously. (bottom) In the SMD simulations where the reaction paused at the enolate intermediate, the evolution of the third CV, $r_{C-[H]}$ was enforced during a subsequent SMD simulation to finalize the C_α protonation.

identified from the 10 independent SMD simulations (Supporting Figure A.7). The first one involves a single water molecule in the deprotonation/protonation events of the S_γ and C_α atoms, respectively. Figure A.7 depicts this direct route followed by the proton during the deprotonation and protonation events. In 5 of the 6 cases where the protonation of

C_α occurred spontaneously, the same proton was transferred from the S_γ atom to a water molecule, and then to the C_α atom of the warhead. The 4 cases where the effective distance $r_{C-[H]}$ was included in SMD correspond to a more indirect route of a proton on its way to the C_α atom. Once the proton is transferred from S_γ , a series of local transfers may lead to the stabilization of a solvated hydronium ion in the QM water molecules. As discussed below, the possibility that this hydronium cation may diffuse away in solution then results in a long-lived anionic enolate intermediate state. A comparison of the 10 pathways generated by the SMD simulations is shown in Figure A.5. These initial pathways do not show a correlation between the 3 reactions. However, since 6 of the 10 pathways exhibit a spontaneous C_α protonation upon covalent bond formation, this suggests a possible correlation between these two reactions. However, the initial reaction pathway from SMD only represents a crude approximate transition reaction path. Results from SMD trajectories may be affected by the nonequilibrium time-dependent forces steering the atoms. To better assess the correlation between microscopic events, it is necessary to go beyond the results from simple SMD.

3.3.3 *Optimized reaction pathway from the string method*

To construct an initial string pathway, a set of 32 configurations spanning the transition were selected from the SMD trajectory. The 32 configurations were chosen to have the same Euclidian distance between adjacent images, calculated in the subspace of the 3 CVs, r_{SC} , $r_{S-[H]}$, and $r_{C-[H]}$, to generate an initial string with equally spaced images. A total of 10 initial string pathways were generated and then optimized in CV space using the string method with swarms-of-trajectories (SMwST).¹³⁰ The evolution of the 3 CVs along the optimized string, shown in Figure 3.8 (top panel), is suggestive of strong correlations between the three chemical reactions of interest. In particular, the deprotonation of the S_γ atom (blue line) only begins after the C_β - S_γ distance is 3.8 Å. The deprotonation of the S_γ atom is not initiated while the ligand is at a larger distance. Once the S_γ atom is

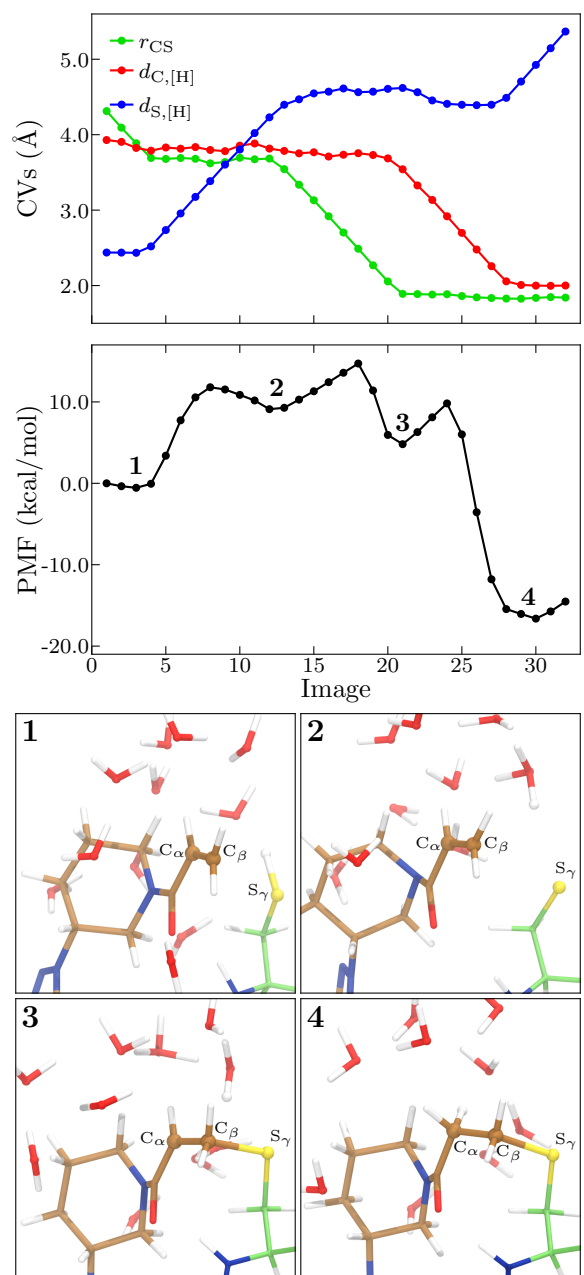


Figure 3.8: Results from the optimized string for the direct pathway. (top) Evolution of the CVs (in Å) and associated PMF (in kcal/mol) along the optimized string. The four metastable states along the reaction corresponding to local free energy minima are indicated in the PMF. (bottom) Typical configurations corresponding to the four metastable states along the reaction: state 1: the associated ibrutinib in the binding pocket with an interatomic distance between the C_β atom (sienna) of the warhead and the S_γ atom (yellow) of Cys481 of 3.7 Å; state 2: thiolate intermediate after deprotonation of the S_γ -Cys481 atom; state 3: enolate intermediate after formation of the the S_γ - C_β covalent bond, state 4: completed reaction after the C_α protonation occurred. Results from the optimized string for the indirect pathway are shown in Supporting Figure A.8.

deprotonated, a hydronium cation hovers in the vicinity of the acrylamide warhead (state 2 in Figure 3.8). This shows that the deprotonation of the cysteine is an integral part of the reaction that occurs in the presence of the ligand rather than an event that takes place independently. Consistent with this observation, Awoonor-Williams and Rowley showed that the noncovalent binding free energy of an ibrutinib-like inhibitor of BTK with a *t*-butyl cyanoacrylamide warhead varied from -7.4 kcal/mol for the neutral (protonated) cysteine to -11.4 kcal/mol for the deprotonated thiolate form.⁸¹ The difference in the calculated noncovalent binding free energy is indicative of the coupling between the deprotonation of Cys481 and the presence of the ligand, suggesting that this step must be treated as part of the full chemical reaction pathway. This suggests that quantifying the pK_a of the cysteine side chains, albeit useful for characterizing the accuracy of the computational models, may be of limited significance in predicting the selectivity of covalent kinase inhibitors.¹²⁵

The product state of the initial deprotonation of S_γ coincides with the shortening of the C_β - S_γ distance (green line), leading progressively to the formation of the covalent bond. The formation of the C_β - S_γ covalent bond appears to trigger the protonation of the C_α atom (red line). Figure 3.8 (central panel) shows the free energy profile calculated along the optimized string. The local free energy wells can be directly related to the main states along the reaction pathway, the unlinked-associated drug (state 1), the thiolate intermediate with the deprotonated S_γ atom of the Cys481 residue (state 2), the anionic enolate intermediate with covalently bound ibrutinib (state 3), and the final product with protonated- C_α (state 4).

Consistent with the Michael addition reaction (Figure 3.2), the pathway from the string shown in Figure 3.8 displays four distinct states with three chemical steps. Based on the PMF shown in Figure 3.8, the relative stability of the metastable intermediate states along the reaction pathway is roughly similar until the final product, which is considerably more stable. It is unclear how this aspect of the results may be affected by the accuracy of the QM method,

as well as the reliance on the one-dimensional reaction pathway (string) in the subspace of the CVs. In a cleverly designed strategy, Awoonor-Williams and Rowley determined the thermodynamic stability of all metastable intermediate states of an ibrutinib-like covalent inhibitor of BTK with a *t*-butyl cyanoacrylamide warhead using a combination of QM/MM simulations together with an alchemical free energy method.⁸¹ Although an alchemical free energy method is not designed to map out an optimal pathway along a multi-step reaction, it could be used to rigorously address issues of thermodynamic stability for all well-identified intermediate chemical states.

To gain further insight into the mechanism of covalent inhibition, we estimate the transition rate from all the steps displayed in Figure 3.8. For the sake of simplicity, the transition rate from the reactant (r) to the product (p) for each of these steps is approximated within the framework of classical transition state theory (TST),^{171;172}

$$k_{r \rightarrow p} = \frac{e^{-W(s^\dagger)/k_B T}}{\int_r e^{-W(s)/k_B T} ds} \langle \dot{s} \theta(\dot{s}) \rangle \quad (3.3)$$

where the reaction coordinate s is the arc length along the string, and $s^\dagger$ is the transition state between the reactant “r” and the product “p”. The quantity $\langle \dot{s} \theta(\dot{s}) \rangle$ represents the average forward velocity of the CVs projected along the string. The transition rate can be expressed as, $k_{r \rightarrow p} = e^{-\Delta W^\dagger/k_B T} \langle \dot{s} \theta(\dot{s}) \rangle / L_r$, where $\Delta W^\dagger = W(s^\dagger) - W(s_r)$ is the activation free energy, s_r represents the position of the minimum in the reactant well, and $L_r = \int_r e^{-[W(s)-W(s_r)]/k_B T} ds$, is the effective length of the reactant well “r”.

The results are summarized in Table 3.1. The average forward velocity of the CVs projected along the string for the C_α - S_γ bond reaction (2-3 and 3-2 transitions) is 2.27. Å·ps⁻¹, matching closely the value from Maxwell’s distribution for the carbon-sulfur reduced mass, $(k_B T(m_C + m_S)/(2\pi(m_C m_S)))^{1/2} = 2.14$ Å·ps⁻¹. The average forward velocity of the CVs projected along the string for the proton transfer reactions (1-2, 2-1, 3-4, and 4-3 transitions) is on the order of 8.70 Å·ps⁻¹, slightly larger than the value from Maxwell’s distribution for

Table 3.1: Estimated rates from transition state theory

Transition ^a <i>i</i> to <i>j</i>	$\Delta W(s^\dagger)$ (kcal/mol)	L_r (Å)	$\langle \dot{s} \theta(\dot{s}) \rangle$ (Å·ps ⁻¹)	k_{ij} (s ⁻¹)
1 to 2	12.35 ±1.3	0.48	8.74	1.84×10^4
2 to 1	2.69 ±1.0	0.42	8.74	2.25×10^{11}
2 to 3	5.61 ±0.6	0.42	2.27	4.42×10^8
3 to 2	9.91 ±0.7	0.19	2.27	7.28×10^5
3 to 4	5.01 ±0.7	0.19	8.68	1.03×10^{10}
4 to 3	26.44 ±1.0	0.34	8.68	1.41×10^{-6}

^aState 1: unlinked-associated inhibitor, state 2: thiolate, state 3: enolate, and state 4: final product.

a single proton, $(k_B T / (2\pi m_H))^{1/2} = 6.31 \text{ Å} \cdot \text{ps}^{-1}$ (using the reduced mass does not change this value significantly). Because TST ignores recrossings and assumes a transmission coefficient of 1, the transition rate estimates are upper bounds.^{172;173} Thus, the approximate values in Table 3.1 are expected to be overestimated. However, the limited accuracy of the semi-empirical QM method used is likely to be of greater importance.^{81;153;174}

One important value from Table 3.1 is the transition rate from the final product to the enolate intermediate (k_{43}), which is extremely slow ($1.41 \times 10^{-6} \text{ s}^{-1}$). Essentially, the covalent inhibition is irreversible according to the calculations. The string reaction pathway displays a direct and highly concerted proton transfer from the S_γ atom of Cys481 to the C_α atom of ibrutinib. In principle, this process is possible, as long as an excess proton remains present inside the kinase binding pocket in the vicinity of the inhibitor. While the attractive electrostatic interactions between the excess proton and the anionic enolate intermediate oppose the dissociation of the excess proton from the binding pocket, the latter should ultimately be able to escape and diffuse into the bulk solvent. The absence of an excess proton in the binding pocket could then give rise to a long-lived enolate intermediate state, as the system must wait for a new proton coming from the bulk solution to complete the reaction. A picture of metastable state 2 in Supporting Figure A.8 hints at the diffusion of a solvated proton away from the binding pocket. However, in the context of the QM/MM simulation, dissociation cannot occur because the excess proton is artificially held captive

by the finite number of water molecules in the QM region. To characterize the escape process, the PMF of a hydronium cation was calculated as a function of its distance to the ligand using MD simulations based on a classical MM force field with the enhanced sampling algorithm WTM-eABF.^{157;158} The calculated PMF is shown in Figure 3.9. Based

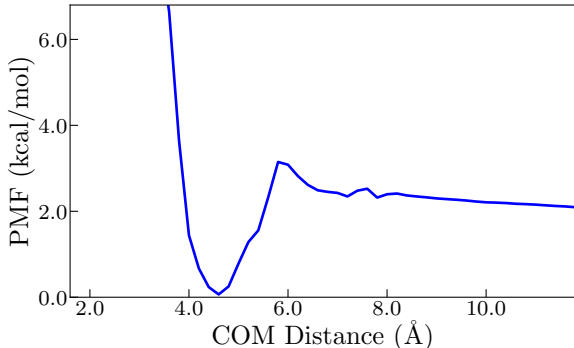


Figure 3.9: PMF for a hydronium cation escaping from the binding pocket into the bulk solvent, leaving a negatively charged enolate intermediate.

on the PMF, the proton escaping from the binding pocket encounters a free energy barrier of about 3.6 kcal/mol. The escape rate of the hydronium cation from the binding pocket to the bulk solution (Figure 3.9) can be estimated in the high friction diffusive limit from the Kramers-Smoluchowski theory,^{175;176}

$$k_{\text{diff}} = \frac{D}{2\pi k_B T} (W''(r_b) |W''(r^\ddagger)|)^{1/2} e^{-\Delta W^\ddagger/k_B T} \quad (3.4)$$

where D is the H^+ diffusivity, $r^\ddagger$ is the transition state (5.3 Å), $W''(r_b)$ is the second derivative in the reactant well, $W''(r^\ddagger)$ is the second derivative at the transition state, and $\Delta W^\ddagger$ is the activation free energy. Using the H^+ diffusivity of $0.93 \text{ Å}^2/\text{ps}$ in bulk water¹⁷⁷ yields an escape rate of about $2.2 \times 10^9 \text{ s}^{-1}$. Assuming that the association rate of a solvated proton with the binding pocket is diffusion-limited, yields a rate on the order of $4.02 \times 10^2 \text{ s}^{-1}$ at a physiological pH of 7.4.

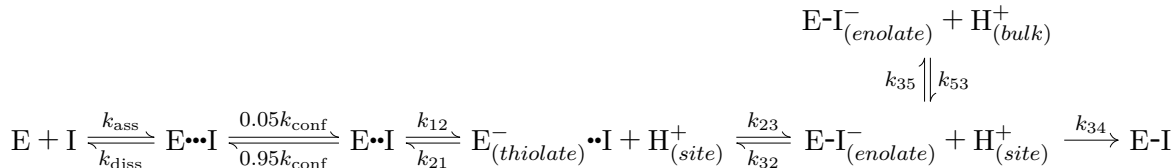
The present results can be compared with a previous QM/MM computational study on covalent inhibition of BTK by ibrutinib based on the density-functional tight-binding

(DFTB3) method.¹²³ Partial free energy surfaces were generated from umbrella sampling QM/MM simulations to determine the reaction pathway, leading to the conclusion that the mechanism for the deprotonation of Cys481 involved a direct proton transfer to the carbonyl oxygen atom of ibrutinib, resulting in a Cys-[S_γ/C]OH⁺ ion pair.¹²³ This led to the proposal that the formation of the C_β-S_γ bond leads to an enol complex and subsequently by a solvent-assisted keto-enol tautomerization step to form the covalent keto product.¹²³ This conclusion stands in contrast with the present results, where the proton leaving the sulfur atom of Cys481 is immediately captured by nearby water molecules in the binding pocket, and the subsequent protonation of the C_α comes also from the solvent. An important factor that may be responsible for the considerable difference in the proposed mechanisms is the selection of the atoms treated as part of the QM region. The previous study included all of the ibrutinib molecule and the side chain of Cys481 but only a single water molecule hydrogen bonded to the carbonyl oxygen of the acrylamide as part of the QM region.¹²³ In the present study, the QM region included the ibrutinib acrylamide warhead and its neighboring pyrrolopyrimidine group, the entire Cys481 residue together with the backbone atoms of Gly480 and Leu482, and also 16 water molecules (within 3.6 Å from the center of the reaction). These water molecules represented at the QM level are critical for the protonation and deprotonation steps observed along the reaction pathway determined from the string method.

3.3.4 *Kinetic implications*

The quantitative results from the string method based on QM/MM simulations reported in Figure 3.8 and Table 3.1 provide considerable insight into the reaction pathway. Nonetheless, it is difficult to appreciate how these results relate to experimental observations. From this point of view, it is important to try putting the present results in a broader context regarding the time course of covalent inhibition of BTK by ibrutinib. The information gained from the

classical MD simulations of the associated and unlinked inhibitor, and the string reaction pathway from QM/MM simulations allow us to go beyond the generic picture that was shown in Scheme 3.1 in representing the covalent inhibition process in some detail.



Scheme 3.2: Detailed binding mechanism informed by molecular dynamics simulations

To make progress, we represent the overall time-course of inhibition with the more elaborate Scheme 3.2, encompassing the microscopic steps of Michael addition suggested from the calculations. The simple association of the ligand to the kinase active site is a reversible process that is similar for non-covalent as well as covalent inhibitors. While the ibrutinib-BTK non-covalent dissociation constant, K_d (*i.e.* K_i), could be calculated using free energy simulations with some efforts,^{81;178} this is unlikely to significantly affect the main conclusions. For the purpose of our numerical simulations of the kinetic model, a K_d of 1 nM was chosen, broadly consistent with available experimental ibrutinib data, reporting values of 10,¹⁷⁹ 1.9,¹⁸⁰ and 0.64 nM.¹⁸¹ A residence time τ_m of 15 minutes was assumed, which is a reasonable estimate compared to similar ligands. Once associated with the binding pocket, the non-covalently bonded ligand undergoes dynamic fluctuations. According to the classical MD simulation of the unlinked-associated inhibitor, about 5% of the configurations yield a warhead that is favorably positioned to allow the subsequent chemical reactions to proceed (although this number may require further attention). The rate of exchange between the loosely-associated and well-positioned ligand, $E \cdots I \rightleftharpoons E \bullet I$, is very fast, and only the relative population of the two states matters (Figure 3.6). This is represented in Scheme 3.2 by a μs -timescale rate constant, k_{conf} , scaled by the equilibrium probability of each conformation.

The process of covalent inhibition of BTK by ibrutinib was simulated within the framework of Scheme 3.2 by solving the ordinary differential equations (ODEs) numerically with

the transition rate constants estimated above. The results are illustrated in Figure 3.10. In the initial phase, the population of the state $E\cdots I$, corresponding to the unlinked ligand

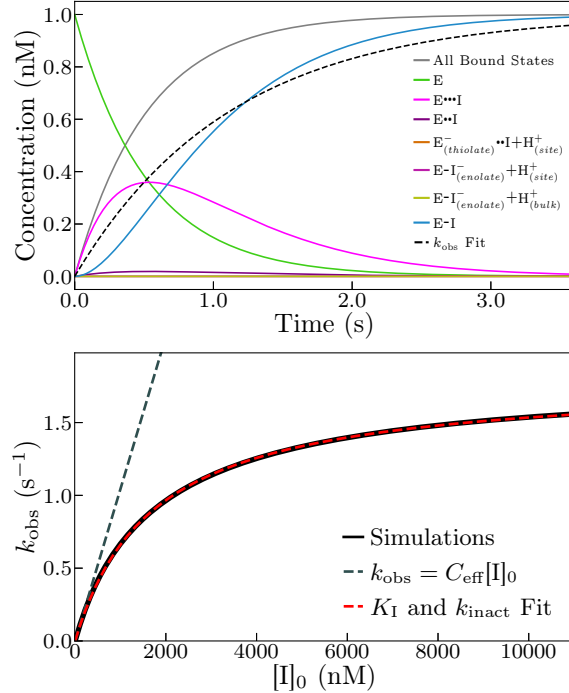


Figure 3.10: Numerical simulation of Scheme 3.2. (top) Time course of the different states when $[I] = 1.72 \times 10^3 \text{ nM} \approx K_I$ and $[E] = 1 \text{ nM}$. The observed rate constant of $[E-I]$, k_{obs} , was 0.90 s^{-1} . Because it is assumed that $[I] \gg [E]$, the inhibitor concentration remains essentially constant with time. (bottom) The variation of k_{obs} over $[I]$ (black solid), the two-step curve $k_{\text{obs}} = k_{\text{inact}}[I]/(K_I + [I])$ (red dashed), and the one-step line $k_{\text{obs}} = C_{\text{eff}}[I]$ (dark-gray dashed). Fitting the two-step curve yielded $K_I \approx 1.72 \times 10^3 \text{ nM}$, $k_{\text{inact}} \approx 1.8 \text{ s}^{-1}$, and $C_{\text{eff}} \approx 1.04 \times 10^{-3} \text{ nM}^{-1}\text{s}^{-1}$.

loosely associated with the binding pocket, transiently rises and then decays over time. The population of the state $E\bullet I$, corresponding to the unlinked ligand with the warhead positioned favorably, also peaks with the state $E\cdots I$ but always remains smaller (roughly 5% of the combined population of the two unlinked states). The observed rate constant, k_{obs} , is obtained by fitting a generic growth factor, $\text{Occ}_{\text{cov}}(t) = [E]_0(1 - e^{-k_{\text{obs}}t})$ to the time-dependent occupancy of the covalently-bound state, $E-I$.^{87;105}

Figure 3.10 (top) exemplifies a lag phase in covalent occupancy due to intermediate steps that are quick to equilibrate yet heavily in favor of non-progressive states. This affects the quality of the fitted $\text{Occ}_{\text{cov}}(t)$, and it may go unnoticed in experiments due to lower

Table 3.2: Experimental estimates of kinetic parameters of BTK inhibition by ibrutinib

K_I (nM)	k_{inact} (s^{-1})	C_{eff} ($\text{nM}^{-1}\text{s}^{-1}$)	Reference
425	9.00×10^{-2}	2.1×10^{-4}	Watterson et al. ¹⁸²
54.2	2.66×10^{-2}	4.8×10^{-4}	Kaptein et al. ¹⁸³
0.95	1.10×10^{-2}	1.2×10^{-2}	Hopper et al. ¹⁸⁴
4.8	1.83×10^{-3}	3.8×10^{-4}	Liang et al. ¹⁸⁵
		3.3×10^{-4}	Toenjes et al. ¹⁸⁶

temporal and state resolution. For systems designed to accumulate in an intermediate state, this deviation may significantly hinder the drug’s evaluation. For example, a well-performing warhead may be overlooked if the same drug has a much higher propensity to exist in an unfavorable conformation compared to its inverse. Despite the deviations, this analysis remains useful for interpreting the rate constants reported herein through an experimental lens.

The variation of k_{obs} as a function inhibitor concentration, $[I]$, is then fit to $k_{\text{obs}} = k_{\text{inact}}[I]/(K_I + [I])$ for the inhibition constant, K_I , and the inactivation rate constant, k_{inact} (Figure 3.10, bottom). A reaction may appear one-step in both non-saturating ($[I] \ll K_I$) and saturating ($[I] \gg K_I$) conditions. These linear approximations are distinguishable by the Y-intercept, where an intercept of 0 yields a slope equivalent to the covalent efficiency, $C_{\text{eff}} = k_{\text{inact}}/K_I$. The latter is often sufficient for evaluating covalent inhibitor potency, especially in preliminary screenings where practical challenges for resolving k_{inact} and K_I are not justified.⁴¹

The simulations of the kinetic model yield an estimated covalent efficiency, C_{eff} , of $1.04 \times 10^{-3} \text{ nM}^{-1}\text{s}^{-1}$, an apparent inhibition constant, K_I , of $1.72 \times 10^3 \text{ nM}$, and an apparent inactivation rate constant, k_{inact} , of 1.8 s^{-1} . While the magnitude of C_{eff} from the present computations is in good agreement with the data reported in Table 3.2, the inactivation rate constant is, depending on the reference, overestimated by a factor of 20,¹⁸² 67,¹⁸³ 163,¹⁸⁴ or 983¹⁸⁵ (the latter appears to be an outlier and will be omitted from the following discussion).

A number of factors could be at the origin of the overestimated inactivation rate constant,

k_{inact} . In particular, the approximate TST transition rates from Table 3.1 assume a transmission coefficient of 1 and are expected to be overestimated.¹⁷² Nonetheless, a survey of computed transmission coefficients in enzymatic reactions indicates that they vary between 0.36 and 0.99, suggesting that these effects are relatively modest.¹⁷³ More importantly, we used a population of 5% for the conformation in which the warhead is positioned favorably to allow the chemistry to proceed based on the analysis of the 6.6 μs classical MD simulation of the unlinked inhibitor associated with the kinase binding pocket. However, as this simulation was started from the coordinates of the crystallographic structure of ibrutinib covalently linked to BTK, it is quite possible that the population of conformations with a favorably positioned warhead was overestimated. Although the dynamics of the unlinked ligand has no bearing on the subsequent chemical reactions described by the QM/MM simulations, the population of conformations with a favorably positioned warhead has a direct impact on the apparent inactivation rate. For instance, the discrepancy with the experimental rate of inactivation is immediately reduced by a factor of 5 if we assume a population of 1%, showing that the apparent inactivation rates can be altered by modulating the conformational dynamics of the unlinked ligand associated with the kinase active site. Finally, a likely factor for the disagreement with experimental data is the height of free energy barriers for the chemical steps determined from the semi-empirical QM method. The free energy barriers bear also some statistical uncertainty due to the finite sampling (Table 3.1). With a population of conformations with a favorably positioned warhead of 5%, the experimental inactivation rate is matched exactly if one assumes a uniform increase of 1.8-3.0 kcal/mol for all the free energy barriers given in Table 3.1, but this is reduced to 0.8-2.1 kcal/mol if one assumes a population of 1% (excluding the results from ref 185). Even though the overall accuracy of the free energy barriers deserves further attention, which specific step may be the most inaccurate is not immediately clear. For example, even though the initial deprotonation of Cys481 is the slowest step along the present pathway, experimental exchange-rate

studies have shown that the equilibrium between the protonated and deprotonated states of solvent-exposed cysteine side chains is fast.^{187–189} Moreover, a similarly small free energy barrier for the chemical reaction leading to the formation of the enolate intermediate complex was obtained using umbrella sampling QM/MM simulations in the case of a *t*-butyl cyanoacrylamide warhead ligand.⁸¹

It is noteworthy that, according to the simulation of the kinetic model, about 82% of covalent inhibition results from reaction steps following the direct pathway, without escape of the excess proton to the bulk solution. Because the bimolecular association rate of a proton within the binding pocket depends on the $[H^+]$ concentration, this suggests the possibility of a long-lived enolate intermediate state at high pH. The existence of direct and indirect routes leading to covalent inhibition, with the excess proton either remaining within the binding site or escaping to the bulk, represents an intriguing departure from the expected mechanism for a Michael addition in solution. Although of a different character, a direct proton transfer mechanism was also proposed in the QM/MM simulation study of Mulholland and co-workers.¹²³ Even though computational studies are burdened by a number of approximations, the concept of a direct proton transfer pathway leading to covalent inhibition within the confine of the kinase binding pocket should be considered carefully.

Lastly, it is worth pointing out that Scheme 3.2 assumes that the initial diffusive encounter and non-covalent association of ibrutinib with the kinase occurs with a protonated Cys481. Because of the elevated pK_a of 7.7, this is the most likely scenario, as the equilibrium population distribution of the thiol and thiolate are about 0.67 and 0.33, respectively. It is possible to expand the kinetic scheme to account for a scenario in the initial encounter involves the thiolate form (see Scheme A.1 in Supporting Information). Solving the ODEs numerically while accounting for this expanded scenario (Supporting Figure A.9) and integrating the net flux for the *thiolate*–*enolate* transitions, $\Phi(t) = P_{(t.)}(t)k_{23} - P_{(e.)}(t)k_{32}$, over the time course of the reaction shows that these ratios remain and that the dominant

reaction pathway involves an encounter with the thiol form to about 67%, and 33% with the thiolate form. The overall kinetic behavior of the model with respect to the final state, E-I, is not greatly affected when including the encounter with the thiolate form. The covalent efficiency C_{eff} , is $1.01 \times 10^{-3} \text{ nM}^{-1}\text{s}^{-1}$, the apparent inhibition constant, K_{I} , is $2.83 \times 10^3 \text{ nM}$, and the apparent inactivation rate constant, k_{inact} , is 2.86 s^{-1} . The kinetic behavior of the total occupancy is identical between the two mechanistic reaction schemes. The two possible reaction pathways are compared via a schematic free energy diagram of all the states and transitions in Supporting Figure A.10. This analysis shows that the dominant reaction pathway corresponds to the encounter and non-covalent association with the thiol form of Cys481.

3.4 Conclusion

A computational approach was used to study the mechanism of covalent inhibition of BTK by ibrutinib. MD simulations based on classical force field indicate that the unlinked inhibitor associated with the kinase binding pocket undergoes considerable fluctuations and that only a small fraction of configurations position the ligand favorably to permit the initial chemical steps toward covalent inhibition. The complete reaction pathway leading to the formation of the covalent bond of the ibrutinib acrylamide warhead with Cys481 via a Michael addition was determined with the string method in collective variables on the basis of QM/MM simulations. The string pathway shows that there is a strong correlation between the covalent bond formation and the protonation/deprotonation events taking place sequentially in the covalent inhibition reaction in a sequence of 3 steps. Two possible mechanisms could affect the deprotonation/protonation events from the thiolate to the enolate intermediate: a highly correlated direct pathway involving one or a few water molecules, and a more indirect pathway involving the escape of H^+ to the bulk. The possibility of a direct proton transfer pathway leading to covalent inhibition within the confine of the kinase binding pocket rep-

resents an intriguing departure from the expected mechanism for a Michael addition in bulk solution. About 82% of covalent inhibition results from reaction steps following the direct pathway, without escape of the excess proton from the binding pocket to the bulk solution, suggesting the possibility of a long-lived enolate intermediate state at high pH. Although the calculated apparent inactivation rate is faster than the value reported from experiment,¹⁸³ the discrepancy is strongly affected by the fluctuations of the unlinked ligand associated with the kinase binding pocket and the probability that the warhead is positioned favorably to allow the chemistry to proceed. The experimental inactivation rate is matched exactly if one assumes a relatively modest uniform increase of 1.8-3.0 kcal/mol or 0.8-2.1 kcal/mol if a population of conformations with a favorably positioned warhead of 5% or 1% is assumed, respectively.

Ultimately, to achieve a quantitative description of covalent inhibition, the present exercise will have to be repeated with more extensive sampling from longer simulations along the optimized path collective variables¹⁹⁰ using QM methods of higher accuracy,^{191;192} more sophisticated treatments of the system's conformational dynamics,¹⁹³⁻¹⁹⁵ and of the transition kinetics.^{172-174;196;197} Despite its limitations, it is hoped that the present effort provides a useful overview of the different microscopic factors affecting the mechanism of covalent inhibition.

CHAPTER 4

MOLECULAR DYNAMICS-GUIDED DESIGN AND CHEMOPROTEOMIC PROFILING OF COVALENT KINASE ACTIVITY PROBES

Reproduced in part with permission from Chakraborty, P.,[†] Carlos, A.,[†] Thomas, T.,[†] Ghaby, K., Fathi, S., Sharma, M., Nguyen, N. K., Farmwald, M., Blachowicz, L., Alcyone, E., Harter, K., Yu, S., Pillai, K. S., Roux, B. & Moellering, R. E. Molecular Dynamics-Guided Design and Chemoproteomic Profiling of Covalent Kinase Activity Probes. *Journal of the American Chemical Society* **148**, 12641–12654 (2026).¹⁹⁸ Copyright 2026 American Chemical Society. The published version is available at <https://doi.org/10.1021/jacs.5c18691>. [†]These authors contributed equally. My main contribution was the kinetic modeling. Chemical syntheses and NMR validation are omitted from the Supplementary Information.

Abstract

Covalent small molecule activity probes can be powerful tools to interrogate protein function in native cellular environments. The design of family-wide activity probes requires an understanding of the molecular sources of general targeting potential and specificity to enable broad targeting of protein family members. Here, we developed and applied a multifaceted docking and molecular dynamics (MD) simulation pipeline to design and test cell-permeable covalent kinase activity probes from a set of hinge-binding pharmacophores. This computationally-guided approach yielded a new cell-active probe, K60P, which targets around 114 kinases across distinct kinase classes in live cells by targeting the conserved active site lysine. Chemoproteomic profiling of this probe and a clinical candidate sharing the same indazole core, KW-2449, identified kinase and non-kinase target profiles that differ from recombinant protein assay profiles, underscoring the utility of native kinase profiling

in situ. Bayesian Metropolis Monte-Carlo (BMMC) sampling modeling applied to single dose- and time- response data identified covalent efficiency and k_{inact} as key parameters for performance of covalent probes. Finally, focused molecular dynamics simulations revealed that K60P, as well as the comparator probe XO44, preferentially engage with target kinases in their active, DFG-in conformations, which is driven by increasing population of reaction-ready small molecule conformation. These results together establish a computational and kinetic modeling framework for designing covalent activity probes and highlight the balance of target selectivity and kinetic efficiency as a key factor in determining their proteome-wide reactivity.

4.1 Introduction

Protein kinases are involved in regulation of diverse cellular processes, including metabolism, cell cycle control, transcription, cell migration and cell death. Aberrant kinase activity contributes to many diseases, most notably cancers, where mutation, translocation, or amplification of specific kinases can contribute to unchecked proliferative signaling. Advances in medicinal chemistry and structure-based design have resulted in the development of small molecule inhibitors and FDA-approved drugs targeting conserved ATP-binding active sites of dozens of kinases for various disease indications. However, the fine-tuning of potent activity over structurally related kinases is an important aspect of kinase inhibitor development. In particular, identification of the on- and off-target interaction across the kinome as well as the broader proteome remains a challenge, and many kinase inhibitors are known to exhibit polypharmacology.^{179;199–203} The disparity between recombinant assays compared to that against native kinase targets in cells is well known. Technologies like KinomeScan,^{202;203} Kinobeads,¹⁷⁹ and KiNativ²⁰⁴ have been widely deployed to identify the target landscape of kinase inhibitors using either recombinant proteins or native proteins in cell lysates. However, these approaches suffer from several drawbacks. For instance, KinomeScan requires the

genetic engineering or purification of recombinant proteins, typically just interrogating the catalytic domain of target kinases and thereby does not replicate native cellular conditions. Kinobeads utilize immobilized kinase inhibitors to scan for native protein kinase binding in cell lysates, and KiNativ employs a reactive ATP-mimetic probe to label active kinases; both technologies perform these measurements in the artificial cell lysate environment. There remains a need for *in situ* profiling of native kinases and streamlined proteomic workflows to profile small molecule probes and potential selective therapeutics.

Covalent small molecules are attractive as both probes and therapeutics due to their potential for increased potencies, improved target residence kinetics, and high selectivity. Recently, several cysteine-targeted covalent drugs have been approved in cancer, including molecules targeting wild-type cysteines adjacent to the ATP binding site in protein kinases^{205;206} or neomorphic mutant cysteines in GTPases like k-RAS.²⁰⁷ Targeting non-catalytic cysteines near the ATP-binding site has served as a general approach to design irreversible kinase probes, including several FDA-approved drugs.^{118;208} Expansion to other nucleophiles like lysine and tyrosine is increasingly gaining popularity to expand the targetable kinome due to the absence of druggable cysteines near the active site of many kinases. Activated esters, vinyl sulfones, sulfonamides, and sulfonyl fluorides have been used to target lysines.²⁰⁹ In particular, sulfonyl fluorides have been widely used to target lysines due to their desired reactivity and hydrolytic stability in biological environments. For example, Taunton and coworkers installed a sulfonyl fluoride warhead on pyrimidine 3-aminopyrazole, which is a promiscuous kinase-binding scaffold, to generate a general lysine-reactive kinase probe (XO44) that can label active kinases directly in cells. XO44 and its fluorosulfonate and aldehyde analogs have been used to profile the catalytic lysines of the kinome,^{210;211} permitting interrogation of hundreds of kinases. However, XO44 targets a subset of the kinome and requires downstream click-chemistry labeling for target detection and quantification, potentially limiting applications like spatial or low-sample-input profiling.^{212–216} Additionally,

XO44 also labels many other proteins owing to its highly reactive sulfonyl fluoride covalent warhead, raising the question of whether less reactive kinase probes could be useful for in-cell or live tissue profiling studies.

Here, we elaborate a virtual screening workflow that combines docking and molecular dynamics (MD) to develop and study several covalent kinase activity probe scaffolds. This work follows considerable recent interest in applying and expanding computational screening methods for the design of covalent ligands.^{64;68} Guided by the computational modeling, we considered several core kinase-targeting scaffolds and ultimately synthesized and tested a novel sulfonylfluoride-modified indazole probe, K60P, that targets the catalytic lysine of kinases, specifically in their active conformations, *in vitro* and in cells. We also synthesized a biotinylated XO44 analog, X4K, to enable click-free in-cell protein labeling and streamline readouts and included it as a comparator in our proteomics and biochemical assays. Our integrated computational, biochemical, and cell-based profiling workflow helped rationalize kinetic determinants of selectivity and potency for specific kinases, including ABL1 as a model kinase.

4.2 Results

4.2.1 Computational design of covalent kinase activity probes identifies

K60P as candidate

When considering core kinase-targeting for the design of cell permeable activity probes, we focused on several hinge-binding scaffolds with reasonably promiscuous *in vitro* kinase profiles. These included amino-pyrazoles, pyridyl-imidazoles, indazoles and others, which are found in pre-clinical and clinical kinase inhibitors (*e.g.*, sunitinib²¹⁷ and KW-2449²¹⁸). To identify permissive sites for covalent warhead introduction around these hinge-binding core scaffolds, we developed a computational workflow comprised of docking followed by MD

simulations. From initial studies and published in vitro kinome profiles,^{202;203} we determined that the compact indazole pharmacophore related to the phase 1 clinical candidate KW-2449 was a suitable starting point for development of pan- or multi-kinase covalent probes. We determined the analog space to screen based on inspection of the predicted bound pose of KW-2449, inferred from its shared scaffold with axitinib, which has been co-crystallized with ABL1.²¹⁹ We also aligned representative members from seven kinase groups against the ABL1/KW-2449 model (Figure 4.1A, Figure B.1A) which revealed conserved alignment of hinge-binding residues and proximity of conserved catalytic lysines to the indazole and amino pyrimidine cores of KW-2449 and XO44, respectively. The alignment also highlighted the need to stay clear of the gatekeeper residue to avoid individual kinase selectivity.²²⁰

We focused on a library of analogs containing a phenylsulfonyl fluoride group introduced at the 4, 5, or 6 positions of the indazole heterocycle connected with linker lengths of 0, 1, or 2 atoms (Figure 4.1B, Figure B.1B). Initial docking was performed as a coarse screen to rapidly eliminate analogs where the addition of the linker and phenyl group introduced steric clashes or positioned the warhead outside of the binding pocket. The receptors for docking were 4WA9 (ABL1 + axitinib), 5K9I (SRC + XO44), and 3G0E (KIT + sunitinib), and were chosen based on ligands of interest and to expand the binding cavity around potential scaffold modification positions.^{210;219;221} Analogs that docked in multiple viable poses across multiple receptors were considered to have the highest potential. From this evaluation, analogs modified at the 5 or 6 position of the indazole, with linker lengths of 0 or 2 atoms were identified as the most promising cores for further study. We also included designs with a retrievable desthiobiotin or biotin tag linked to the piperazine of the KW-2449 pharmacophore for further computational screening, since neither tag interfered with target binding.

We subjected 12 analogs to further scrutiny by conducting unbiased MD simulations of their covalently linked complex, further expanding the set of analogs to account for the

possibility of attaching the warhead at the ortho-, meta-, or para- positions of the phenyl ring. Each analog was initially aligned to the crystallographic pose of axitinib in ABL1 kinase (PDB ID: 4WA9, chain B). A covalent bond was then introduced in the force field between the conserved Lys271 in ABL1 and the electrophilic warhead sulfur for each analog. The systems were equilibrated with the ligand’s indazole nitrogens and protein restrained, prior to being released for 100 ns of unbiased MD simulation with explicit solvent. The structural strain of the covalently linked pose was evaluated over the final 10 ns of MD as a combination of two metrics: 1) the position of the ligand indazole core relative to the axitinib defined pose; 2) any perturbation of the lysine backbone relative to the protein crystal structure (Figure 4.1B). We assumed that both metrics are reporters of the same strain within the covalently linked structure, with a high score in either indicating that the ligand in the kinase active site would be unlikely to adopt a suitable reaction-ready conformation with the catalytic lysine before covalent engagement. We also scored the candidates based on the minimum warhead-S to lysine-N distance between a conformational search of the linker and warhead of each analog and to the crystallographic lysine-N positions of structures 4WA9 and 5K9I with no consideration for sterics as a “naïve” structural metric (Figure 4.1B table). With the exception of K50P and K62O, all ligands maintained key hydrogen bond interactions to the hinge, although many did so from more strained conformations (Figure B.1C). Though several promising scaffolds were found, K60P was selected for synthesis based on the metrics as the best performing analog overall (Figure 4.1C–E). We also synthesized an analog, K50P, which showed a less favorable core displacement after bond formation with Lys271, but was still predicted to bind reversibly (Figure 4.1C–E, Figure B.22).

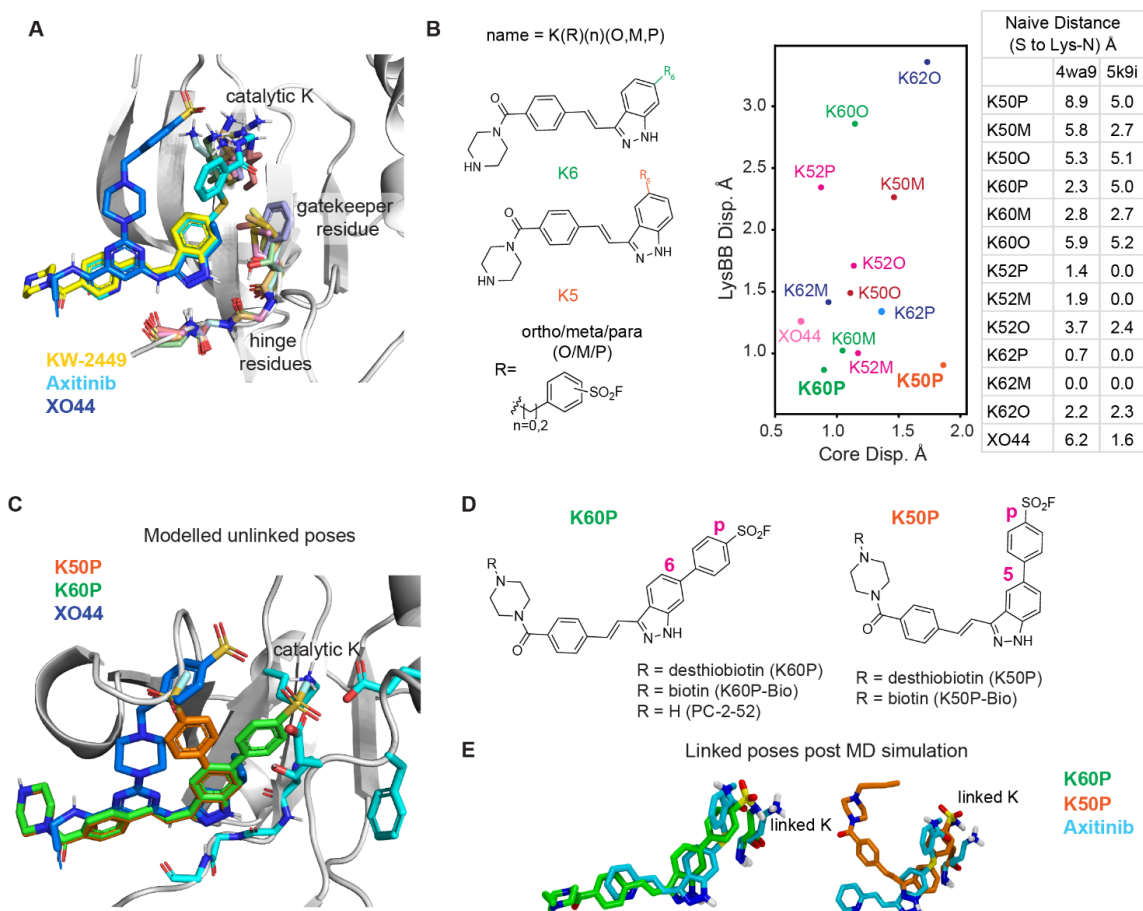


Figure 4.1: Computation-guided design for covalent, conserved lysine-targeting kinase probes. (A) Overlaid binding poses of indicated ligands within the conserved ATP-binding pocket defined by features shared between representative kinases CAMK1D, AKT1, MAP2K1, ALK2, ABL1, CDK1, CK1 α . (B) Left: probe library scaffold structure and naming convention. Middle: lysine backbone and core displacements (in Å) by each probe analog. Right: minimum achievable distance between the SO₂F warhead and the crystallographic pose of the lysine in 4WA9 (ABL1) and 5K9I (SRC) for each anchored analog. (C) Modeled unlinked poses of K60P (green), K50P (orange), and XO44 (blue) in ABL1 kinase, illustrating warhead positioning. (D) Chemical structures of K60P and K50P. (E) Final linked poses of K60P (left, green), K50P (right, orange), and linked lysine (yellow) after Molecular Dynamics (MD) simulations (100 ns). The crystallographic pose of Axitinib (cyan) and lysine is shown for comparison.

4.2.2 Proteome-wide profiling reveals kinase and non-kinase targets of K60P and KW-2449

We next explored the utility of K50P and K60P as cell-permeable kinase activity probes by probing the proteome-wide labeling of their desthiobiotin-labeled versions. Both probes showed dose-dependent covalent labeling of endogenous protein targets, as visualized by

streptavidin immunoblotting of desthiobiotin-tagged proteins in cell lysates and intact cells (Figure 4.2A). These data confirmed the suitable stability and cell permeability for *in situ* profiling of these probes. In accordance with our modeling predictions, K60P exhibited more robust labeling of the proteome than K50P in western blot-based assays. We therefore proceeded with K60P for further downstream analyses to identify the probe-labeled proteins. To identify the targets of K60P quantitatively, we synthesized the biotin-labeled version of K60P, K60P-Bio, and performed LC-MS/MS analysis using stable isotope labeling by amino acids in cell culture (SILAC) for quantitation of probe-labeled proteins (Figure 4.2B). We treated K562, HeLa, and MCF7 cell lines with K60P-Bio prior to enriching probe-labeled proteins on streptavidin beads. K60P-Bio enriched a total of 114 protein kinases in these three cell lines combined (Figure 4.2C–D), including known targets of the parent scaffold KW-2449 such as ABL1, AURKA and B, PAK2, BMP2K, RIPK1, TNIK, DYRK1A, and CSNK2A1 and A2.^{179;203;222} For ABL1, because SILAC quantification relied on peptides shared by BCR-ABL and WT ABL, we could not conclusively determine whether enrichment was biased for the oncogenic protein, however, ABL1 was only enriched in K562 cells where around 80% of ABL1 is BCR-ABL1.^{223;224} We also observed enrichment of kinases not previously annotated as targets of KW-2449, including YES1, LIMK1, TAOK1, 2, 3, CDK1, 2, 4, 5, 6, and RSK1 and 2. Several metabolic kinases, including HK1 and 2, DTYMK, AGK, ADK, CKB, and PANK4 were also enriched. Intriguingly, the kinome coverage of K60P was reduced compared to a biotinylated version of XO44, X4K, that we synthesized and profiled in K562 cells (Figure 4.2C, Figure B.3). X4K enriched roughly twice as many kinases in this context alongside a much larger total number of targets (~1800) compared to K60P-Bio.

K60P is closely related to the parent indazole, KW-2449, which was assessed in the clinic for targeting FLT3 and potentially ABL1, two oncogenic tyrosine kinases. Despite this structural origin, K60P demonstrated less engagement of TK and TKL kinase sub-families

relative to X4K (Figure B.3). Common targets between the two probes in these families included YES, LYN, BTK, CSK, PYK2, LIMK, RIPK1 and ABL1. By contrast, K60P engaged considerably more kinases in the STE, CK, AGC and CAMK sub-families, which extends the coverage of the kinome across these two cell permeable probes (Figure 4.2D). Beyond kinases, approximately 30% of the proteins labeled by both probes were annotated as ‘nucleotide binding’ from gene ontology analyses (Figure 4.2E). Additional classes of K60P-Bio and X4K-enriched proteins included those involved in protein synthesis, nucleic acid binding, and cytoskeletal proteins (Figure 4.2E, Figure B.4). This likely suggests that these off-targets are highly abundant proteins, rich in solvent-exposed lysines and tyrosines, which are non-specifically labeled by the electrophilic sulfonyl fluoride warhead, regardless of the probe scaffold.²²⁵

To further explore the specificity of kinase and other target enrichments at the protein level, we performed competitions between a non-biotinylated K60P analog, PC-2-52, and the previously validated kinase probe XO44. Five-fold excess PC-2-52 or XO44 were pre-incubated with K562 cell lysates prior to treatment with 10 μ M K60P-Bio. Both PC-2-52 and XO44 competed with K60P-Bio enrichment of many known kinase targets of its indazole progenitor KW-2449, including AURKB, CDK6, TNIK, CSNK2A2 indicating high confidence hits (Figure 4.3A–C, Figure B.5A–C). Gene Ontology ‘Molecular Function’ (GOMF) bioinformatic enrichment analysis showed kinases to be the major class of proteins competed, confirming that K60P-Bio has high specificity for kinases (Figure 4.3B). More than 60 kinases enriched by K60P-Bio were significantly competed by XO44 in K562 cells, including notable shared targets like YES1, DYRK1A, TK1, CSNK1A, CDK1,5 indicating high confidence hits (Figure 4.3C). XO44 also competed many other non-kinase off-targets of K60P-Bio, confirming that they are common off-targets of reactive sulfonyl fluoride probes (Figure B.5C). The inverse experiment, where PC-2-52 was used to compete with X4K-enriched protein targets also demonstrated significant competition with approximately 90 enriched protein kinases

(Figure B.6). Collectively, these dual-competition studies support the covalent, active-site targeted engagement of both K60P and XO44 (and its biotinylated analog X4K) in lysates and live cells.

Because K60P engaged a distinct subset of kinase and non-kinase targets in cells, we sought to use it in a competitive format to identify targets of the reversible parent scaffold KW-2449. SILAC-labeled K562 cells were pre-treated with 2 μ M KW-2449 or DMSO for 1 hour prior to K60P-Bio treatment, target enrichment, and analysis by LC-MS/MS. NTRK, TNIK, and AURKB,^{179;226} which are known targets of KW-2449, were significantly competed under these conditions. Several novel kinase targets, including RPS6KB2, MAP3K20, GSK3A, VRK2, as well as the metabolic kinases AGK and HK2, were significantly competed (Figure 4.3D–E). These data suggest that KW-2449 may exert considerable background target engagement of housekeeping enzymes under physiologically relevant conditions. Gene Ontology ‘Cellular Component’ (GOCC) bioinformatic enrichment analysis of competed KW-2449 targets indicated strong enrichment of proteins with reported localization to mitochondria, including several kinases (HK2, RAF1, VRK2 and AGK, (Figure B.7A), which could be related to the piperazine linker in this molecule and many other small molecule kinase inhibitors. There were numerous non-kinase targets that were significantly competed by KW-2449, about 30% of these included several GDP/GTP binding proteins such as DNM1, RAB proteins, and the immunomodulatory protein STING1 (Figure B.7B). This could potentially be attributed to binding site structural similarity, perhaps indicating a degree of molecular crosstalk that would be missed in recombinant kinase assays. Notably, ABL1 or BCR-ABL was not significantly competed by KW-2449, despite it being previously reported as a high-affinity and -value clinical target for development.²²⁶ We noted that competition between a covalent probe and a non-covalent competitor with the same pharmacophore likely biases toward capture of higher affinity targets and may miss other transient interaction partners. To explore this and confirm this was not unique to K60P en-

gagement of BCR-ABL1/ABL1, we performed a similar competition with KW-2449 against X4K-dependent protein enrichment in live K562 cells. Of the approximately 200 kinases enriched by X4K in these cells, KW-2449 showed significant competition with approximately 30, including previous annotated targets like BMP2K, NTRK1, AURKB and RIPK1, as well as target kinases not previously associated with KW-2449, such as several MAPK-family kinases, STK16, and ROCK2 (Figure 4.3E, Figure B.8).^{179;211} Intriguingly, BCR-ABL1 was enriched by X4K but not significantly competed by KW-2449, confirming the results from K60P competition experiments. In contrast, AURKB and C were significantly competed by KW-2449. In cell viability assays with K562 cells, KW-2449 showed potent GI₅₀ values of $\sim 0.4 \mu\text{M}$, but the response was majorly cytostatic, with viability plateauing at 20% residual signal (Figure B.9). This is consistent with prior reports showing strong AURK-driven G2/M arrest in K562 cells and comparatively weaker BCR-ABL pathway suppression and apoptosis upon KW-2449 treatment, with strong BCR-ABL dephosphorylation and cell death occurring mainly in combination settings.^{218;227} Together, these data suggest that the Aurora kinases and the other kinases competed in these assays are likely higher-affinity cellular targets of KW-2449, and that despite its annotation in some previous literature, KW-2449 is not a high affinity ligand for endogenous BCR-ABL or ABL1. This conclusion agrees with published bead-based profiling of native proteins in lysates,¹⁷⁹ and relatively low potency in cell based assays against BCR-ABL1 activity, but runs counter to work showing that KW-2449 potentially targets the catalytic domain of ABL1.²⁰²

4.2.3 Covalent probes display specific modes of active site labeling despite disparate kinetics for ABL1, a model target kinase

To understand the mechanisms underlying active site engagement and covalent modification by K60P and close analogs, we performed biochemical and biophysical studies with recombinant ABL1, a reported target of the progenitor scaffold and a focus of our computational

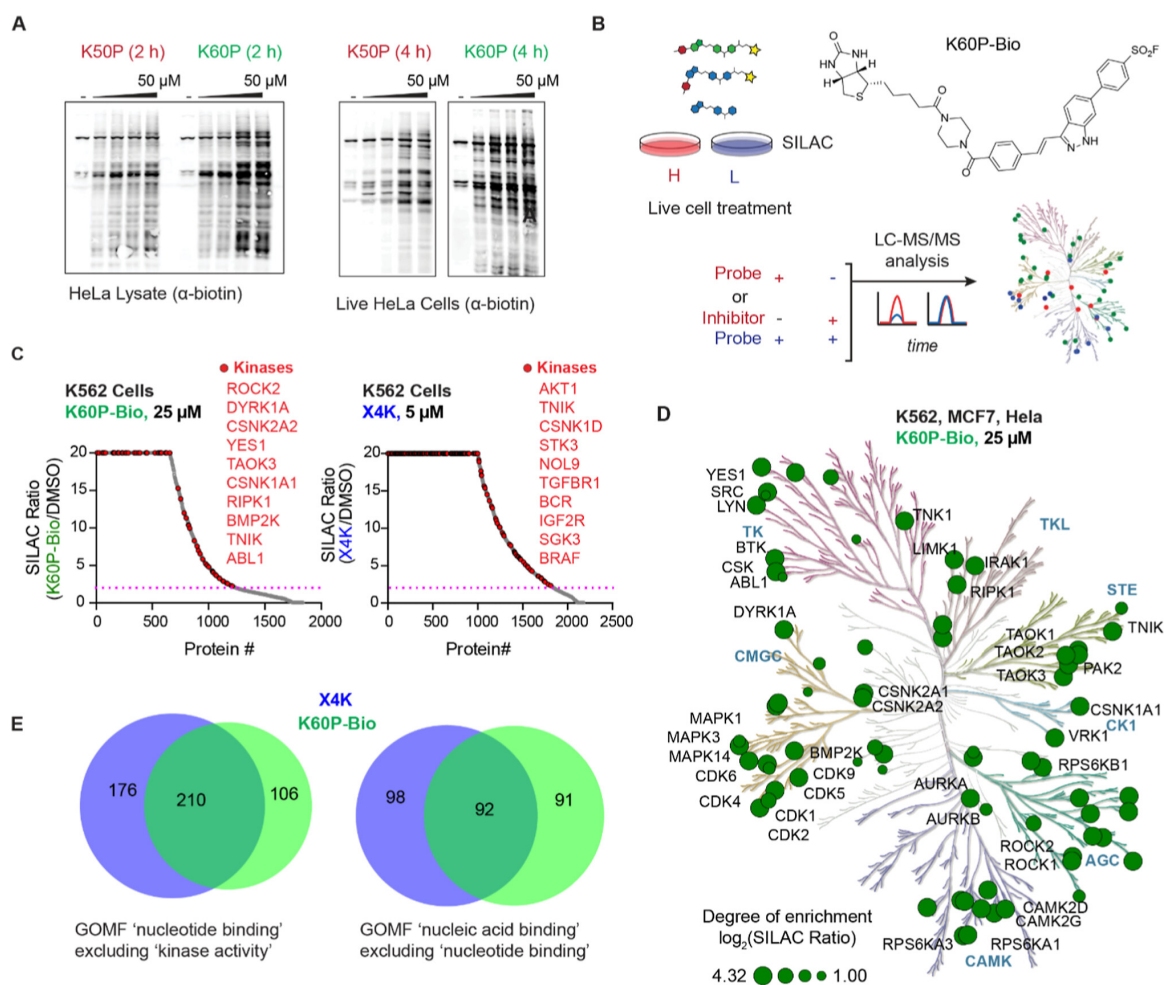


Figure 4.2: Proteome-wide labeling by K60P-Bio in lysates and live cells. (A) Representative Western blots demonstrate broad labeling of proteins in lysates (left) and in live cells (right) by K60P and K50P. (B) Workflow of kinase probe SILAC-MS/MS profiling experiments to identify proteins labeled by K60P-Bio (structure shown). (C) Waterfall plots of kinases labeled by K60P-Bio or X4K in live K562 cells treated with either K60P-Bio (25 μ M for 4 hours, left) or X4K (5 μ M for 1 hour, right). Control cells were treated with DMSO. Cutoff: SILAC Ratio ≥ 2 is shown as dotted line; $n = 2$ (D) Dendrogram showing kinases enriched by K60P-Bio in K562, HeLa, and MCF7 cell lines, $n = 2$ for each cell line. (E) Venn diagram of shared and unique nucleotide- and nucleic acid-binding proteins labeled by K60P-Bio and X4K in live K562 cells.

screening effort. We performed dose- and time-dependent labeling studies with recombinant ABL1 (residues 46-515) and K60P, K50P, and X4K. Western blot analyses confirmed sub-micromolar covalent labeling of ABL1 by K60P ($EC_{50} = 0.6 \mu$ M) and X4K ($EC_{50} = 0.7 \mu$ M), and less potent engagement by K50P ($EC_{50} = 2.9 \mu$ M) (Figure 4.4A, Figure B.10). Preincubation with PC-2-52 resulted in complete competition with both K60P and X4K with sub-micromolar IC_{50} values (0.5–1 μ M), suggesting that the predominant covalent labeling

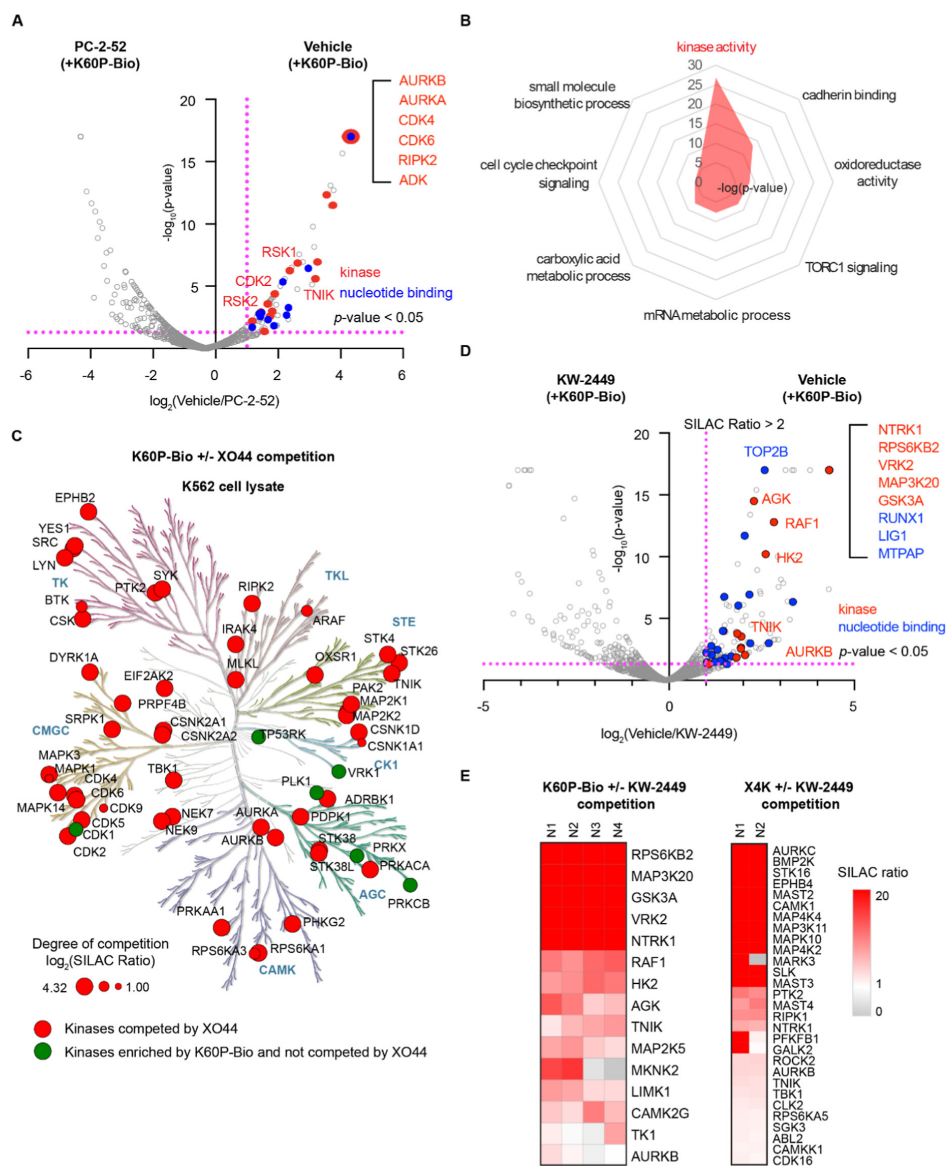


Figure 4.3: Competition with K60P-Bio reveals its specificity for kinases and known novel targets of KW-2449 in live cells. (A) Volcano plot of streptavidin-enriched SILAC ratios for K562 cell lysate treated with 50 μM PC-2-52 (2 hours) followed by with 10 μM K60P-Bio (1 hour). Cutoffs: competition SILAC Ratio ≥ 2 , $p < 0.05$; $n = 2$. (B) Gene Ontology Molecular Function analysis showing kinases as the major classes of proteins significantly competed by PC-2-52. (C) Dendrogram showing kinases that are both enriched by K60P-Bio and competed by XO44. (D) Volcano plot of streptavidin-enriched SILAC ratios for K562 cells treated with 2 μM KW-2449 (1 hour) followed by with 25 μM K60P-Bio (2 hours). Cutoffs: competition SILAC Ratio ≥ 2 , $p < 0.05$; $n = 4$. (E) Heatmap of kinases competed by over 2-fold in competition of KW-2449 (2 μM) against K60P-Bio (left) and X4K (right).

was occurring within the active site of ABL1 (Figure 4.4B, Figure B.11). Time-dependent studies showed X4K labeled recombinant protein to apparent saturation within minutes, which was much faster than either K60P or K50P (Figure 4.4C, Figure B.10). For the com-

petition of PC-2-52 with X4K where PC-2-52 was present during probe labeling, there was complete loss of X4K labeling at $\geq 1 \mu\text{M}$ PC-2-52, despite X4K being used in excess at $5 \mu\text{M}$ (Figure 4.4B). This indicated that PC-2-52 effectively outcompeted X4K for active site engagement. Although X4K may exhibit rapid covalent labelling once bound, its effective site engagement was surprisingly less efficient in these direct competition studies. These results were unexpected given the similar arylsulfonylfluoride warhead and molecular dynamics scores between K60P and XO44 for ABL1. This kinetic advantage could contribute to the apparent increase in kinome labeling by X4K, as well as higher labeling of non-kinase targets in the proteome.

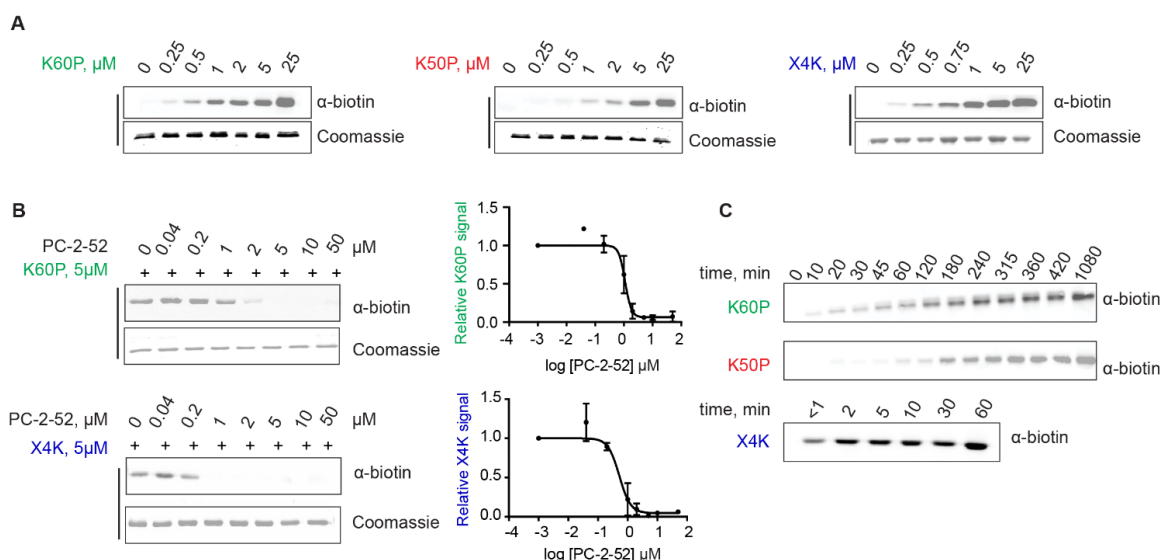


Figure 4.4: K60P, K50P, and X4K covalently label recombinant ABL1 with distinct kinetic profiles. (A) Streptavidin immunoblot showing dose-dependent labeling of the recombinant ABL1 protein with K60P (left), K50P (middle), and X4K (right) (0 – $25 \mu\text{M}$, 1 hour), $n = 2$. Coomassie stain was used as loading control. (B) Preincubation of recombinant ABL1 with PC-2-52 (non-biotinylated analog of K60P) dose-dependently decreases recombinant ABL1 labeling by both K60P (top) and X4K (bottom), $n = 2$. (C) Streptavidin immunoblot showing time-dependent labeling of recombinant ABL1 with K60P and K50P ($10 \mu\text{M}$, up to 18 hours) and X4K ($10 \mu\text{M}$, up to 1 hour), $n = 2$.

To confirm and compare the site(s) of covalent labeling by K60P and X4K, we performed focused proteomic mapping to detect and quantify the covalent adducts between each probe across the recombinant ABL1 surface. Recombinant ABL1 was incubated with each probe at $10 \mu\text{M}$ for 1 hour, followed by digestion and processing for peptide-level detection of

covalent adducts across lysine and tyrosine residues in the protein. We observed direct modification of the catalytic target lysine, K271, as the dominant adduct by both peptide spectral matches (PSMs) and ion intensity for both K60P and X4K (Figure 4.5A–B, Figure B.12). Two other sites of modification were also observed for K60P: Y226, located in the SH2 kinase linker, and Y393 in the activation loop (A loop), though these peptides were observed at lower relative ion intensities (Figure 4.5A, Figure B.13). By contrast, X4K labeling was observed at seven unique sites within ABL1, including the aforementioned K271, Y226, and Y393, as well as Y115, Y215, Y139, and K262 (Figure 4.5B–C, Figure B.13). Inspection of available ABL1 structures indicated that the phenolic oxygens in Y226 and Y393 do not come into reasonable proximity with the bound sulfonylfluoride warhead, however two possible configurations with distance of less than 5 Å between Tyr393 and the K60P warhead location could be found within a 400 μ s MD dataset of apo-ABL1 seeded from ABL1 crystal structures that was previously reported by Meng *et al.* (Figure B.14).²²⁸ Sulfonylfluorides react more readily with tyrosine than lysine²²⁹ but collectively this data suggests that such poses are short-lived and relatively rare. Another possibility supported by this data involves covalent labeling of the observed tyrosines *in trans* on another bound-ABL1 monomer, which should bring these tyrosines into close active site proximity due to their role as primary sites of *trans* auto-phosphorylation.^{230;231} Additionally, there are multiple other surface-exposed tyrosines with comparable or higher solvent-exposed surface area than Y226 and Y393, which also should generate reasonable tryptic peptides for detection by mass spectrometry (Figure B.15, Table B.1).²³² However, these sites were not observed to be modified by K60P or at comparably much lower levels for the more reactive X4K, indicating that the labeling of Y393 and Y226 may not be merely off-target non-specific labeling of exposed surface residues. This provides a tantalizing possibility that the phosphorylatable tyrosines of an active kinase monomer could be labeled by a probe reversibly bound to a second kinase monomer, through a similar mechanism (Figure 4.5D). Indeed, reversibly bound K60P places

the reactive warhead in proximity of the catalytic lysine as well as in the vicinity of the active site entrance involved in phosphotransfer to these tyrosine targets of auto-phosphorylation (Figure 4.5E). This mechanism would support the possibility of covalent targeting of kinase substrates *in trans* (Figure 4.5D–E); additional focused work to elucidate the mechanism and generality underlying these observations is needed.

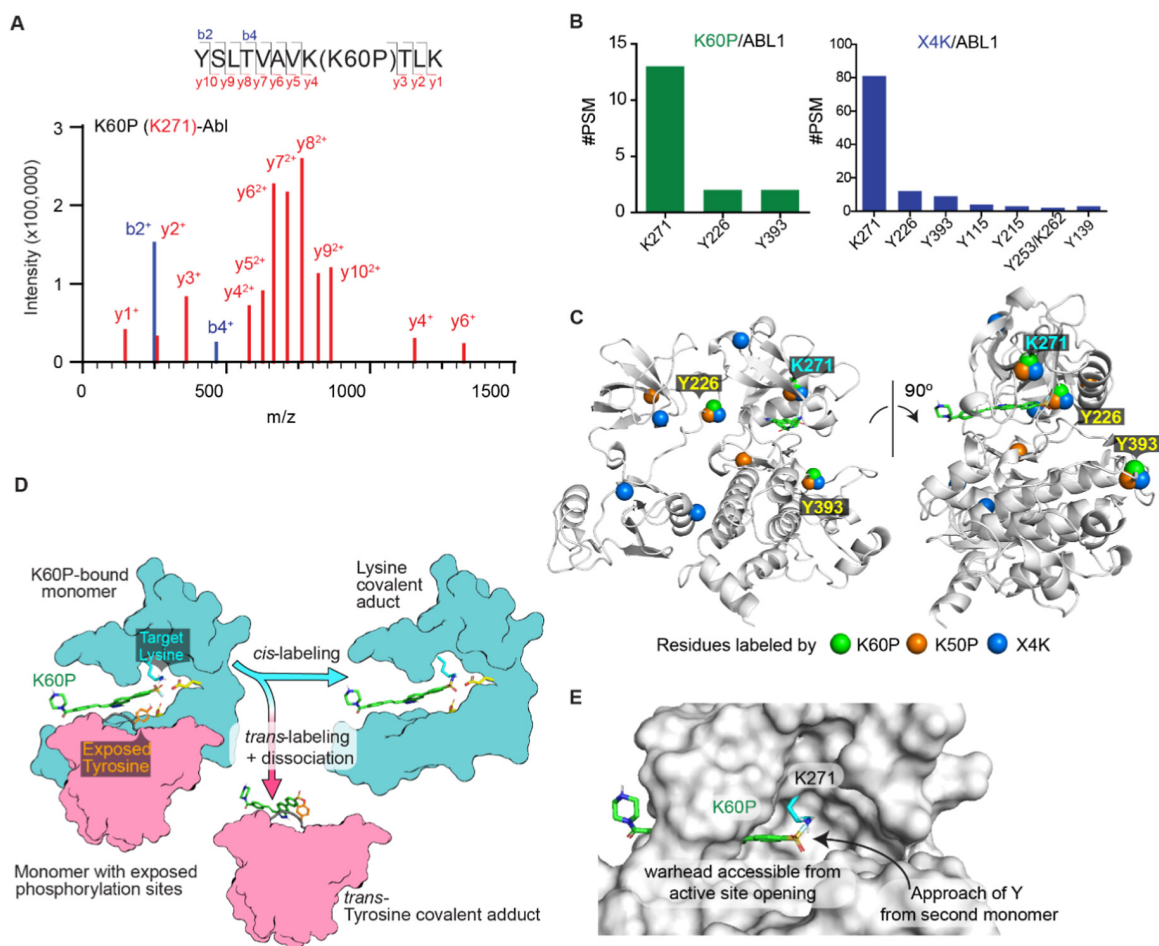
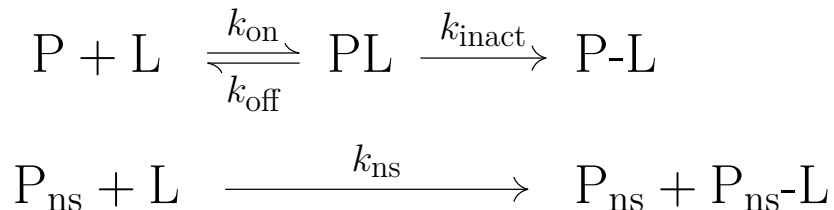


Figure 4.5: K60P and X4K covalently label distinct sites in ABL1 with *cis*-labeling of conserved lysine and a proposed *trans*-labeling of phosphorylatable tyrosines. (A) MS/MS spectrum of trypsinized ABL1/K60P adduct showing the labeled K271. (B) ABL1 sites labeled by K60P (green, left) and X4K (blue, right). Bar graphs show the number of peptide spectral matches (#PSMs) for each site. (C) ABL1 structure showing the positions of residues labeled by K60P (green), K50P (orange), and X4K (blue) relative to K60P docked in the active site (green). K271, Y226, and Y393 are labeled by all probes. (D) Proposed model for *cis*-labeling of K271 and *trans*-labeling of Y226/Y393 in ABL1 by K60P. Structural model showing K60P docked in the active site of one ABL1 monomer (teal), either reacting with K271 of the same monomer (*cis*) or with exposed tyrosine phosphorylation sites Y226/Y393 on a neighboring monomer (pink, *trans*). (E) K60P docked into ABL1 active site showing that the probe is positioned to access K271 within the same monomer, while its reactive group remains exposed at the active site opening for approach of neighboring monomer tyrosines.

4.2.4 A BMMC kinetic modeling framework captures key determinants of covalent probe efficiency

Given the disparate kinetics of covalent and kinome-wide labeling of these model sulfonylfluoride probes, as well as the large differences in probe performance in recombinant assays, we sought to rationalize probe performance by quantitatively characterizing the kinetics underlying probe reactivity with ABL1 and other kinases. Conventional dose-rate fitting can implicitly subsume rapidly fluctuating microscopic intermediates (like reaction-ready conformations) and obscure the true determinants of covalent potency. We therefore explored a wide range of possible kinetic parameters to fit experimental covalent labeling data within a Bayesian statistical modeling framework using extensive Metropolis Monte-Carlo sampling²³³ (BMMC). In this BMMC procedure, experimental time- and dose-dependent probe labeling data with recombinant ABL1 were fit using parameterized mass-action simulations of irreversible binding between protein P and ligand L (Scheme 1).³⁷



Scheme 4.1: Schematic basis of numerical kinetic simulations.

An initial least-squares (LS) fit normalized the experimental labeling intensities to the fraction of covalently-bound protein and provided a starting point for the BMMC. Sampled parameters included the apparent rate constants of association (k_{on}), dissociation (k_{off}), inactivation (k_{inact}), and nonspecific-binding (k_{ns} , e.g., non-specific labeling of surface residues). To focus the sampling on the equilibrium binding affinity ($K_{\text{d}}=k_{\text{off}}/k_{\text{on}}$), for K60P and K50P, the k_{on} was fixed (diffusion-limited) while for X4K, k_{off} was fixed since k_{on} appeared to be rate-limiting. The median (or fixed value) of each parameter together with the 95% highest density interval is tabulated in Table B.2. To relate the fitted rate con-

starts with familiar empirical constants of irreversible kinetics, the affinity, the inactivation constant ($K_I=(k_{\text{off}}+k_{\text{inact}})/k_{\text{on}}$), and the covalent efficiency ($C_{\text{eff}}=k_{\text{inact}}/K_I$) were calculated for each BMMC sample and are also included in the table.^{13;87;105} Simulated dose- and time-responses using reasonable sets of sampled rate constants are shown in Figure 4.6. The predicted rate constants in the regime of the K50P and K60P indicate a relatively slow k_{inact} (Figures B.16–B.18), and thus the rate of inactivation (*i.e.*, $k_{\text{inact}} \times [\text{PL}]$), depends on both k_{inact} and the affinity ($K_d=k_{\text{off}}/k_{\text{on}}$). As such, K50P and K60P maintain an effective equilibrium of the reversible binding step and fall in line with the rapid-equilibrium approximation that is fundamental to conventional empirical models.¹⁰⁴ In contrast, the BMMC sampling suggests that X4K displays a relatively fast k_{inact} , which precludes the equilibrium of the reversible bound state. For X4K, the reversible binding affinity is not resolved by the measurements as the inactivation occurred on a timescale comparable to or faster than reversible binding. Furthermore, the dose-response data for X4K does not appear to reflect a saturation curve of the reversibly associated state, PL, which hinders the precision of rate constant estimations and is explained further in the Supplemental Kinetic Modeling Discussion. Notwithstanding the uncertainty, the modeled lower bound of sampled k_{inact} for X4K is over 20 times greater than the upper bound for K60P and the inherent affinity, as fit by $k_{\text{off}}/k_{\text{on}}$ parameters, does not explain the observed differences. The estimated potency (*i.e.*, covalent efficiency), which describes the proportionality of the rate constants that was conserved across variations of K50P and regimes of X4K, provides a more reliable comparison between the probes, showing that X4K has the highest modeled potency ($C_{\text{eff}} \approx 3.3\text{--}13.5 \times 10^{-4} \mu\text{M}^{-1}\text{s}^{-1}$), followed by K60P ($\approx 0.69\text{--}2.9 \times 10^{-4} \mu\text{M}^{-1}\text{s}^{-1}$) and then K50P ($\approx 0.16\text{--}0.53 \times 10^{-4} \mu\text{M}^{-1}\text{s}^{-1}$). This approach suggested that complex descriptors of covalent probe potency and kinetics could be predicted with straightforward dose- and time-dependent protein engagement data. For comparison, we experimentally determined k_{inact}/K_I of K60P, K50P and X4K with more elaborate time- and dose-response series using

recombinant ABL1 protein. Notably, experimental $k_{\text{inact}}/K_{\text{I}}$ values were $0.9 \times 10^{-4} \mu\text{M}^{-1}\text{s}^{-1}$ for K60P and $0.23 \times 10^{-4} \mu\text{M}^{-1}\text{s}^{-1}$ for K50P, aligning with the modeled ranges from BMMC (Figures B.19–B.20, Table B.3). The $k_{\text{inact}}/K_{\text{I}}$ for X4K was more challenging to determine using these experimental assays due to rapid saturation of labeling with recombinant ABL1, although these studies supported the conclusion that X4K displays a micromolar K_{I} with a considerably higher k_{inact} compared to K60P or K50P (Table B.2). Indeed, these data align with all previous experiments in cell lysates, live cells and recombinant protein that place the covalent efficiency of X4K and its progenitor XO44 10-20 fold higher than the K60P comparator that displays a similar K_{I} , underscoring the importance of adopting reaction-ready conformations — captured by k_{inact} , for productive covalent inactivation as a leading parameter for suitable covalent activity probes or targeted inhibitor development (Figure 4.2C, Figure 4.4A–C, Table B.2).

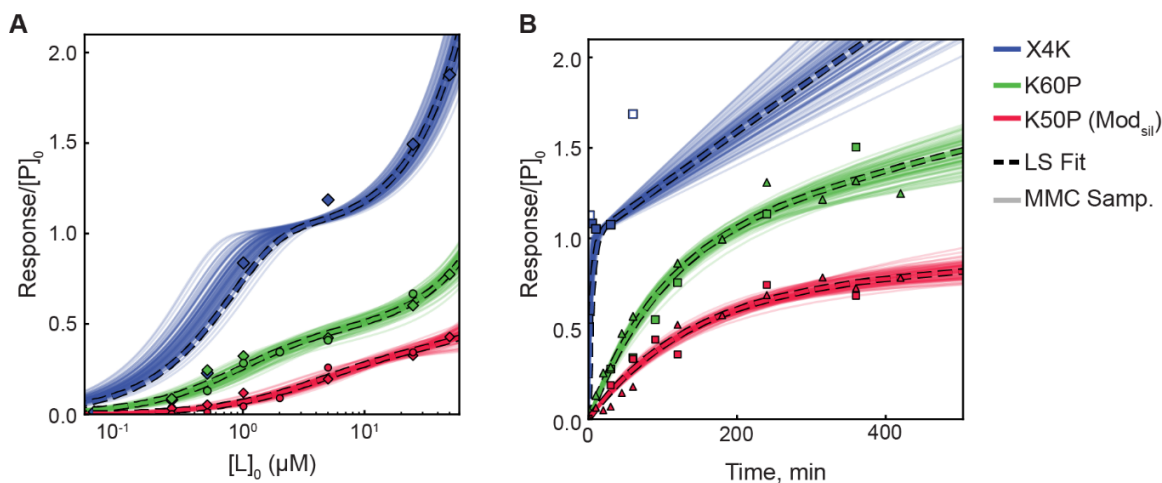


Figure 4.6: Kinetic modeling reveals complex kinetics of ABL1 labeling by K60P, K50P, and X4K. Experimental and simulated dose-responses (A) and time-responses (B) of K60P (green), K50P (red), and X4K (blue) binding to ABL1. The experimental data are scaled according to the least-squares (LS) minimization (dashed lines), and the families of solid lines correspond to the kinetic simulations of every 10 BMMC samples whose rate constants exist within their respective 95% highest density intervals. Simulated response is the sum of the specifically- and nonspecifically-bound covalent complexes.

*4.2.5 Molecular dynamics modeling rationalizes “active” kinase
conformation preference and inherent differences in covalent probe
kinetic efficiency*

Kinetic modeling indicated that the slower rate-limiting processes for K60P and K50P compared to X4K likely occur post-association within the kinase active site. To account for the differences in k_{inact} , we considered two factors: 1) the differences in the intrinsic reactivity of the warhead leading to the formation of the covalent bond; 2) the likelihood of a reaction-ready arrangement within the binding pocket, in which the target lysine is deprotonated and the ligand warhead is in close-proximity. To assess the intrinsic chemical reactivity of the different warheads to form the covalent bond with the lysine, we conducted labeling experiments of X4K and our probes with bovine serum albumin (BSA) as a model non-specific target protein. This experiment indicated ~ 1.7 -fold higher labeling intensity for X4K compared to K60P over the same treatment time (Figure B.21). This was in close agreement with the non-specific labeling rates estimated from the kinetic modeling (Table B.2). However, Gilbert et al.²²⁹ explored the kinetics of a series of XO44 analogs and showed that a <100 -fold difference in intrinsic reactivity did not correlate directly with significant changes in on-target labeling in model systems. Based on these observations, we conclude that differences in intrinsic reactivity are likely to partly account for the slower rate of labeling for K60P relative to X4K, and intrinsic reactivity of the latter likely contributes to off-target binding profile within the proteome.

Another requirement of the reaction-ready arrangement is the deprotonation of the lysine that enables a nucleophilic attack on the warhead. The likelihood of the lysine being neutral is expected to be influenced by its potential salt-bridges to both the DFG-motif-Asp and α C-helix-Glu. The presence of these salt-bridges has been predicted to result in an increase of two $\text{p}K_{\text{a}}$ units,²³⁴ which is enough to decrease the likelihood of a reaction-ready configuration with a neutral lysine by a factor of 100, and hence alter the apparent rate k_{inact} . The

K60P warhead is positioned much closer to the salt-bridge partners than that of XO44 and is more likely to affect, and be affected by, the dynamics of the associated structural components, particularly of the DFG-motif, suggesting that kinase activity state could be a major contributor to probe labeling. To explore the dynamics of the ligand and binding pocket prior to the formation of the reactive complex, and to ascertain any sensitivity to the protein conformation, we conducted a series of unbiased, 4 μ s MD simulations of both K60P and XO44 probes reversibly associated but before covalent engagement (unlinked) with ABL1 and SRC kinases, including DFG-in and DFG-out conformations. The results of the simulations are summarized in Figure 4.7, as Empirical Cumulative Distribution Function (ECDF) plots (Figure B.22) and discussed further in the supplementary molecular modeling discussion (Section B.4). Across these simulations, we observed increased reaction-ready proximity (warhead-S to lysine-N distances of < 4.5 Å) in the DFG-in conformations for both K60P and XO44, relative to the DFG-out conformations (Figure 4.7A). This preference can be explained by the observation that the DFG-out phenylalanine largely blocks proximity between the lysine and warhead of both K60P and XO44 when bound to the DFG-out conformation of ABL1 (Figure 4.7B), suggesting that each probe is more likely to label kinases in an ‘active’ conformation. Comparing the dynamics of ABL1 and SRC in the DFG-in conformation revealed that, while the K60P warhead maintains overall proximity to the catalytic lysine, it exhibits a reduced probability of achieving a reaction-ready proximity relative to XO44. This appears to be because of over-confinement of the warhead within the narrow kinase cleft, which restricts lysine rotameric flexibility and the approach geometry required for nucleophilic attack (Figure 4.7C).

The dynamics also inform us about the likelihood of a neutral reaction-ready lysine through the dependence of its pK_a on the protein conformation. Increased sampling of the DFG-out conformation, which may have a lower lysine pK_a with ligand bound, is a plausible mechanism to increase the likelihood of reaction-ready lysine.²³⁴ Both ligands showed

a preference for the DFG-in conformation, but XO44 has been co-crystallized with other kinases in the DFG-out conformation,²¹⁰ and in ABL1 the extended P-loop conformation that it binds to is associated with the inactive state.²³⁵ The K60P complex may also increase the likelihood of reaction-ready lysine by another mechanism, as simulations of the DFG-in conformation showed it could break the Lys to α C-helix-Glu salt-bridge by inducing the α C-helix-out conformation (Figure 4.7D), and also to weaken through sterics the salt-bridge between the DFG-Asp and the lysine. Other chemical steps may also be relevant,¹⁰⁷ a second deprotonation step of the intermediate complex also has potential to be rate-limiting and may be more easily deprotonated for XO44 where the larger unlinked warhead to lysine distance results in the reaction intermediate being further away from the salt-bridge partners. There are multiple possible reaction pathways, which could be investigated via QM/MM simulations.⁸⁴

In summary, the overall slower labeling observed for K60P compared to X4K is likely a combination of both unexpected but lower intrinsic reactivity, and lower sampling of reaction-ready warhead-lysine proximities. But these differences are relatively small compared to the differences in modeled k_{inact} . It is likely that the exceptional kinetic performance of X4K (notably over very similar analogs²²⁹) is due to subtle differences in the kinetics of nucleophilic attack by Lys271 and the subsequent steps of deprotonation and fluoride expulsion. These findings also align with the notion that kinetically tuned electrophilic probes can show higher on-target labeling that is resistant to quenching by high abundance off-target proteins in whole proteome or live cells. Overall, the apparent preference for DFG-in kinase conformations by both probes supports their utility as activity probes in cellular settings and warrants more global investigation across diverse kinases.

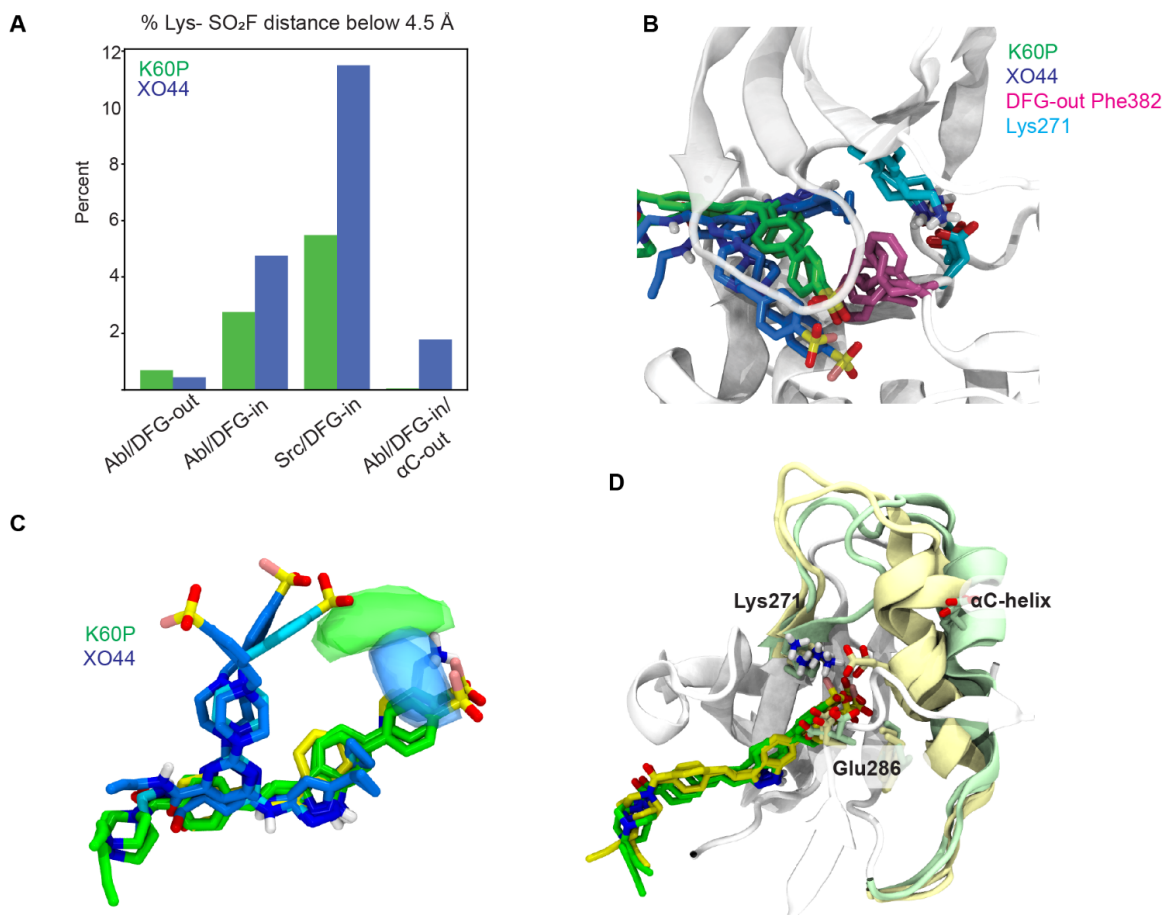


Figure 4.7: Molecular dynamics reveals preference of both K60P and X4K for DFG-in active conformation of kinases. (A) Estimation of the reaction-ready population in each system, determined by percentage of frames with < 4.5 Å lysine to warhead-S distance in each $4 \mu\text{s}$ simulation, excluding the first 100 ns. (B) Final structures of K60P (green) and XO44 (blue) in Abl/DFG-out simulations (both replicates overlaid). The DFG-out phenylalanine (pink) fences the warhead of both ligands away from the lysine (cyan, salt-bridging with DFG-Asp) and displaces the K60P core. (C) Positions of probes and the differing regions of highest lysine occupancy (blobs) for K60P (green) and XO44 (blue) for each of 2 replicates of ABL/DFG-in simulations at $\sim 4 \mu\text{s}$. Expected poses for KW-2449 and lysine (yellow), and XO44 (cyan) are shown for reference. (D) Breaking of the Lys-Glu salt bridge and αC -helix adopting the out conformation in K60P ABL/DFG-in simulations. Two replicates are shown at both 250 ns (yellow) and at 500 ns (green).

4.3 Discussion

Here, we aimed to apply docking and molecular dynamics simulations to identify new covalent kinase activity probe scaffolds and to understand the contributors to probe specificity and kinetic efficiency. We focused on the hinge-binding indazole pharmacophore found in the clinical candidates KW-2449 and axitinib, under the premise that these molecules have a relatively promiscuous *in vitro* kinase inhibition profile and thus could serve as general

kinome-targeted activity probes in lysates, live cells and tissues. Through a computational docking approach, we identified a subset of phenylsulfonylfluoride containing analogs for further study, including a cell-active probe, K60P. Direct target-engagement and competition profiling proteomics demonstrated that K60P and its progenitor KW-2449 broadly target kinases as well as targets in related nucleotide-binding pathways. These targets and profiles could be missed by focused recombinant profiling assays. This point is further emphasized by the fact that KW-2449 only weakly inhibited labeling of native BCR-ABL1 in cells by either K60P or X4K, despite being stated as a clinical development target. This discrepancy can be attributed to the competition with intracellular ATP in live cells as well as the difference between recombinant catalytic domains and full-length, native proteins in cellular environments.²³⁶ These results indicate that, despite prior *in vitro* profiling of KW-2449 in the catalytic domain and our initial docking and molecular dynamics suggesting nanomolar potency and favorable binding of KW-2449 and related analogs toward ABL1, the effective affinity for native ABL1 is relatively weak in the high nanomolar to low micromolar range. In contrast, Aurora kinases B and C were strongly competed in our competition profiles. Our viability assays and prior reports show that the dominant anti-proliferative effect in K562 cells is caused by G2/M arrest at lower concentrations consistent with AURK inhibition whereas BCR-ABL inhibition occurs mainly at higher doses. These data suggest our assays preferentially captured the higher-affinity cellular targets of KW-2449 and more accurately represent the target profile of KW-2449 at low therapeutically relevant concentrations. These data add weight to the general consensus supporting the implementation of probe selectivity profiling beyond biochemical assays throughout the therapeutic discovery pipeline.^{25;237}

Cellular profiling and computational modeling studies demonstrated that K60P engages more than 100 kinases across three model cell lines and though it is not as promiscuous as XO44, K60P expands coverage across several kinase sub-families. While broad-spectrum kinase probes like XO44 are valuable for mapping inhibitor target landscapes, more selective

cell-permeable kinase probes like K60P-Bio may be better suited for applications such as biomarker discovery and pathway studies, where off-target engagement of multiple kinases can make results hard to interpret. Greater selectivity for target kinase can also improve the signal-to-noise, especially important for low-abundance targets. Together, these probes expand the scope of targeted kinases by cell-permeable kinase probes (Figure 4.2). Our biochemical, proteomics and kinetic modeling studies make a clear case that the rapid X4K labeling of ABL1 relative to K60P is due to a faster k_{inact} , and not a reversible binding affinity step. Here, we suggest that increased kinetic efficiency of covalent labeling increases the utility of kinase probes across diverse kinases, as seen here between K60P (lower kinetic efficiency, lower global kinome labeling) and XO44 (high kinetic efficiency, increased global kinome labeling). The increased performance of kinetically efficient probes likely involves less competition and absorption by abundant off-target proteins sites in the proteome, resulting in increased ‘on-target’ protein labeling within the kinome. The general design strategy for cell-permeable kinome probes has been to choose promiscuous scaffolds with inherently high affinity across their targets and functionalising them with a general reactive warhead. Here, we conclude that beyond suitable non-covalent affinity and warhead reactivity, it is critical to also optimize other components of k_{inact} for increasing the residence time of a ligand ‘reaction ready’ conformation to develop promiscuous or class-specific probes. We found assessing strain in the covalently linked pose to provide an efficient estimate of the ligand-conformation component of k_{inact} during screening, however, subsequent analysis in this study suggested that components of k_{inact} related to deprotonation and protein conformation were more likely to be rate-limiting and thus more predictive, albeit requiring significantly greater computational resources to screen.

We also demonstrated for the first time that BMCC fitting of low complexity dose- and time-dependent covalent labelling data can be used to accurately predict more complex descriptors of covalent inhibitor potency and inhibitor efficiency. Our results comparing K60P,

K50P and X4K demonstrate that IC_{50}/EC_{50} data commonly used to describe covalent inhibitors and probes fall short of describing target-specific covalent potency. Instead, C_{eff} and k_{inact} are the most important terms to optimize for both promiscuous activity probes and targeted covalent inhibitor development. We believe this computational kinetic modeling framework could streamline use in other chemical proteomic and covalent inhibitor development campaigns. We implemented molecular dynamics simulations to rationalize the potential source of improved k_{inact} for XO44 (or its biotinylated probe analog, X4K) relative to K60P, despite their similar K_I and covalent warheads. These studies suggested that both probes prefer the active, DFG-in conformations of ABL1 and SRC, a preference that is likely similar for other kinases in the proteome and underlines differences between native proteome and *in vitro* profiling studies. These data supports their utility for comparative kinase profiling in cell and patient samples, inhibitor selectivity profiling, as well as more focused spatial and single-cell activity profiling studies.^{212–215} Finally, we posit that the combination of computational docking, molecular dynamics and kinetic BMMC simulation described here can be used more generally to identify suitable ligand scaffolds for synthesis and testing, followed by optimization by kinetic modeling to identify more efficient paths to reaction-ready conformations.

CHAPTER 5

DETERMINANTS OF COVALENT INHIBITOR RESIDENCE TIME REVEALED BY MOLECULAR AND KINETIC SIMULATIONS

Abstract

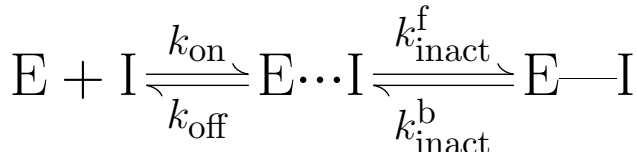
Covalent inhibition proceeds through a hierarchy of steps occurring across widely disparate timescales, including noncovalent association, conformational fluctuations within the bound complex, and covalent bond formation. For reversible covalent ligands, the nascent covalent bond may subsequently break, permitting dissociation of the ligand from its target. Because most experimental observables reflect the kinetics of an inherently nonequilibrium process, an atomistic understanding of covalent inhibition is essential for extracting mechanistic insight from experimental data and guiding the rational design of new covalent inhibitors. Here are laid out the elements of an integrated Bayesian strategy in which multiple sources of time-course data are fit together with detailed mass-action kinetic models using results predicted by atomistic simulations as priors. As an illustrative example, we examine three ibrutinib-derived reversible covalent inhibitors of Bruton’s tyrosine kinase (BTK) that exhibit markedly different residence times, ranging from 1 to 200 hours, despite only minor structural differences. For each ligand, the conformational fluctuations of both the covalently linked and unlinked complexes were characterized by molecular dynamics simulations, the potential of mean force over the initial deprotonation of the retro-Michael reaction was estimated from QM/MM simulations, and the noncovalent binding affinity was determined from absolute binding free energy calculations based on alchemical molecular dynamics simulations. The results of these atomistic calculations were then incorporated into a detailed mass-action kinetic model to fit time-course experimental data from multiple sources. Careful analysis shows that no single factor is dominant and that different microscopic steps may

be responsible for the observed residence time of the different reversible inhibitors.

5.1 Introduction

Covalent inhibitors couple molecular recognition to covalent bond formation, allowing potency and selectivity to arise from conditional reactivity rather than noncovalent affinity alone.^{10;33;41} While the end goal is clear, the mechanistic leverage that makes covalent inhibition attractive as a pharmacological strategy complicates its design. Covalent inhibition is controlled by a hierarchy of steps including diffusional encounter, noncovalent association, conformational rearrangements of the enzyme–inhibitor complex, covalent bond formation, and possibly bond cleavage. These processes can occur on vastly disparate timescales, spanning nanoseconds for fast conformational fluctuations to days for slow covalent bond turnover.²³⁸

The standard model of reversible covalent inhibition is represented by Scheme 5.1. First,



Scheme 5.1: Traditional two-step kinetic scheme of reversible covalent inhibition.

the solvated enzyme and inhibitor may undergo bimolecular association with the rate constant k_{on} into the noncovalent complex $\text{E} \cdots \text{I}$. The noncovalent complex may either dissociate into the solvated state with the rate constant k_{off} or form the covalent adduct $\text{E} \text{---} \text{I}$ with the forward inactivation rate constant $k_{\text{inact}}^{\text{f}}$. In reversible systems, the nascent covalent bond can break, returning the system to the noncovalent state with the backward inactivation rate constant $k_{\text{inact}}^{\text{b}}$. In reality, it is important to realize that all rates in this Scheme are effective or apparent, as they are expected to hide significant microscopic details. For example, the noncovalent complex $\text{E} \cdots \text{I}$ may undergo rapid internal dynamics between reactive and nonreactive poses.⁸⁴

A practical consequence of covalent binding is that most experimental readouts are actually reporting on various aspects of nonequilibrium behavior. For an idealized assay probing the action of excess covalent inhibitor with a protein target, the long-time equilibrated outcome is $\sim 100\%$ depletion of the free protein, and metrics derived from the equilibrium distribution of bound and unbound states are therefore not informative.¹⁰⁵ The metrics that remain useful, *e.g.*, time-dependent IC_{50} shifts, k_{inact}/K_I , and residence times, are obtained by fitting analytical expressions derived from simplified kinetic schemes to a biochemical observable measured during the relaxation to equilibrium.⁸⁷ These expressions describe the time- and dose-dependence of inhibition with a few apparent parameters which, due to the simplification, absorb contributions from multiple microscopic processes. The resulting kinetic degeneracy makes it difficult to diagnose which microscopic step limits, differentiates, or contributes to the observed behavior.³⁷

Structural modifications may improve potency by decelerating dissociation of the non-covalent complex, increasing the probability of reactive conformations, lowering an energy barrier in the covalent action, or (in reversible systems) suppressing bond cleavage. The optimal design intervention depends on the magnitude, determinants, and overall contribution of the timescale associated with each stage. All these issues become especially important in the rational design and optimization of targeted covalent inhibitors, whose mode of action is sensitive to both thermodynamic and kinetic factors. A fundamental understanding of the process of covalent inhibition at the atomic level is required to make full use of available experimental data, and ultimately enable a rational design strategy for the discovery of new covalent inhibitors.

These challenges motivate an integrated strategy in which multiple sources of time-course data are fit together with detailed mass-action kinetic models that include states and rate constants predicted by atomistic simulations. Atomistic characterization of the covalent step via quantum mechanical/molecular mechanical (QM/MM) and alchemical free energy meth-

ods has been applied across diverse warhead classes and targets,^{59;64;239} and mass-action fitting of enzymatic time courses is the mechanistically explicit, albeit highly parameterized, strategy from which more practical models were derived.^{35;240;241} This decomposition is essential for rational optimization.

Protein kinases provide a clinically relevant testbed for such an approach. Kinases are implicated in many malignancies and inflammatory disease, motivating sustained efforts in drug discovery.^{17;117} Their binding sites exhibit recurrent reactive residues (notably cysteines and lysines) and conserved architecture for the endogenous small-molecule ligand, ATP, enabling both irreversible and reversible covalent small-molecule strategies.¹¹⁸ One particular success story is ibrutinib, an irreversible covalent inhibitor of Bruton’s tyrosine kinase (BTK) approved for B-cell malignancies.^{18;84;114;123}

The present study focuses on three ibrutinib-derived reversible covalent BTK inhibitors developed by Bradshaw et al.³⁶ (compounds 1–3 therein; hereafter 4A9, 4B9, and 4C9). Despite small structural differences, the estimated residence times differ markedly: 0.53, 91, and >200 hours, respectively. Building on a previous investigation into the irreversible binding mechanism of ibrutinib to BTK,⁸⁴ these three compounds are particularly interesting for a computational study. They share the cyanoacrylamide-based warhead which is the canonical reversible Michael acceptor for cysteine-targeted inhibition.²⁷ Further optimization of related molecules produced rilzabrutinib, recently FDA-approved for immune thrombocytopenia.^{242;243} Importantly, the availability of experimental inhibitor dissociation data enables direct interrogation of residence time, an experimentally accessible parameter that more closely reflects *in vivo* efficacy than do single-time-point potency or covalent efficiency ($k_{\text{inact}}/K_{\text{I}}$) measurements.²⁴⁴ Furthermore, reversibility in covalent inhibition is an attractive feature that offers a tunable compromise between durability and safety, mitigating liabilities associated with permanent modification while retaining the possibility of prolonged occupancy.^{53;245} Despite the kinetic and thermodynamic studies of cyanoacrylamide–thiol

reactions,^{238;246} including an absolute binding free energy profile for BTK-4C9,⁸¹ an explanation for the prolonged and differential residence times of these ligands is still lacking.

The following investigation is organized around where in the unbinding process the microscopic timescales of these ligands are expected to differ. The design hypothesis attributes the differential residence time to steric gating of the initial proton-abstraction in the retro-Michael elimination by the variably-sized terminal alkyl group.³⁶ Therefore, for each ligand, the conformational distribution of the covalent complex is estimated from unbiased molecular dynamics simulations, and the abstraction energy barrier is computed via nonequilibrium QM/MM simulations with that conformation restrained. The same alkyl groups were observed in a crystal structure to form hydrophobic contacts with the protein,³⁶ and we expected this would affect the noncovalent off-rate. The noncovalent binding affinity K_d , as we predict via absolute binding free energy calculations, is a more practical proxy for k_{off} under the assumption of a shared k_{on} . For simplicity, we also assume that the reaction rates for the Michael addition and remaining steps of the retro-Michael elimination are shared across the series. The chemistry is likely identical, and in all three cases the reactive geometries are shown to be very similar. We incorporated these calculations into a detailed kinetic scheme, where the relevance of each molecular insight in a full unbinding event is evaluated through mass-action simulations fit to the experimental data.

5.2 Methods

5.2.1 Atomic models

Three reversible BTK inhibitors, 4A9, 4B9 and 4C9, taken from the study of Bradshaw et al.³⁶, were compared with the irreversible inhibitor ibrutinib.⁸⁴ Those compounds, depicted in Figure 5.1, all target the non-catalytic Cys481 via a thiol-Michael addition reaction. High resolution crystallographic structures are available for 4C9 (PDB ID: 4YHF³⁶) and

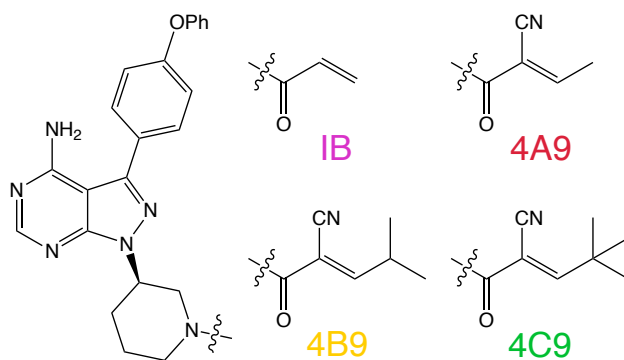


Figure 5.1: The chemical structures of the warheads and shared scaffold of irreversible inhibitor ibrutinib (IB) and reversible inhibitors 4A9, 4B9, and 4C9.

ibrutinib (PDB ID: 5P9J¹³²). 11 additional covalent inhibitors were examined to broaden the analysis. In total, 15 kinase-inhibitor complexes (Supporting Table C.1) were constructed from published crystal structures of covalently linked adducts for molecular dynamics (MD) simulations. Missing non-terminal backbone segments were rebuilt with MODELLER,²⁴⁷ and the N-terminal “AKDAW” motif was included when absent from the deposited coordinates to preserve possible functional contacts. Side chains of modeled residues were placed with SCWRL4.²⁴⁸ Each system was solvated with TIP3P water¹³⁹ and 0.15 M KCl in a cubic box with a 12 Å buffer using CHARMM-GUI.^{135;136}

Ligand coordinates for both the linked (covalent adduct with target residue, acetyl and N-methylamide caps) and unlinked (pre-reactive warhead) states were generated with the PyMOL Builder module.²⁴⁹ Both states were parameterized with CGenFF.^{140–142} Protein parameters were taken from CHARMM36M.^{137;250} A CHARMM patch connecting the ligand to the target residue was constructed by replacing CHARMM36 atom types and charges of the affected residue atoms with the corresponding CGenFF linked-state parameters. The small charge discrepancy between the unlinked and linked topologies was distributed over polarizable ligand atoms distal to the warhead. Selected dihedrals were reparameterized against relaxed potential-energy scans at the MP2/6-31G* level.²⁵¹ PSF files were assembled with psfgen,¹⁵⁰ and ParmED²⁵² was used to convert the CHARMM topology to AMBER format.

Classical simulations

Unbiased MD was performed with the AMBER 24 engine^{253;254} at 310 K under a Monte Carlo barostat (1 bar). Free energy calculations were performed with NAMD 3¹⁵² and the colvars module¹⁵⁶ at 300 K under the Langevin piston barostat (1 atm). Unless otherwise stated, simulations used a 12 Å real-space cutoff with force-based switching at 10 Å, particle-mesh Ewald electrostatics, SHAKE²⁵⁵ constraints on bonds to hydrogen, a 2 fs time step, and Langevin dynamics ($\gamma = 1 \text{ ps}^{-1}$). For the unbiased MD, each system was first minimized for 5,000 cycles (2,500 steepest descent, 2,500 conjugate gradient) with positional restraints of $1.0 \text{ kcal mol}^{-1} \text{ Å}^{-2}$ on protein backbone heavy atoms (C_α , C, O, N), ligand heavy atoms, and heavy atoms of all ordered waters within 10 Å of the ligand. The system was then equilibrated in the NVT ensemble over 25 ps (1 fs time step) with $10 \text{ kcal mol}^{-1} \text{ Å}^{-2}$ restraints on the same mask. Five successive 2 ns NPT stages then relaxed the restraints through 10.0, 5.0, 2.5, 1.0, and $0.5 \text{ kcal mol}^{-1} \text{ Å}^{-2}$. Unrestrained production was propagated in NPT for 4–8 μs per system.

QM/MM simulations

QM/MM simulations were carried out with AMBER 24. The QM region comprised the cysteine target residue (Cys481), the ligand warhead and scaffold, and 20 water molecules selected by proximity ($\leq 5.5 \text{ Å}$) to the S_γ , C_α , and C_β atoms at the start of the equilibrated BTK–4A9 structure. The QM subsystem was treated at the PM6 semi-empirical method²⁵⁶ because of its computational efficiency. The MM remainder used the CHARMM36M, CGenFF, and TIP3P potential. QM–MM electrostatic interactions were computed with PME. SHAKE was disabled for QM atoms. Upper-wall restraints ($r_{\text{wall}} = 7.0 \text{ Å}$, $\kappa = 2.0 \text{ kcal mol}^{-1} \text{ Å}^{-2}$) on the distances between each QM water oxygen and the center of mass of S_γ , C_α , and C_β prevented evaporation of the QM solvent shell. To minimize systematic differences across ligands, the three systems were prepared sequentially: 4A9 was equilibrated with QM/MM,

mutated to 4B9 (PyMOL), re-equilibrated, mutated to 4C9, and re-equilibrated. 99 independent steered MD replicates were launched from each equilibrated endpoint (Langevin thermostat, $\gamma = 1 \text{ ps}^{-1}$, $T = 300 \text{ K}$).

5.2.2 Free energy landscape of C_α -H bond cleavage

The initiation of the retro-Michael elimination (C_α -H bond cleavage) was modeled with QM/MM steered molecular dynamics (SMD) in AMBER 24 coupled to PLUMED²⁵⁷ for the non-steering restraints. The coordinate r was constructed to reflect the fact that the protons in the QM region are indistinguishable particles, following the protocol of ref 84. A generalized mean distance between one atom A and a group of N protons was defined as,

$$r_{\text{A-[H]}} = \left(\frac{1}{N} \sum_{i=1}^N \frac{1}{\|\mathbf{r}_\text{A} - \mathbf{r}_i\|^n} \right)^{-1/n} \quad (5.1)$$

This expression treats all the protons as identical particles but weights more heavily the one at the shortest distance from atom A. As a result, this strategy avoids pre-selection of a proton acceptor and permits the reaction to choose the lowest-energy pathway. In the present work, a value of 6 was used for n , and atom A corresponded to the acrylamide C_α . The pulling path spanned from 2.2 Å (bonded) to 3.5 Å (abstracted) over 10 ps with a 0.4 fs time step, using a harmonic restraint of $200 \text{ kcal mol}^{-1} \text{ Å}^{-2}$. Two heavy-atom dihedrals (Supporting Figure C.2) connecting the carbonyl and the alkyl capping group were restrained with harmonic potentials ($\kappa = 0.15 \text{ kcal mol}^{-1} \text{ deg}^{-2}$ to allow a standard deviation of 2°) to hold the warhead in the retro-Michael-competent orientation ($\phi \approx 180^\circ$).

The potential of mean force (PMF) $W(r)$ for deprotonation was reconstructed from the ensemble of nonequilibrium work trajectories via the weighted-histogram estimator from

$$W(r) = -\beta^{-1} \ln \frac{\sum_t \frac{\langle \delta(r-r_t) \exp(-\beta w_t) \rangle}{\langle \exp(-\beta w_t) \rangle}}{\sum_t \frac{\exp(-\beta u(r,t))}{\langle \exp(-\beta w_t) \rangle}} \quad (5.2)$$

where $\beta = (k_B T)^{-1}$ is the inverse temperature, $\langle \dots \rangle$ denotes an average over all SMD replicates, r_t and w_t are the stochastic coordinate value and accumulated work at time t , $\delta(r - r_t)$ in practice is a histogram bin filter for some value r along the PMF, and $u(r, t) = \frac{1}{2} \kappa [r - r_0(t)]^2$ is the harmonic bias with time-dependent center $r_0(t)$ and spring constant κ . The time dependent nonequilibrium work w_t is given by

$$w_t = \int_0^t \frac{\partial}{\partial t'} \frac{1}{2} \kappa [r - r_0(t')]^2 dt' \quad (5.3)$$

$$w_t = \int_0^t \kappa [r - r_0(t')] \dot{r}_0(t') dt' \quad (5.4)$$

5.2.3 Noncovalent binding free energy

The noncovalent binding affinity of four BTK inhibitors (4A9, 4B9, 4C9, IB) was calculated using the alchemical double-decoupling route, following the BFEE2 protocol based on classical MD.²⁵⁹ This computes the absolute standard binding free energy $\Delta G_{\text{bind}}^{(\circ)}$ at a 1 M reference concentration via the potential energy $U(\lambda)$, where λ is a thermodynamic coupling parameter such that $U(0)$ corresponds to the ligand fully decoupled from its environment and $U(1)$ to the ligand fully coupled.

The reference bound-state structure for each ligand was taken from unlinked equilibrium MD simulations as the centroid of the dominant cluster of five coordinates: two warhead-scaffold dihedrals, linking-atom distance, ligand center-of-mass deviation from the crystal pose, and the distance between the ligand carbonyl oxygen and the backbone nitrogen of

the target cysteine. The equilibrated complex (100 ns for the ligand–protein system and 20 ns for the ligand-only system) served as the starting point for all alchemical stages. The reversible coupling (forward) and decoupling (backward) of the inhibitor to and from the solvated kinase (bound state) and the bulk (unbound state) were conducted using stratified free energy perturbation (FEP)²⁶⁰. The number of λ values (50–200) and sampling per λ (2–8 ns, including 0.2–1 ns equilibration) were adjusted on a per-system basis until convergence of each direction and adequate overlap between λ values of opposite directions and overlap between sequential λ values in one direction. Data from forward and backward simulations were combined via the Bennett acceptance ratio (BAR)²⁶¹ to obtain the associated free energy difference.

During the bound FEP simulations, the ligand was restrained to the bound reference conformation, orientation, and position using seven coordinates: a symmetry-corrected RMSD of the ligand heavy atoms ($\kappa = 10 \text{ kcal mol}^{-1} \text{ \AA}^{-2}$), three Euler angles and two polar angles describing the ligand orientation relative to the protein ($\kappa = 0.1 \text{ kcal mol}^{-1} \text{ deg}^{-2}$ each), and the protein–ligand center-of-mass separation ($\kappa = 10 \text{ kcal mol}^{-1} \text{ \AA}^{-2}$). Only the RMSD restraint was applied for the unbound FEP simulations. The free energy cost associated with the restraints was evaluated by progressively reducing the force constants to zero using thermodynamic integration (TI)²⁶² bidirectionally. For both forward and backward bound simulations, 101 values of λ were sampled each for 4 ns. 201 values of λ sampled for 2 ns each were used for the unbound TI. The free energy contribution accounting for the reorientation and translation of the unbound inhibitor was calculated analytically.²⁶³

Because the bound-state decoupling step exhibited persistent hysteresis between the forward and backward simulations that could not be fully mitigated with additional sampling, the free-energy bounds from each direction were propagated as uncertainties on the final ABFE error estimate.

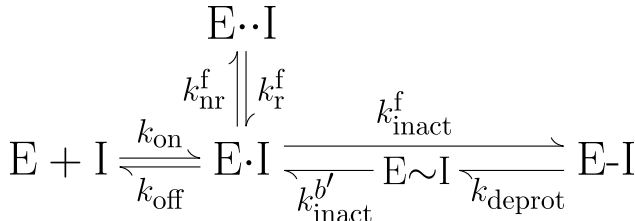
Since intramolecular non-bonded ligand interactions were preserved during the FEP sim-

ulations (`alchDecouple on`), the absolute solvation free energy could be obtained with only one additional calculation: a bidirectional TI in vacuum to evaluate the free energy cost of imposing the RMSD restraint on the isolated ligand. This vacuum TI used 201 values of λ for 2 ns each, with PME disabled and nonbonded cutoffs set effectively to infinity.

5.2.4 Kinetic modeling

Biochemical assay data measuring over time the dissociation of 4A9, 4B9, and 4C9 from BTK were digitized from Supplementary Figure 5 in ref 36. Each time course was independently fit to obtain experimental residence times, τ^{exp} , using a single-exponential decay function, % BTK Bound = $Y_0 \exp(-t/\tau^{\text{exp}})$. Differently from the original analysis, the fitting here assumed that the total occupancy goes to near-zero, and therefore did not include a parameter for a plateau.

Following the molecular modeling, the process depicted by the simple and generalizable Scheme 5.1 is resolved into the more complex and bespoke kinetic Scheme 5.2. It is important to emphasize that deliberate choices are made when decomposing kinetic schemes, and the choices herein are informed by the atomistic simulations of these particular ligands. For example, these ligands in reality may associate into the nonreactive or reactive state, but we have chosen to define the scheme such that it only binds into the reactive state. This approximation is allowed by the overwhelming presence of the reactive state in the unlinked MD simulations and greatly reduces the complexity of the resolved scheme. In previous works on ibrutinib binding kinetics⁸⁴ and a theoretical demonstration of rapid-intermediate gating³⁷, the choice to bind only in the nonreactive pose was more suitable.



Scheme 5.2: Covalent inhibition resolved for 4A9, 4B9, and 4C9 from atomistic simulations.

Scheme 5.2 was simulated in the deterministic limit as a system of ordinary differential equations using the in-house Python package GEKiM³⁷. Global fitting of the simulated and experimental total occupancy curves was first done with a local minimization using LMFIT²⁶⁴ with the Levenberg-Marquardt least-squares (LS) algorithm.²⁶⁵ A wide range of possible parameters was then explored with Bayesian Metropolis Monte Carlo (BMMC) sampling.^{198;233} The sampled parameters were $\log_{10} k_{\text{on}}$, $\log_{10} k_{\text{inact}}^{\text{f}}$, $\log_{10} k_{\text{inact}}^{\text{b'}}$, and each ligand-specific $\log_{10} K_{\text{d}}$. Rate constants and dissociation constants were sampled in $\log_{10}$ space to ensure uniform multiplicative exploration across orders of magnitude. The bounds of each K_{d} were set from the ABFE error. The minimum bound of k_{on} was set to $1 \times 10^4 \text{ (Ms)}^{-1}$, informed by a survey of binding kinetics and the expectation that the association is not conformationally gated.²⁶⁶

Per-point weights were constructed as the product of a temporal-density correction, $w_t = \sqrt{\Delta t_i}$, and an inverse-occupancy term, $w_y = 1/(y_i + 0.1)$, and normalized per ligand. Early timepoints for 4B9 and 4C9 were upweighted to anchor the simulated curves to the initial occupancy plateau, which is well-constrained experimentally but otherwise underweighted by the inverse-occupancy correction.

The Metropolis acceptance criterion used the sum of squared residuals between a simulated time course and the experimental data, with temperature tuned to yield a range of fitted curves that reasonably spans the experimental spread (true experimental uncertainties could not be recovered). Six independent chains were initially perturbed around the LS minimum, run for 1×10^6 steps, and thinned for every 1000 samples (discarding the burn-in period) to reduce the effect of autocorrelation.

5.3 Results and discussion

5.3.1 Dynamics of associated inhibitor

Linked inhibitor

The covalent bond cleavage between the target cysteine and the three cyanoacrylamide-based inhibitors occurs through a retro-Michael elimination. A critical first and necessary step of this reaction is the abstraction of the proton on the warhead C_α . This process is most feasible when the dihedral angle ϕ between that proton and the carbonyl oxygen is approximately 270° (Figure 5.2, top).³⁶ However, in equilibrium simulations of the linked state, we observed no sampling near this region, ruling out rapid conformational gating as a contributing factor to the prolonged residence time.

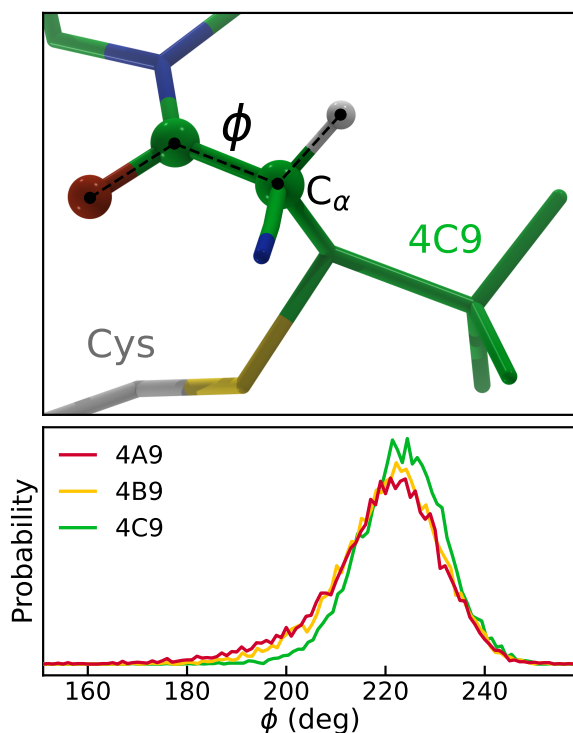


Figure 5.2: (top) Definition of the dihedral ϕ ($O=C-C_\alpha-H$) governing the geometric competence of the proton abstraction, illustrated on the 4C9-Cys covalent adduct. (bottom) Equilibrium distributions of ϕ in the linked simulations of 4A9, 4B9, and 4C9.

To estimate general differences in the energy landscape and kinetics of C_α deprotonation,

a nonequilibrium steered QM/MM simulation was used to examine the proton abstraction step. The steering coordinate was the generalized mean distance between the C_α and all hydrogen atoms in the QM region (eq 5.1). All systems contained 20 QM waters. As shown in Figure 5.3 (top), the barrier heights for 4B9 and 4C9 are indistinguishable, whereas the barrier for 4A9 is ~ 2.4 kcal/mol lower.

Because the proton was abstracted in the direction opposite to the nucleophilic sulfur and further blocked from the sulfur by the cyanoacrylamide group (Figure 5.2, top), we reasoned that a water would be the likely acceptor. We therefore examined the availability of water as a possible acceptor (*i.e.*, the average number of waters in the first solvation shell of the C_α) along the steering coordinate. As shown in Figure 5.3 (bottom), the qualitative picture matches that of the PMF. The first-shell population for 4B9 and 4C9 contains about the same number of waters on average (~ 1.1 – 1.2), which contrasts the markedly higher number of waters available for 4A9 (~ 1.4 – 1.5). This excess is already present in the reactant state and persists through the transition state, consistent with steric exclusion by the isopropyl and *tert*-butyl capping groups of 4B9 and 4C9 restricting solvent access to the C_α independently of the torsional gate. The collapse of all three curves at the upper end of the generalized mean distance reflects only that the first shell is essentially defined to be empty at those values.

This differential hydration offers a plausible explanation for the difference in the energy barrier. Beyond direct contribution to proton transfer, a drier first shell may also raise the cost of stabilizing the resulting carbanion, further disfavoring the transition state for the bulkier inhibitors.

For the sake of simplicity, the transition rate constant for deprotonation from the protonated (prot) state is approximated within the framework of classical transition state theory (TST),^{171;172;267}

$$k_{\text{deprot}}^{\text{TST}} = \frac{e^{-\beta W(r^\ddagger)}}{\int_{\text{prot}} e^{-\beta W(r)} dr} \langle \dot{r} \theta(\dot{r}) \rangle_{r^\ddagger} \quad (5.5)$$

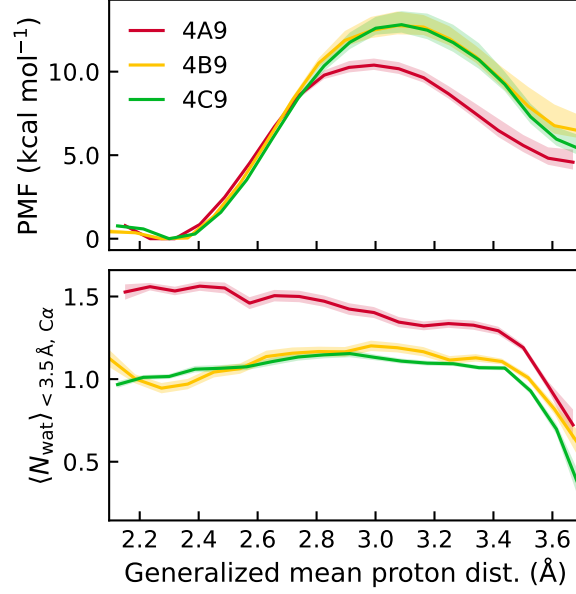


Figure 5.3: (top) Potentials of mean force (PMFs) and (bottom) mean number of QM waters in the first solvation shell of the C_α as a function of the generalized mean proton distance for each ligand. Shaded regions are pointwise 95% bootstrap confidence intervals from 1000 resamplings of the 99 SMD replicates per system (top) and the standard error of the mean across replicates (bottom).

where $r^\ddagger$ is the transition state between the states “prot” and “deprot”, and the quantity $\langle \dot{r} \theta(\dot{r}) \rangle_{r^\ddagger}$ represents the average forward velocity of r at the transition state using the Heaviside function $\theta(\dot{r})$ to select positive values. The transition rate constant can be expressed as, $k_{\text{deprot}}^{\text{TST}} = e^{-\beta \Delta W^\ddagger} \langle \dot{r} \theta(\dot{r}) \rangle_{r^\ddagger} / L_{\text{prot}}$, where $\Delta W^\ddagger = W(r^\ddagger) - W(r_{\text{prot}})$ is the activation free energy, r_{prot} represents the position of the minimum in the protonated well, and $L_{\text{prot}} = \int_{\text{prot}} e^{-\beta[W(r) - W(r_{\text{prot}})]} dr$, is the effective length of the protonated well. The average forward velocity was calculated by finite difference from a 50 ps trajectory harmonically restrained ($\kappa = 800 \text{ kcal mol}^{-1} \text{ \AA}^{-2}$) at the transition state $r^\ddagger$. This simulation was launched from the SMD replicate that used the lowest amount of nonequilibrium work to reach $r^\ddagger$.

Table 5.1: Estimated rate constants from transition state theory

	$\Delta W^\ddagger$ (kcal/mol)	L_{prot} (Å)	$\langle \dot{r} \theta(\dot{r}) \rangle_{r^\ddagger}$ (Å/ps)	$k_{\text{deprot}}^{\text{TST}}$ (s ⁻¹)
4A9	10.4 ^{+0.6} _{-0.4}	0.15	17.7	3.21×10^6
4B9	12.8 ^{+0.4} _{-0.3}	0.14	14.9	5.55×10^4
4C9	12.8 ^{+1.1} _{-0.5}	0.17	16.2	4.54×10^4

The results are summarized in Table 5.1, along with the 95% bootstrap confidence interval of $\Delta W^\dagger$. The average forward velocity is 15–18 Å/ps, almost three times the value from Maxwell’s distribution for a single proton, $(k_B T / (2\pi m_H))^{1/2} = 6.31$ Å/ps. Using the reduced mass does not change this value significantly. It can be shown that the time derivative of the generalized mean distance r defined by eq 5.1 is increased when the weighted sum is dominated by one or a few short distances. Examination of the QM proton density as a function of distance from the C_α during the transition-state simulation confirms this regime: only ~ 11 – 12% of the density is closer than $\langle r \rangle$.

The timescales of deprotonation predicted by the direct transition rates, $\tau_{\text{deprot}}^{\text{TST}}$, are 3.11×10^{-7} , 1.80×10^{-5} , and 2.20×10^{-5} s, for 4A9, 4B9, and 4C9, respectively. Clearly, the kinetics of this step does not make a difference on the overall residence time. Because TST ignores recrossings and assumes a transmission coefficient of 1, the transition rate estimates are upper bounds.^{172;173} Additionally, the semi-empirical QM method used can underestimate energy barriers.^{84;256;268} Thus, the approximate rate constants in Table 5.1 are expected to be overestimated, but only by 3–4 kcal/mol.

Unlinked inhibitor

The bound-but-unlinked conformational ensemble warrants examination in its own right. Conventional design heuristics rank unlinked inhibitor poses by the spatial proximity of the electrophilic warhead to the target nucleophile.^{68;69;79;80;82;83} While geometric competence is necessary, it is not sufficient when the reactive pose is transient and in rapid equilibrium with nonreactive conformations.^{37;70} Prior work on BTK–ibrutinib, where the reactive geometry was predicted to be a minority state of the unlinked ensemble, established that this regime is not a hypothetical concern for ligands of similar structure.⁸⁴ We therefore examined the unlinked ensembles of 4A9, 4B9, 4C9, and ibrutinib bound to BTK with unbiased classical MD simulations.

To assess the conformational landscapes of each ligand, two-dimensional potentials of mean force (PMFs) were computed over two general coordinates using MD simulations based on a classical force field: (i) the linking-atom distance between the warhead electrophile and the target nucleophile, and (ii) the no-fit RMSD of the ligand center of mass relative to the crystallographic pose with the protein pocket aligned. The first coordinate reports geometric competence of the warhead for covalent action, while the second captures scaffold translation independent of warhead rotation.

As shown in Figure 5.4, the PMFs of the three cyanoacrylamide-based inhibitors are qualitatively distinct from that of acrylamide-based ibrutinib. The BTK–ibrutinib ensemble was partitioned between a reactive and nonreactive basin, with only $\sim 3\%$ of sampled configurations satisfying the reactive-geometry criterion, consistent with prior work.⁸⁴ In contrast, 4A9, 4B9, and 4C9 appeared to mostly sample a single reactive basin. Conformational states were further resolved with the addition of two warhead–scaffold dihedral angles, ϕ_A and ϕ_B , (Supporting Figure C.1) and clustered with HDBSCAN²⁶⁹. This decomposition exposed multiple reactive and nonreactive conformational substates for the three cyanoacrylamide-based inhibitors, with summed reactive populations of 93%, 97%, and 98% for 4A9, 4B9, and 4C9 (Figure 5.4, bottom). The unlinked reactive-pose probability p_r^f is therefore effectively saturated for these reversible inhibitors and cannot account for differential residence time. This bias comes almost entirely from the intramolecular dynamics of the ligand. The cyano group appears necessary to stabilize the dihedral ϕ_A around -30° , a reorientation of the entire warhead that places the electrophile in range of the target nucleophile.

Interestingly, the hydrogen bond between the warhead amide carbonyl and BTK cysteine backbone nitrogen is hardly present in the unlinked state for 4A9 (7%), 4B9 (3%) and 4C9 (1%), but is almost entirely present for unlinked ibrutinib (96%). This hydrogen bond was suggested by Bradshaw et al. to potentially stabilize the covalent complex after observing it in the crystal structure.³⁶ It is almost always present in the linked states of all three

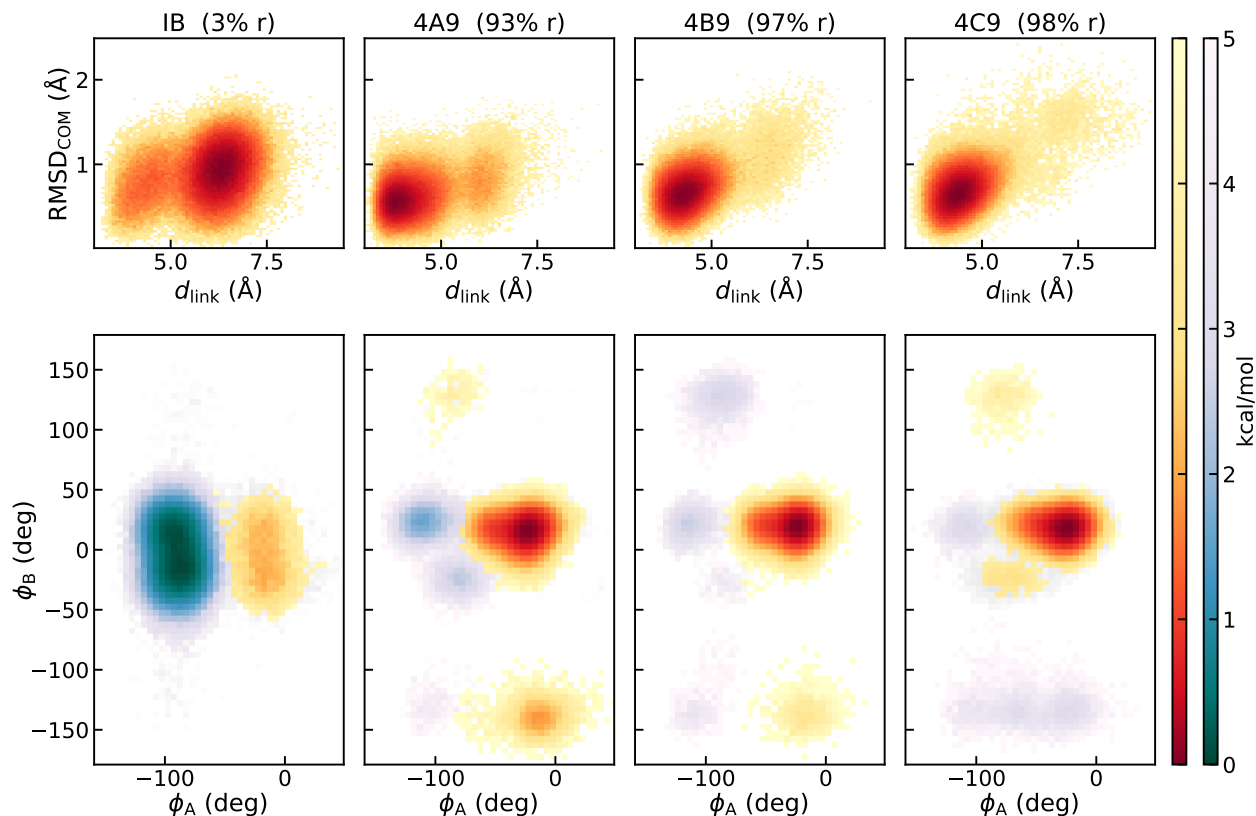


Figure 5.4: (top) Free energy landscapes (red–yellow) as a function of linking-atom distance (d_{link}) and center-of-mass deviation (RMSD_{COM}) for the unlinked ensembles of ibrutinib (IB), 4A9, 4B9, and 4C9, bound to BTK. (bottom) Free energy landscapes over the dihedrals ϕ_A and ϕ_B . Each bin is colored by its cluster class: reactive (red–yellow), nonreactive (blue–green), or HDBSCAN noise (gray). A conformation is deemed reactive (r) if the cluster’s centroid has a linking-atom distance under 4.5 Å.

reversible ligands (>99%), but only present for 24% of the linked BTK–ibrutinib simulation.

The stark difference in conformational gating of the forward reaction between structurally-similar ligands motivated a broader survey of a diverse set of kinase–inhibitor pairs, analyzing the PMFs over the two general coordinates used initially (Supporting Figure C.3). Across the three ibrutinib complexes, the warhead is most stable in BTK and progressively destabilized in EGFR and MKK7, recapitulating the rank order of experimentally observed potencies^{180;181;270;271} and indicating that ibrutinib’s reactive-pose suppression is not a property of the warhead alone. The remaining panel spanned qualitatively diverse regimes. Several systems remained near a single conformation throughout the multi-microsecond trajectories, while others adopted multiple well-resolved states separated by 1–3 kcal/mol barriers. For

example, EGFR–afatinib sampled multiple basins dominated by translational displacement of the scaffold. In contrast, c-SRC–DGY-06-116 was stable around a single conformation, but the target cysteine on the mobile P-loop was rarely near the warhead. Proximity-based ranking of docked poses would correctly identify the reactive configurations of the three cyanoacrylamide-based inhibitors (4A9, 4B9, and 4C9) but rank ibrutinib incorrectly and other inhibitors whose reactive geometries are minority states (*e.g.* DGY, 5YZ, and PF), consistent with the broader point that geometric competence of a static pose does not imply a populated reactive state. The saturation of the unlinked reactive-pose probability p_{r}^{f} for 4A9, 4B9, and 4C9 is therefore a property of these compounds in this pocket, rather than a feature that can be assumed *a priori*.

5.3.2 Noncovalent association

The first step in Scheme 5.1 corresponds to the noncovalent association of the inhibitor with the binding pocket without the formation of a covalent bond. The crystal structure of the BTK–4C9 complex revealed hydrophobic contacts between the *tert*-butyl capping group and surrounding protein residues,³⁶ suggesting that noncovalent affinity contributes to the differential residence time. To quantify this effect, the absolute standard binding free energies (ABFE) $\Delta G_{\text{bind}}^{(\circ)}$ of 4A9, 4B9, 4C9, and ibrutinib to BTK were computed using the alchemical route with the BFEE2 protocol.²⁵⁹ The results are given in Table 5.2.

Table 5.2: Computed absolute binding and solvation free energies for the major unlinked poses of BTK inhibitors.

	$\Delta G_{\text{bind}}^{(\circ)}$ (kcal/mol)	ΔG_{solv} (kcal/mol)
4A9	-7.9 ± 1.8	-27.9 ± 0.1
4B9	-12.9 ± 1.7	-27.5 ± 0.5
4C9	-15.1 ± 2.8	-25.0 ± 0.1
IB	-14.7 ± 0.7	-22.5 ± 0.1

The solubility ΔG_{solv} is anticorrelated to the binding free energy of the three reversible inhibitors. Increasing the steric bulk of the terminal alkyl group increases the binding affinity

and destabilizes the solvated state. The cyanoacrylamide warhead, being substantially more polar than the acrylamide of ibrutinib, is more favorably solvated by 2.5–5.4 kcal/mol across the series, implying a larger desolvation cost upon binding. Since 4C9 achieves a binding free energy on par with ibrutinib, it is likely that the larger desolvation cost is offset by more favorable interactions between the large hydrophobic capping group and the binding pocket.

The calculated noncovalent K_d values of the three reversible ligands span several orders of magnitude: 1.6×10^3 , 0.43, and 0.011 nM for 4A9, 4B9, and 4C9, respectively (conservative bounds: $83\text{--}3.2 \times 10^4$, 0.026–7.2, and $8.9 \times 10^{-5}\text{--}1.2$ nM). The trend is consistent with the hypothesis that larger alkyl substituents enhance the noncovalent binding affinity. However, comparing these values to a theoretical binding affinity $K_d^{\text{exp}} = (k_{\text{on}} \tau^{\text{exp}})^{-1}$ for the whole unbinding process calculated as a function of total residence time and using a generic value of $1 \times 10^5 \text{ (Ms)}^{-1}$ for k_{on} shows that noncovalent dissociation alone does not account for the observed timescale of unbinding (Figure 5.5).

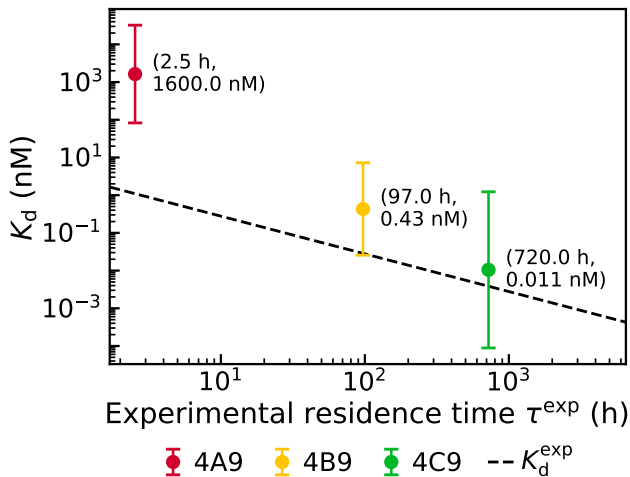


Figure 5.5: Noncovalent dissociation constant K_d from absolute binding free energy calculations vs. experimental residence time τ^{exp} from single-exponential fits of the BTK unbinding data for 4A9, 4B9, and 4C9. Vertical bars are the conservative K_d bounds from the ABFE hysteresis. The dashed line is $K_d^{\text{exp}} = (k_{\text{on}} \tau^{\text{exp}})^{-1}$ using a generic value of $1 \times 10^5 \text{ (Ms)}^{-1}$ for k_{on} .

The computed affinities, together with the unlinked dynamics, suggest that noncovalent behavior of these inhibitors occupy a qualitatively different regime than their irreversible counterpart, ibrutinib (Supporting Figure C.4). Despite ibrutinib’s calculated K_d

of ~ 0.02 nM (0.0068–0.062 nM) reflecting very tight noncovalent binding, the probability of the reactive pose is only $\sim 3\%$. Strong noncovalent affinity thus does not guarantee efficient covalent engagement. In contrast, the reversible inhibitors are all almost entirely poised to react once bound, but their affinities vary dramatically. Favorable reactive geometry thus does not guarantee tight noncovalent engagement.

5.3.3 Kinetic modeling

Taken individually, the computational results are suggestive but provide an incomplete picture. To develop a picture that integrates all the different pieces of information across the various stages, the experimental data were fit altogether with mass-action simulations of a detailed kinetic scheme (Scheme 5.2) tailored for these three ligands.

In the first steps of this scheme, the solvated enzyme and inhibitor undergo bimolecular association with the rate constant k_{on} into the reactive noncovalent complex, which can dissociate with the rate constant k_{off} . The noncovalent complex is partitioned between reactive (E·I) and nonreactive (E··I) poses with rapid conformational rate constants k_{r}^{f} and k_{nr}^{f} . Since the unlinked ligands exist almost entirely in reactive state, the association into the nonreactive state can be ignored for simplicity. The reactive pose forms the covalent adduct with the forward inactivation rate constant $k_{\text{inact}}^{\text{f}}$. The proton-abstracted intermediate (E~I), formed from the reactive covalent adduct with the deprotonation rate constant k_{deprot} , then completes the backwards reaction with the rate constant $k_{\text{inact}}^{\text{b'}}$.

Most of the rate constants were constrained by the results from the molecular modeling. Conformational transition rate constants (k_{r}^{f} and k_{nr}^{f} for the unlinked ensemble) were assigned under a rapid equilibrium approximation by fixing the sum $k_{\text{r}} + k_{\text{nr}}$ at 10^8 s^{-1} and partitioning the forward and backward rates according to the equilibrium populations extracted from the MD trajectories, *i.e.*, $k_{\text{r}} = p_{\text{r}}(k_{\text{r}} + k_{\text{nr}})$. The deprotonation rate constant k_{deprot} was estimated from the PMF of proton abstraction. The noncovalent dissociation

rate constant k_{off} was related to k_{on} and the ABFE-derived K_{d} with $k_{\text{off}} = K_{\text{d}} k_{\text{on}}$. The three remaining rate constants, k_{on} (bimolecular association), $k_{\text{inact}}^{\text{b}'}$ (the remainder of the retro-Michael reaction after proton abstraction), and $k_{\text{inact}}^{\text{f}}$ (the Michael addition), were fit freely. These were shared across the ligands, because we do not expect them to be much different from each other and sharing them significantly restricts the parameter space.

The simulated protocol reproduced the experimental protocol: an incubation phase ($[E]_0 = 0.45 \mu\text{M}$, $[I]_0 = 1.5 \mu\text{M}$, $t_{\text{inc}} = 30 \text{ min}$) followed by setting k_{on} to 0 and monitoring of the fraction of occupied enzyme over 120 hours. Including the incubation step was critical because 4A9 does not reach full occupancy in that time, allowing for estimates of k_{on} and $k_{\text{inact}}^{\text{f}}$. Distributions of possible parameter values were obtained with Bayesian Metropolis Monte Carlo (BMMC) sampling of parameter space. The resulting time courses are compared to the experimental data and single-exponential fit in Figure 5.6. Decompositions of all enzyme-containing species during both the incubation and post-incubation phases are shown in Supporting Figure C.5. The full set of free and fixed parameters is given in Supporting Table C.2, and the trajectories of the free parameters are shown in Supporting Figure C.6.

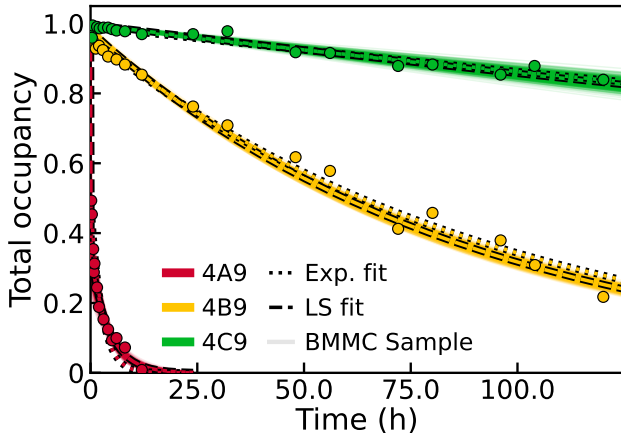


Figure 5.6: Fraction of total BTK occupancy over time for 4A9, 4B9, and 4C9. Points are the experimental data digitized from ref 36. Dotted lines are single-exponential fits to those points, and dashed lines are the least-squares (LS) fit of mass-action Scheme 5.2. 500 randomly chosen BMMC samples are shown as transparent solid lines.

To better understand the timescales associated with the noncovalent and covalent states

in the coarse-grained sense, Scheme 5.2 was reduced to an apparent two-step model (Scheme 5.1) that generalizes the kinetics for reversible covalent inhibition. The apparent noncovalent off-rate and forward inactivation rate constants were obtained under a rapid-equilibrium approximation by scaling the intrinsic rate constants by the unlinked reactive-pose probability, $k_{\text{off}}^{\text{app}} = p_{\text{r}}^{\text{f}} k_{\text{off}}^{\text{f}}$ and $k_{\text{inact}}^{\text{f,app}} = p_{\text{r}}^{\text{f}} k_{\text{inact}}^{\text{f}}$.³⁷ The apparent backwards inactivation rate constant was obtained as the series combination of effective timescales of the deprotonation and remaining chemical steps, $1/k_{\text{inact}}^{\text{b,app}} = 1/(k_{\text{deprot}} + 1/k_{\text{inact}}^{\text{b'}})$. The overall apparent residence time of Scheme 5.1 was calculated analytically using the known expression for induced-fit models,²⁷²

$$\begin{aligned}\tau^{\text{app}} &= \frac{k_{\text{off}}^{\text{app}} + k_{\text{inact}}^{\text{f,app}} + k_{\text{inact}}^{\text{b,app}}}{k_{\text{off}}^{\text{app}} k_{\text{inact}}^{\text{b,app}}} \\ &= \frac{1}{k_{\text{inact}}^{\text{b,app}}} + \frac{1}{k_{\text{off}}^{\text{app}}} + \frac{k_{\text{inact}}^{\text{f,app}}}{k_{\text{off}}^{\text{app}} k_{\text{inact}}^{\text{b,app}}}\end{aligned}\quad (5.6)$$

Depending on the system, any of the three terms may dominate the apparent residence time τ^{app} . The most naive interpretations of residence data commonly focus on the first term, $1/k_{\text{inact}}^{\text{b,app}}$, which is the timescale for breaking the covalent linkage between the ligand and the kinase. Evidently, if this is the slowest process in all the microscopic steps, then it does indeed dominate the apparent residence time. The second term, $1/k_{\text{off}}^{\text{app}}$, is the timescale for the dissociation of the noncovalently bound ligand and dominates if the breaking of the covalent bond is fast and negligible. The third and most salient term, $k_{\text{inact}}^{\text{f,app}}/(k_{\text{off}}^{\text{app}} k_{\text{inact}}^{\text{b,app}})$, represents the timescale associated with covalent reengagements. Its contribution is dissected after a brief account of the first two terms for each ligand.

The 95% highest density intervals (HDIs) from the BMCMC sampling of the contributing timescales are reported in Table 5.3. Despite broad distributions in the sampled parameters, robust qualitative conclusions can be drawn about the origins of the prolonged lengths and differences in the overall residence times. For 4A9, the noncovalent dissociation timescale

Table 5.3: The 95% highest density interval from BMCMC simulations of the timescales of reduced rate constants compared to the Scheme 5.1 residence time (τ^{app}) and single-exponential residence time (τ^{exp}).

	$1/k_{\text{off}}^{\text{app}}$ (h)	$1/k_{\text{inact}}^{\text{b,app}}$ (h)	$k_{\text{inact}}^{\text{f,app}}/(k_{\text{off}}^{\text{app}} k_{\text{inact}}^{\text{b,app}})$ (h)	τ^{app} (h)	τ^{exp} (h)	$\langle n \rangle + 1$
	min – max	min – max	min – max	min – max		min – max
4A9	0.000042 – 0.13	3.2 – 5.4	0.00059 – 0.37	3.4 – 5.5	2.5	1.0 – 1.1
4B9	7.4 – 23	3.2 – 5.4	63 – 78	85 – 93	97	14 – 22
4C9	56 – 180	3.2 – 5.4	440 – 660	560 – 780	720	95 – 180

($1/k_{\text{off}}^{\text{app}} < 0.2$ h across the 95% HDI) is negligible compared to the covalent bond lifetime ($1/k_{\text{inact}}^{\text{b,app}} \approx 3\text{--}5$ h). The apparent residence time ($\tau^{\text{app}} \approx 3\text{--}5$ h) is almost entirely determined by the rate of covalent bond cleavage. Once the bond breaks, the inhibitor escapes the pocket almost instantaneously. In contrast, the noncovalent dissociation timescales of 4B9 ($1/k_{\text{off}}^{\text{app}} \approx 7.4\text{--}23$ h) and 4C9 ($1/k_{\text{off}}^{\text{app}} \approx 56\text{--}180$ h) are very long. This is likely due to underestimation of $1/k_{\text{inact}}^{\text{b,app}}$. A deeper investigation is ultimately required to find the exact differences in the kinetics of the chemical reaction, but principles of prolonged residence time are still unraveled below.

The fitted rate constants can be benchmarked against comparable values in the literature. The association rate constant of rilzabrutinib, a reversible BTK inhibitor with a similar scaffold and cyanoacrylamide-based warhead, was determined from TR-FRET measurements to be relatively slow at $5.1 \pm 2.1 \times 10^4 \text{ (Ms)}^{-1}$.²⁴² Our fitted k_{on} (95% HDI: $0.1\text{--}6.4 \times 10^5 \text{ (Ms)}^{-1}$), which came entirely from kinetic inference of the incubation period yet played an important role in identifying k_{off} , meets the expectation that 4A9, 4B9, and 4C9 also do not exist on the higher end of the distribution of kinase-inhibitor association rates.²⁶⁶ Similarly, our fitted $k_{\text{inact}}^{\text{f}}$ ($1.4\text{--}5.7 \times 10^{-4} \text{ s}^{-1}$) was not informed by any molecular modeling and still falls in the expected regime of being rate-limiting over noncovalent association. However, it is likely an underestimate. Similar BTK inhibitors from the same study show $k_{\text{inact}} \sim 10^{-3}\text{--}10^{-2} \text{ s}^{-1}$.³⁶ The conclusions are not hostage to the fitted range. The BMCMC posterior shows a positive rank correlation between $k_{\text{inact}}^{\text{f}}$ and $k_{\text{off}}^{\text{app}}$ ($\rho = 0.63, 0.79, 0.54$ for 4A9, 4B9, 4C9). A higher $k_{\text{inact}}^{\text{f}}$ would therefore be partially absorbed by a higher $k_{\text{off}}^{\text{app}}$, yet still lengthen

the covalent reengagements timescale. Precise estimation of the individual rate constants ultimately requires incorporation of covalent occupancy time-courses at multiple doses to resolve ligand-specific $k_{\text{inact}}^{\text{f}}$ values.

As mentioned, the apparent residence time of a generalized reversible covalent inhibitor (Scheme 5.1) is a series combination that includes not only the mean timescales of singular noncovalent and covalent disengagements, but also that of covalent reengagements, $k_{\text{inact}}^{\text{f,app}}/(k_{\text{off}}^{\text{app}} k_{\text{inact}}^{\text{b,app}})$. The contributions of the covalent and noncovalent off-rates are therefore less straightforward than simple proportions of the apparent residence time. This is clarified by considering a refactored form of τ^{app} re-written to distinguish the timescales for a single noncovalent dissociation and multiple covalent cleavages,

$$\tau^{\text{app}} = (1 + \langle n \rangle) \frac{1}{k_{\text{inact}}^{\text{b,app}}} + \frac{1}{k_{\text{off}}^{\text{app}}} \quad (5.7)$$

Here, the prefactor $(1 + \langle n \rangle) = (1 + k_{\text{inact}}^{\text{f,app}}/k_{\text{off}}^{\text{app}})$ corresponds to the mean number of inactivation events before unbinding (see Appendix A). The leading 1 accounts for the initial inactivation event that places the system in the covalently-bound state E—I, while the ratio $k_{\text{inact}}^{\text{f,app}}/k_{\text{off}}^{\text{app}}$ is the odds of inactivation vs. dissociation evaluated per visit to the noncovalent complex E...I. Interestingly, the apparent residence time expressed in the form of eq 5.7 makes it clear that the mean number of inactivation steps per association event represents a critical design feature that can be dialed to modulate the behavior of this class of reversible inhibitors. The smaller the noncovalent off-rate, the higher the probability of re-inactivation. The smaller the covalent off-rate, the more each re-inactivation contributes to the overall timescale. It is through these covalent reengagements that the apparent residence times of 4B9 and 4C9 receive their largest contribution (Table 5.3). By our estimates, 4B9 and 4C9 inactivate on average 13–21 and 94–175 times per binding event, whereas 4A9 inactivates a negligible amount of times (9.4×10^{-5} –0.095). This analysis suggests that only 4A9

behaves according to the typical view of covalent inhibition, where a relatively rapid and constant exchange of noncovalent association and dissociation consistently offers transient and seldom-taken opportunities to covalently bind. Instead, 4B9 and 4C9 appear to adopt a quasi-equilibrium with respect to the covalent linkage and cleavage.

The apparent residence time can also be refactored in a different form to highlight the implications of this atypical balance,

$$\begin{aligned}\tau^{\text{app}} &= \frac{1}{k_{\text{inact}}^{\text{b,app}}} + (1 + K_{\text{inact}}) \frac{1}{k_{\text{off}}^{\text{app}}} \\ &= \frac{1}{k_{\text{inact}}^{\text{b,app}}} + \frac{1}{p_{\text{off}}^{\text{app}} k_{\text{off}}^{\text{app}}}\end{aligned}\tag{5.8}$$

where $K_{\text{inact}} = (k_{\text{inact}}^{\text{f,app}}/k_{\text{inact}}^{\text{b,app}})$ is the equilibrium constant of the covalent linkage, and $p_{\text{off}}^{\text{app}} = k_{\text{inact}}^{\text{b,app}}/(k_{\text{inact}}^{\text{f,app}} + k_{\text{inact}}^{\text{b,app}})$ is the probability to find the associated ligand in the unlinked configuration. Here, the apparent timescale for the dissociation of the noncovalently bound ligand is modulated by the equilibrium constant of forming the covalent bond itself, a perspective that is uncommon perhaps somewhat surprising. If individually the rates $k_{\text{inact}}^{\text{b,app}}$ and $k_{\text{inact}}^{\text{f,app}}$ are relatively fast, then K_{inact} becomes the relevant variable and the thermodynamics of the covalent bond become an exploitable design feature that can be dialed to optimize the apparent residence time.

5.4 Conclusion

Here we carried out atomistic calculations on the linked and unlinked inhibitor, deprotonation, and binding free energy. Because our goal was to focus primarily on the steps expected to play the most important role on the residence time, all the microscopic steps of Scheme 5.2 were not simulated explicitly. For example, we did not characterize the forward Michael reaction, though our previous study of ibrutinib shows how this can be done. Furthermore, we did not calculate the dissociate rate of the unlinked ligand. While this is not conceptually

complex, this can be a challenging calculation though there are promising enhanced sampling methods to do this efficiently. In principle, atomic calculations could be carried out for all the steps to include in our integrated strategy.

Still, without the full molecular basis, we were able to use a kinetic fitting protocol with the experimental data digitized from Bradshaw et al.³⁶ to deduce a broader conclusion: prolonged residence time is not the property of any single apparent rate constant. For the two well-performing inhibitors, it is an emergent consequence of repeated covalent reengagement, in which the reformation of the broken bond outpaces escape from the pocket. The methodology thereby identifies which kinetic steps are worth modulating and predicts how modulating them will affect the overall kinetic behavior. The theoretical analysis provides a deeper understanding of how residence time is governed by the dissociation rate, forward rate of inactivation, and backward rate of inactivation that is applicable to any reversible covalent inhibitor.

Appendix A: mean number of inactivation attempts

For a bound Markovian ligand traveling in Scheme 5.1, each E...I sojourn leads to either inactivation with $p = k_{\text{inact}}^{\text{f,app}} / (k_{\text{off}}^{\text{app}} + k_{\text{inact}}^{\text{f,app}})$ or dissociation with $q = 1 - p = k_{\text{off}}^{\text{app}} / (k_{\text{off}}^{\text{app}} + k_{\text{inact}}^{\text{f,app}})$. The mean number of inactivations $\langle n \rangle$ during one binding event is given by the geometric series over sequences of n E...I sojourns,

$$\begin{aligned}
 \langle n \rangle &= \sum_{n \geq 0} n (p^n q) = p \sum_{n \geq 0} n p^{n-1} q = p \frac{\partial}{\partial p} \left(\sum_{n \geq 0} p^n \right) q \\
 &= p \frac{\partial}{\partial p} \left(\frac{1}{1-p} \right) q = p \frac{1}{(1-p)^2} q = \frac{p}{q} \\
 &= \frac{k_{\text{inact}}^{\text{f,app}}}{k_{\text{off}}^{\text{app}}}
 \end{aligned} \tag{A.1}$$

CHAPTER 6

CONCLUSION AND OUTLOOK

This thesis was motivated by a difficulty that is not, on inspection, any single hard problem. Designing covalent inhibitors rationally is hard because the process is governed across a wide range of scales. Those that understand covalent inhibition through experimental binding curves of homemade molecules and those who understand it through simulated motion of individual atoms have partial views of the picture that do not overlap without a comprehensive, unified framework.

We started from these two sources of information, and what came into view was how kinetics can be used as a language to help them communicate. These two sources can inform each other. Microscopic data constrain the detailed macroscopic fit and the macroscopic data reveal which microscopic details matter. The role of the framework is to hand the medicinal chemist a “control panel” with dials and buttons that connect microscopic chemical features to some experimental outcome.

6.1 Summary of dissertation

Chapter 2 introduced the principle of rapid intermediates and the mass-action machinery used throughout. It shows that an apparent rate of inactivation can be dominated by a microscopic factor that has nothing to do with the reactivity of the warhead. The apparent inactivation efficiency must therefore be deconvolved into its chemical and conformational contributions before any design conclusion is drawn from it. A relationship between the EC_{50} and kinetic parameters was derived to show explicitly the convoluted impact that k_{off} , k_{inact} , and conformational probabilities can have on dose-dependency. This relationship also provides the basis leveraged in Chapter 4, that dose-response data measured at an appropriate time point can efficiently constrain kinetic parameters.

Chapter 3 built, from the bottom-up, macroscopic kinetic parameters for ibrutinib binding Bruton’s tyrosine kinase. A binding pathway parameterized by molecular dynamics reproduced the experimentally measured covalent efficiency once its microscopic steps were assembled into a kinetic model and coarse-grained. This agreement was obtained using estimated values for k_{on} and k_{off} . While the good agreement is welcome, one may note that the covalent efficiency is a comparatively forgiving quantity to predict because it depends on ratios of rate constants rather than their absolute values. When integrating the microscopic steps in this way, the dynamics can be benchmarked to experiment, pathways can be separated and interrogated, and the effect of rapid conformational gating on the overall kinetics is quantified from first principles rather than assumed to be a hypothetical concern.

Chapter 4 entailed a collaborative effort for designing novel kinase probes that spans from proteomics to molecular modeling. My contribution was the kinetic modeling of biochemical *in vitro* data, which advances the evaluation of kinetics from experimental assays. The binding of irreversible probes to the kinase ABL1 was fit using mass-action simulations and a Bayesian Metropolis Monte Carlo protocol. This approach extracts kinetic parameters more accurately and from less data than conventional analyses, and separates specific binding from nonspecific and apparent sub-stoichiometric factors. This approach also recovers the reversible affinity directly from the data when reversible binding maintains a rapid equilibrium during the hours it takes to covalently inactivate all enzymes, a regime that conventional empirical models already assume. After a pilot assay, the protocol can be run in reverse to identify the experimental conditions that will most efficiently reduce parameter degeneracy and uncertainty in a second round of measurements.

Chapter 5 examined the molecular and kinetic determinants of differences in residence time across a set of ibrutinib-derived reversible covalent inhibitors that differ only slightly in structure. The residence times of those compounds span more than two orders of magnitude, and the atomistic calculations yielded only a partial answer as to why. The conceptual payoff

came through the lens of the coarse-grained, two-step binding mechanism. The residence time of the shortest-lived inhibitor is almost entirely limited by the rate at which its covalent bond breaks, whereas the two longer-lived inhibitors obtain most of their residence time from repeated covalent reengagement rather than one long-lived covalent bond. The theoretical analysis emphasized how residence time for a reversible covalent inhibitor is determined jointly by the rates of dissociation, covalent bond formation, and covalent bond cleavage. This result applies to any reversible covalent inhibitor.

6.2 Implications for covalent inhibitor design

An effective design strategy requires answering two questions. The first is which molecular feature to change in order to produce a desired effect. The second is which chemical modification actually changes that molecular feature. That is, the designer has many dials available, but cannot determine from the observed outcome alone which combination is responsible for it.

A medicinal chemist modifies a covalent inhibitor and measures the resulting change in potency. The measured change should constrain the next modification, but on its own the data does not identify the microscopic mechanism at the origin of the change. The apparent rate of inactivation, for example, has both a chemical contribution from the intrinsic reactivity of the warhead and a conformational contribution from the probability of a reactive geometry, and a change in the apparent rate can arise from either. As shown in Chapter 2, increasing the conformational contribution allows the same apparent inactivation rate to be reached with a less reactive warhead, which is expected to reduce off-target reactivity and the associated toxicity. This decomposition also weighs on how selectivity is understood. Selectivity in covalent inhibition is usually attributed to a warhead of low intrinsic reactivity positioned near a nucleophile that is present in the target but absent from most off-target proteins.¹⁰ The present results show that the probability of a reactive geometry can itself

contribute to selectivity, since a warhead that is rarely presented for reaction in an off-target site will react there slowly even when the same nucleophile is available. Optimizing the conformational contribution can therefore improve selectivity while also allowing a less reactive and safer warhead. Because standard ways of measuring covalent occupancy do not distinguish the two contributions, a design decision based solely on these measurements is made without knowledge of which contribution it will affect.

The same issue is more pronounced for reversible covalent inhibitors. Design efforts (and post hoc rationale) typically concentrate on the stability of the covalent complex. Chapter 5 showed instead that the prolonged residence time of a reversible inhibitor emerges from the compounding effect of several rate constants that belong to different processes and act in opposite directions. The regime in which a ligand sits with respect to the balance of these rates determines which of them limits the residence time and which can be productively modified. A strategy that considers only the stability of the covalent bond might be pushing at the point of already-diminished returns, missing out on the multiplicativity of the timescales of noncovalent dissociation and covalent bond cleavage.

A reasonable concern is that the microscopic mechanisms involved in covalent inhibition vary enough from system to system that a single, generalized framework cannot meaningfully apply across them. In the ideal case, a covalent inhibitor should be as potent and selective as possible, binding with a dose-efficient strength to only a single target in the proteome. To this end, one aims to optimize properties that boil down to the noncovalent affinity and the kinetics of the covalent step. The difficulty is not that these properties differ between molecules, but that the experimental observable is a nonlinear combination of them. Two molecules can produce the same enzyme-binding curve for different microscopic reasons, *i.e.*, by relying on different mechanisms. This distinguishes covalent inhibition from most problems in molecular biophysics, where a simulated quantity corresponds directly to a well-defined measurement, such as the current through an ion channel, a nuclear magnetic

resonance relaxation time, or some structural order parameter. For covalent inhibition, the experimentally measured quantity is a complex combination of many quantities that each requires their own substantial computation effort. This dissertation’s framework is useful because it relates the observable in terms of these underlying quantities. A structural modification can then be directed at the microscopic step responsible for a given behavior rather than chosen from the aggregate outcome.

The field has not reached a consensus on how best to develop a covalent therapeutic.³³ That said, the strategies of starting from an established noncovalent scaffold and starting from a screened covalent fragment seem to complement each other more than they compete. The best option would depend, for instance, on whether the goal is to target known pockets or discover new ones. Still, some believe neither strategy works and any success is just a matter of luck. Perhaps we already agree that the convolution and degeneracy of microscopic effects make covalent inhibition nearly impossible to understand with phenomenological data alone. Where we differ is that this dissertation treats the problem as one requiring a “computational microscope” rather than luck.

This framework can be immediately applied to the growing number of covalent inhibitors for which both structural and kinetic data are available. Resolved kinetic models can be constructed to explain why structurally similar molecules behave differently, as was done for the three reversible inhibitors in Chapter 5. The explanations obtained in this way can then serve as Bayesian priors for the prospective design of new molecules.

More than a set of results, the objective of this dissertation is to provide a roadmap. It identifies several kinetic regimes in which the dominant determinant of potency or residence time is different, and in which the same structural modification therefore acts through a different mechanism. Starting from a detailed understanding of a single molecule, the property that must be changed to achieve a given effect can be identified. The behavior of related molecules can be predicted from molecular modeling without reconstructing the full kinetic

description for each new compound. The framework itself does not depend on the specific warhead chemistry or on the identity of the target. Kinases and their inhibitors were used as a model system to unravel these principles, but the analysis applies to any covalent inhibitor whose binding can be written as a kinetic scheme.

6.3 Refining the kinetic model

6.3.1 *Model fidelity and its limits*

It is important to recognize the limitations of the proposed framework. Many careful-but-necessary approximations are made in modeling potential energies, sampling alternative paths, and deriving rate constants. Therefore, the information derived from molecular dynamics is indeed best considered Bayesian priors rather than ground truth. The same goes for crystallographic structures. In practice, a structure often functions as a form of risk management rather than a specific answer. If the lead molecule of a drug development initiative is found to cause some adverse effect, a purely phenotypic approach would have to restart with far less information than if they had a structure. A structure can at least indicate which changes are likely to modulate an important interaction. A set of molecular dynamics simulations, like a static structure, is not guaranteed to inspire the perfect modification, but it does provide a much more thorough understanding. Therefore, it is significantly more valuable when a large investment is at stake.

The present framework is most useful once a small number of lead compounds exist and have been characterized experimentally. The experimental data contain the parameter space and reduce the amount of computation required to determine what controls the overall kinetics. Furthermore, it is much easier to make statements about the relative differences between structurally-similar molecules than the absolute behavior of one without any context. In Chapter 4, the fits to XO44 and K50P *in vitro* binding data were largely possible

because of the parameter constraints imposed by the data for K60P binding, for which the doses and time-points happened to suit better. In Chapter 5, the molecular factors governing the unlinked conformational reactivities or the noncovalent binding free energies could be inferred through structural comparisons across the series of cyanoacrylamide-based inhibitors and their less complex parent inhibitor, ibrutinib. That being said, the more thorough investigation of ibrutinib alone in Chapter 3 yielded a greater understanding of the binding process, revealing factors such as the solvent-mediated nature of the chemical reaction and the conformational gate that comes before.

The game we play in using this framework is finding the simplest kinetic scheme that is not simplistic. A scheme that more precisely reflects reality may not be any more informative than a simpler version of the same binding process that better communicates the salient point. This is especially true in a drug design pipeline, where the cost of added complexity must be justified by the engineering question it aims to answer. This is less true for scientific endeavors where the entire point is a strict description of some natural phenomenon. A drug candidate designed today could be obsolete tomorrow, but the mechanism by which one endogenous biomolecule reacts with another will be relevant for a very long time.

6.3.2 *Resolving the covalent step*

The most concrete extension of this work is exemplified by a study of the reversible inactivation pathway of the three reversible cyanoacrylamide-based inhibitors studied in Chapter 5. The analysis was guided by the molecular reasoning of the medicinal chemists who developed the compounds, and this reasoning did not fully account for the differences in residence time. The conclusions drawn from detailed atomistic calculations were coarse-grained and limited in quantitative accuracy. The forward and backward inactivation rates were not characterized. Moreover, they are not expected to be the same for the three ligands. The backward inactivation is initialized by the deprotonation of the inhibitor C $_{\alpha}$. It was thought

that the rate of this deprotonation is slow and ligand-specific, and therefore contributes to the differences in residence time. While the deprotonation rate is ligand-specific, it is always too fast to be rate-limiting in any capacity. Instead, the rate-limiting step of the reverse reaction appears to be cleavage of the covalent bond between the S_γ atom of the target cysteine and the C_β of the inhibitor. This tracks with previous quantum-mechanical studies of acrylamide-thiol reactions with and without a cyano group stabilizing the anionic form of the C_α .²⁴⁶ It may be that the larger substituents on the better-performing ligands increase the torsional strain and displace the warhead from the geometry required to break this bond, where the leaving sulfur atom must rotate out of the plane of the acceptor. Characterizing the chemical step explicitly for each ligand, rather than treating them identically, is the most direct way to improve the fidelity of the kinetic modeling.

The quantum-mechanical pathways, although improved relative to previous work and optimized for the dominant route, were branched only where a specific question required it, such as the dependence of the reaction on pH for BTK and ibrutinib in Chapter 3. Such questions are not difficult to address once the main reaction pathway has been established, but doing so requires care, and probably the exploitation of rapid local equilibria. Good practice is to check, with every added detail, that a numerical simulation of the more complex kinetic scheme produces nearly the same change in an observable as the more coarse-grained one. This check also applies in the opposite direction, when compressing a detailed scheme. A large source of intuition for the regimes and sensitivities of a given binding process comes from testing how changes in kinetic parameters, experimental conditions, and network topology push the boundaries of approximations and break the equivalence between simple and complex schemes.

The heuristics of conformational reactivity can also be refined if the situation calls for it. For determining whether a given conformational state is reactive, the distance between the reacting atoms is often an adequate metric, and a simple cutoff is usually sufficient.

For reversible ligands, reactivity may depend on more continuous quantities. The orbital overlap that determines the optimal geometry for the reverse reaction, for example, makes the conformational reactivity a function of a dihedral angle (ϕ), varying approximately as $\sin \phi$ rather than as a binary indicator. Different warheads introduce further differences. Two warheads may have different rate-limiting steps for the same overall reaction, such as the retro-Michael reaction for an acrylamide and a cyanoacrylamide.²⁴⁶ They may also begin the reaction from different conformations, as suggested by the unbiased simulations of both unlinked and linked states in Chapter 5 where the cyanoacrylamide-based inhibitors and acrylamide-based ibrutinib reside in opposing conformations. As a result, a single enhanced-sampling protocol may not apply to two slightly different warheads that undergo the same general class of reaction.

The investigation of the alternative path in Chapter 3 points to important additional considerations. A multi-step pathway that avoids a slow downstream step at the cost of a higher initial barrier can impact the macroscopic kinetics. Yet, the current suite of enhanced sampling methods is best suited for characterizing the energy landscape of one kinetically-trapped pathway at a time. This is especially relevant in studying the chemical reaction. The process of modeling these expensive quantum-mechanical steps is a science like any other, and it should draw from previous knowledge. Unbiased MD in the unlinked and linked states, as well as prior study of the general chemistry, already indicate which conformations are populated and how the reaction is likely to proceed. A reaction coordinate can also be designed so that the system finds the lowest-energy route on its own, as is done in Chapter 3 by using the distance between a carbon to a group of indistinguishable protons instead of just one. Because the rate through a path depends exponentially on the barrier, a path even modestly lower in energy dominates, so the enumeration of alternative pathways does not need to be exhaustive in order to be quantitatively useful. Exploratory enhanced sampling can find paths by walking several coordinates at once. These methods still have blind spots,

and an imperfect choice of coordinates gives a barrier that does not necessarily describe the true minimum free energy between two states, inviting hysteresis and poor convergence from slow orthogonal degrees of freedom.

Transition path theory (TPT) is a tantalizing option that can be applied directly to molecular dynamics.²⁷³ This framework characterizes the transition from one state to another with the committor, an ideal reaction coordinate that measures progress by probability of outcome rather than geometry, and from it the total rate over all reactive pathways. Sampling this ensemble of pathways directly, through transition path sampling or reconstructing it from a Markov state model (MSM), would address this problem without presupposing a dominant route.^{274;275} None of this is currently feasible for a bond-breaking step. The committor is a function of the entire configuration space (when applied to molecular dynamics data, this is the set of all possible atomic positions) that requires the reactive barriers to be visited repeatedly. It is unclear if the amount of sampling required to estimate the committor, or the network of an MSM, can be achieved with a quantum-mechanical level of theory. An MSM partitions configuration space into states connected by transition probabilities and characteristic timescales. MSMs themselves are master equation models of the kind used throughout this dissertation. These tools are better used for mechanisms with high network complexity and fast sampling of configuration space. This is the case for protein folding mechanisms sampled classically²⁷⁶, but not for covalent binding mechanisms.

6.3.3 *State representation*

Some key simplifications were used with even the most detailed of kinetic schemes found in this dissertation, and refining the quantitative outcomes would require relaxing these simplifications. For one, noncovalent association led explicitly into one of two conformational possibilities: a nonreactive or reactive state. In Chapter 2, the enzyme and inhibitor were considered to bind into a nonreactive conformation, illustrating in a theoretical way the im-

pact of conformational gating. Chapters 3 and 5 modeled noncovalent binding into either the nonreactive and reactive conformations, respectively, because one or the other was almost entirely dominant in the unlinked molecular simulations. Since these conformations exist in a rapid equilibrium, the choice essentially becomes a matter of clarity, granted the apparent dissociation rate is scaled by the right probability. If the conformational distribution is more evenly split between different noncovalently-bound states, it may be desirable to incorporate each conformation's association/dissociation explicitly. For conformations that are kinetically trapped rather than rapidly equilibrated, one would need to take care in partitioning a computed binding free energy across these states. The binding free energy computed for one state can be converted for another state from the ratio of state probabilities.

Only one protein conformation was considered in the kinetic modeling, although both small side-chain rearrangements, such as the DFG-in and DFG-out states for kinases, and larger changes between active and inactive conformations could be relevant. This is also true for protonation states, particularly for reactions targeting a lysine. Only the kinetic model in Chapter 3, Supplementary Scheme A.1, considered the pathway of binding BTK where the cysteine was already deprotonated.

The use of continuous rather than discrete states is another possibility for further refining a kinetic model. Some states are better described by a continuum. This would apply for the continuous function of reactivity mentioned in the previous subsection, as well as when the kinetics depend on, for example, the degree of folding of an RNA molecule or the rotational state of a molecular motor. The general framework here can accommodate such descriptions. It is independent of the type of observable, provided that the kinetic scheme is specified correctly and the relationship between the simulation time course and the measured signal is defined. A FRET signal that correlates with a mass-action probability distribution as a function of FRET distance could be fit in this way.

6.4 Extending beyond *in vitro* kinetics

6.4.1 Pharmacokinetics

The general idea of the present work, to seek a better representation of the macroscopic behavior of a system by relying on a detailed kinetic framework incorporating information from atomistic modeling and computations, can be extended to the full biological context of the system of interest. For example, the behavior of a covalent inhibitor *in vivo* does not begin or end with the generic binding scheme, $E + I \rightleftharpoons E \cdot I \longrightarrow E-I$. An orally administered drug passes through the gut, the intestinal lining, and the liver, before it is either excreted in the bile or put into systemic circulation. It distributes across the intravascular and extravascular space, interacting with plasma proteins, crossing the capillary endothelium into the interstitial fluid and permeating into circulatory cells, the endothelium itself, or the parenchymal cells of whatever tissue it reaches. Intracellular drug targets can be located in several different compartments with their own barriers of entry. When the drug is finally in close proximity, molecular engagement ($E + I \rightleftharpoons E \cdot I$) must still compete with endogenous substrates also trying to bind the target and with active efflux transporters trying to remove the drug from the cell. The drug, or more likely a metabolite,²⁷⁷ is eventually eliminated through the kidney or, by way of the liver, the bile, with a fraction reabsorbed through enterohepatic recirculation.

Metabolites are an important point of study with the present framework. Because a large fraction of an orally administered compound, such as ibrutinib,²⁷⁸ can be chemically altered during first pass through the gut lining and liver, the behavior of its metabolite(s) may be more relevant than that of a parent molecule. The framework is well-suited for comparing the two. A comparison between a parent molecule and its metabolites is more reliable than the absolute characterization of a single molecule, because the systematic errors in the potential energy approximations largely cancel. The comparison also makes it straightforward to

isolate the few binding steps that differ between the two molecules, which can be modeled explicitly with molecular dynamics while the remaining parameters are shared and fit freely.

Every kinetic model in this dissertation is reductionist in that it describes a homogeneous solution containing one type of enzyme and one type of inhibitor. The molecules are well-mixed, present in enormous copy number, and reach each other by rapid Brownian diffusion. This is the right description of the biochemical assays the models were fit to. However, nearly every stage of the pharmacokinetic process described above violates at least one of these conditions.³²

A drug does not diffuse from the gut to a tumor. It is carried there by flowing blood with a rate set by perfusion. This is a hydrodynamic process and requires the addition of an explicit transport term in the kinetic model of this stage.^{279;280} While diffusing through a tissue, the concentration gradient is made irregular by the irregular layers of cells. More accurate descriptions of this stage, when mixing is incomplete and concentrations vary over space, are possible with reaction-diffusion models that track concentrations as a field rather than as one number.²⁸¹

Diffusion inside a cell is also not what it is in water. The cytosol is crowded with macromolecules and partitioned by membranes and filaments, so a molecule explores space more slowly than Fick's law predicts.¹⁰⁰ A reaction depletes its immediate surroundings faster than they are replenished, so the apparent rate "constant" is not constant and instead decays with time. Fractal kinetics accommodates this by making the rate coefficients explicitly time-dependent.²⁸²

Low copy number is another problem that must be taken into account when modeling *in cellulo* kinetics.¹⁰⁰ An inhibitor at a low concentration in a small compartment may amount to a handful of molecules. Inhibition is then a sequence of rare discrete events competing against the synthesis of the target, and whether there is an apparent effect may be decided by the chance of many overlapping target-inhibitor bindings rather than by

the mean concentration of the bound complex. Therefore, when sensitivity to stochastic fluctuations is a problem or perhaps a deliberate design principle of the system in question, the deterministic mass-action limit of the chemical master equation may fail to capture the meaningful kinetic features.²⁸³ One must instead use stochastic methods of simulating the chemical master equation. Where roughly a single inhibitor is searching an organelle for its target, one should use Monte Carlo kinetics with the Gillespie algorithm.²⁸⁴ For intermediate copy numbers above roughly 100 that one could expect in the cytosol, the chemical Langevin equation captures the fluctuations at a much lower cost.²⁸⁵

Some stages are slow aggregates of many steps that are not worth resolving individually. For example, absorption and distribution mainly manifest in the overall kinetics as a lag before the drug appears in the next compartment. Delay differential equations represent these lag phases as explicit time delays,²⁸⁶ whereas mass-action ordinary differential equations would require a chain of intermediate states and the parameters that accompany them.

The process by which a drug reaches a target is a stochastic, spatially-extended, multi-compartment process. No single type of kinetic simulation engine is appropriate for all of it. Each method carries its own assumptions and computational cost, and the modular design of the kinetic modeling software developed for the present work allows them to be combined within a single scheme.

Pharmacokinetics is a convenient source of examples, but the general point goes beyond it. The use of mass-action, let alone the empirical models derived from it, should be audited whenever the data come from anything but a test-tube reaction. Identifying which assumption fails for a given set of data, and choosing the representation that avoids it, is part of building a kinetic analysis protocol.

6.4.2 *Systems biology*

The approach developed here is applicable beyond covalent inhibition. Most biological systems operate across disparate length and time scales. Their mechanisms often must be inferred from noisy data that reflect the quantity of interest only indirectly with limited resolution in both scales. Kinetic analysis of complicated networks is a common undertaking in systems biology research, where the central-dogma machinery of *E. coli*²⁸⁷ or the whole-cell model of *M. genitalium*²⁸⁸ are among the more astounding network reconstructions. Although molecular modeling is not routinely involved in these effects, it has an important role to play.

In most complex kinetic models, the topology of the network can be specified and analyzed, but the individual rate constants are neither measured nor identifiable from the available data. Even the coarse-grained parameters they are folded into are often “sloppy.”²⁸⁹ However, molecular dynamics can be used to estimate these rate constants, and a sensitivity or identifiability analysis of the network can determine which of them most affect a given question and which experiment would rigorously resolve a degeneracy.

As the accessibility of powerful computer hardware grows, so does the number of research projects who use molecular dynamics. However, these are largely independent investigations motivated by a molecular understanding of one small piece of a living cell. The questions are rarely derived from observations of the larger network in which the particular piece is just a cog in the wheel. Rarely, too, do these investigations compute kinetic parameters that could slot into a larger kinetic scheme, even though methods for doing so are landmark achievements in statistical mechanics. The broader importance of these parameters is lost when there is no centralized effort to integrate them and provide insight for the next iteration of experiments or simulations.

A “complete” mechanistic and quantitative picture of any lifeform is most satisfactory when it resolves the entire hierarchy of events that underlie its behavior at a level of detail

that no single phenomenological model can achieve. This goal, whether it be for a whole cell or single signaling pathway, could sustain a multidisciplinary research program that combines molecular biological experiment and biophysical simulation.

As an illustration, how effectively an adhesion G protein coupled receptor searches the extracellular space, how an adhesion-derived signal is distributed across canonical and non-canonical pathways, and how alternative splicing supports the probabilities of these pathways, are all questions of kinetics that fold into a multiscale model. The molecular modeling supplies the microscopic rate constants that quantify the inherent likelihoods of different events. These probabilities, together with the concentrations of species involved, inform the transition rates of the kinetic model. The phenomenological data provides the macroscopic observables the model must reproduce.

Curated repositories of kinetic parameters and assembled models already exist.^{290;291} They are archives, though, aggregating heterogeneous literature values measured under incommensurable conditions, and they offer no way to identify which unmeasured rate constant most limits a given prediction, nor to derive missing constants from a consistent physical model. What is absent is a directed and collective effort in which a shared kinetic scaffold of a complex biological system is populated incrementally rather than in isolation. Such a framework would let contributors identify which transitions remain unparameterized, claim the ones suited to their expertise, constrain the larger scheme with deposited rate constants, and expose the many layers of operation that direct future research.

6.5 Concluding Remarks

This thesis is a proof of principle. It shows that detailed atomistic calculations can be incorporated into kinetic models and, in this way, compared directly against experimental measurements. This direct comparison is the central feature of the approach. The kinetic theory that makes it possible is more than a century old and was developed when the gov-

erning equations could not be integrated numerically, which is why the early treatments of enzyme kinetics and their successors relied on simplification and analytical approximations.³¹ The novelty is that the microscopic quantities entering the kinetic model can now be computed with atomistic simulation. The framework can also be used in the reverse direction, taking a set of plausible parameter values and selecting the experimental conditions that will most improve the next set of measurements.

After solving a long sequence of theoretical and technical problems in the course of a project like this one, it is easy to lose sight of the bigger picture. Every field, every perspective, every level of description is one piece of a larger puzzle. Empirical understandings of selectivity and potency would come from experiments, such as chemoproteomics, whereas theoretical understandings would come from modeling, such as molecular simulations. For most of the history of covalent inhibitor design, every piece but the theoretical ones have been laid in place, waiting for them to drop in and reveal the final image. Often they have not, and the missing pieces were either deemed unnecessary and left out, or a guess of what they would have shown is drawn straight onto the table in their place. In these cases, the experiments were not guided by physics-based input.

Part of the reason is that what is wanted of computation — or any form of mathematical modeling since the pre-computer days of molecular medicine^a — is the inverse problem, namely which molecule should be made to produce a desired behavior. There is no general way to solve it. Computation runs in the forward direction instead, predicting what a given molecule does. The practical route is to iteratively modify the molecule when it falls short of what is wanted and predict again, which requires the prediction to yield the same quantity the chemist is trying to improve. Existing methods can perturb a structure and compute the corresponding change in free energy, or simulate different molecules to characterize a chemical

a. For example, Torsten Teorell wrote: “It might very well be possible, by aid of relations derived, to find out which is the best way to dose and to administer a drug, or even how to *change* the drug, in order to ascertain a definite time-concentration relation and thereby a desired therapeutical effect.” This is from 1937, in what is arguably the first landmark pharmacokinetics paper.²⁷⁹

mechanism. However, the resulting thermodynamic parameters and individual microscopic steps do not necessarily correlate with the observed overall behavior and can be misleading on their own. Combining them in a kinetic network that is grounded with experimental data makes them useful for design.

A theoretical framework should also be intuitive. It should not require an esoteric understanding of its internal details in order to be applied or communicated. Expressing simulation results in terms of macroscopic observables makes them accessible to the experimentalists who use them. Furthermore, as Chapter 4 demonstrated, a framework that is fed by multiple disciplines is more likely to yield methodological improvements, since each perspective could ask for information from the others in a way that breaks routine.

More generally, this framework is one contribution to the broader problem of inferring mechanism and identifying the determinants that matter when the available data are incomplete, noisy, or indirect, or systems that operate across a wide range of scales. Covalent inhibition is a difficult example of this problem, because its experimental observables reflect so many underlying processes. The fact that these observables can nevertheless be interpreted in microscopic terms suggests that the same can be done for other systems whose behavior is driven by microscopic dynamics that are not directly observed.

APPENDIX A

SUPPLEMENTARY INFORMATION FOR CHAPTER 3

A.1 Structural dynamics

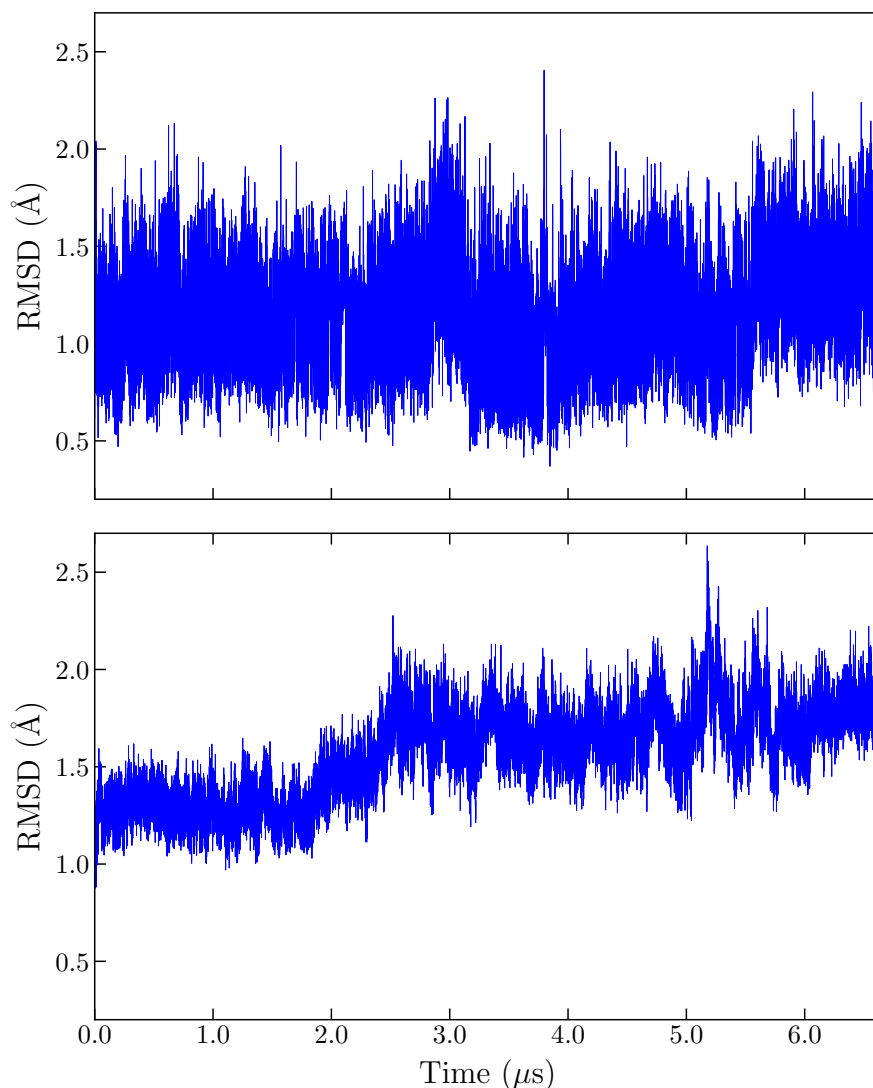


Figure A.1: (top) RMSD of the inhibitor's heavy atoms from its crystallographic covalently-linked pose (PDB ID: 5P9J¹³²) calculated for the 6.6 μs classical MD simulation of the covalently-unlinked inhibitor. RMSD was calculated post-alignment of the binding pocket protein backbone C_α 's (within 10 $\text{\AA}$ of the inhibitor) to the crystallographic structure. The values were corrected with respect to symmetry-equivalent atoms in the inhibitors. (bottom) RMSD of the binding pocket backbone C_α 's from the crystallographic pose.

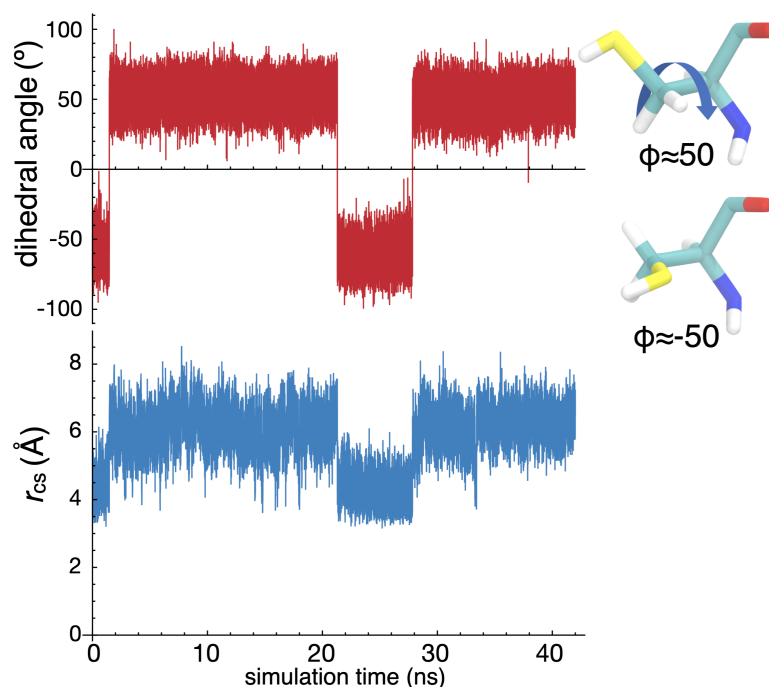


Figure A.2: Fluctuations during the 6.6 μ s MD simulation with the of the unlinked inhibitor associated in the BTK binding pocket. The time evolution of the χ_1 side-chain dihedral angle of the residue Cys481 and interatomic distances between the C_β atom (cyan) of the acrylamide warhead and the S_γ atom (yellow) of Cys481 is shown for the initial 40 ns of the simulation.

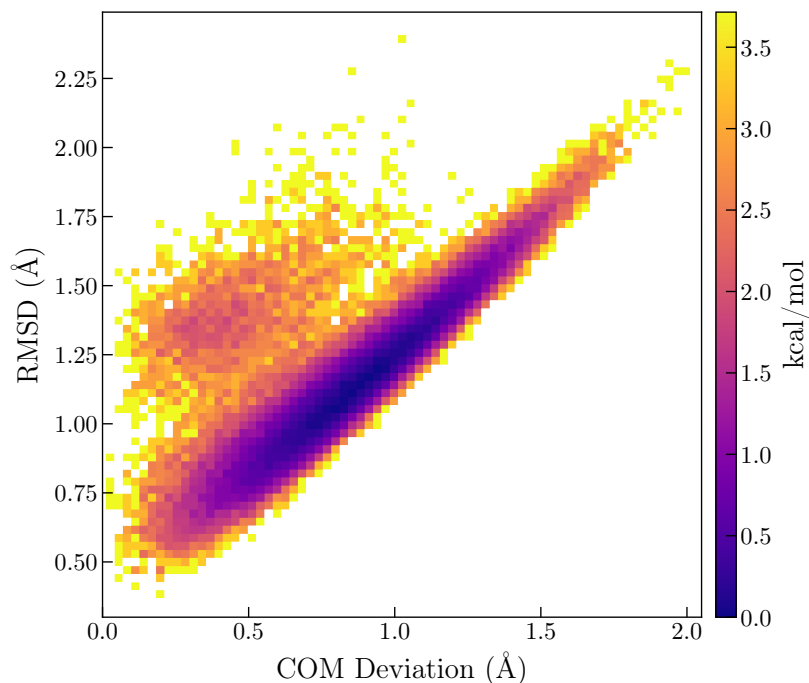


Figure A.3: 2D-PMF calculated from the 6.6 μ s classical MD simulation as a function of the (x -axis) COM and the (y -axis) RMS deviations of the inhibitor's heavy atoms from its crystallographic binding pose (PDB ID: 5P9J¹³²). Both deviations were calculated post-alignment of the binding pocket backbone C_α 's (within 10 Å of the inhibitor) to the crystallographic structure.

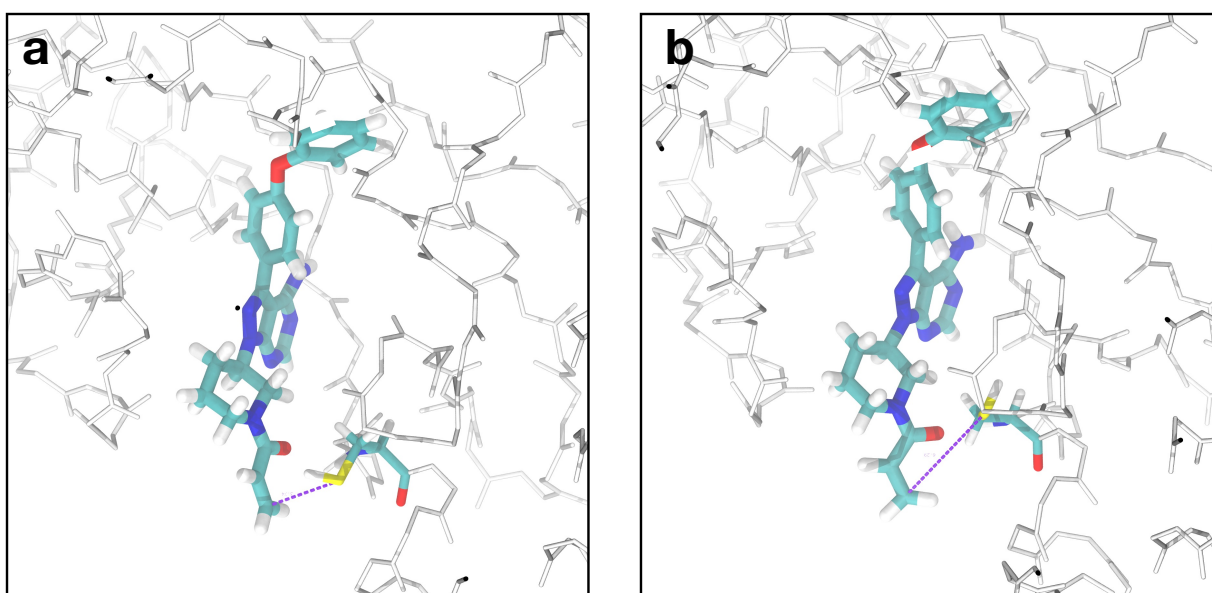


Figure A.4: Conformations observed during the 6.6 μ s MD simulation with the of the unlinked inhibitor associated in the BTK binding pocket. In (a) the interatomic distances between the C_{β} atom (cyan) of the warhead and the S_{γ} atom (yellow) of Cys481 is 6.1 Å and in (b) the distance is 3.8 Å. See also Figure 3.6 in the main text.

A.2 Reaction pathways

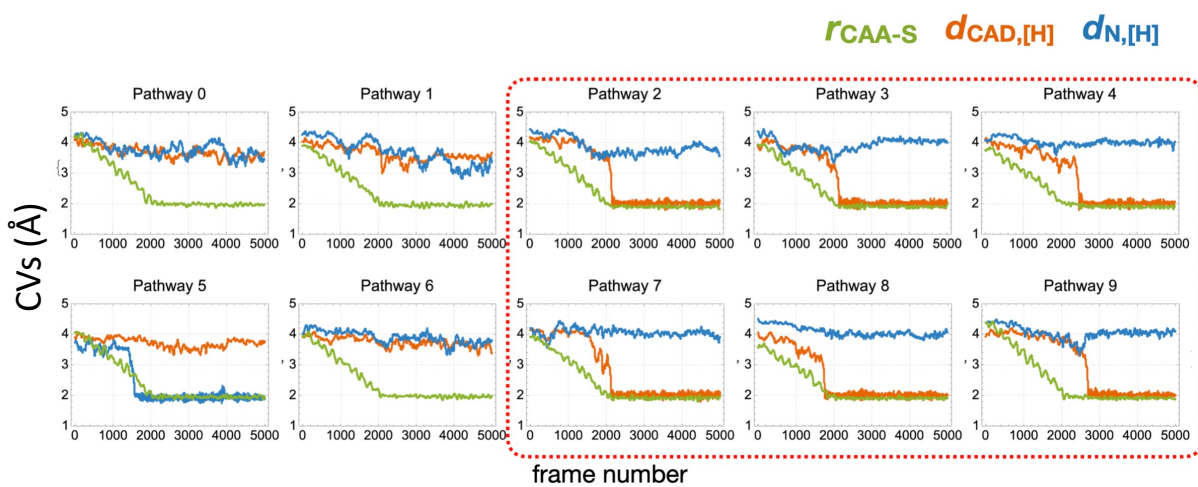


Figure A.5: All 10 initial pathways generated from SMD trajectories.

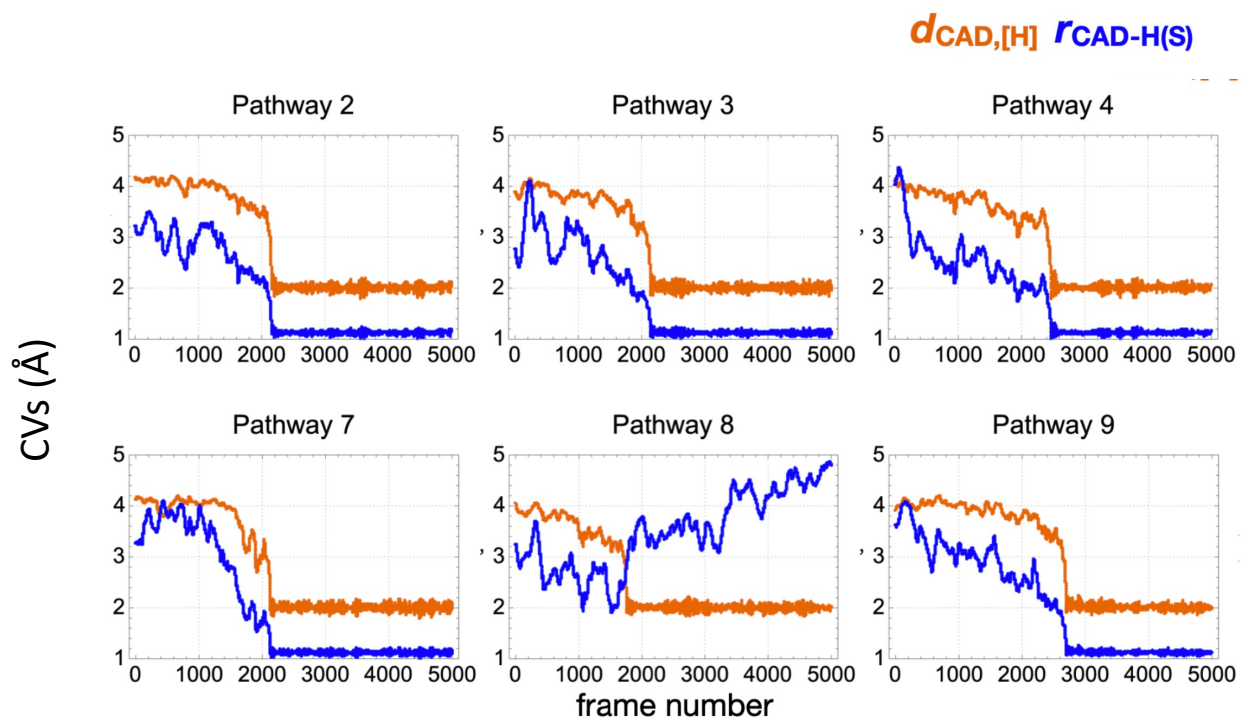


Figure A.6: Completing the reaction for the indirect pathway with SMD.

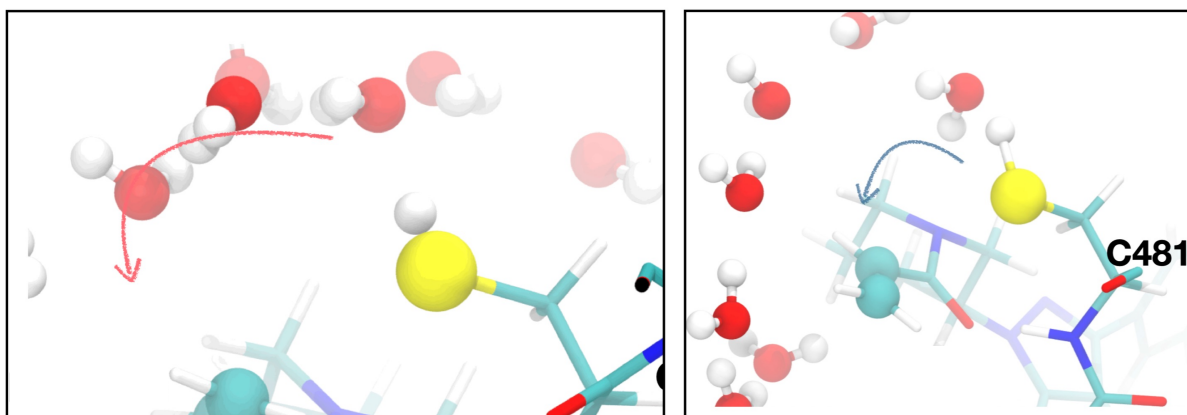


Figure A.7: Indirect (left) and direct (right) pathway observed from SMD trajectories. Carbon is cyan, sulfur is yellow, oxygen is red, nitrogen is blue, and hydrogen is white.

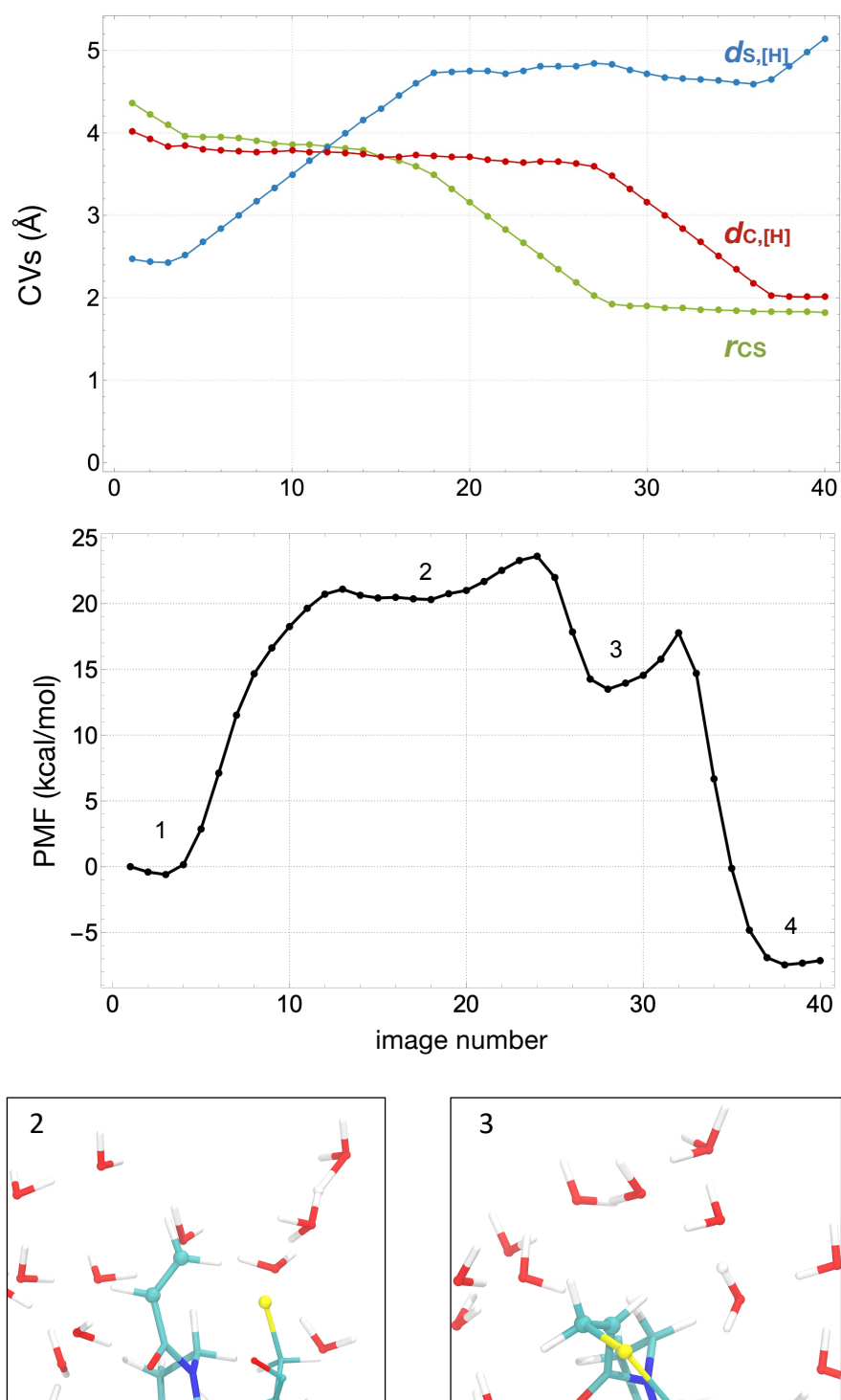
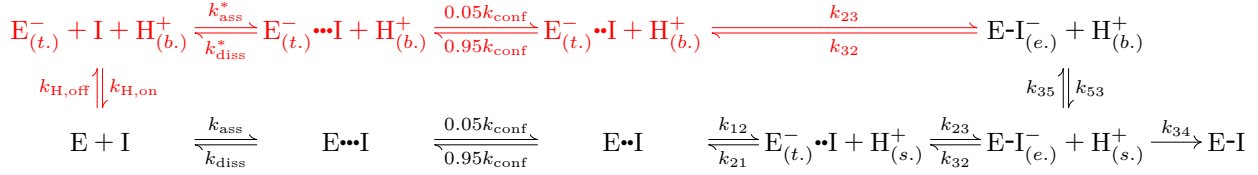


Figure A.8: Evolution of the CVs and PMF along the optimized string of the indirect path (top). Instantaneous snapshots along the string corresponding to metastable states 2 and 3 (bottom). Carbon is cyan, sulfur is yellow, oxygen is red, nitrogen is blue, and hydrogen is white.

A.3 Expanded kinetic scheme

Scheme A.1 includes the possibility of a diffusive encounter with a kinase containing a deprotonated target cysteine (thiolate form).



Scheme A.1: Detailed mechanism including the diffusive encounter with the thiolate form.

The clockwise and anticlockwise product of the transition rate constants around the ring must obey the condition,

$$1 = \frac{k_{\text{ass}}}{k_{\text{diss}}} \frac{0.95k_{\text{conf}}}{0.05k_{\text{conf}}} \frac{k_{12}}{k_{21}} \frac{k_{23}}{k_{32}} \frac{k_{35}}{k_{53}} \frac{k_{32}}{k_{23}} \frac{0.05k_{\text{conf}}}{0.95k_{\text{conf}}} \frac{k_{\text{diss}}^*}{k_{\text{ass}}^*} \frac{k_{\text{H,on}}}{k_{\text{H,off}}} = \frac{1}{k_{\text{diss}}} \frac{k_{12}}{k_{21}} k_{35} k_{\text{diss}}^* \frac{1}{k_{\text{H,off}}} \quad (\text{A.1})$$

where we assumed that $k_{\text{ass}} = k_{\text{ass}}^*$ and $k_{\text{H,on}} = k_{53}$. Solving for k_{diss}^* yields $1.22 \times 10^{-3} \text{ s}^{-1}$, which is in very close agreement to $k_{\text{diss}} = 1.11 \times 10^{-3} \text{ s}^{-1}$. The numerical simulations of Scheme A.1 shown in Figure A.9 yields an estimated covalent efficiency, C_{eff} , of $1.01 \times 10^{-3} \text{ nM}^{-1}\text{s}^{-1}$, an apparent inhibition constant, K_{I} , of $2.83 \times 10^3 \text{ nM}$, and an apparent inactivation rate constant, k_{inact} , of 2.86 s^{-1} .

Comparing the two possible mechanism is visually more compelling via a free energy diagram of all the states and transitions, as shown in Figure A.10.

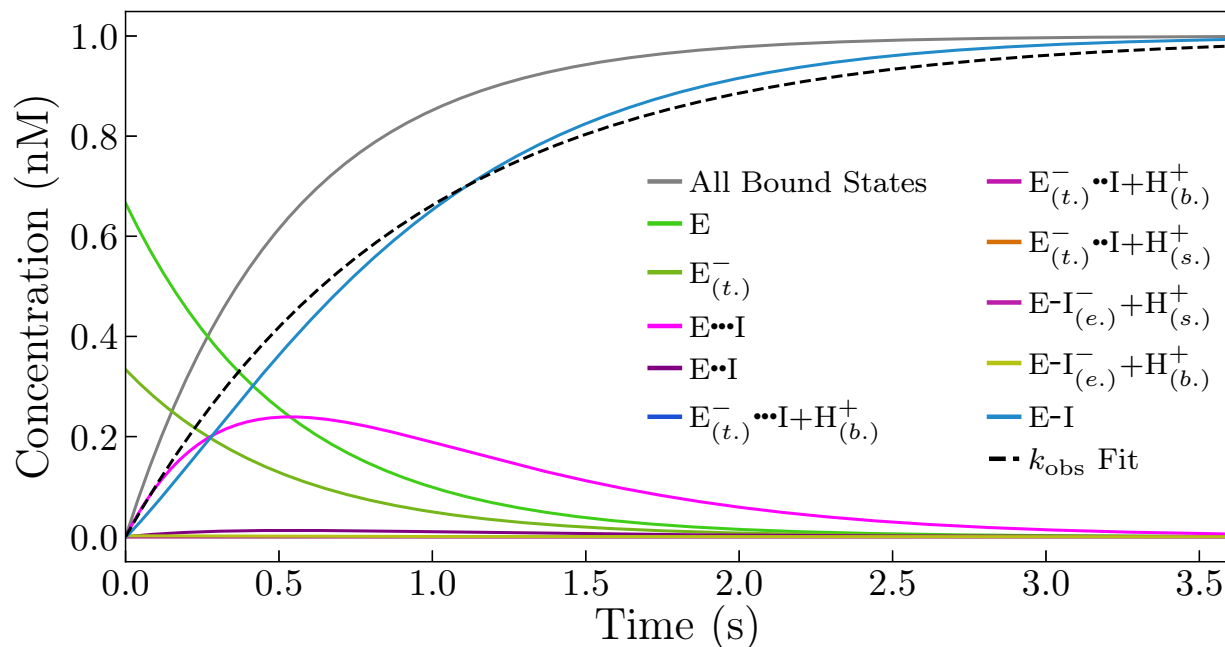


Figure A.9: Time course of Scheme A.1. The initial concentrations used for Figure 3.10 (top), $[I] = 1.72 \times 10^3$ nM and $[E] + [E_{(t.)}^-] = 1$ nM, yielded $k_{\text{obs}} = 1.09 \text{ s}^{-1}$. The total occupancy (gray) is identical between Schemes 3.2 and A.1.

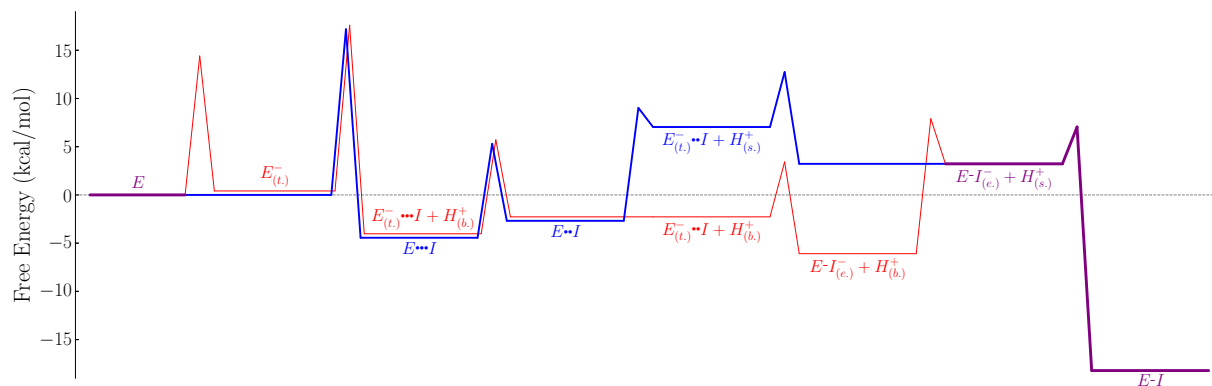


Figure A.10: Schematic free energy profile of the thiol (blue) and thiolate (red) reaction pathways corresponding to Scheme A.1. The bimolecular association rate constants, k_{ass} and k_{ass}^* , are multiplied by the initial concentration of inhibitor, $[I] = 1.72 \times 10^3$ nM.

APPENDIX B

SUPPLEMENTARY INFORMATION FOR CHAPTER 4

B.1 Supplementary figures and tables

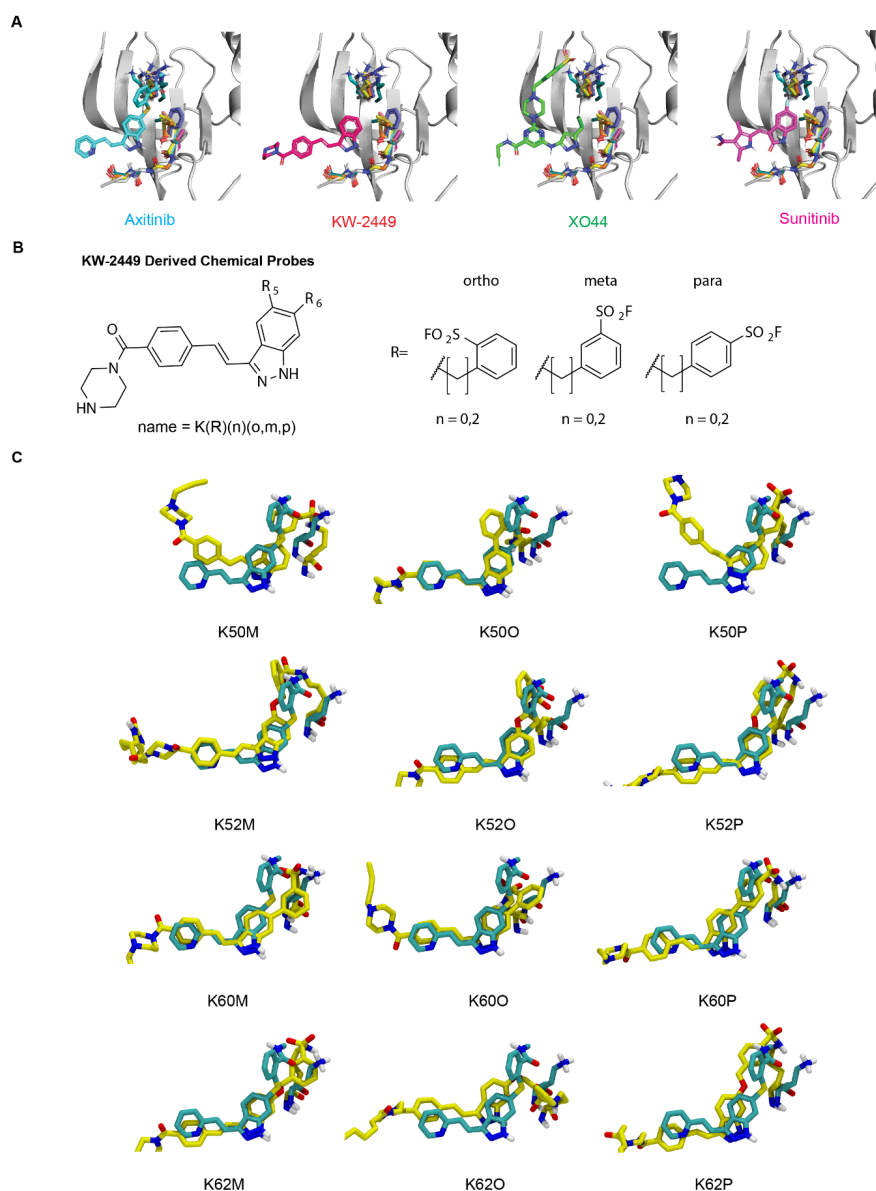


Figure B.1: Modeled poses of starting scaffolds and final linked poses of screened compounds. (A) Overlaid binding poses of relevant ligands (shown separately) alongside overlaid conserved binding pocket features from each of 7 kinase families (B) Naming convention for KW-2449 scaffold probes in molecular dynamics screening. (C) Final structures of KW-2449 scaffold probes and lysine in MD screening simulations (yellow) after 100 ns, overlaid with co-crystallized poses of axitinib and lysine from PDB ID: 4WA9 (cyan).

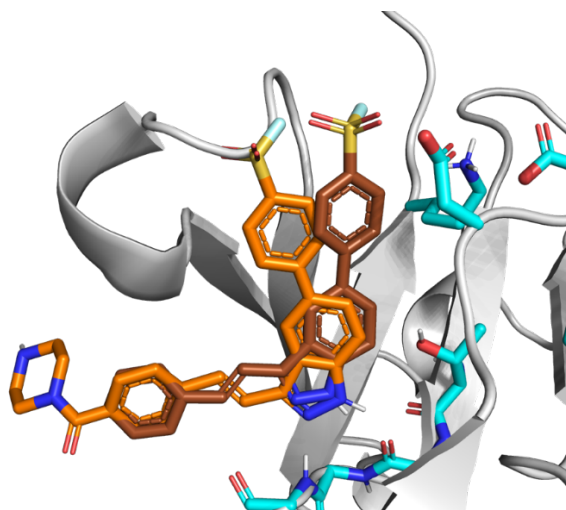


Figure B.2: Alternative alkene rotation for K5* series compounds. Modeled pose of K50P (orange) aligned to axitinib in PDB ID: 4TWP, and alternative rotamer of the ring-connecting alkene (brown). The backbone of kinase hinge residues, gatekeeper residue, conserved lysine, and its salt-bridge partners are shown in cyan.

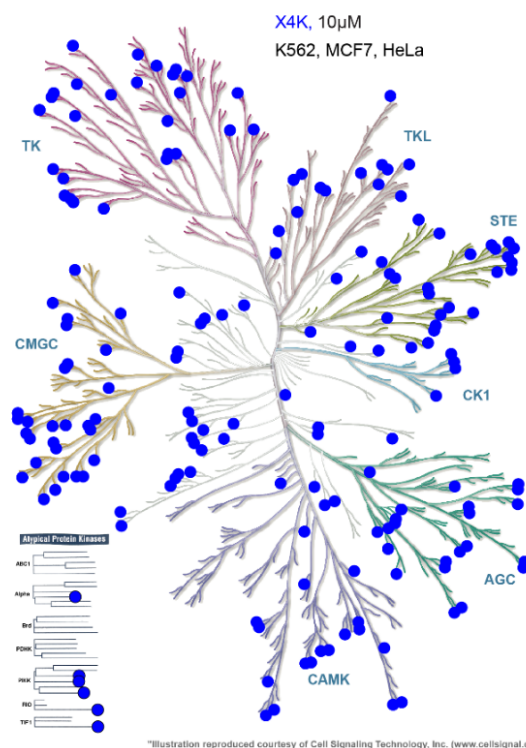


Figure B.3: Kinome scope of X4K. Dendrogram representation of the kinases targeted by X4K in K562, HeLa, and MCF7 cell lines shown as blue spheres.

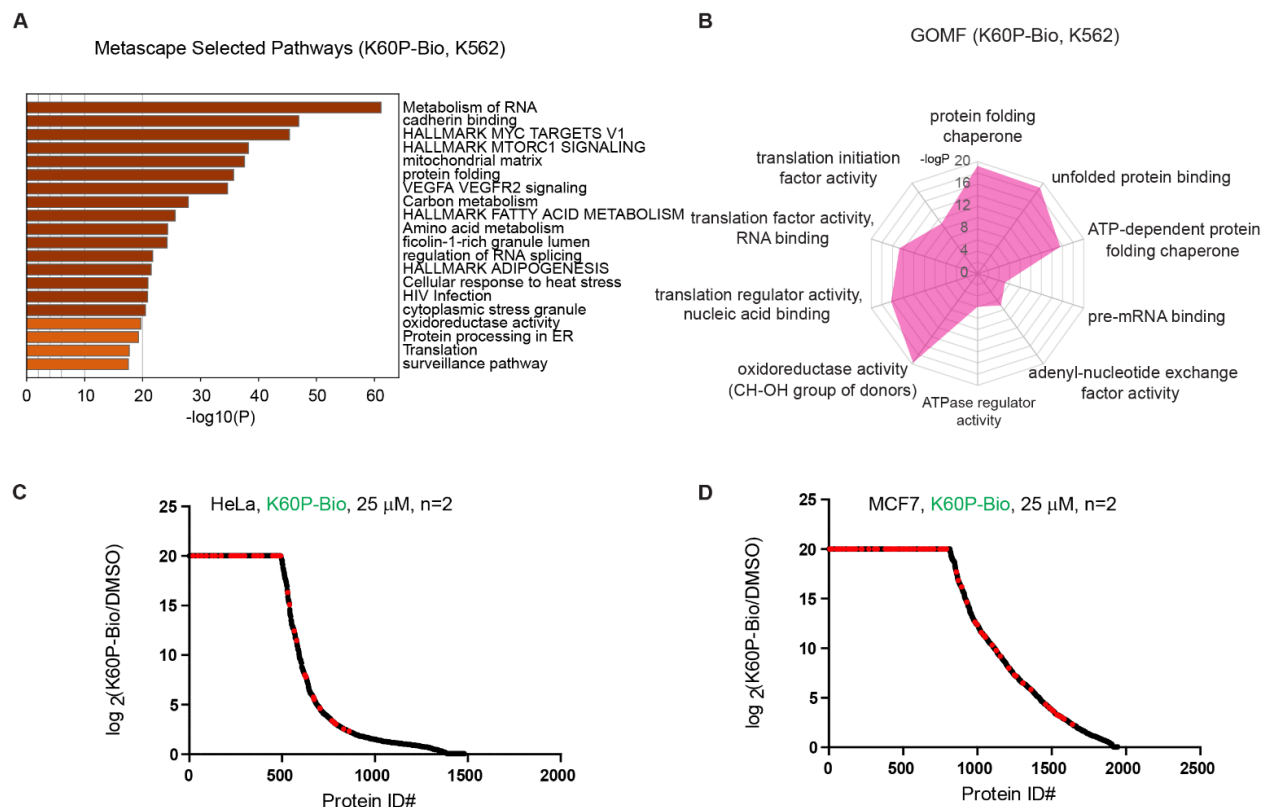


Figure B.4: K60P-Bio labels kinases and other classes of proteins across different cell lines. (A) Metascape analysis of proteins enriched more than 2-fold by K60P-Bio in K562 cells. (B) Gene Ontology Molecular Function analysis of proteins enriched more than 2-fold by K60P-Bio in K562 cells. (C–D) Waterfall plot of proteins enriched by K60P-Bio (25 μ M, 4 hours) in live HeLa (C) and MCF7 (D) cells. Kinases are shown as red dots. Control cells were treated with DMSO. Cutoff: SILAC ratio ≥ 2 ; n=2.

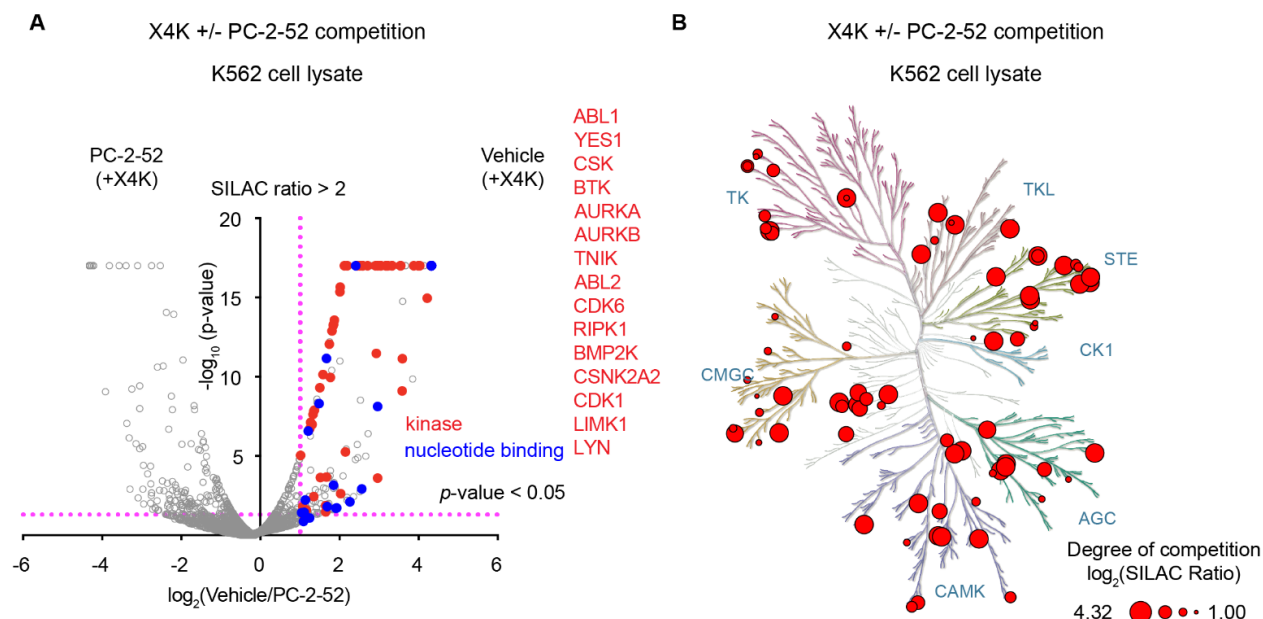


Figure B.6: (A) Volcano plot of streptavidin-enriched SILAC ratios for competition of PC-2-52 with X4K in K562 cell lysate. Red dots are kinases and blue dots are nucleotide-binding proteins. Cutoffs: competition SILAC ratio ≥ 2 , $p < 0.05$; $n = 2$. (B) Dendrogram showing kinases competed by PC-2-52 from X4K. Red dots representing kinases are scaled according to their SILAC ratios.

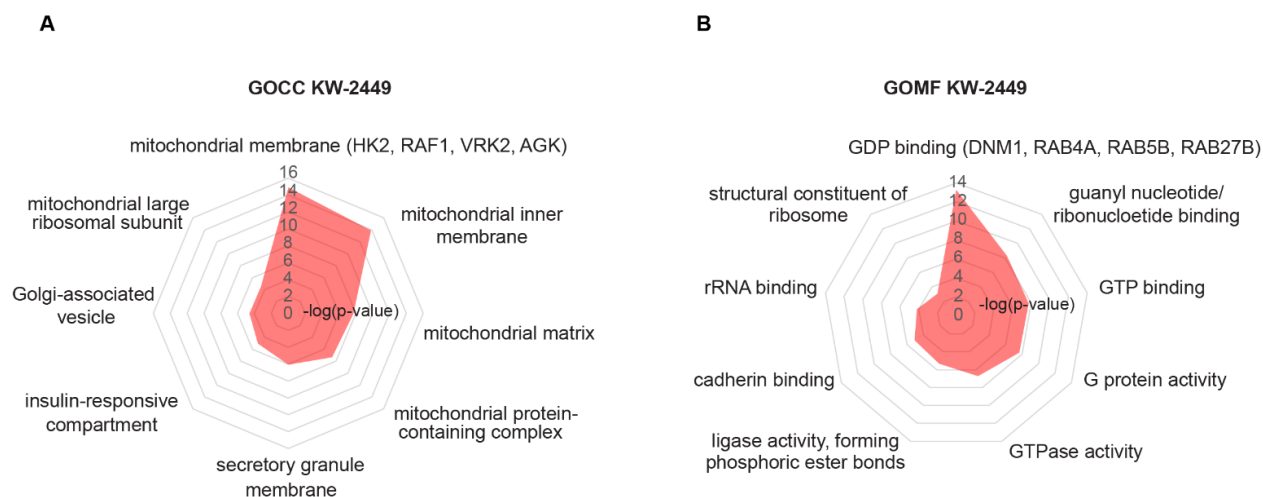


Figure B.7: Gene Ontology cellular component (A) and molecular function (B) bioinformatic enrichment analysis of all protein targets significantly competed by KW-2449 from K60P-Bio (SILAC ratio ≥ 2 , $p < 0.05$).

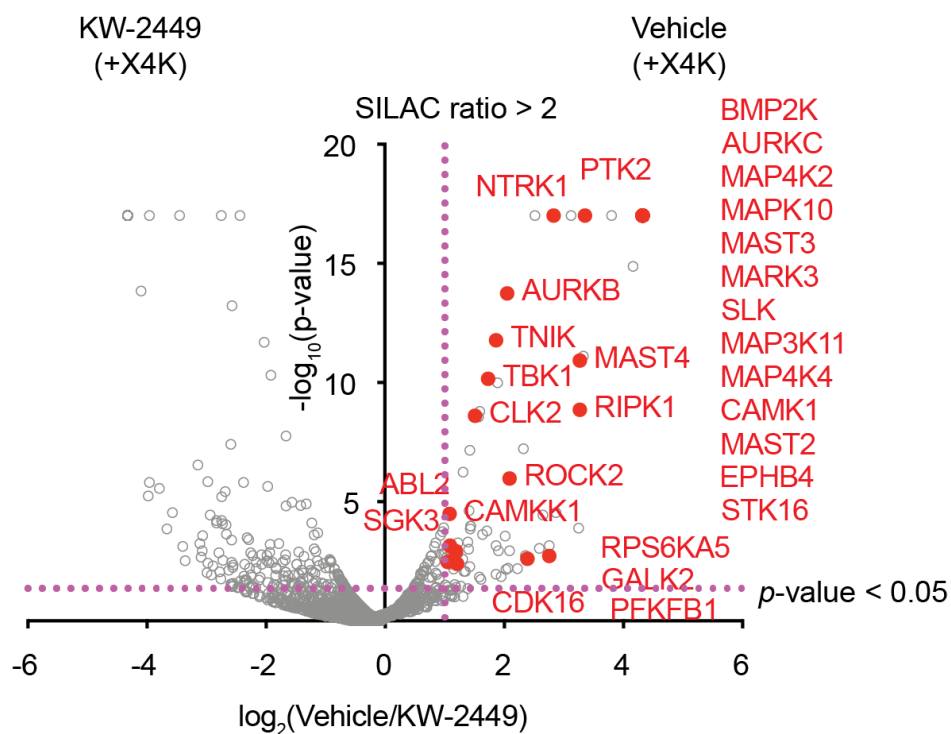


Figure B.8: Volcano plot of streptavidin-enriched SILAC ratios for K562 live cells treated with 2 μM KW-2449 (1 hour) followed by 5 μM X4K (1 hour). Cutoffs: competition SILAC ratio ≥ 2 , $p < 0.05$; $n = 2$.

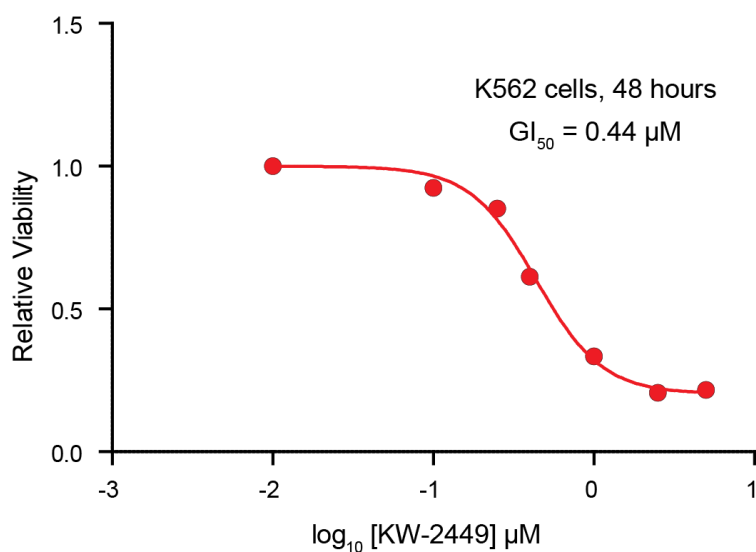


Figure B.9: 2-day CTG assay viability curve of KW-2449 in K562 cells; $n = 3$.

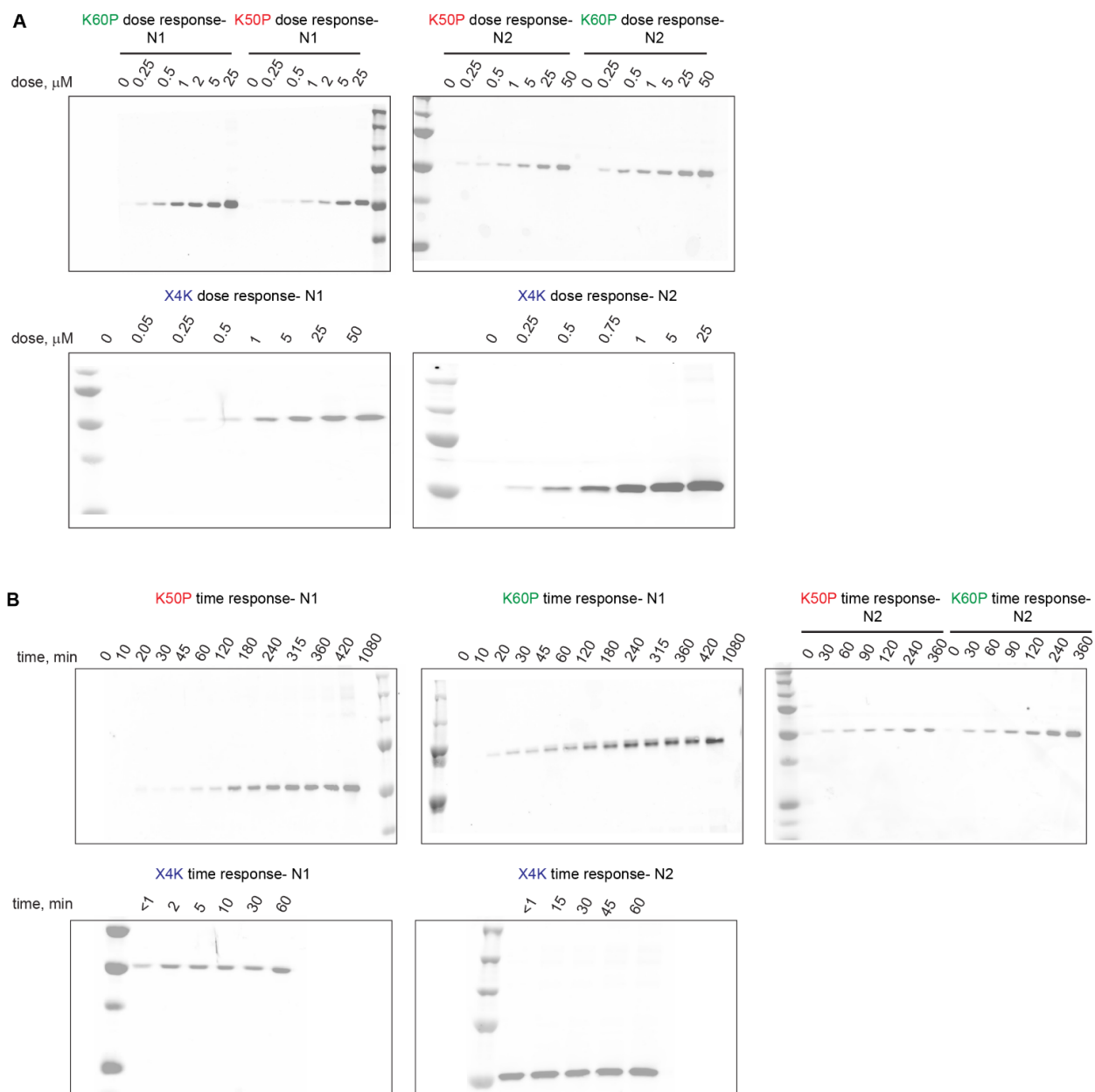


Figure B.10: Uncropped blots for dose- and time-responses of ABL1 labeling by K60P, K50P, and X4K. Bioreplicates of Figure 4.4A and 4C: dose response (A) and time response (B) of ABL1 labeling by K60P, K50P, and X4K.

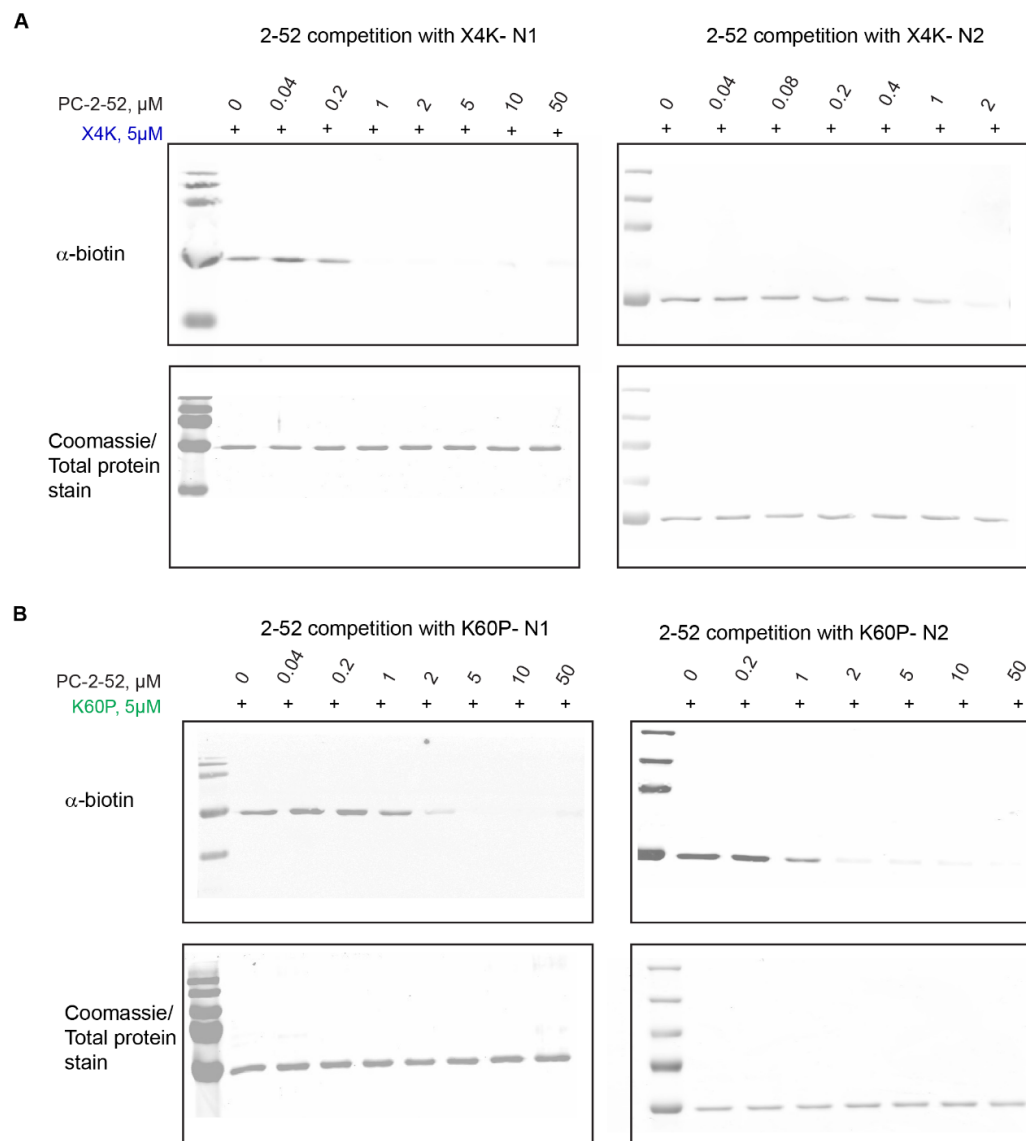


Figure B.11: Uncropped blots for competition of PC-2-52 with X4K and K60P. Bioreplicates of Figure 4.4B: competition of PC-2-52 with X4K (A) and with K60P (B) for ABL1 labeling.

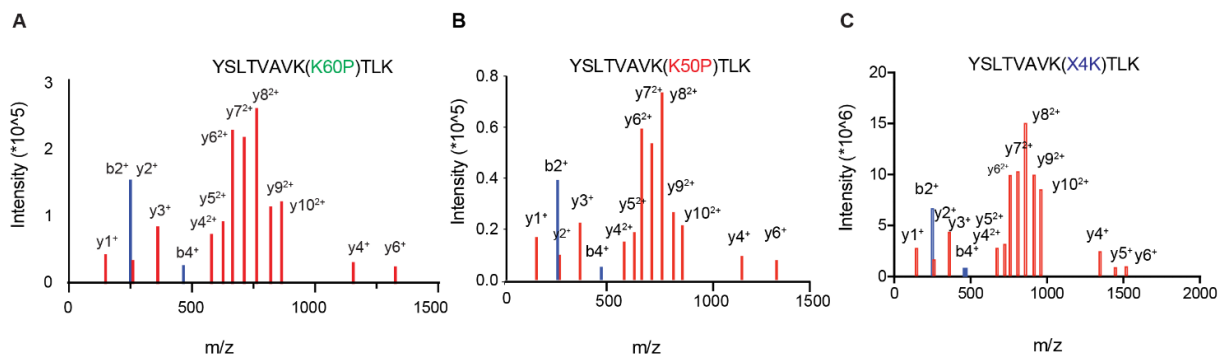


Figure B.12: K60P, K50P, and X4K covalently label the conserved K271 of ABL1. MS/MS spectra of K271-labeled peptides for K60P (A), K50P (B), and X4K (C).

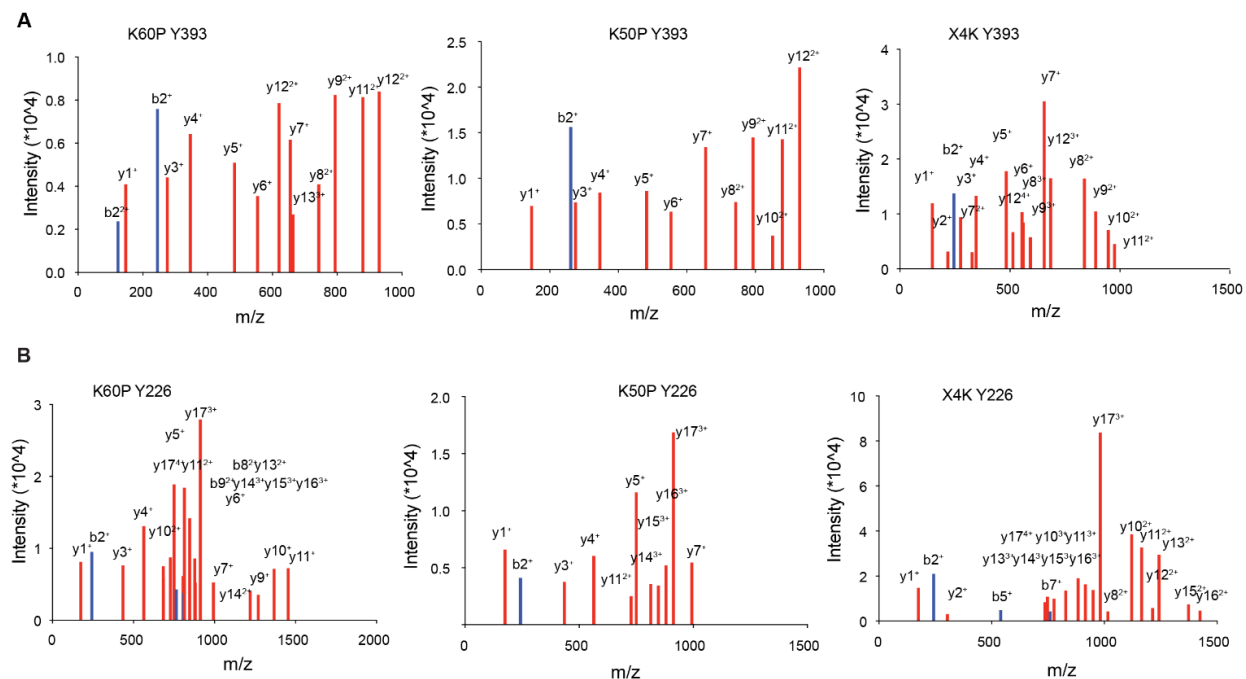


Figure B.13: K60P, K50P, and X4K covalently label Y393 and Y226 in ABL1. MS/MS spectra of Y393 (A) and Y226 (B) labeled peptides for K60P (left), K50P (middle), and X4K (right).

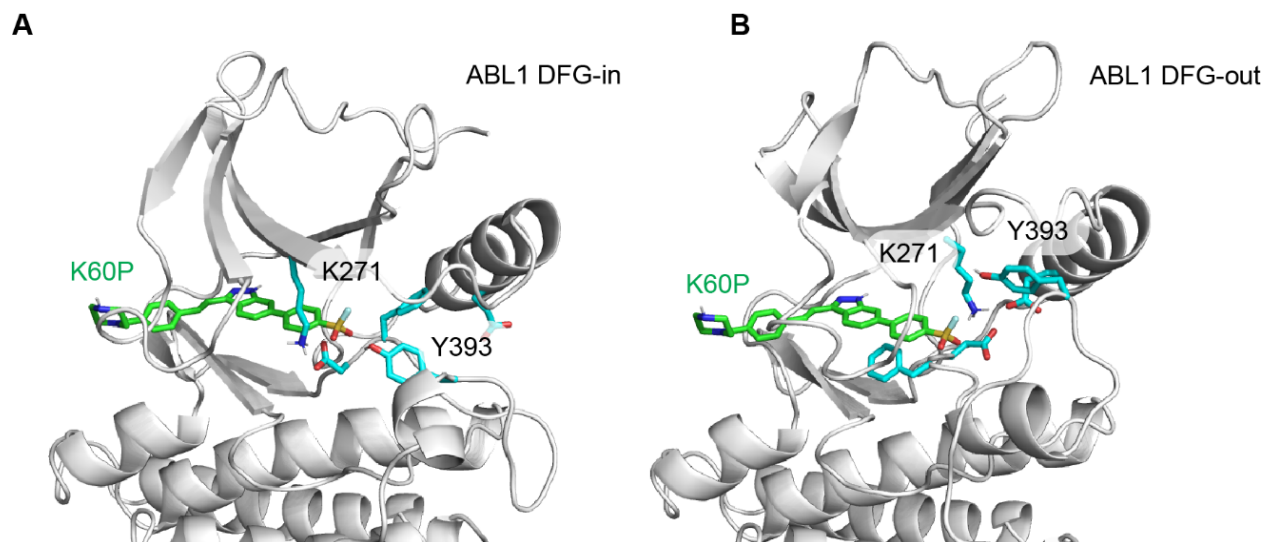


Figure B.14: Conformations of apo-ABL1 identified with Y393 closest to K60P. K60P is modeled in green with DFG-in (A) and DFG-out (B) conformations. K271 and Y393 are shown in cyan.

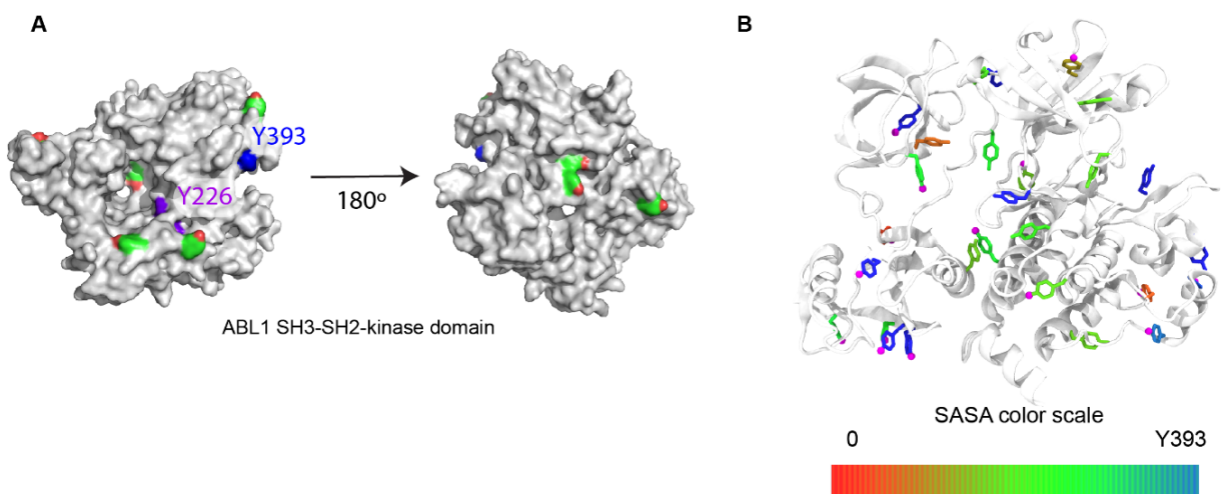


Figure B.15: (A) Surface map of SH3-SH2-kinase domain of ABL1 (PDB ID: 1OPK) showing surface exposed tyrosines. Y393 is shown in blue, Y226 is shown in purple, and the other tyrosines are shown with side chain color coded by atoms (C: green, O: red). (B) SH3-SH2-kinase domain of ABL1 (PDB ID: 1OPK) tyrosine residues colored according to solvent accessible surface area (SASA) of their sidechains from red (0) to Y393 or equivalent (blue). Tyrosine -OH groups coordinating with crystallographic waters are indicated by pink spheres.

Table B.1: Solvent-exposed tyrosines in the SH3-SH2-kinase domain of ABL1 with their sidechain solvent-accessible surface areas (SASA, calculated from PDB structure 1OPK) and their peptide sequences. Tyrosines are numbered according to PDB structure 4WA9. Important tyrosines are in bold.

Tyr	SASA ($\text{\AA}^2$)	Peptide sequence
93	98.7	VLGYNHNGEWCEAQTK
185	41.3	LYVSSESR
215	101.0	FNTLAELVHHHSTVADGLITTLHYPAPK
226	38.4	NKPTVYGVSPNYDK
232	156.9	NKPTVYGVSPNYDK
320	83.9	EPPFYIITEFMTYGNLLDYLR
326	32.6	EPPFYIITEFMTYGNLLDYLR
342	39.2	QEVNAVVLlyMATQISSAMEYLEK
393	72.2	LMTGDTYTAHAGAK
413	123.6	WTAPESLAYNK

Table B.2: The estimated median and 95% highest density interval ($^{\text{upper}}$ _{lower}) of parameters from BMCMC simulations.

	K50P		K60P		X4K
	Mod _{del}	Mod _{sil}			
k_{on} (($\mu\text{M s}$) ⁻¹)	1.82×10^{-2} *	1.82×10^{-2} *	1.82×10^{-2} *	1.26×10^{-3} (ND)	5.20×10^{-3} 4.56×10^{-4}
k_{off} (s ⁻¹)	7.62×10^{-2} 1.50×10^{-1} 2.29×10^{-2}	1.17×10^{-1} 2.25×10^{-1} 5.20×10^{-2}	2.10×10^{-2} 4.54×10^{-2} 5.80×10^{-3}	8.95×10^{-2} *	
K_{d} (μM)	4.19×10^0 8.27×10^0 1.26×10^0	6.43×10^0 1.24×10^1 2.86×10^0	1.15×10^0 2.50×10^0 3.19×10^{-1}	7.13×10^1 (ND)	1.51×10^2 7.89×10^0
k_{inact} (s ⁻¹)	1.19×10^{-4} 1.64×10^{-4} 7.40×10^{-5}	2.06×10^{-4} 2.82×10^{-4} 1.41×10^{-4}	1.91×10^{-4} 2.45×10^{-4} 1.40×10^{-4}	1.34×10^{-1} (ND)	3.12×10^{-1} 5.99×10^{-3}
K_{I} (μM)	4.19×10^0 8.28×10^0 1.26×10^0	6.44×10^0 1.24×10^1 2.87×10^0	1.16×10^0 2.51×10^0 3.27×10^{-1}	1.86×10^2 (ND)	5.85×10^2 8.45×10^0
C_{eff} (($\mu\text{M s}$) ⁻¹)	2.82×10^{-5} 5.03×10^{-5} 1.55×10^{-5}	3.18×10^{-5} 5.32×10^{-5} 1.70×10^{-5}	1.63×10^{-4} 2.92×10^{-4} 6.85×10^{-5}	6.43×10^{-4} (ND)	1.35×10^{-3} 3.32×10^{-4}
k_{ns} (($\mu\text{M s}$) ⁻¹)	5.69×10^{-7} 1.43×10^{-6} 1.46×10^{-9}	5.38×10^{-7} 1.53×10^{-6} 1.85×10^{-9}	1.65×10^{-6} 2.22×10^{-6} 1.11×10^{-6}	5.14×10^{-6}	6.86×10^{-6} 3.36×10^{-6}
k_{del} (s ⁻¹)	1.12×10^{-4} 1.71×10^{-4} 5.44×10^{-5}				
Frac _{sil}		2.79×10^{-1} 4.21×10^{-1} 1.28×10^{-1}			
λ_{BR1}	1.56×10^0	1.46×10^0	6.65×10^{-1}	5.93×10^{-1}	
λ_{BR2}	1.37×10^0	1.28×10^0	8.01×10^{-1}		
λ_{BR3}	2.40×10^0	2.34×10^0	1.29×10^0	5.33×10^{-1}	
λ_{BR4}	3.05×10^0	2.92×10^0	1.50×10^0		

Parameters without a range were fixed during BMCMC. Values marked with an asterisk (*) were also fixed during LS minimization. The K_{d} , K_{I} , and C_{eff} were not explicit parameters and were instead calculated analytically per sample. Parameters whose bounds could not be resolved are marked Not Determined (ND), with the median from BMCMC sampling in parentheses. The uncertainty behind these parameters is discussed in the supplementary kinetic modeling discussion, Section B.6.

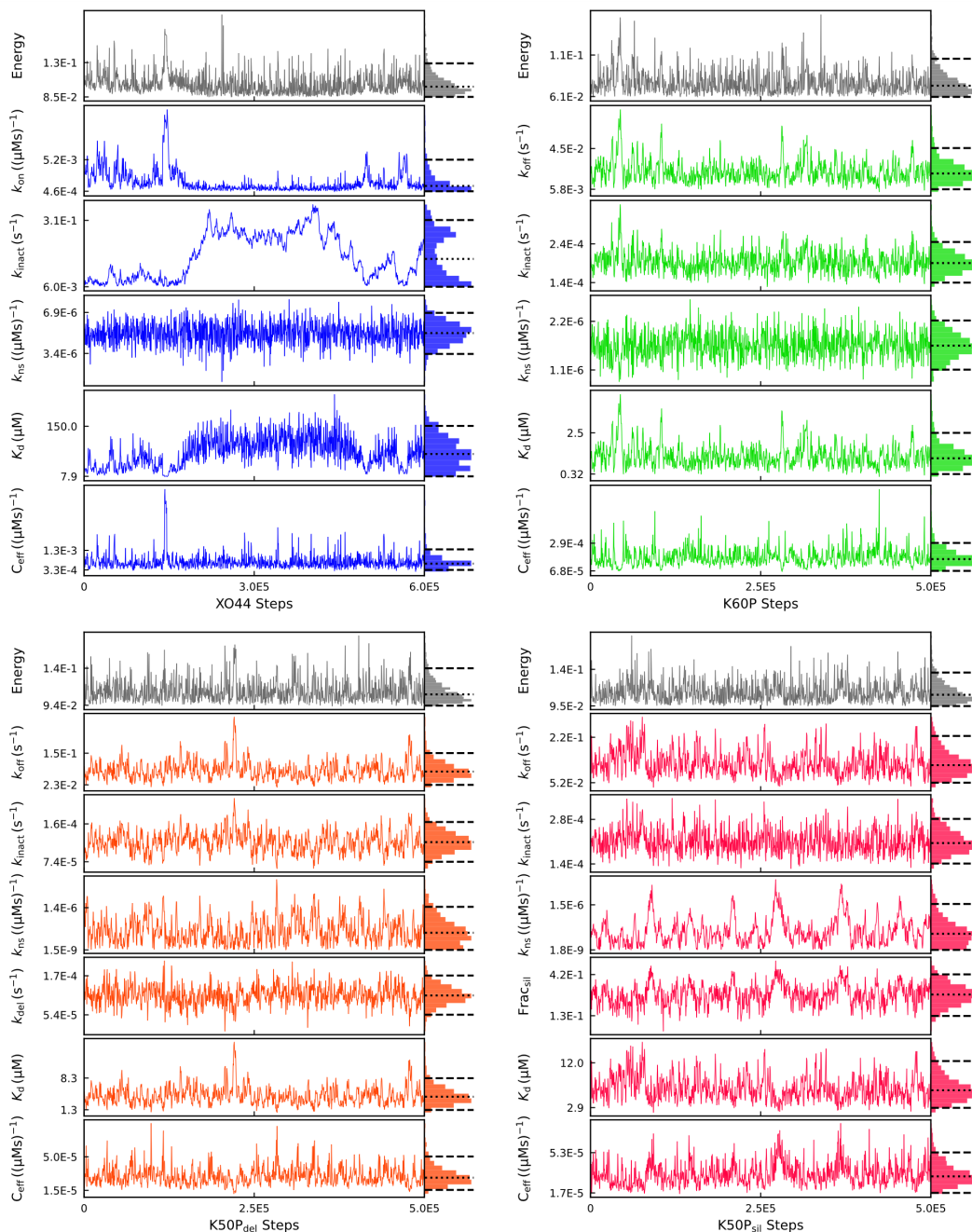


Figure B.16: Time-evolution of the effective energy and variable parameters during the BMMC simulations. Samples were collected every 500 steps to account for autocorrelation of parameters. The equilibrium non-covalent dissociation constant ($K_d = k_{\text{off}}/k_{\text{on}}$) and covalent efficiency ($C_{\text{eff}} = k_{\text{inact}}/K_I$, where $K_I = (k_{\text{off}} + k_{\text{inact}})/k_{\text{on}}$) were calculated at each sample for K60P (green), K50P (orange, red), and X4K (blue). The distributions of each time evolution are shown to the right of each plot. The median of each parameter is marked with a dotted line, and the bounds of the 95% highest density interval are marked with a dashed line.

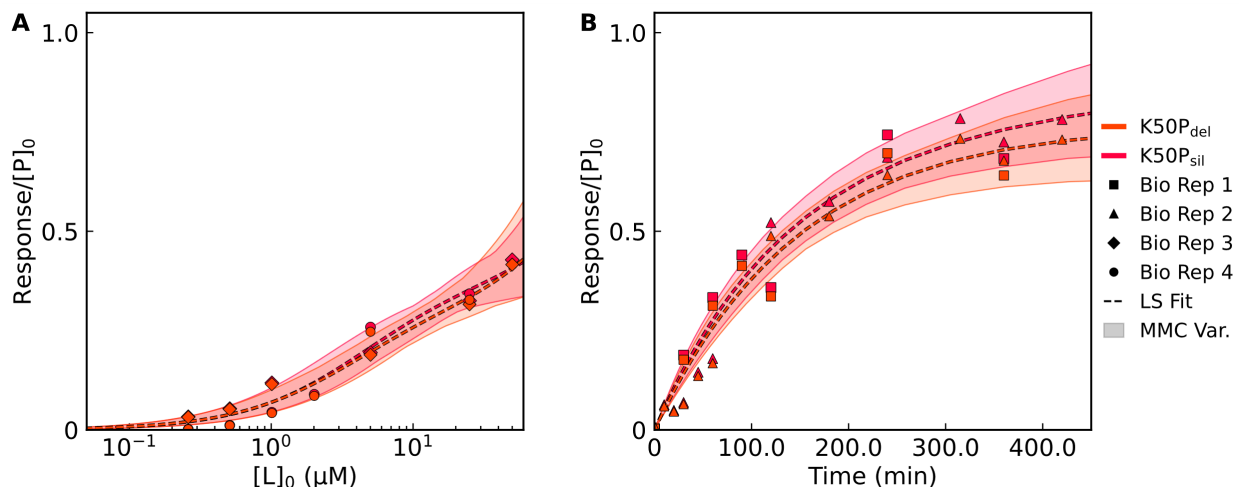


Figure B.17: The least-squares (LS) fit and Bayesian Metropolis Monte Carlo (BMMC) variation of dose-responses (A) and time-responses (B) for each K50P modified scheme. The samples have been filtered so that each kinetic parameter falls within its respective 95% highest density interval. Simulated “response” is the sum of the specifically- and nonspecifically-bound covalent complexes.

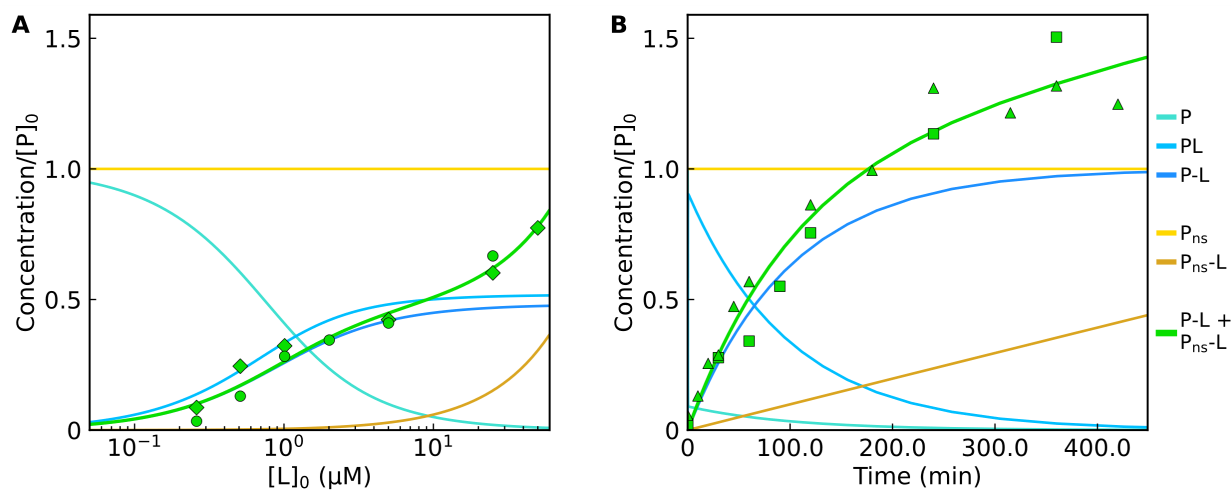


Figure B.18: Evolution of species over time during a simulated dose-response (A) and time-response (B) using the best parameter values from the least-squares fit for K60P. The normalized concentrations of all Scheme 1 species, the normalized sum of both covalently bound species, and the normalized experimental data are shown.

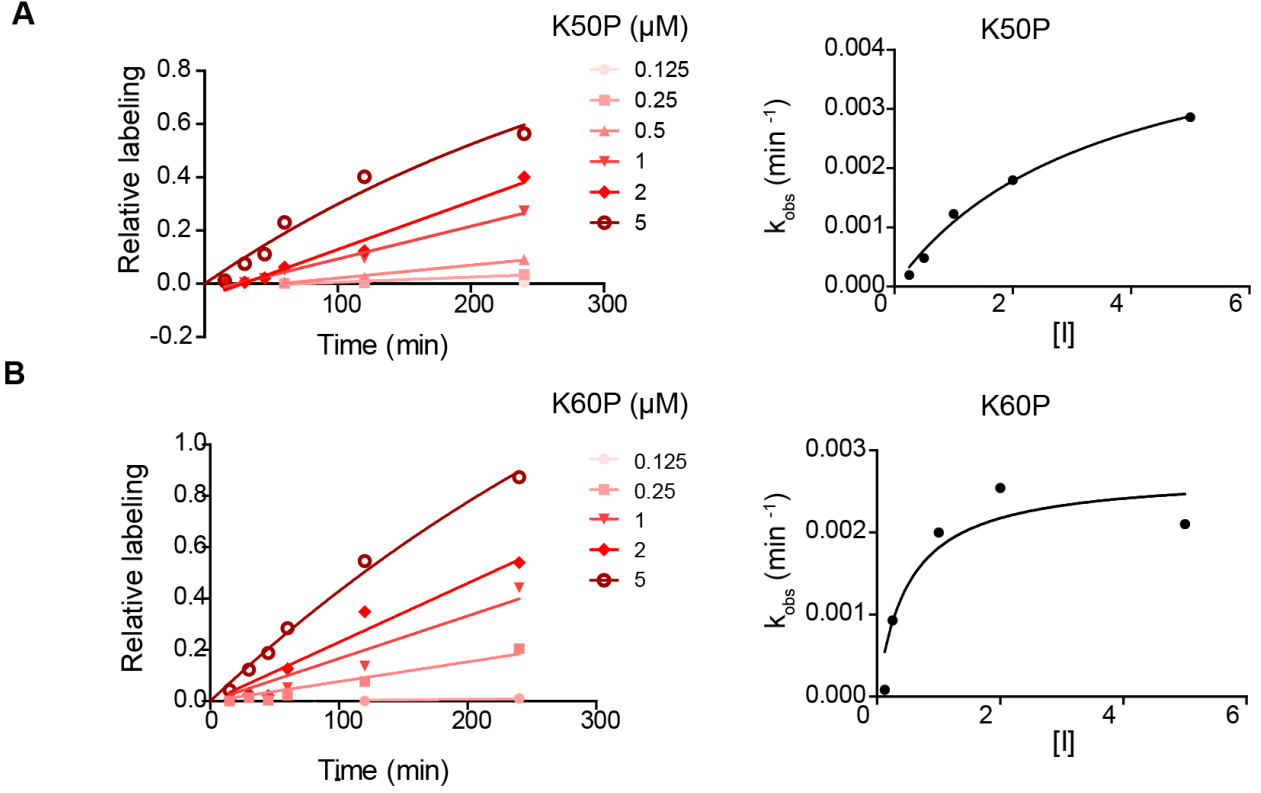


Figure B.19: Experimental determination of k_{inact}/K_I , k_{inact} , and K_I for K50P (A) and K60P (B). Left: time responses at different doses of probe. Right: plot of k_{obs} versus $[I]$.

Table B.3: Experimentally determined k_{inact}/K_I , k_{inact} , and K_I for K60P and K50P along with their ranges from BMMC simulations. The ranges from the Mod_{sil} simulations are considered for K50P.

	Experimental value	Range from BMMC sampling
K60P		
k_{inact}	$0.4 \times 10^{-4} \text{ s}^{-1}$	$(1.40\text{--}2.45) \times 10^{-4} \text{ s}^{-1}$
K_I	$0.49 \text{ } \mu\text{M}$	$0.33\text{--}2.51 \text{ } \mu\text{M}$
k_{inact}/K_I	$0.9 \times 10^{-4} \text{ } \mu\text{M}^{-1} \text{ s}^{-1}$	$(0.68\text{--}2.92) \times 10^{-4} \text{ } \mu\text{M}^{-1} \text{ s}^{-1}$
K50P		
k_{inact}	$0.81 \times 10^{-4} \text{ s}^{-1}$	$(1.41\text{--}2.82) \times 10^{-4} \text{ s}^{-1}$
K_I	$3.46 \text{ } \mu\text{M}$	$2.87\text{--}12.4 \text{ } \mu\text{M}$
k_{inact}/K_I	$0.23 \times 10^{-4} \text{ } \mu\text{M}^{-1} \text{ s}^{-1}$	$(0.17\text{--}0.53) \times 10^{-4} \text{ } \mu\text{M}^{-1} \text{ s}^{-1}$

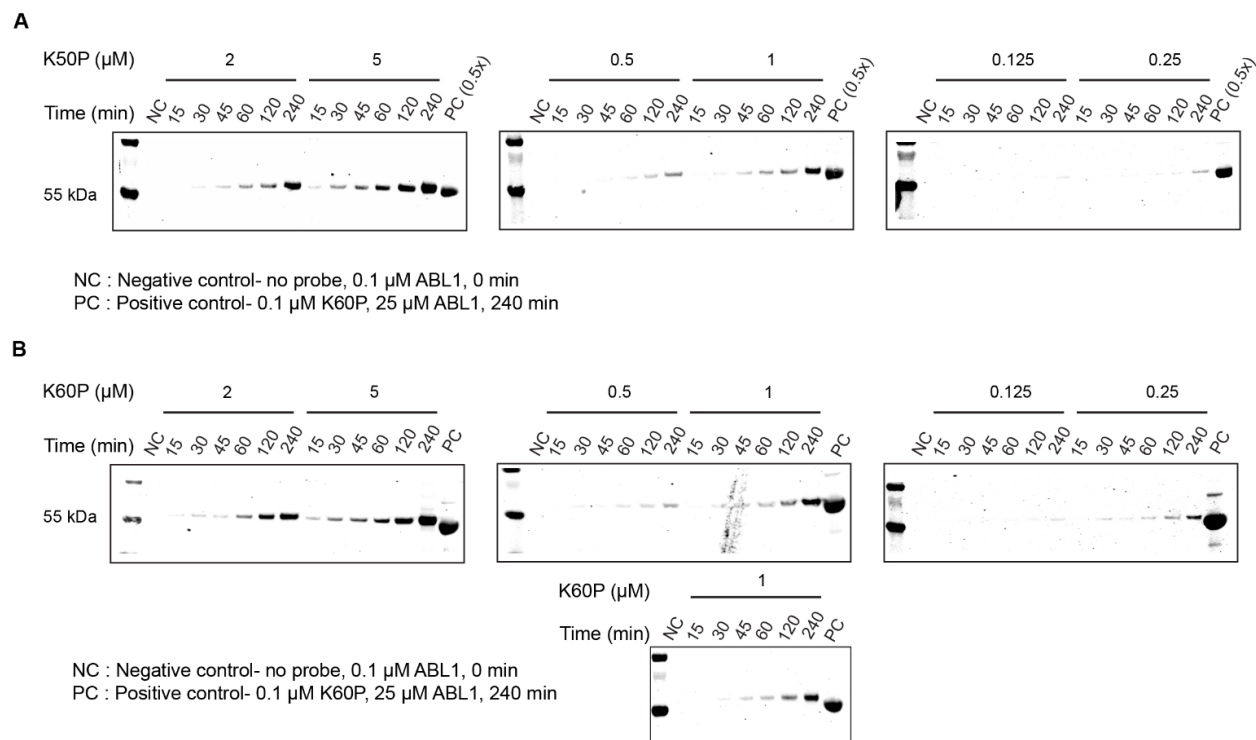


Figure B.20: Time responses of different doses of K50P (A) and K60P (B) with 0.1 μM recombinant ABL1. Negative control (NC) was no-probe, 0 min sample. Positive control (PC) was 0.1 μM K60P, 25 μM ABL1, 240 min-treated sample. For K50P, half the volume of PC was loaded (denoted PC (0.5x)). All lane signal intensities were normalized to PC intensity.

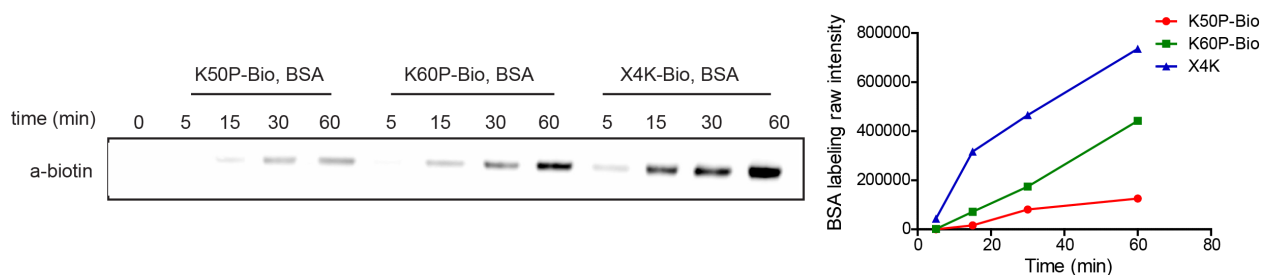


Figure B.21: X4K, K60P, and K50P exhibited different inherent reactivities. Rate of labeling of non-specific target protein BSA with K60P, K50P, and X4K shows intrinsic reactivity in order X4K > K60P > K50P.

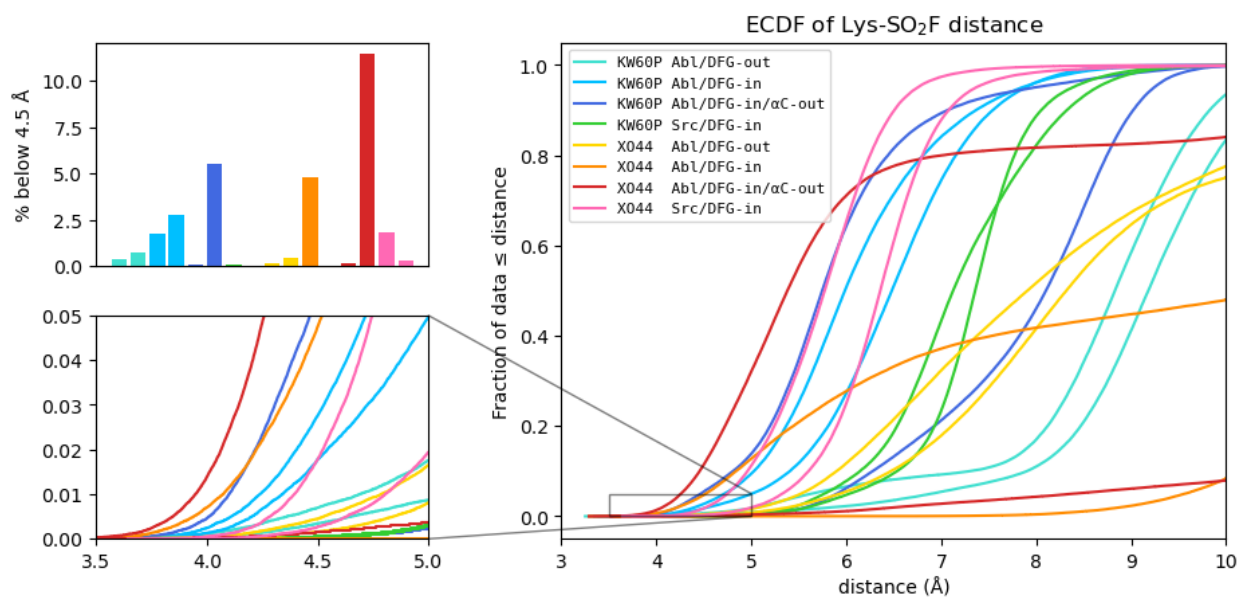


Figure B.22: Empirical Cumulative Distribution Function (ECDF) of the lysine to warhead-S distance in each 4 μ s simulation, excluding the first 100 ns. Duplicates with slight tweaks to starting conformation are shown as the same color.

B.2 Methods for biology

B.2.1 Cell culture

K562, HeLa, and MCF7 cells were propagated in RPMI (Corning) supplemented with 10% fetal bovine serum (FBS, Corning) and 1% penicillin/streptomycin (Gibco). All cell lines were grown at 37 °C in a 5% CO₂ humidified incubator. All cell lines tested negative for mycoplasma using Lonza MycoAlert PLUS Detection Kit.

B.2.2 In vitro recombinant Abl labeling assays

Protein purification

The c-Abl three-domain construct (SH3-SH2-kinase; residues 46-515, human Abl 1a numbering) was coexpressed in *E. coli* BL21(DE3) cells with YopH phosphatase and purified as described in Seeliger et al.²⁹² with the following modifications. Cleared cell lysate in 50 mM Tris (pH 8.0), 500 mM NaCl, 5% glycerol, 25 mM imidazole (buffer A) was loaded onto Ni-NTA agarose resin (Qiagen) equilibrated with buffer A. The resin was washed with 15 bed volumes of buffer A and eluted with 3 bed volumes of buffer B (buffer A plus 0.5 M imidazole). The His tag was cleaved from the eluted kinase by incubation with 1 mg of TEV protease per 25 mg of kinase while dialyzing against 20 volumes of 20 mM Tris (pH 8.0), 100 mM NaCl, 5% glycerol, 1 mM DTT for 16 h at 4°C with a 13 kDa molecular weight cutoff (MWCO) membrane. In order to separate the kinase from the cleaved His tag, TEV protease (which contains a His tag) and YopH phosphatase (which binds nonspecifically to the Ni-NTA resin), the dialysate was passed over Ni-NTA agarose resin equilibrated with buffer A. The flow-through contained pure kinase protein with the His tag removed. Concentration of the purified kinase was determined by absorbance spectroscopy at 280 nm using the calculated extinction coefficient of 60,550 M⁻¹ cm⁻¹.

The purified protein was buffer exchanged to PBS using 30kDa MWCO Millipore Amicon

centrifugal filter (3000×g, 10 min, 4°C, 8 times) to remove tris in the protein storage buffer from interacting with probe. The resulting protein concentration was measured using BCA assay and adjusted to make 0.025 mg/mL protein stock.

Dose response assays

For dose response, 20 μ L of protein stock was added to each sample followed by addition of 5 μ L of probe stock (5×). The samples were incubated for 1 hour at 37 °C. The samples were then diluted with 4× Laemmli buffer containing 50 mM dithiothreitol (DTT) as reducing agent, heated to 95 °C for 5 minutes, and cooled to room temperature. The samples were then resolved on a 10% SDS-PAGE gel and transferred to nitrocellulose membranes using standard western blotting techniques. Membranes were blocked with 2% BSA in TBS containing 0.1% Tween-20 (TBST) and probed with streptavidin-800 (Licor IRDye[®] 800CW, Licor Odyssey CLx imager) to detect probe labeling. Coomassie staining was done as loading control for the gels. Blot intensities were quantified using ImageJ and normalized in Microsoft Excel.

Time response assays

For time response, 20 μ L of protein stock was added to each sample followed by addition of 5 μ L of probe stock (5×, 50 μ M) and incubated at 37 °C. At the end of each time point, the respective sample was diluted with 4× Laemmli buffer containing 50 mM dithiothreitol (DTT) as reducing agent and heated to 95 °C for 5 minutes. The samples were then then probed for streptavidin signal according to western blot protocol mentioned above. Coomassie staining was done as loading control for the gels.

Competition assays with PC-2-52

For the competition of PC-2-52 with K60P and X4K, 20 μ L of protein stock was added to each sample followed by addition of 2.5 μ L of PC-2-52 stock (9×) and incubated at 37 °C

for 1 hour. Then, 2.5 μL of probe ($10\times$, 50 μM) of K60P (1 hour) or X4K (30 minutes). The sample was then diluted with $4\times$ Laemmli buffer containing 50 mM dithiothreitol (DTT) as reducing agent and heated to 95 $^{\circ}\text{C}$ for 5 minutes. The samples were then then probed for streptavidin signal according to western blot protocol mentioned above. Coomassie staining was done as loading control for the gels.

Kinetics assays

For K50P and K60P, 40 μL of protein stock was added to each sample tube. This was followed by the addition of $5\times$ probe stock to give final probe concentrations of 0.125, 0.25, 0.5, 1, 2, 5 μM and final protein concentration of 0.1 μM . Samples were incubated at 37 $^{\circ}\text{C}$ for 15, 30, 45, 60, 120 and 240 minutes. At the end of each timepoint, the samples were diluted with $4\times$ Laemmli buffer containing 50 mM dithiothreitol (DTT) as reducing agent and heated to 95 $^{\circ}\text{C}$ for 5 minutes. The samples were then then probed for streptavidin signal according to western blot protocol mentioned above. All signal intensities were normalized to reference sample containing 0.1 μM K60P, 10 μM protein, representing ‘maximum signal’. The response data at lower concentrations were fit using linear fits and the response data at higher concentrations were fit using one-phase association fit (shared plateau constrained to > 0) in GraphPad Prism. The k_{obs} obtained from the fits were plotted against respective probe concentrations $[\text{I}]$ and fit to the equation $k_{\text{obs}} = (k_{\text{inact}}[\text{I}]) / (K_{\text{I}} + [\text{I}])$ to obtain k_{inact} and K_{I} values.

B.2.3 BSA labeling assays

20 μL of a BSA stock solution (0.025 mg/mL) was added to each sample followed by addition of 5 μL of probe stock (50 μM) of K50P-Bio, K60P-Bio, or X4K. The samples were incubated for 5, 15, 30, and 60 minutes at 37 $^{\circ}\text{C}$. The samples were then diluted with $4\times$ Laemmli buffer containing 50 mM dithiothreitol (DTT) as reducing agent, heated to 95 $^{\circ}\text{C}$ for 5 minutes,

and cooled to room temperature. The samples were then resolved on a 10% SDS-PAGE gel and transferred to nitrocellulose membranes using standard western blotting techniques. Membranes were blocked with 2% BSA in TBS containing 0.1% Tween-20 (TBST) and probed with streptavidin-800 (Licor IRDye[®] 800CW, Licor Odyssey CLx imager) to detect probe labeling. Coomassie staining was done as loading control for the gels. Blot intensities were quantified using ImageJ.

B.2.4 K50P and K60P probe dosing on live cells and lysates for immunoblot analysis

Wildtype HeLa cells were seeded in 24-well plate at 0.1×10^6 cells per well in 1 mL of RPMI media (Corning) supplemented with 10% fetal bovine serum (Corning) and 1% penicillin/streptomycin (Gibco) 24 hour before experimentation. Cells were treated with 0–50 μ M of probe in 500 μ L of serum-free, phenol-red-free RPMI media (Gibco) for 4 hr at 37 °C. The cells were then washed in 500uL PBS twice and lysed in ice-cold 1x RIPA buffer supplemented with protease inhibitor (Roche) and 1 mM DTT. Cells were allowed to lyse while shaking on ice for 20 min. The lysates were collected and insoluble debris was cleared using centrifugation. The supernatant was diluted in 4 $\times$ Laemmli buffer containing 50 mM dithiothreitol (DTT) as a reducing agent. SDS-PAGE samples were prepared by heating the mixture at 95 °C for 5 minutes, then cooling to room temperature. The samples were resolved on a 10% SDS-PAGE gel and transferred to nitrocellulose membranes using standard western blotting techniques. Membranes were blocked with 2% BSA in TBS containing 0.1% Tween-20 (TBST) and probed with streptavidin-800 (Licor IRDye[®] 800CW, Licor Odyssey CLx imager) to detect biotin/desthiobiotin labeling. Blot intensities were quantified using ImageJ and normalized in Microsoft Excel.

For lysate dosing studies, clarified HeLa lysates were treated with 0–50 μ M probe for 2 hours at 37 °C before preparing samples for immunoblotting as described above.

B.2.5 Viability assay

5000 cells were seeded in a 96-well plate on day 1 in 50 μ L media. On day 2, 25 μ L of 3 $\times$ KW-2449 stock was added to the cells and the cells were incubated in the incubator (37 $^{\circ}$ C, 5% CO₂) for 48 hours. Next, the plate was removed from the incubator and brought to room temperature. 75 μ L of CTG reagent from the Promega CellTiter-Glo Luminescent Cell Viability Assay kit (#G7572) was added into each well. The plate was covered with aluminum foil and mixed on a shaker for 2 minutes. The plate was allowed to stand for 10 minutes at room temperature, after which luminescence was recorded.

B.2.6 ABL1 site of labeling assays

For mapping the site of labeling of the probes in ABL1, the protein was buffer-exchanged to PBS as mentioned above and diluted to a concentration of 0.2 mg/mL protein in PBS. 250 μ L of the protein was taken and treated with 10 μ M probe for 1 h at 37 $^{\circ}$ C. Excess probe was quenched with 10 mM Tris base and 1 $\times$ 8 M urea (250 μ L) was added. Then, DTT aqueous solution (5 μ L, 1 M) was added, followed by incubation at 65 $^{\circ}$ C for 15 min. Iodoacetamide aqueous solution (40 μ L, 0.5 M) was added to the sample, followed by further incubation for 30 min in the dark. The sample was then transferred to a 30kDa MWCO Millipore Amicon centrifugation filter unit and washed twice with 1 mL 2 M urea in 25 mM ammonium bicarbonate solution. The sample volume was brought up to 300 μ L with 2 M urea in 25 mM ammonium bicarbonate. CaCl₂ (3 μ L, 100 mM) was added to the sample followed by 1 μ g of sequencing-grade trypsin. The samples were digested overnight at 37 $^{\circ}$ C while shaking. Following trypsinization, supernatant was collected, acidified with HPLC grade formic acid (2% final, pH 2–3), and peptides were then desalted on ZipTip C18 tips (100 μ L, Millipore), dried under vacuum, resuspended with LC-MS grade water (Sigma Aldrich), and then lyophilized.

Lyophilized peptides were dissolved in LC-MS/MS Buffer (H₂O with 0.1% formic acid,

LC-MS grade, Sigma Aldrich) for proteomic analysis. Proteomic methods reported are adopted from our previous reports.²⁹³ LC-MS/MS analysis for proteomics samples was performed with an UltiMate 3000 RSLCnano System (Thermo Fisher Scientific) using an Acclaim PepMap RSLC C18 column ($75\ \mu\text{m} \times 15\ \text{cm}$, $2\ \mu\text{m}$, $100\ \text{\AA}$, Thermo Fisher Scientific) with an in-line Acclaim PepMap 100 C18 trap column ($75\ \mu\text{m} \times 2\ \text{cm}$, $3\ \mu\text{m}$, $100\ \text{\AA}$, Thermo Fisher Scientific) heated to $45\ ^\circ\text{C}$. The LC system was coupled to an Orbitrap Exploris 480 and Nanospray Flex Ion Source with stainless steel emitter tip (Thermo Fisher Scientific). Mobile phase A was composed of H_2O supplemented with 0.1% formic acid, and mobile phase B was composed of CH_3CN supplemented with 0.1% formic acid. The instrument was run at $0.3\ \mu\text{L min}^{-1}$ with 2-hour gradients. MS/MS spectra were collected for the entirety of the gradient using a data-dependent, 2-second cycle time setting with the following details: full MS scans were acquired at a resolution of 120,000, scan range of 380 m/z to 1,500 m/z, maximum IT of 25 ms, normalized AGC target of 300% and data collection in profile mode. MS2 scans were performed by high-energy collision dissociation (HCD) fragmentation with a resolution of 15,000, normalized AGC target of 50%, maximum IT of 50 ms, HCD collision energy of 30% and data collection in centroid mode. The isolation window for precursor ions was set to 1.6 m/z. Peptides with a charge state of 1, 7+ and unassigned were excluded, and dynamic exclusion was set to 40 seconds. The RF lens % was set to 40 with a spray voltage value of 2.0 kV and an ionization chamber temperature of $300\ ^\circ\text{C}$.

Data were processed using the SEQUEST HT search engine node within the Proteome Discoverer 3.0 software package. Data were searched using a concatenated target/decoy UniProt database of the human proteome with isoforms. Digest enzyme specificity was set to trypsin with up to two missed cleavages allowed, and peptide length was set to between 6 and 144 residues. Precursor mass range was set to 350–6500. Precursor mass tolerance was set to 10 ppm, and fragment mass tolerance was set to 0.02 Da. Up to 4 dynamic modifications were allowed per peptide, including K60P/K50P (+666.262) or X4K (+858.353) modification

on lysine or tyrosine, oxidized methionine (+15.9949), N-terminal acetylation (+42.0106), N-terminal Met-loss (−131.0405) and N-terminal Met-loss + acetylation (−89.0299). Cysteine carboxyamidomethylation (+57.0215) was set as a static modification. A minimum of two peptides, with a minimum length of 6, was required for protein identification, and false discovery rate (FDR) was determined using Percolator with FDR rate set at 1%.

B.2.7 SILAC cell culture methods and proteomic sample preparation

SILAC labeling was performed by growing cells for at least five passages in lysine- and arginine- free SILAC medium (RPMI, Invitrogen) supplemented with 10% dialyzed fetal calf serum (GeminiBio) and 1% Pen/Strep. “Light” and “heavy” media were supplemented with natural lysine and arginine (0.1 mg/mL) for “light”, and ^{13}C -, ^{15}N -labeled lysine and arginine (0.1 mg/mL) for “heavy”, respectively.

B.2.8 Sample preparation and streptavidin enrichment of biotin-labelled proteins in live-cell experiments

SILAC quantitative proteomics was performed with “heavy” and “light” labeled cells. For K562 cells, SILAC-labeled cells were treated at 1×10^6 cells/mL density in T75 flasks. HeLa and MCF7 cells were grown to confluency and treated in a 10 cm dish. Cells were incubated with DMSO alone (light cells) or chemical probe (25 μM , heavy cells) for 4 hours (for K60P-Bio) or 1 hour (for X4K) in serum-free SILAC RPMI for probe no-probe studies. For inhibitor competition studies, cells were pretreated with inhibitor (KW-2449, 2 μM , light cells) or with vehicle (DMSO, 2 μM , heavy cells) for 1 hour before probe treatment of both heavy and light with probe (25 μM) for 2 hours. After incubation, excess probe was removed and supplanted with media for 10 minutes.

Following aspiration of media, cells were collected and then lysed in RIPA lysis buffer (50 mM Tris, 150 mM NaCl, 1% Triton X-100, 0.5% deoxycholate, pH 7.4) supplemented with

EDTA-free complete protease inhibitor (Roche) and 1 mM DTT, at 4 °C. After sonication, insoluble debris was cleared by centrifugation (17,000 g, 15 min). BCA assay was performed to normalize Heavy and Light protein concentrations to ~1 mg/ml. Streptavidin agarose beads (50 μ L slurry, Pierce) were washed twice with RIPA buffer, and each cell lysate was separately incubated with the beads with rotation overnight at 4 °C. The beads were subsequently washed five times with 0.5 mL of RIPA lysis buffer containing 1 mM DTT, combined together, four times with 0.5 mL PBS, and two times with 2 M Urea in 25 mM ammonium bicarbonate. 500 μ L of 6 M Urea in 50 mM ammonium bicarbonate was then added to the beads, and samples were reduced on resin by TCEP (10 mM final) with orbital shaking for 20 minutes at 65 °C. Samples were then alkylated by adding iodoacetamide (20 mM final), covered from the light and with orbital shaking, for 40 minutes at 37 °C. The streptavidin agarose beads were collected, washed once with 2 M Urea in 25 mM ammonium bicarbonate, and the buffer exchanged to 2 M Urea in 25 mM ammonium bicarbonate supplemented with 1 mM CaCl_2 . Enriched proteins were digested on-bead by the incubation of 2 μ g sequencing grade trypsin overnight at 37 °C. Following trypsinization, supernatant was collected by passing through a Micro Bio-Spin column (Bio-Rad) into a fresh LoBind tube, acidified with HPLC grade formic acid (2% final, pH 2-3), and peptides were then desalted on ZipTip C18 tips (100 μ L, Millipore), dried under vacuum, resuspended with LC-MS grade water (Sigma Aldrich), and then lyophilized. Lyophilized peptides were dissolved in LC-MS/MS Buffer (H_2O with 0.1% formic acid, LC-MS grade, Sigma Aldrich) for proteomic analysis.

B.2.9 Sample preparation and streptavidin enrichment of biotin-labelled proteins in lysate experiments

SILAC quantitative proteomics was performed with “heavy” and “light” labeled K562 cells. 10^7 cells were lysed in 500 μ L of DPBS per sample per channel. For competition experiments,

light cells were pretreated with inhibitor PC-2-52 or XO44 (2.5 μ L, 10 mM) and heavy cell lysate were pretreated with vehicle (DMSO, 2.5 μ L) for 2 hours at 37 °C. This was followed by treatment of both light and heavy cell lysate with K60P-Bio (1 μ L, 5 mM) or X4K (2.5 μ L, 1 mM) for 1 hour at 37 °C.

Following treatment, cell lysates were processed for MS/MS analyses as reported before. Briefly, heavy and light samples were first combined in methanol (2 mL) and chloroform (0.5 mL), vortexed, and then DPBS (1 mL) was added. The resulting cloudy suspension was vortexed and centrifuged (5,000g, 15 min, 4 °C). The organic and aqueous phases were aspirated, leaving a protein disc at the interface. This disc was washed with ice-cold methanol (3 $\times$ 1 mL) by sonication. The organic phase was removed by aspiration. The pellet was dissolved in freshly prepared urea (500 μ L, 6 M in DPBS). TCEP (5 μ L, 1 M) and the sample was incubated at 37 °C with shaking for 30 min. Iodoacetamide (36 μ L, 300 mM) was then added and the mixture was incubated in the dark for 30 min at room temperature. Proteomes were diluted with SDS (140 μ L, 10% in DPBS) followed by DPBS (5.5 mL), then incubated with DPBS-washed Streptavidin agarose beads (50 μ L slurry, Pierce) and rotated at room temperature for 90 min. Beads were subsequently pelleted by centrifugation (1,400g, < 1 min) and washed sequentially with 0.25% SDS in DPBS (3 $\times$ 10 mL), DPBS (3 $\times$ 10 mL), and ddH₂O (3 $\times$ 10 mL). They were then transferred to a Protein LoBind tube (Eppendorf). Enriched proteins were digested on-bead by the incubation of 2 μ g sequencing grade trypsin overnight at 37 °C in 2 M urea in DPBS with 1 mM CaCl₂. Following trypsinization, supernatant was collected by passing through a Micro Bio-Spin column (Bio-Rad) into a fresh LoBind tube, acidified with HPLC grade formic acid (2% final, pH 2–3), and peptides were then desalted on ZipTip C18 tips (100 μ L, Millipore), dried under vacuum, resuspended with LC-MS grade water (Sigma Aldrich), and then lyophilized. Lyophilized peptides were dissolved in LC-MS/MS Buffer (H₂O with 0.1% formic acid, LC-MS grade, Sigma Aldrich) for proteomic analysis.

B.2.10 LC-MS/MS acquisition and analysis

Proteomic methods reported are adopted from our previous reports.²⁹³ LC-MS/MS analysis for proteomics samples was performed with an UltiMate 3000 RSLCnano System (Thermo Fisher Scientific) using an Acclaim PepMap RSLC C18 column ($75\ \mu\text{m} \times 15\ \text{cm}$, $2\ \mu\text{m}$, $100\ \text{\AA}$, Thermo Fisher Scientific) with an in-line Acclaim PepMap 100 C18 trap column ($75\ \mu\text{m} \times 2\ \text{cm}$, $3\ \mu\text{m}$, $100\ \text{\AA}$, Thermo Fisher Scientific) heated to $45\ ^\circ\text{C}$. The LC system was coupled to an Orbitrap Exploris 480 and Nanospray Flex Ion Source with stainless steel emitter tip (Thermo Fisher Scientific). Mobile phase A was composed of H_2O supplemented with 0.1% formic acid, and mobile phase B was composed of CH_3CN supplemented with 0.1% formic acid. The instrument was run at $0.3\ \mu\text{L min}^{-1}$ with 2-hour gradients. MS/MS spectra were collected for the entirety of the gradient using a data-dependent, 2-second cycle time setting with the following details: full MS scans were acquired at a resolution of 120,000, scan range of 380 m/z to 1,500 m/z , maximum IT of 25 ms, normalized AGC target of 300% and data collection in profile mode. MS2 scans were performed by high-energy collision dissociation (HCD) fragmentation with a resolution of 15,000, normalized AGC target of 50%, maximum IT of 50 ms, HCD collision energy of 30% and data collection in centroid mode. The isolation window for precursor ions was set to 1.6 m/z . Peptides with a charge state of 1, 7+ and unassigned were excluded, and dynamic exclusion was set to 40 seconds. The RF lens % was set to 40 with a spray voltage value of 2.0 kV and an ionization chamber temperature of $300\ ^\circ\text{C}$.

Data were processed using the SEQUEST HT search engine node within the Proteome Discoverer 3.0 software package. Data were searched using a concatenated target/decoy UniProt database of the human proteome with isoforms. Digest enzyme specificity was set to trypsin with up to two missed cleavages allowed, and peptide length was set to between 6 and 144 residues. Precursor mass range was set to 350–6500. Precursor mass tolerance was set to 10 ppm, and fragment mass tolerance was set to 0.02 Da. Up to 4 dynamic modifica-

tions were allowed per peptide, including heavy lysine (+8.0142), heavy arginine (+10.0083), oxidized methionine (+15.9949), N-terminal acetylation (+42.0106), N-terminal Met-loss (-131.0405) and N-terminal Met-loss + acetylation (-89.0299). Cysteine carboxyamidomethylation (+57.0215) was set as a static modification. A minimum of two peptides, with a minimum length of 6, was required for protein identification, and false discovery rate (FDR) was determined using Percolator with FDR rate set at 1%. Before quantification, chromatographic alignment was performed, with a maximum retention time difference of 10 min allowed, a mass tolerance of 10 ppm and a minimum signal/noise threshold of 5 required for feature mapping. SILAC ratios were determined using precursor-based quantification in a protein abundance manner based on peak intensity without normalization or scaling using a maximum ratio of 20. Proteins considered as enriched exhibited a probe dependent median SILAC ratio greater than 2 across two biological replicates combined (for Figure 4.2) and across 4 biological replicates combined (for Figure 4.3). Kinome Dendrograms were produced using <http://www.kinhub.org/kinmap> and Gene Ontology analysis was done using Metascape.

B.3 Supplementary molecular modeling methodology

Visualization of systems was performed using VMD 1.9.4a57¹⁵⁰ or Pymol 2.5 [The PyMOL Molecular Graphics System, Version 2.5 Schrödinger, Inc.]. Analysis was performed with MDTraj 1.9.8²⁹⁴ and graphs were generated with matplotlib 3.6.2.²⁹⁵

We initially considered 3 ligands of interest - KW-2449, sunitinib, and XO44 - and the receptors for docking were selected from structures containing these or similar ligands (Figure B.1). Crystal structures for sunitinib and XO44 were already available, and crystal structures of axitinib were identified as analogous to KW-2449. The selected structures for docking were 5K9I (Src+XO44), 4WA9 (Abl+axitinib), and 3G0E (KIT+sunitinib). The structures 5K9I and 4WA9 exhibited open binding pockets suitable for docking, the 3G0E binding pocket was

much tighter and the A-loop after the DFG phenyl was removed from the 3G0E structure prior to docking.

We determined the analogue space to screen based on examination of the distances between the scaffold and conserved lysine, as well as by examining the sterically similar group on Axitinib. Modifications on the core ring with linker lengths of 0, 1, 2 between the scaffold and phenylsulfonyl fluoride warhead were considered.

Docking experiments were performed with only the attached phenyl ring, and not the fluorosulphonyl group of the warhead. This allowed for the rapid elimination of modification positions that caused steric clashes, whilst minimizing the potential for false negatives due to the inflexibility of the receptor backbone during docking or any clash between the warhead leaving group and lysine. Docking was performed with Smina 2020.12.10.²⁹⁶ Proteins were prepared for docking using AutoDockTools within the MGLTools software suite 1.5.7.²⁹⁷ Docking was confined to a box of 30x30x30 Å centered on the binding site. The protein backbone was fixed during docking, but all sidechains facing into the binding pocket were considered flexible. We initially generated 100 poses for each ligand, with the restraint that all poses were > 0.2 Å apart and increased the exhaustiveness of the search to ensure conformations were sampled thoroughly. Poses in which the core-NH was > 2 Å from the crystallographic pose, or that would direct the warhead out of the pocket, were then discarded. Generating multiple poses in multiple receptors was used as the primary indication of probe viability, probes with many generated poses were reasoned to be more likely to form an unstrained covalent bond once the fluorosulfonyl was attached to the ortho-, meta-, or para- positions. Modification positions that did not produce any viable poses, or only produced poses in 1 receptor were eliminated.

The initial template for MD simulations was constructed with CHARMM-GUI^{134;135;147} with the protein from the 4WA9 (chain B, Abl + axitinib) crystal structure. Systems were constructed by replacing the protein in the template system after alignment to its kinase

hinge (resid 316 to 318) with proteins from PDB IDs: 4WA9 (chain B), 4TWP (chain B), 2G1T (chain A), 5K9I (chain A). Ligands were aligned to the pharmacophore of axitinib in the template structure. For systems with ligands modified from 5-position of the KW-2449 scaffold (K5*), the initial ligand conformation prior to alignment was adjusted to an alternate rotation of the ring-connecting alkene relative to the axitinib structure (Figure B.2). This conformation better directs the warhead towards Lys271 and is necessary to avoid clashes with the curled Abl P-loop. Both alkene rotations engage in the same hydrogen bonds and were observed to freely exchange during simulations, a QM dihedral scan (not shown) predicted both rotations to have within 0.3 kcal/mol energy in solution. For systems of XO44 in proteins from PDB IDs: 4WA9 or 4TWP, the curled P-loop conformations were replaced by alignment with extended conformations from PDB ID: 6BL8²²⁸ to avoid clashes with the ligand. The T315I gatekeeper mutation in 4TWP was reverted to wild-type. Missing residues in the A-loop of 5K9I (chain A) (and residues up to 2 adjacent) were modeled with MODELLER.²⁴⁷ Minor extensions were modeled for protein termini in Abl so that all proteins were the same length (resid 233-508). Clashing water molecules (< 1 Å from ligand or protein) were shifted to the bulk solvent. All histidines were protonated HSD except His361/375 in Abl and His384/492 in Src which were protonated as HSE.

The linked MD simulations for screening were performed with NAMD 3.0a13²⁹⁸ for ortho/meta/para warhead positions for all molecules that passed the initial screening and were determined to be synthetically viable: modified at the 5 or 6 position and with a linker length of 0 or 2. Ortho-, meta-, and para- variants were generated for each chosen modification-position and linker combination. The systems were equilibrated, as per CHARMM-GUI NAMD output, with protein heavy atoms restrained using a force constant of 1.0 kcal/mol for the backbone and 0.5 kcal/mol for the sidechain. The ligand core was restrained only by position restraints on the indazole nitrogens, and Lys271 was not restrained. For the linked XO44 simulation the ligand core restraints also included the full ring system except

the warhead ring, and the modeled P-loop residues were also unrestrained. The initial NVT equilibration of 250 ps was followed by an additional 1 ns NPT equilibration with the same restraints. Production runs were performed for 100 ns. RMSD calculations for the core displacement, and lysine backbone displacement metrics were calculated on the final 10 ns of simulation frames aligned by all C_α to the 4WA9 crystal structure chain B.

The unlinked simulations to investigate the labelling rate were performed using the CUDA version of Amber 20.^{143;144;254} The systems were converted to amber format using CHAMBER²⁹⁹ within ParmEd.²⁵² The systems were equilibrated, as per CHARMM-GUI Amber output but the ligands were restrained (XO44 heavy atoms, K60P indazole), and modeled protein residues were unrestrained. Production runs were performed for 4 μ s.

Force field parameters for each ligand were obtained for the CGenFF 4.4 force field^{140;300} using Paramchem.^{141;142} Parameters were calculated for the unlinked ligand, as well as for a linked form where the fluorine was replaced with a nitrogen and attached lysine sidechain (CB as a methyl group). The bonded terms around the covalent bond were a good match to the model compound (MBSM) used in the development of the CGenFF force field,³⁰⁰ and we reproduced the potential energy scans in MD and QM (not shown). The dihedral parameter between the aromatic and piperazine rings of XO44 selected by Paramchem was a poor analogy and the dihedral ‘NG2R62 CG2R64 NG301 CG331’ was used instead. The resulting parameters were applied to the unlinked ligand and lysine system as a patch using NAMD’s psfgen. The parameters for the charge of the warhead fluorine and S-F bond length were incorrect and were re-fit. The charges of the warhead and adjacent phenyl ring were fit using FFParam³⁰¹ according to the CGenFF parameterization guidelines, using Gaussian²⁵¹ and OpenMM³⁰² for the required QM and MD calculations respectively. The S-F bond-length was set to 1.62 Å as per the QM optimization.

B.4 Supplementary molecular modeling discussion

When conducting MD simulations of the unlinked ligand to investigate the labelling rate, the simulations were performed using protein conformations taken from X-ray structures of the kinase co-crystallized with axitinib (Abl) and XO44 (Src): Abl DFG-out (Abl/DFG-out, 4WA9), Abl DFG-in (Abl/DFG-in, 4TWP), Src DFG-in (Src/DFG-in, 5K9I), and Abl DFG-in/ α C-helix-out (Abl/DFG-in/ α C-helix-out, 2G1T). This resulted in a total of 2 duplicates of 4 systems for each of the 2 ligands ($16 \times 4 \mu\text{s}$ simulations).

The DFG-out simulations for both ligands had the lowest sampling of reaction-ready proximities. In this conformation, π -stacking interactions between the warhead ring and the phenylalanine sidechain of the DFG motif effectively displaced the warhead, disrupting the core hydrogen bonds and fencing it away from the target lysine (Figure 4.2C, 2D). This suggests that both ligands have a preference for reversible binding to the DFG-in conformation of Abl kinase.

In the Abl DFG-in simulations XO44 outperformed K60P in spite of unstable interactions with the Abl P-loop. The XO44 warhead was observed to flip into and out of the binding pocket, resulting in closer proximities than in Src DFG-in where the warhead position remained stable. In 1 replicate of each DFG-in Abl-XO44, the P-loop curled up preventing the warheads re-entry. K60P was stable in the DFG-in simulations, but we observed that the warhead ring tended to cause the lysine to be displaced away from the warhead towards the P-loop. The Lys to DFG-in-Asp salt-bridge strained to form over the top of the ligand, inducing the scaffold to angle slightly deeper into the cleft than our predicted poses and more frequently adopt the 180° rotated alkene position (Figure B.2).

All K60P simulations resulted in the α C-helix adopting the out conformation after a few 100 ns, likely due to charge-repulsion from the warhead, thus preventing the formation of a salt-bridge between the lysine and α C-helix-Glu. For the XO44 simulations the α C-helix remained in its initial conformation. The XO44 crystal structures are in the α C-helix-out

conformation, but it is unclear if this only occurs post-reaction.

B.5 Supplementary kinetic modeling methodology

To characterize the kinetics of ABL1-probe binding, the experimental data were fit with numerical simulations of irreversible kinetic schemes. The basic kinetic scheme underlying all systems is defined with two-step inactivation and one-step nonspecific binding (Scheme 1). The two-step inactivation comprises a reversible non-covalent association/dissociation step followed by an irreversible covalent reaction step. The protein species that exists solely for non-specific binding, P_{ns} , is unconsumed to approximate the slow saturation of nonspecific binding sites.

Ordinary differential equations (ODEs) were constructed for each transition in the scheme and integrated together numerically to yield the time evolution of each species concentration. These simulations used the same basic conditions as the experiments. The two independent free protein species, P and P_{ns} , had an initial concentration of $0.357 \mu\text{M}$. Time-responses used an initial probe concentration of $10 \mu\text{M}$. Dose-responses used a timepoint of one hour.

The time- and dose-responses on their own do not offer much value for fitting unnormalized labeling data, but they complement each other well when used simultaneously. The fitting procedure, which is done for each probe individually, leverages this idea by considering the residuals between all corresponding experimental and simulated data at once. The simulations are parameterized with the rate constants for each kinetic transition and a normalizing factor (λ) for each experimental data set. Therefore, a reasonable set of rate constants must yield a reasonable fit to every replicate of both time- and dose-responses. The simulated response is 1) the sum of the concentrations of the specifically- and nonspecifically-bound covalent complexes, $[P\text{-}L]$ and $[P_{\text{ns}}\text{-}L]$, 2) normalized by the initial protein concentration, and 3) multiplied by the λ parameter for the corresponding experiment. Note that when plotting and analyzing the outcome of the fitting procedure, the simulated response is left normal-

ized to initial protein concentration, and the experimental datapoints are divided by λ as an effective normalization. The residuals were minimized using LMFIT²⁶⁴ with the Levenberg-Marquardt least-squares (LS) algorithm²⁶⁵ and Trust Region Reflective method.³⁰³

Initial guesses and bounds of k_{on} and k_{off} were informed by a comprehensive kinetic survey between kinases and competitive ligands.²⁶⁶ To efficiently focus the sampling on the binding affinity, ($K_d=k_{\text{off}}/k_{\text{on}}$), with minimal degeneracy, either k_{on} or k_{off} must be fixed. Fixing k_{on} is more ideal, since it is largely diffusion limited and differences in the ligands are more likely to be represented by differences in the k_{off} . This was a viable strategy for K50P and K60P, but X4K required a fixed k_{off} . X4K was possibly rate-limited by k_{on} , so the rest of the kinetic parameters and affinity would be largely dependent on the fixed value of k_{on} . The k_{on} was fixed for the KW probes at 100 times less than the value for KW-449/ABL1, $1.82 \times 10^{-2} \mu\text{M}^{-1}\text{s}^{-1}$. For X4K, k_{off} was fixed according to analogue VX-680, $8.95 \times 10^{-2} \text{s}^{-1}$. We discarded the 2- and 60-minute timepoints in the X4K time-response experiment, because no reasonable fits to the dose-response could be produced otherwise.

The best parameters from the LS minimization were used to initialize Metropolis Monte-Carlo (BMMC) sampling, which used the same fitting procedure to calculate the chi-square value from the residuals as an “energy” for the purposes of acceptance. The normalization factors were held constant in addition to the parameters that were already fixed in the LS fit. The step size for each variable parameter was chosen uniformly between +/- 5% of the respective best-fit value. To reduce the effect of autocorrelation, only every 500th step in the BMMC chain was kept. The presented range of each variable parameter is the highest density interval for 95% of its BMMC distribution.³⁰⁴

Although the absolute scale of labeling data with respect to total protein concentration is unknown, the relative scaling between the KW probes was approximated as an inter-ligand scaling factor from the one time-response (Bio Rep 1) and two dose-responses (Bio Reps 3 and 4) that captured K50P and K60P together on the same gel. This inter-ligand scaling

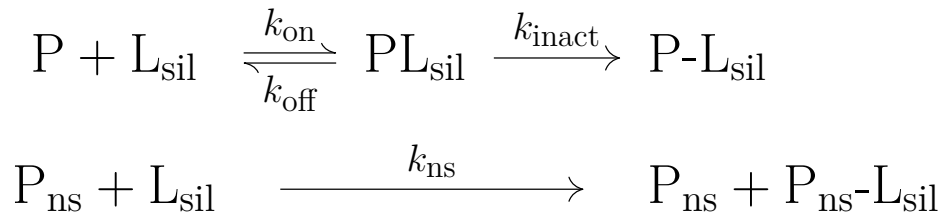
factor for each replicate was calculated as the K60P:K50P ratio of the final response. Probe K60P was fit independently, yielding a standard error for each normalization factor that was calculated as the square root of the product of the diagonal elements of the covariance matrix and the reduced chi-square. The initial guess of the respective K50P normalization factor was the K60P best fit multiplied by the inter-ligand scalar. The bounds were the initial guess +/- two scaled standard errors. The propagation of uncertainty and the flexibility of inter-ligand scaling, though crude, enabled the K50P fitting to utilize the relative scaling observed in all three Bio Reps.

Following the observations of relative scaling between the KW probes, the kinetic scheme of K50P was modified to account for sub-stoichiometry. Two modified versions of the scheme were created to allow the fraction of specific binding to plateau below 100% of the initial protein concentration, Mod_{del} (Scheme B.1) and Mod_{sil} (Scheme B.2). Mod_{del} introduces the parameter k_{del} , *i.e.* the rate constant of K50P “deletion” (del) and models a ligand-limited plateau. Mod_{sil} introduces the parameter Frac_{sil}, *i.e.* the fraction of “silent” (sil) ligand and models a protein-limited plateau. The term “silent” stems from the scenario where a portion of the total ligand population is consuming protein without contributing to the observable. Note that the measured effect of Mod_{sil} would be equivalent had the silent population been protein instead, effectively modeling the scenario where the protein is the affected species, *e.g.*, if the ligand induces and traps a silencing and rapidly-equilibrating conformational change in the protein. The sub-stoichiometric modeling is elaborated upon in the supplementary kinetic modeling discussion.



Scheme B.1: Additional transition to express sub-stoichiometric covalent occupancy in terms of ligand “deletion”.

Note that the original ligand concentration is distributed over two populations, L and L_{sil}, where the proportion of silent ligand is defined as the parameter, Frac_{sil}.



Scheme B.2: Additional transitions to express sub-stoichiometric covalent occupancy in terms of a “silent” ligand population.

A Jupyter notebook for kinetic modeling is accessible at <https://github.com/RouxLab/covalent-kinase-activity-probes-2025>.

B.6 Supplementary kinetic modeling discussion

Early in the study, the experimental time-dependent data displayed a striking difference in unnormalized intensity between K50P and K60P. By the six-hour endpoint, K60P exceeded twice the intensity of K50P, which had reached an apparent plateau. Since the specific binding of both irreversible probes should plateau at 100% of the protein concentration, we reasoned that the difference in intensity could be caused by greater nonspecific binding by K60P or an unknown mechanism causing K50P to plateau sub-stoichiometrically. Reasonable estimates of the probes’ nonspecific-binding rate constants could be fitted to the dose-responses. Yet, the nonspecific binding only accounted for about half of the difference in probe intensity. The remaining half was achieved by inducing either ligand-limited or protein-limited sub-stoichiometry with one of two generic modifications, Mod_{del} or Mod_{sil} , to the base kinetic model of K50P. Mod_{del} introduces a one-step transition that deletes (del) the ligand and lowers its effective concentration over time (Scheme B.1). Mod_{sil} introduces a silent (sil) fraction of the ligand population that can covalently bind protein without producing signal (Scheme B.2), akin to the transducer ratio for the operational model of agonism.³⁰⁵ Both modifications caused the same apparent behavior (Figure B.17).

The models of sub-stoichiometry and their associated parameters are not representative

of any specific mechanism. They instead represent generic mechanisms that act on the simulated response to align with empirical observations and cover a broad range of possibilities. For example, Mod_{sil} covers covalent unbinding within the quenching period of the assay, and Mod_{del} covers hydrolysis of the sulfonylfluoride warhead. An important purpose of these modifications is to demonstrate how estimations of canonical kinetic features, such as the affinity and the rate constant of inactivation, are dependent on the model of stoichiometry. The fitted values of K_{d} and k_{inact} were mostly unaffected when the kinetic analysis included Mod_{sil} but were approximately halved when it included Mod_{del} . These features are especially susceptible to complex versions of Mod_{del} . As an illustrative yet unlikely example, modeling depletion of productive ligand via protein-mediated hydrolysis at the binding site lowers the estimated K_{d} by about an order of magnitude, since the complete depletion within the experimental time frame depends on the quality of reversible binding.

In any case, the inactivation rate constants, k_{inact} , for the KW ligands were consistent with previously reported values for XO44 analogues (k_{obs} ranging from 3×10^{-4} to $8 \times 10^{-4} \text{ s}^{-1}$)²²⁹ and to a similarly positioned probe (ABP4) binding to ABL1 ($k_{\text{inact}} = 0.778 \times 10^{-4} \text{ s}^{-1}$).³⁰⁶ The median k_{inact} values for K50P_{sil} and K60P were nearly identical despite their empirical differences. However, the median nonspecific-binding rate constant, k_{ns} , for K60P was about three times larger than that of both K50P variants, consistent with the qualitative trends observed in nonspecific reactivity assays (Figure B.21).

Unlike the KW probes, the k_{inact} of X4K could be fast enough to compete with k_{on} , (and the fixed k_{off}) and prevent an effective equilibrium of association and dissociation. Consequently, the apparent K_{d} is less meaningful, particularly in the commonly-sampled space where k_{inact} greatly exceeds k_{off} . It is also possible that the k_{on} is entirely rate-limiting, in which case the upper bound of k_{inact} is indeterminable from the labeling data. Yet, this is not the only factor of ambiguity surrounding the X4K analysis. In the dose-response, X4K appeared to covalently bind 100% of ABL1 within the incubation time at doses too low

to rapidly saturate the reversibly associated state and yield $k_{\text{on}} \times [\text{L}] \gg k_{\text{inact}}$. Therefore, unlike the KW probes, the unique role of k_{inact} in modulating the saturation-driven plateau of covalent occupancy could not be leveraged for fitting k_{inact} .³⁷ With only the EC_{50} to modulate, the k_{inact} and k_{on} were highly correlative in the fitting. In turn, the sampled values of k_{inact} were often also dependent on the assumed value of k_{off} ($8.95 \times 10^{-2} \text{ s}^{-1}$). Still, as pointed out in the main text, it is clear qualitatively that the X4K labeling is faster than that of K50P and K60P due to a faster k_{inact} . This increased reactivity is also reflected in the estimated nonspecific rate constant k_{ns} , which has a median approximately three times greater for X4K than K60P.

APPENDIX C

SUPPLEMENTARY INFORMATION FOR CHAPTER 5

Table C.1: Kinase structures with covalent inhibitors. 4A9, 4B9, and 4C9 are reversible.^{36;119;132;133;210;270;307–315}

PDB ID	Kinase	Ligand	Site
	BTK	4A9	Cys 481
	BTK	4B9	Cys 481
4YHF	BTK	4C9	Cys 481
5P9J	BTK	Ibrutinib	Cys 481
5YU9	EGFR (T790M)	Ibrutinib	Cys 797
	EGFR (T790M)	4C9	Cys 797
6YG2	MKK7	Ibrutinib	Cys 218
4G5J	EGFR	Afatinib	Cys 797
6E6E	c-SRC	DGY-06-116	Cys 280
5TOZ	JAK3	PF-06651600	Cys 909
4GS6	TAK1	5Z-7-Oxozeaenol	Cys 174
5K9I	c-SRC	probe O44	Lys 295
5U8L	EGFR	probe O44	Lys 745
7FIC	AURKA	5YZ	Lys 162
	c-SRC	5YZ	Lys 295

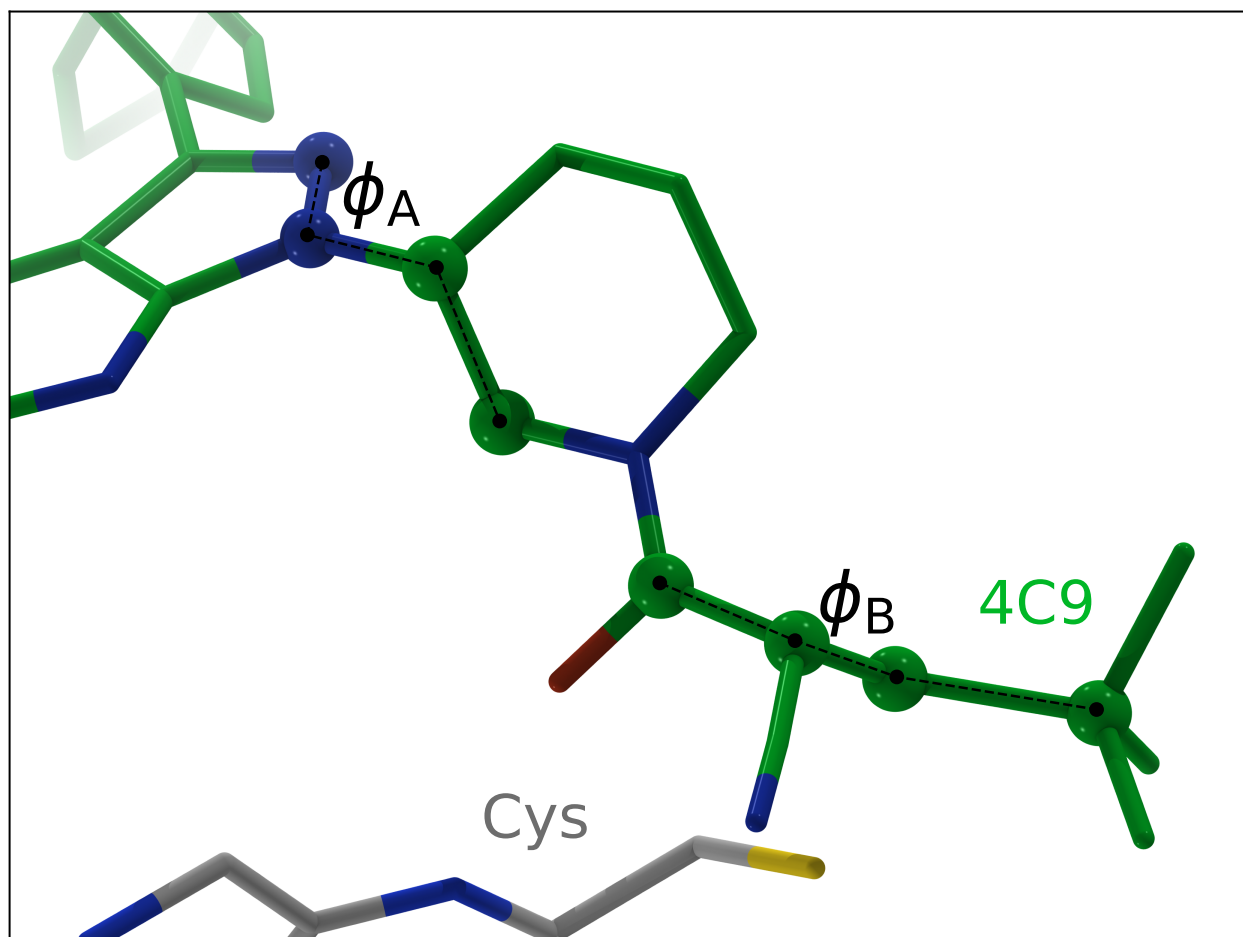


Figure C.1: The two dihedrals, ϕ_A and ϕ_B , used to resolve the conformational states in the unlinked ensemble, depicted on 4C9.

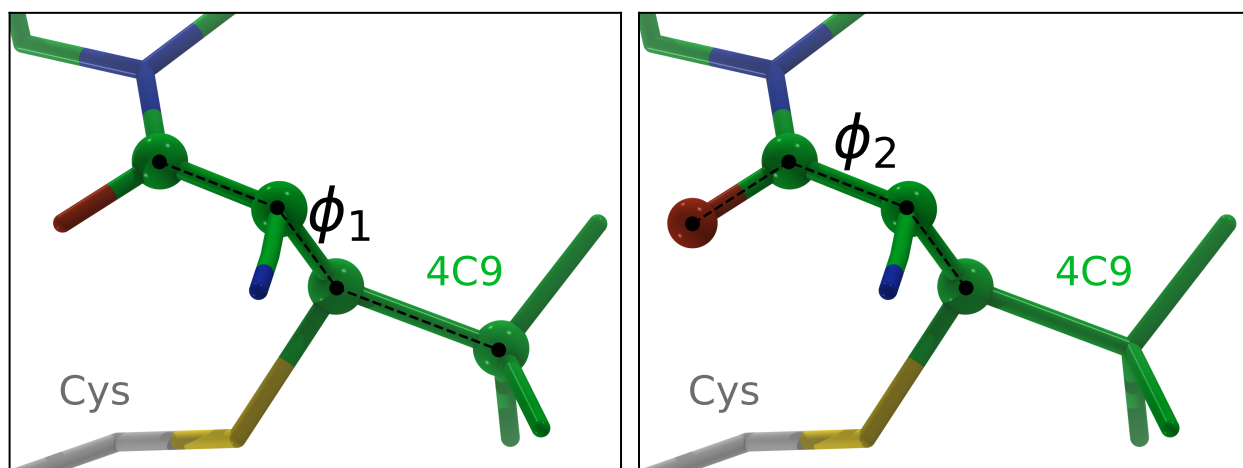


Figure C.2: The two dihedrals, ϕ_1 (left) and ϕ_2 (right), restrained during the steered proton-abstraction simulations, depicted on 4C9.

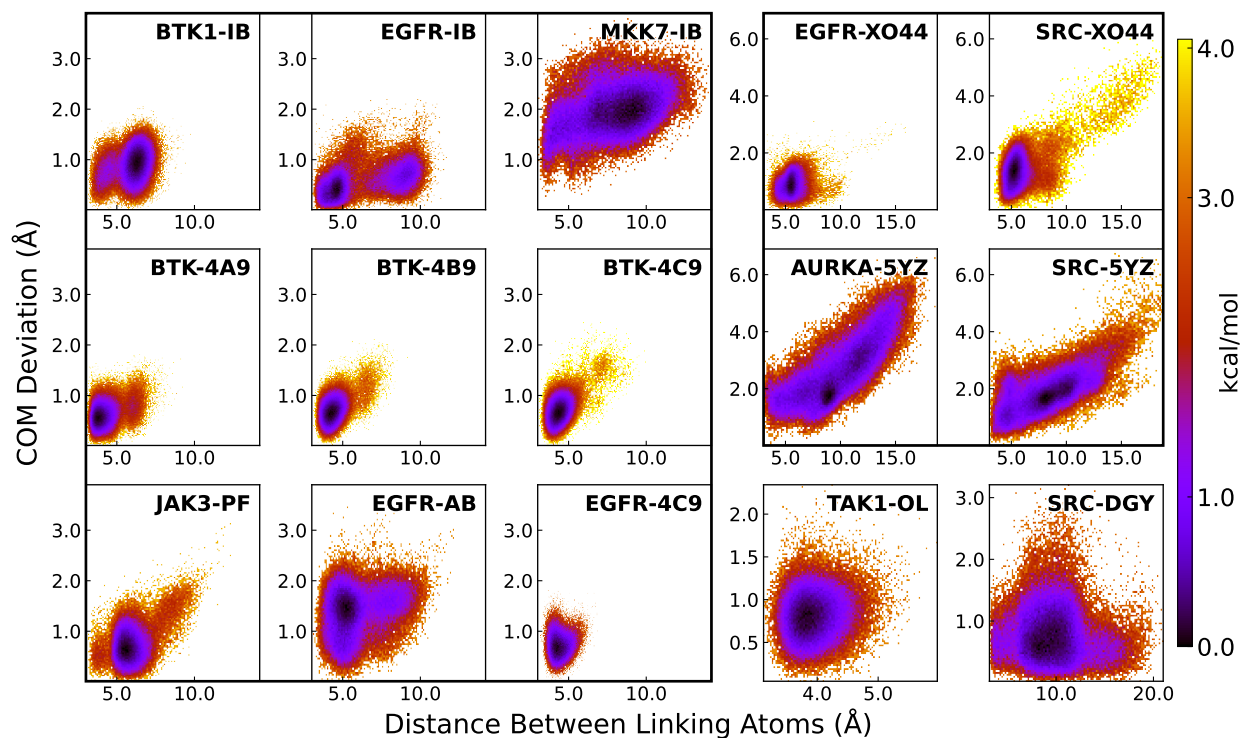


Figure C.3: Two-dimensional PMFs over the distance between the linking atoms and RMSD of the ligand center of mass with respect to the crystal pose. Similar systems are grouped.

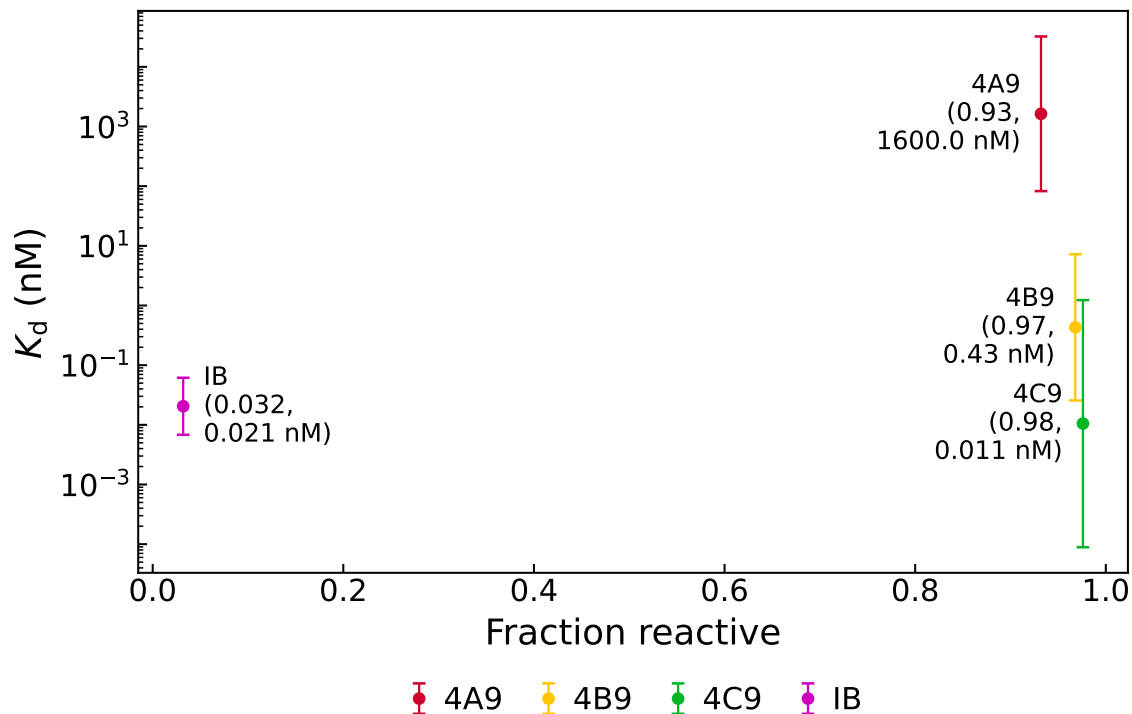


Figure C.4: Fraction of the unlinked ensemble spent in a reactive state vs. the noncovalent dissociation constant (K_d) for irreversible ibrutinib (IB) and reversible 4A9, 4B9, and 4C9.

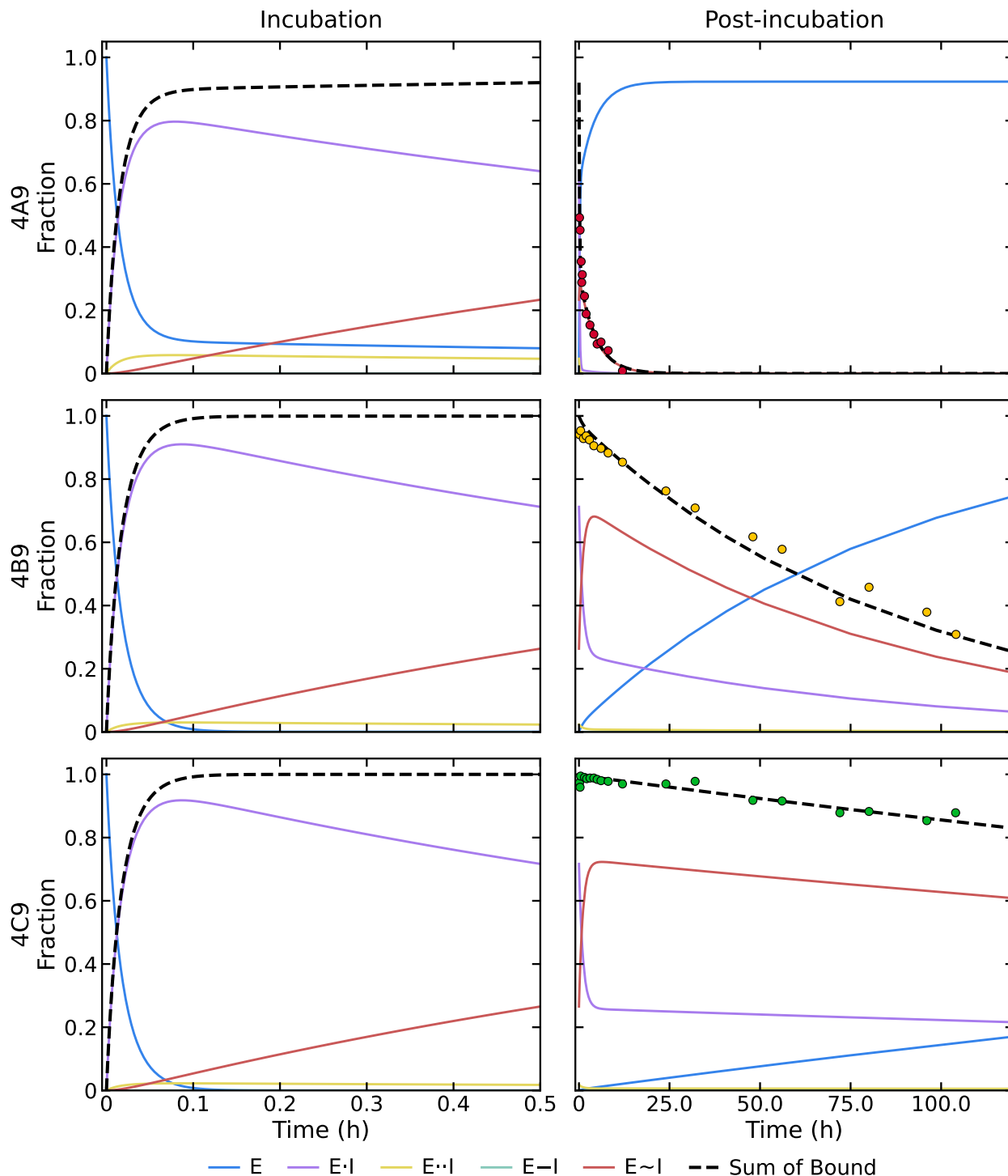


Figure C.5: Demonstrative time courses of all enzyme-containing species during the incubation (left) and post-incubation (right) phases of the experimental protocol from ref 36 for 4A9 (top), 4B9 (middle), and 4C9 (bottom). The simulations use the parameter set with the lowest energy from the post-burned/thinned BMMC samples ($k_{\text{on}} = 1.19 \times 10^4 \text{ M}^{-1}\text{s}^{-1}$, $k_{\text{inact}}^{\text{f}} = 1.95 \times 10^{-4} \text{ s}^{-1}$, $k_{\text{inact}}^{\text{b}'} = 6.96 \times 10^{-5} \text{ s}^{-1}$, $K_{\text{d}}^{4\text{A}9} = 134 \text{ nM}$, $K_{\text{d}}^{4\text{B}9} = 1.04 \text{ nM}$, $K_{\text{d}}^{4\text{C}9} = 0.137 \text{ nM}$). Colored solid lines are individual species. The dashed black line is the sum of all bound species. Points on the right column are the experimental occupancy data.

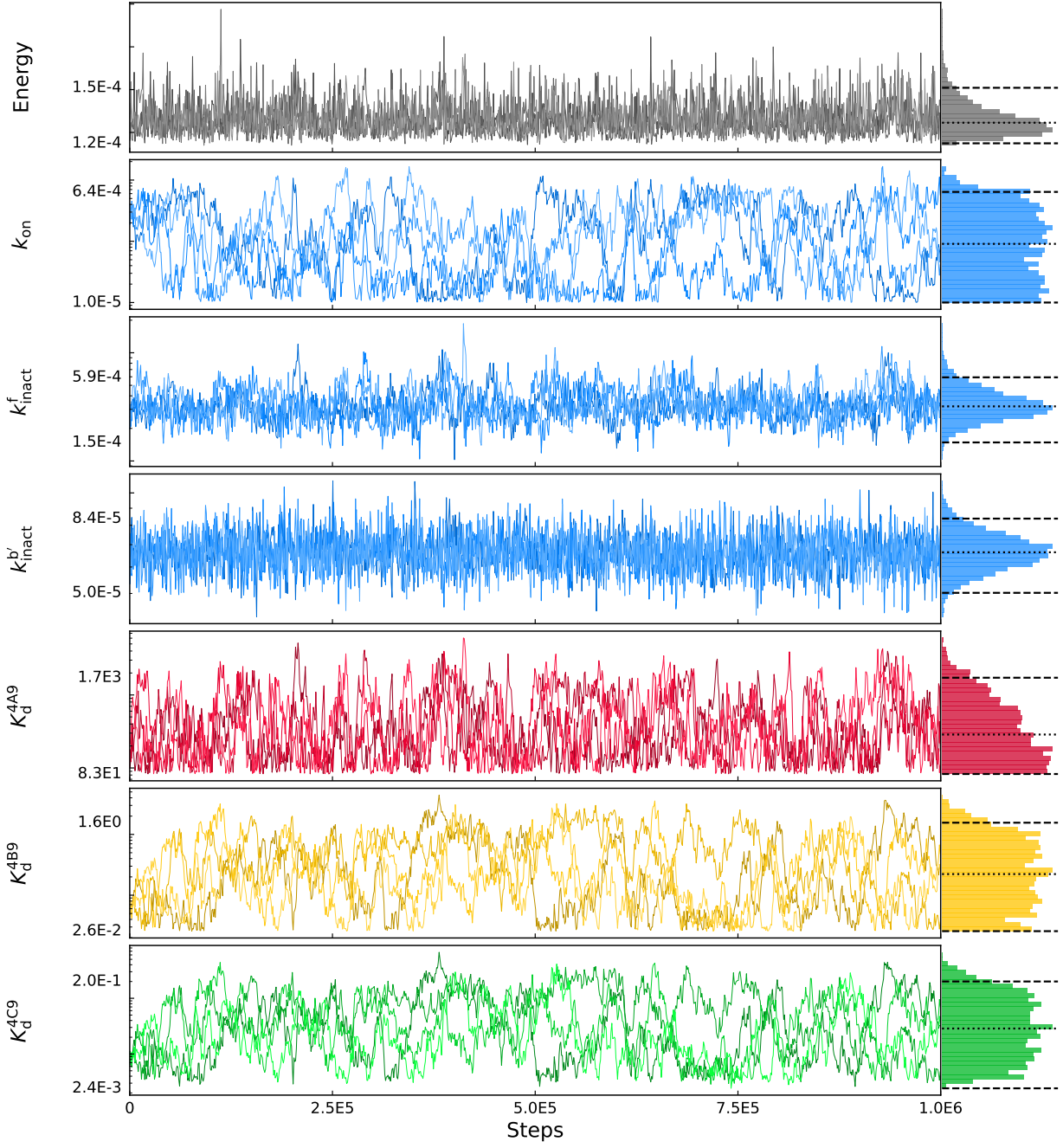


Figure C.6: Time evolution of the energy and free parameters during the BMMC sampling for each chain. The dissociation and rate constants are plotted on a $\log_{10}$ scale. In the right margin are histograms of the combined sample set. The median is represented with a dotted line, and the 95% HDI is represented with the dashed line and y -axis tick labels.

Table C.2: The medians and 95% highest density bounds ($^{\text{upper}}$ _{lower}) of parameters used in BMCMC simulations.

	4A9		4B9		4C9	
K_d (nM)	2.87×10^2	$\frac{1.72 \times 10^3}{8.26 \times 10^1}$	2.23×10^{-1}	$\frac{1.56 \times 10^0}{2.56 \times 10^{-2}}$	2.87×10^{-2}	$\frac{2.00 \times 10^{-1}}{2.42 \times 10^{-3}}$
p_r^f	0.932	*	0.968	*	0.976	*
p_r^b	1.00	*	1.00	*	1.00	*
$\Delta W_{\text{deprot}}^\dagger$ (kcal/mol)	10.4	*	12.8	*	12.8	*
k_{on} ($\text{M}^{-1}\text{s}^{-1}$)			9.08×10^4	$\frac{6.41 \times 10^5}{1.00 \times 10^4}$		
k_{inact}^f (s^{-1})			3.20×10^{-4}	$\frac{5.93 \times 10^{-4}}{1.48 \times 10^{-4}}$		
$k_{\text{inact}}^{b'}$ (s^{-1})			6.64×10^{-5}	$\frac{8.38 \times 10^{-5}}{5.02 \times 10^{-5}}$		

Shared parameters are displayed spanning all three ligand columns. Parameters marked with an asterisk (*) were fixed during both least-squares minimization and BMCMC sampling.

REFERENCES

- [1] Ehrlich, P. Chemotherapeutics: Scientific Principles, Methods, and Results. *The Lancet* **182**, 445–451 (1913).
- [2] Deng, Y., Beahm, D. R., Ionov, S. & Sarpeshkar, R. Measuring and Modeling Energy and Power Consumption in Living Microbial Cells With a Synthetic ATP Reporter. *BMC Biology* **19**, 101 (2021).
- [3] Milo, R., Jorgensen, P., Moran, U., Weber, G. & Springer, M. BioNumbers—the Database of Key Numbers in Molecular and Cell Biology. *Nucleic Acids Research* **38**, D750–D753 (2010).
- [4] Bianconi, E., Piovesan, A., Facchin, F., Beraudi, A., Casadei, R., Frabetti, F., Vitale, L., Pelleri, M. C., Tassani, S., Piva, F., Perez-Amodio, S., Strippoli, P. & Canaider, S. An Estimation of the Number of Cells in the Human Body. *Annals of Human Biology* **40**, 463–471 (2013).
- [5] Milo, R. What Is the Total Number of Protein Molecules Per Cell Volume? A Call to Rethink Some Published Values. *BioEssays* **35**, 1050–1055 (2013).
- [6] Smyth, D. G., Stein, W. H. & Moore, S. The Sequence of Amino Acid Residues in Bovine Pancreatic Ribonuclease: Revisions and Confirmations. *Journal of Biological Chemistry* **238**, 227–234 (1963).
- [7] Maier, T., Leibundgut, M. & Ban, N. The Crystal Structure of a Mammalian Fatty Acid Synthase. *Science* **321**, 1315–1322 (2008).
- [8] Beck, M., Schmidt, A., Malmstroem, J., Claassen, M., Ori, A., Szymborska, A., Herzog, F., Rinner, O., Ellenberg, J. & Aebersold, R. The Quantitative Proteome of a Human Cell Line. *Molecular Systems Biology* **7**, MSB201182 (2011).
- [9] Robertson, J. G. Enzymes as a Special Class of Therapeutic Target: Clinical Drugs and Modes of Action. *Current Opinion in Structural Biology* **17**, 674–679 (2007).
- [10] Singh, J., Petter, R. C., Baillie, T. A. & Whitty, A. The Resurgence of Covalent Drugs. *Nat Rev Drug Discov* **10**, 307–317 (2011).
- [11] Roth, G. J., Stanford, N. & Majerus, P. W. Acetylation of Prostaglandin Synthase by Aspirin. *Proceedings of the National Academy of Sciences* **72**, 3073–3076 (1975).
- [12] Fleming, A. On the Antibacterial Action of Cultures of a *Penicillium*, With Special Reference to Their Use in the Isolation of *B. Influenzæ*. *British Journal of Experimental Pathology* **10**, 226–236 (1929).
- [13] Kitz, R. & Wilson, I. B. Esters of Methanesulfonic Acid as Irreversible Inhibitors of Acetylcholinesterase. *Journal of Biological Chemistry* **237**, 3245–3249 (1962).

- [14] Tipper, D. J. & Strominger, J. L. Mechanism of Action of Penicillins: A Proposal Based on Their Structural Similarity to Acyl-D-alanyl-D-alanine. *Proceedings of the National Academy of Sciences* **54**, 1133–1141 (1965).
- [15] Walsh, C. Suicide Substrates: Mechanism-Based Enzyme Inactivators. *Tetrahedron* **38**, 871–909 (1982).
- [16] Fry, D. W., Bridges, A. J., Denny, W. A., Doherty, A., Greis, K. D., Hicks, J. L., Hook, K. E., Keller, P. R., Leopold, W. R., Loo, J. A., McNamara, D. J., Nelson, J. M., Sherwood, V., Smaill, J. B., Trumpp-Kallmeyer, S. & Dobrusin, E. M. Specific, Irreversible Inactivation of the Epidermal Growth Factor Receptor and erbB2, by a New Class of Tyrosine Kinase Inhibitor. *Proceedings of the National Academy of Sciences* **95**, 12022–12027 (1998).
- [17] Honigberg, L. A., Smith, A. M., Sirisawad, M., Verner, E., Loury, D., Chang, B., Li, S., Pan, Z., Thamm, D. H., Miller, R. A. & Buggy, J. J. The Bruton Tyrosine Kinase Inhibitor PCI-32765 Blocks B-cell Activation and Is Efficacious in Models of Autoimmune Disease and B-cell Malignancy. *Proceedings of the National Academy of Sciences* **107**, 13075–13080 (2010).
- [18] Byrd, J. C., Furman, R. R., Coutre, S. E., Flinn, I. W., Burger, J. A., Blum, K. A., Grant, B., Sharman, J. P., Coleman, M., Wierda, W. G., Jones, J. A., Zhao, W., Heerema, N. A., Johnson, A. J., Sukbuntherng, J., Chang, B. Y., Clow, F., Hedrick, E., Buggy, J. J., James, D. F. & O'Brien, S. Targeting BTK With Ibrutinib in Relapsed Chronic Lymphocytic Leukemia. *New England Journal of Medicine* **369**, 32–42 (2013).
- [19] Ostrem, J. M. L., Peters, U., Sos, M. L., Wells, J. A. & Shokat, K. M. K-Ras(G12C) Inhibitors Allosterically Control GTP Affinity and Effector Interactions. *Nature* **503**, 548–551 (2013).
- [20] Loll, P. J., Picot, D. & Garavito, R. M. The Structural Basis of Aspirin Activity Inferred From the Crystal Structure of Inactivated Prostaglandin H2 Synthase. *Nature Structural Biology* **2**, 637–643 (1995).
- [21] Manning, G., Whyte, D. B., Martinez, R., Hunter, T. & Sudarsanam, S. The Protein Kinase Complement of the Human Genome. *Science* **298**, 1912–1934 (2002).
- [22] Liu, Q., Sabnis, Y., Zhao, Z., Zhang, T., Buhrlage, S. J., Jones, L. H. & Gray, N. S. Developing Irreversible Inhibitors of the Protein Kinase Cysteinome. *Chemistry & Biology* **20**, 146–159 (2013).
- [23] Marino, S. M. & Gladyshev, V. N. Cysteine Function Governs Its Conservation and Degeneration and Restricts Its Utilization on Protein Surfaces. *Journal of Molecular Biology* **404**, 902–916 (2010).
- [24] Weerapana, E., Wang, C., Simon, G. M., Richter, F., Khare, S., Dillon, M. B. D., Bachovchin, D. A., Mowen, K., Baker, D. & Cravatt, B. F. Quantitative Reactivity Profiling Predicts Functional Cysteines in Proteomes. *Nature* **468**, 790–795 (2010).

- [25] Moellering, R. E. & Cravatt, B. F. How Chemoproteomics Can Enable Drug Discovery and Development. *Chemistry & Biology* **19**, 11–22 (2012).
- [26] Backus, K. M., Correia, B. E., Lum, K. M., Forli, S., Horning, B. D., González-Páez, G. E., Chatterjee, S., Lanning, B. R., Teijaro, J. R., Olson, A. J., Wolan, D. W. & Cravatt, B. F. Proteome-Wide Covalent Ligand Discovery in Native Biological Systems. *Nature* **534**, 570–574 (2016).
- [27] Serafimova, I. M., Pufall, M. A., Krishnan, S., Duda, K., Cohen, M. S., Maglathlin, R. L., McFarland, J. M., Miller, R. M., Frödin, M. & Taunton, J. Reversible Targeting of Noncatalytic Cysteines With Chemically Tuned Electrophiles. *Nature Chemical Biology* **8**, 471–476 (2012).
- [28] London, N., Miller, R. M., Krishnan, S., Uchida, K., Irwin, J. J., Eidam, O., Gibold, L., Cimermančič, P., Bonnet, R., Shoichet, B. K. & Taunton, J. Covalent Docking of Large Libraries for the Discovery of Chemical Probes. *Nature Chemical Biology* **10**, 1066–1072 (2014).
- [29] Lonsdale, R., Burgess, J., Colclough, N., Davies, N. L., Lenz, E. M., Orton, A. L. & Ward, R. A. Expanding the Armory: Predicting and Tuning Covalent Warhead Reactivity. *Journal of Chemical Information and Modeling* **57**, 3124–3137 (2017).
- [30] Palermo, G., Branduardi, D., Masetti, M., Lodola, A., Mor, M., Piomelli, D., Cavalli, A. & De Vivo, M. Covalent Inhibitors of Fatty Acid Amide Hydrolase: A Rationale for the Activity of Piperidine and Piperazine Aryl Ureas. *Journal of Medicinal Chemistry* **54**, 6612–6623 (2011).
- [31] Michaelis, L. & Menten, M. L. Die Kinetik Der Invertinwirkung. *Biochem. z* **49**, 352 (1913).
- [32] Copeland, R. A. *Evaluation of Enzyme Inhibitors in Drug Discovery: A Guide for Medicinal Chemists and Pharmacologists* (Wiley, 2013), 1 edn.
- [33] Boike, L., Henning, N. J. & Nomura, D. K. Advances in Covalent Drug Discovery. *Nat Rev Drug Discov* **21**, 881–898 (2022).
- [34] Karplus, M. & Petsko, G. A. Molecular Dynamics Simulations in Biology. *Nature* **347**, 631–639 (1990).
- [35] Briggs, G. E. & Haldane, J. B. S. A Note on the Kinetics of Enzyme Action. *Biochemical Journal* **19**, 338–339 (1925).
- [36] Bradshaw, J. M., McFarland, J. M., Paavilainen, V. O., Bisconte, A., Tam, D., Phan, V. T., Romanov, S., Finkle, D., Shu, J., Patel, V., Ton, T., Li, X., Loughhead, D. G., Nunn, P. A., Karr, D. E., Gerritsen, M. E., Funk, J. O., Owens, T. D., Verner, E., Brameld, K. A., Hill, R. J., Goldstein, D. M. & Taunton, J. Prolonged and Tunable Residence Time Using Reversible Covalent Kinase Inhibitors. *Nature Chemical Biology* **11**, 525–531 (2015).

- [37] Ghaby, K. & Roux, B. Kinetic Modeling of Covalent Inhibition: Effects of Rapidly Fluctuating Intermediate States. *Journal of Chemical Information and Modeling* **65**, 8730–8739 (2025).
- [38] Baker, B. R. Factors in the Design of Active-Site-Directed Irreversible Inhibitors. *Journal of Pharmaceutical Sciences* **53**, 347–364 (1964).
- [39] Proudfoot, A. T. & Wright, N. Acute Paracetamol Poisoning. *BMJ* **3**, 557–558 (1970).
- [40] Mah, R., Thomas, J. R. & Shafer, C. M. Drug Discovery Considerations in the Development of Covalent Inhibitors. *Bioorganic & Medicinal Chemistry Letters* **24**, 33–39 (2014).
- [41] Bauer, R. A. Covalent Inhibitors in Drug Discovery: From Accidental Discoveries to Avoided Liabilities and Designed Therapies. *Drug Discovery Today* **20**, 1061–1073 (2015).
- [42] González-Bello, C. Designing Irreversible Inhibitors—Worth the Effort? *ChemMedChem* **11**, 22–30 (2016).
- [43] Zhang, T., Hatcher, J. M., Teng, M., Gray, N. S. & Kostic, M. Recent Advances in Selective and Irreversible Covalent Ligand Development and Validation. *Cell Chemical Biology* **26**, 1486–1500 (2019).
- [44] Ghosh, A. K., Samanta, I., Mondal, A. & Liu, W. R. Covalent Inhibition in Drug Discovery. *ChemMedChem* **14**, 889–906 (2019).
- [45] Krantz, A. A Classification of Enzyme Inhibitors. *Bioorganic & Medicinal Chemistry Letters* **2**, 1327–1334 (1992).
- [46] Potashman, M. H. & Duggan, M. E. Covalent Modifiers: An Orthogonal Approach to Drug Design. *Journal of Medicinal Chemistry* **52**, 1231–1246 (2009).
- [47] Smith, A. J. T., Zhang, X., Leach, A. G. & Houk, K. N. Beyond Picomolar Affinities: Quantitative Aspects of Noncovalent and Covalent Binding of Drugs to Proteins. *Journal of Medicinal Chemistry* **52**, 225–233 (2009).
- [48] Johnson, D. S., Weerapana, E. & Cravatt, B. F. Strategies for Discovering and Derisking Covalent, Irreversible Enzyme Inhibitors. *Future Medicinal Chemistry* **2**, 949–964 (2010).
- [49] Kalgutkar, A. S. & Dalvie, D. K. Drug Discovery for a New Generation of Covalent Drugs. *Expert Opinion on Drug Discovery* **7**, 561–581 (2012).
- [50] Niphakis, M. J. & Cravatt, B. F. Enzyme Inhibitor Discovery by Activity-Based Protein Profiling. *Annual Review of Biochemistry* **83**, 341–377 (2014).

- [51] Visscher, M., Arkin, M. R. & Dansen, T. B. Covalent Targeting of Acquired Cysteines in Cancer. *Current Opinion in Chemical Biology* **30**, 61–67 (2016).
- [52] Baillie, T. A. Targeted Covalent Inhibitors for Drug Design. *Angewandte Chemie International Edition* **55**, 13408–13421 (2016).
- [53] Bandyopadhyay, A. & Gao, J. Targeting Biomolecules With Reversible Covalent Chemistry. *Current Opinion in Chemical Biology* **34**, 110–116 (2016).
- [54] De Cesco, S., Kurian, J., Dufresne, C., Mittermaier, A. K. & Moitessier, N. Covalent Inhibitors Design and Discovery. *European Journal of Medicinal Chemistry* **138**, 96–114 (2017).
- [55] Jackson, P. A., Widen, J. C., Harki, D. A. & Brummond, K. M. Covalent Modifiers: A Chemical Perspective on the Reactivity of α , β -Unsaturated Carbonyls With Thiols via Hetero-Michael Addition Reactions. *Journal of Medicinal Chemistry* **60**, 839–885 (2017).
- [56] Lagoutte, R., Patouret, R. & Winssinger, N. Covalent Inhibitors: An Opportunity for Rational Target Selectivity. *Current Opinion in Chemical Biology* **39**, 54–63 (2017).
- [57] Lu, S. & Zhang, J. Designed Covalent Allosteric Modulators: An Emerging Paradigm in Drug Discovery. *Drug Discovery Today* **22**, 447–453 (2017).
- [58] Mukherjee, H. & Grimster, N. P. Beyond Cysteine: Recent Developments in the Area of Targeted Covalent Inhibition. *Current Opinion in Chemical Biology* **44**, 30–38 (2018).
- [59] Lonsdale, R. & Ward, R. A. Structure-Based Design of Targeted Covalent Inhibitors. *Chemical Society Reviews* **47**, 3816–3830 (2018).
- [60] Faridoon, Ng, R., Zhang, G. & Li, J. J. An Update on the Discovery and Development of Reversible Covalent Inhibitors. *Medicinal Chemistry Research* **32**, 1039–1062 (2023).
- [61] Hoyt, K. W., Urul, D. A., Ogboo, B. C., Wittlinger, F., Laufer, S. A., Schaefer, E. M., May, E. W. & Heppner, D. E. Pitfalls and Considerations in Determining the Potency and Mutant Selectivity of Covalent Epidermal Growth Factor Receptor Inhibitors: Miniperspective. *Journal of Medicinal Chemistry* **67**, 2–16 (2024).
- [62] Müller, G., Sennhenn, P. C., Woodcock, T. & Neumann, L. The 'Retro-Design' Concept for Novel Kinase Inhibitors. *IDrugs* **13**, 457–466 (2010).
- [63] Lu, H. & Tonge, P. J. Drug-Target Residence Time: Critical Information for Lead Optimization. *Current Opinion in Chemical Biology* **14**, 467–474 (2010).
- [64] Awoonor-Williams, E., Walsh, A. G. & Rowley, C. N. Modeling Covalent-Modifier Drugs. *Biochim Biophys Acta Proteins Proteom* **1865**, 1664–1675 (2017).

- [65] Tuley, A. & Fast, W. The Taxonomy of Covalent Inhibitors. *Biochemistry* **57**, 3326–3337 (2018).
- [66] Martin, J. S., MacKenzie, C. J., Fletcher, D. & Gilbert, I. H. Characterising Covalent Warhead Reactivity. *Bioorganic & Medicinal Chemistry* **27**, 2066–2074 (2019).
- [67] Abdeldayem, A., Raouf, Y. S., Constantinescu, S. N., Moriggl, R. & Gunning, P. T. Advances in Covalent Kinase Inhibitors. *Chemical Society Reviews* **49**, 2617–2687 (2020).
- [68] Bianco, G., Goodsell, D. S. & Forli, S. Selective and Effective: Current Progress in Computational Structure-Based Drug Discovery of Targeted Covalent Inhibitors. *Trends in Pharmacological Sciences* **41**, 1038–1049 (2020).
- [69] Yoshimori, A., Miljković, F. & Bajorath, J. Approach for the Design of Covalent Protein Kinase Inhibitors via Focused Deep Generative Modeling. *Molecules* **27**, 570 (2022).
- [70] Hasan, M. N., Ray, M. & Saha, A. Tools for Modeling Covalent Modification of Proteins: A Review on Computational Covalent Drug Discovery. *The Journal of Physical Chemistry B* **127**, 9663–9684 (2023).
- [71] Schaefer, D. & Cheng, X. Recent Advances in Covalent Drug Discovery. *Pharmaceuticals* **16**, 663 (2023).
- [72] Raouf, Y. S. Covalent Inhibitors: To Infinity and Beyond. *Journal of Medicinal Chemistry* **67**, 10513–10516 (2024).
- [73] Fradera, X., Kaur, J. & Mestres, J. Unsupervised Guided Docking of Covalently Bound Ligands. *Journal of Computer-Aided Molecular Design* **18**, 635–650 (2004).
- [74] Toledo Warshaviak, D., Golan, G., Borrelli, K. W., Zhu, K. & Kalid, O. Structure-Based Virtual Screening Approach for Discovery of Covalently Bound Ligands. *Journal of Chemical Information and Modeling* **54**, 1941–1950 (2014).
- [75] Zhu, K., Borrelli, K. W., Greenwood, J. R., Day, T., Abel, R., Farid, R. S. & Harder, E. Docking Covalent Inhibitors: A Parameter Free Approach to Pose Prediction and Scoring. *Journal of Chemical Information and Modeling* **54**, 1932–1940 (2014).
- [76] Scholz, C., Knorr, S., Hamacher, K. & Schmidt, B. DOCKTITE-A Highly Versatile Step-by-Step Workflow for Covalent Docking and Virtual Screening in the Molecular Operating Environment. *Journal of Chemical Information and Modeling* **55**, 398–406 (2015).
- [77] Bianco, G., Forli, S., Goodsell, D. S. & Olson, A. J. Covalent Docking Using Autodock: Two-Point Attractor and Flexible Side Chain Methods. *Protein Science* **25**, 295–301 (2016).

- [78] Moitessier, N., Pottel, J., Therrien, E., Englebienne, P., Liu, Z., Tomberg, A. & Corbeil, C. R. Medicinal Chemistry Projects Requiring Imaginative Structure-Based Drug Design Methods. *Accounts of Chemical Research* **49**, 1646–1657 (2016).
- [79] Yu, H. S., Gao, C., Lupyan, D., Wu, Y., Kimura, T., Wu, C., Jacobson, L., Harder, E., Abel, R. & Wang, L. Toward Atomistic Modeling of Irreversible Covalent Inhibitor Binding Kinetics. *Journal of Chemical Information and Modeling* **59**, 3955–3967 (2019).
- [80] Mihalovits, L. M., Ferenczy, G. G. & Keserű, G. M. Affinity and Selectivity Assessment of Covalent Inhibitors by Free Energy Calculations. *Journal of Chemical Information and Modeling* **60**, 6579–6594 (2020).
- [81] Awoonor-Williams, E. & Rowley, C. N. Modeling the Binding and Conformational Energetics of a Targeted Covalent Inhibitor to Bruton’s Tyrosine Kinase. *Journal of Chemical Information and Modeling* **61**, 5234–5242 (2021).
- [82] Endres, E., Yuan Chen, N. & Sottriffer, C. MD-Based Assessment of Covalent Inhibitors in Noncovalent Association Complexes: Learning From Cathepsin K as a Test Case. *Journal of Chemical Information and Modeling* **63**, 3186–3197 (2023).
- [83] Zhou, Y., Yu, H., Vind, A. C., Kong, L., Liu, Y., Song, X., Tu, Z., Yun, C., Smail, J. B., Zhang, Q.-W., Ding, K., Bekker-Jensen, S. & Lu, X. Rational Design of Covalent Kinase Inhibitors by an Integrated Computational Workflow (Kin-Cov). *Journal of Medicinal Chemistry* **66**, 7405–7420 (2023).
- [84] Barragán, A. M., Ghaby, K., Pond, M. P. & Roux, B. Computational Investigation of the Covalent Inhibition Mechanism of Bruton’s Tyrosine Kinase by Ibrutinib. *Journal of Chemical Information and Modeling* **64**, 3488–3502 (2024).
- [85] Liu, Y., Tan, J., Hu, S., Hussain, M., Qiao, C., Tu, Y., Lu, X. & Zhou, Y. Dynamics Playing a Key Role in the Covalent Binding of Inhibitors to Focal Adhesion Kinase. *Journal of Chemical Information and Modeling* **64**, 6053–6061 (2024).
- [86] Mayhew, B. S., Jones, D. R. & Hall, S. D. An In Vitro Model for Predicting In Vivo Inhibition of Cytochrome P450 3A4 by Metabolic Intermediate Complex Formation. *Drug Metabolism and Disposition* **28**, 1031–1037 (2000).
- [87] Mons, E., Roet, S., Kim, R. Q. & Mulder, M. P. C. A Comprehensive Guide for Assessing Covalent Inhibition in Enzymatic Assays Illustrated With Kinetic Simulations. *Current Protocols* **2**, e419 (2022).
- [88] Klinman, J. P. & Matthews, R. G. Calculation of Substrate Dissociation Constants From Steady-State Isotope Effects in Enzyme-Catalyzed Reactions. *Journal of the American Chemical Society* **107**, 1058–1060 (1985).
- [89] Teslenko, V. I. & Kapitanchuk, O. L. Dynamics of Transient Processes in Irreversible Kinetic Models. *Ukrainian Journal of Physics* **57** (2012).

- [90] Bisi, M. & Loy, N. Kinetic Models for Systems of Interacting Agents With Multiple Microscopic States. *Physica D: Nonlinear Phenomena* **457**, 133967 (2024).
- [91] Estevez, J. & Vilanova, E. Model Equations for the Kinetics of Covalent Irreversible Enzyme Inhibition and Spontaneous Reactivation: Esterases and Organophosphorus Compounds. *Critical Reviews in Toxicology* **39**, 427–448 (2009).
- [92] Krippendorff, B.-F., Neuhaus, R., Lienau, P., Reichel, A. & Huisinga, W. Mechanism-Based Inhibition: Deriving KI and kinact Directly From Time-Dependent IC50 Values. *SLAS Discovery* **14**, 913–923 (2009).
- [93] Sliwoski, G., Kothiwale, S., Meiler, J. & Lowe, E. W. Computational Methods in Drug Discovery. *Pharmacological Reviews* **66**, 334–395 (2014).
- [94] Kuzmič, P. Determination of kinact and KI for Covalent Inhibition Using the Omnia©R Assay. BioKin Technical Note TN-2015-02, BioKin Ltd., Watertown MA, [Online] (2015).
- [95] Kuzmič, P., Solowiej, J. & Murray, B. W. An Algebraic Model for the Kinetics of Covalent Enzyme Inhibition at Low Substrate Concentrations. *Analytical Biochemistry* **484**, 82–90 (2015).
- [96] Kuzmič, P. A Steady-State Algebraic Model for the Time Course of Covalent Enzyme Inhibition (2020). ChemRxiv:chemrxiv.12445394.v3.
- [97] Kuzmič, P. A Two-Point IC50 Method for Evaluating the Biochemical Potency of Irreversible Enzyme Inhibitors (2022). BioRxiv:2020.06.25.171207.
- [98] Ahmad, K., Rizzi, A., Capelli, R., Mandelli, D., Lyu, W. & Carloni, P. Enhanced-Sampling Simulations for the Estimation of Ligand Binding Kinetics: Current Status and Perspective. *Frontiers in Molecular Biosciences* **9**, 899805 (2022).
- [99] Mader, L. K. & Keillor, J. W. Fitting of kinact and KI Values From Endpoint Pre-incubation IC50 Data. *ACS Medicinal Chemistry Letters* **15**, 731–738 (2024).
- [100] Srinivasan, B. A Guide to Enzyme Kinetics in Early Drug Discovery. *The FEBS Journal* **290**, 2292–2305 (2023).
- [101] Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., van der Walt, S. J., Brett, M., Wilson, J., Millman, K. J., Mayorov, N., Nelson, A. R. J., Jones, E., Kern, R., Larson, E., Carey, C. J., Polat, İ., Feng, Y., Moore, E. W., VanderPlas, J., Laxalde, D., Perktold, J., Cimrman, R., Henriksen, I., Quintero, E. A., Harris, C. R., Archibald, A. M., Ribeiro, A. H., Pedregosa, F., van Mulbregt, P. & SciPy 1.0 Contributors. SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nature Methods* **17**, 261–272 (2020).

- [102] Hindmarsh, A. C. ODEPACK, a Systematized Collection of ODE Solvers. In Stepleman, R. S. (ed.) *Scientific Computing*, 55–64 (North-Holland, Amsterdam, 1983).
- [103] Petzold, L. Automatic Selection of Methods for Solving Stiff and Nonstiff Systems of Ordinary Differential Equations. *SIAM Journal on Scientific and Statistical Computing* **4**, 136–148 (1983).
- [104] Cornish-Bowden, A. Validity of a 'Steady-State' Treatment of Inactivation Kinetics. *European Journal of Biochemistry* **93**, 383–385 (1979).
- [105] Strelow, J. M. A Perspective on the Kinetics of Covalent and Irreversible Inhibition. *SLAS Discovery* **22**, 3–20 (2017).
- [106] Hill, A. V. A New Mathematical Treatment of Changes of Ionic Concentration in Muscle and Nerve Under the Action of Electric Currents, With a Theory as to Their Mode of Excitation. *The Journal of Physiology* **40**, 190–224 (1910).
- [107] Arafet, K., Scalvini, L., Galvani, F., Martí, S., Moliner, V., Mor, M. & Lodola, A. Mechanistic Modeling of Lys745 Sulfonylation in EGFR C797S Reveals Chemical Determinants for Inhibitor Activity and Discriminates Reversible From Irreversible Agents. *Journal of Chemical Information and Modeling* **63**, 1301–1312 (2023).
- [108] Cornish-Bowden, A. & Koshland Jr, D. E. Diagnostic Uses of the Hill (Logit and Nernst) Plots. *Journal of Molecular Biology* **95**, 201–212 (1975).
- [109] Buchler, N. E. & Louis, M. Molecular Titration and Ultrasensitivity in Regulatory Networks. *Journal of Molecular Biology* **384**, 1106–1119 (2008).
- [110] Jaynes, E. T. Information Theory and Statistical Mechanics. *Physical Review* **106**, 620–630 (1957).
- [111] Yang, S., Blachowicz, L., Makowski, L. & Roux, B. Multidomain Assembled States of HCK Tyrosine Kinase in Solution. *Proceedings of the National Academy of Sciences* **107**, 15757–15762 (2010).
- [112] Schwartz, P. A., Kuzmic, P., Solowiej, J., Bergqvist, S., Bolanos, B., Almaden, C., Nagata, A., Ryan, K., Feng, J., Dalvie, D., Kath, J. C., Xu, M., Wani, R. & Murray, B. W. Covalent EGFR Inhibitor Analysis Reveals Importance of Reversible Interactions to Potency and Mechanisms of Drug Resistance. *Proceedings of the National Academy of Sciences* **111**, 173–178 (2014).
- [113] Zhao, M., Wu, S., Zhou, Q., Vivona, S., Cipriano, D. J., Cheng, Y. & Brunger, A. T. Mechanistic Insights Into the Recycling Machine of the SNARE Complex. *Nature* **518**, 61–67 (2015).
- [114] Pan, Z., Scheerens, H., Li, S. J., Schultz, B. E., Sprengeler, P. A., Burrill, L. C., Mendonca, R. V., Sweeney, M. D., Scott, K. C., Grothaus, P. G., Jeffery, D. A., Spoerke,

- J. M., Honigberg, L. A., Young, P. R., Dalrymple, S. A. & Palmer, J. T. Discovery of Selective Irreversible Inhibitors for Bruton’s Tyrosine Kinase. *ChemMedChem* **2**, 58–61 (2007).
- [115] Dubovsky, J. A., Beckwith, K. A., Natarajan, G., Woyach, J. A., Jaglowski, S., Zhong, Y., Hessler, J. D., Liu, T. M., Chang, B. Y., Larkin, K. M., Stefanovski, M. R., Chappell, D. L., Frissora, F. W., Smith, L. L., Smucker, K. A., Flynn, J. M., Jones, J. A., Andritsos, L. A., Maddocks, K., Lehman, A. M., Furman, R., Sharman, J., Mishra, A., Caligiuri, M. A., Satoskar, A. R., Buggy, J. J., Muthusamy, N., Johnson, A. J. & Byrd, J. C. Ibrutinib Is an Irreversible Molecular Inhibitor of ITK Driving a Th1-Selective Pressure in T Lymphocytes. *Blood* **122**, 2539–2549 (2013).
- [116] Advani, R. H., Buggy, J. J., Sharman, J. P., Smith, S. M., Boyd, T. E., Grant, B., Kolibaba, K. S., Furman, R. R., Rodriguez, S., Chang, B. Y., Sukbuntherng, J., Izumi, R., Hamdy, A., Hedrick, E. & Fowler, N. H. Bruton Tyrosine Kinase Inhibitor Ibrutinib (PCI-32765) Has Significant Activity in Patients With Relapsed/Refractory B-cell Malignancies. *Journal of Clinical Oncology* **31**, 88–94 (2013).
- [117] Pal Singh, S., Dammeijer, F. & Hendriks, R. W. Role of Bruton’s Tyrosine Kinase in B Cells and Malignancies. *Molecular Cancer* **17**, 1–23 (2018).
- [118] Barf, T. & Kaptein, A. Irreversible Protein Kinase Inhibitors: Balancing the Benefits and Risks. *Journal of Medicinal Chemistry* **55**, 6243–6262 (2012).
- [119] Blair, J. A., Rauh, D., Kung, C., Yun, C. H., Fan, Q. W., Rode, H., Zhang, C., Eck, M. J., Weiss, W. A. & Shokat, K. M. Structure-Guided Development of Affinity Probes for Tyrosine Kinases Using Chemical Genetics. *Nature Chemical Biology* **3**, 229–238 (2007).
- [120] Warshel, A. & Levitt, M. Theoretical Studies of Enzymic Reactions - Dielectric, Electrostatic and Steric Stabilization of Carbonium-Ion in Reaction of Lysozyme. *Journal of Molecular Biology* **103**, 227–249 (1976).
- [121] Field, M. J., Bash, P. A. & Karplus, M. A Combined Quantum Mechanical and Molecular Mechanical Potential for Molecular Dynamics Simulations. *Journal of Computational Physics* **11**, 700–733 (1990).
- [122] Melo, M. C. R., Bernardi, R. C., Rudack, T., Scheurer, M., Riplinger, C., Phillips, J. C., Maia, J. D. C., Rocha, G. B., Ribeiro, J. V., Stone, J. E., Neese, F., Schulten, K. & Luthey-Schulten, Z. NAMD Goes Quantum: An Integrative Suite for Hybrid Simulations. *Nature Methods* **15**, 351–354 (2018).
- [123] Voice, A. T., Tresadern, G., Twidale, R. M., van Vlijmen, H. & Mulholland, A. J. Mechanism of Covalent Binding of Ibrutinib to Bruton’s Tyrosine Kinase Revealed by QM/MM Calculations. *Chemical Science* **12**, 5511–5516 (2021).

- [124] Johansson, M. H. Reversible Michael Additions: Covalent Inhibitors and Prodrugs. *Mini-Reviews in Medicinal Chemistry* **12**, 1330–1344 (2012).
- [125] Smith, J. M. & Rowley, C. N. Automated Computational Screening of the Thiol Reactivity of Substituted Alkenes. *Journal of Computer-Aided Molecular Design* **29**, 725–735 (2015).
- [126] E, W., Ren, W. & Vanden-Eijnden, E. String Method for the Study of Rare Events. *Physical Review B* **66**, 052301 (2002).
- [127] E, W., Ren, W. & Vanden-Eijnden, E. Finite Temperature String Method for the Study of Rare Events. *The Journal of Physical Chemistry B* **109**, 6688–6693 (2005).
- [128] Maragliano, L., Fischer, A., Vanden-Eijnden, E. & Ciccotti, G. String Method in Collective Variables: Minimum Free Energy Paths and Isocommittor Surfaces. *The Journal of Chemical Physics* **125**, 24106 (2006).
- [129] E, W., Ren, W. & Vanden-Eijnden, E. Simplified and Improved String Method for Computing the Minimum Energy Paths in Barrier-Crossing Events. *The Journal of Chemical Physics* **126**, 164103 (2007).
- [130] Pan, A. C., Sezer, D. & Roux, B. Finding Transition Pathways Using the String Method With Swarms of Trajectories. *The Journal of Physical Chemistry B* **112**, 3432–3440 (2008).
- [131] Zhu, F. & Hummer, G. Pore Opening and Closing of a Pentameric Ligand-Gated Ion Channel. *Proceedings of the National Academy of Sciences* **107**, 19814–19819 (2010).
- [132] Bender, A. T., Gardberg, A., Pereira, A., Johnson, T., Wu, Y., Grenningloh, R., Head, J., Morandi, F., Haselmayer, P. & Liu-Bujalski, L. Ability of Bruton’s Tyrosine Kinase Inhibitors to Sequester Y551 and Prevent Phosphorylation Determines Potency for Inhibition of Fc Receptor but Not B-Cell Receptor Signaling. *Molecular Pharmacology* **91**, 208–219 (2017).
- [133] Marcotte, D. J., Liu, Y. T., Arduini, R. M., Hession, C. A., Miatkowski, K., Wildes, C. P., Cullen, P. F., Hong, V., Hopkins, B. T., Mertsching, E., Jenkins, T. J., Romanowski, M. J., Baker, D. P. & Silvian, L. F. Structures of Human Bruton’s Tyrosine Kinase in Active and Inactive Conformations Suggest a Mechanism of Activation for TEC Family Kinases. *Protein Science* **19**, 429–439 (2010).
- [134] Jo, S., Kim, T., Iyer, V. G. & Im, W. CHARMM-GUI: A Web-Based Graphical User Interface for CHARMM. *Journal of Computational Chemistry* **29**, 1859–1865 (2008).
- [135] Lee, J., Cheng, X., Swails, J. M., Yeom, M. S., Eastman, P. K., Lemkul, J. A., Wei, S., Buckner, J., Jeong, J. C., Qi, Y., Jo, S., Pande, V. S., Case, D. A., Brooks, C. L. I. I. I., MacKerell, A. D. J., Klauda, J. B. & Im, W. CHARMM-GUI Input Generator for NAMD, GROMACS, AMBER, OpenMM, and CHARMM/OpenMM

- Simulations Using the CHARMM36 Additive Force Field. *Journal of Chemical Theory and Computation* **12**, 405–413 (2016).
- [136] Kim, S., Lee, J., Jo, S., Brooks III, C. L., Lee, H. S. & Im, W. CHARMM-GUI Ligand Reader and Modeler for CHARMM Force Field Generation of Small Molecules. *Journal of Computational Chemistry* **38**, 1879–1886 (2017).
- [137] MacKerell, A. D. J., Bashford, D., Bellot, M., , J.D. Evanseck, R. L. D., Field, M. J., Fischer, S., Gao, J., Guo, H., S. Ha, D. J.-M., Kuchnir, L., , F.T.K. Lau, K. K., Mattos, C., Michnick, S., Ngo, T., , B. Prodhom, D. T. N., Reiher III, W. E., Roux, B., , J. Smith, M. S., Stote, R., Straub, J., Watanabe, M., & M. Karplus, J. W.-K. All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins. *The Journal of Physical Chemistry B* **102**, 3586–3616 (1998).
- [138] Best, R. B., Zhu, X., Shim, J., Lopes, P. E. M., Mittal, J., Feig, M. & MacKerell Jr, A. D. Optimization of the Additive CHARMM All-Atom Protein Force Field Targeting Improved Sampling of the Backbone Φ , Ψ and Side-Chain $\chi(1)$ and $\chi(2)$ Dihedral Angles. *Journal of Chemical Theory and Computation* **8**, 3257–3273 (2012).
- [139] Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., & M. L. Klein, R. W. I. Comparison of Simple Potential Functions for Simulating Liquid Water. *The Journal of Chemical Physics* **79**, 926–935 (1983).
- [140] Vanommeslaeghe, K., Hatcher, E., Acharya, C., Kundu, S., Zhong, S., Shim, J., Darian, E., Guvench, O., Lopes, P., Vorobyov, I. & Mackerell Jr., A. D. CHARMM General Force Field: A Force Field for Drug-Like Molecules Compatible With the CHARMM All-Atom Additive Biological Force Fields. *Journal of Computational Chemistry* **31**, 671–690 (2010).
- [141] Vanommeslaeghe, K. & MacKerell, A. D. Automation of the CHARMM General Force Field (CGenFF) I: Bond Perception and Atom Typing. *Journal of Chemical Information and Modeling* **52**, 3144–3154 (2012). PMID: 23146088.
- [142] Vanommeslaeghe, K., Raman, E. P. & MacKerell, A. D. Automation of the CHARMM General Force Field (CGenFF) II: Assignment of Bonded Parameters and Partial Atomic Charges. *Journal of Chemical Information and Modeling* **52**, 3155–3168 (2012). PMID: 23145473.
- [143] Götz, A. W., Williamson, M. J., Xu, D., Poole, D., Le Grand, S. & Walker, R. C. Routine Microsecond Molecular Dynamics Simulations With AMBER on GPUs. 1. Generalized Born. *Journal of Chemical Theory and Computation* **8**, 1542–1555 (2012).
- [144] Salomon-Ferrer, R., Götz, A. W., Poole, D., Le Grand, S. & Walker, R. C. Routine Microsecond Molecular Dynamics Simulations With AMBER on GPUs. 2. Explicit Solvent Particle Mesh Ewald. *Journal of Chemical Theory and Computation* **9**, 3878–3888 (2013).

- [145] Le Grand, S., Goetz, A. W. & Walker, R. C. SPFP: Speed Without Compromise-a Mixed Precision Model for GPU Accelerated Molecular Dynamics Simulations. *Computer Physics Communications* **184**, 374–380 (2013).
- [146] Essmann, U., Perera, L., Berkowitz, M. L., Darden, T., & LG Pedersen, H. L. A Smooth Particle Mesh Ewald Method. *The Journal of Chemical Physics* **103**, 8577–8593 (1995).
- [147] Brooks, B. R., Brooks, C. L., Mackerell, A. D., Nilsson, L., Petrella, R. J., Roux, B., Won, Y., Archontis, G., Bartels, C., Boresch, S., Caffisch, A., Caves, L., Cui, Q., Dinner, A. R., Feig, M., Fischer, S., Gao, J., Hodoscek, M., Im, W., Kuczera, K., Lazaridis, T., Ma, J., Ovchinnikov, V., Paci, E., Pastor, R. W., Post, C. B., Pu, J. Z., Schaefer, M., Tidor, B., Venable, R. M., Woodcock, H. L., Wu, X., Yang, W., York, D. M. & Karplus, M. CHARMM: The Biomolecular Simulation Program. *Journal of Computational Chemistry* **30**, 1545–1614 (2009).
- [148] Hopkins, C. W., Le Grand, S., Walker, R. C. & Roitberg, A. E. Long-Time-Step Molecular Dynamics Through Hydrogen Mass Repartitioning. *Journal of Chemical Theory and Computation* **11**, 1864–1874 (2015).
- [149] Andersen, H. C. Rattle: A “Velocity” Version of the Shake Algorithm for Molecular Dynamics Calculations. *Journal of Computational Physics* **52**, 24–34 (1983).
- [150] Humphrey, W., Dalke, A. & Schulten, K. VMD – Visual Molecular Dynamics. *Journal of Molecular Graphics* **14**, 33–38 (1996).
- [151] Jiang, W., Phillips, J. C., Huang, L., Fajer, M., Meng, Y., Gumbart, J. C., Luo, Y., Schulten, K. & Roux, B. Generalized Scalable Multiple Copy Algorithms for Molecular Dynamics Simulations in NAMD. *Computer Physics Communications* **185**, 908–916 (2014).
- [152] Phillips, J. C., Hardy, D. J., Maia, J. D. C., Stone, J. E., Ribeiro, J. V., Bernardi, R. C., Buch, R., Fiorin, G., Hénin, J., Jiang, W., McGreevy, R., Melo, M. C. R., Radak, B. K., Skeel, R. D., Singharoy, A., Wang, Y., Roux, B., Aksimentiev, A., Luthey-Schulten, Z., Kalé, L. V., Schulten, K., Chipot, C. & Tajkhorshid, E. Scalable Molecular Dynamics on CPU and GPU Architectures With NAMD. *The Journal of Chemical Physics* **153**, 044130 (2020).
- [153] Stewart, J. J. P. Optimization of Parameters for Semiempirical Methods VI: More Modifications to the NDDO Approximations and Re-Optimization of Parameters. *Journal of Molecular Modeling* **19**, 1–32 (2013).
- [154] Stewart, J. J. P. Special Issue - MOPAC - A Semiempirical Molecular-Orbital Program. *Journal of Computer-Aided Molecular Design* **4**, 1–45 (1990).
- [155] Carvalho Maia, J. D., Urquiza Carvalho, G. A., Manguiera, J. C. P., Santana, S. R., Formiga Cabral, L. A. & Rocha, G. B. GPU Linear Algebra Libraries and GPGPU

- Programming for Accelerating MOPAC Semiempirical Quantum Chemistry Calculations. *Journal of Chemical Theory and Computation* **8**, 3072–3081 (2012).
- [156] Fiorin, G., Klein, M. L. & Henin, J. Using Collective Variables to Drive Molecular Dynamics Simulations. *Molecular Physics* **111**, 3345–3362 (2013).
- [157] Fu, H., Shao, X., Cai, W. & Chipot, C. Taming Rugged Free Energy Landscapes Using an Average Force. *Accounts of Chemical Research* **52**, 3254–3264 (2019).
- [158] Fu, H., Chen, H., Wang, X., Chai, H., Shao, X., Cai, W. & Chipot, C. Finding an Optimal Pathway on a Multidimensional Free-Energy Landscape. *Journal of Chemical Information and Modeling* **60**, 5366–5374 (2020).
- [159] Darve, E., Rodríguez-Gómez, D. & Pohorille, A. Adaptive Biasing Force Method for Scalar and Vector Free Energy Calculations. *The Journal of Chemical Physics* **128**, 144120 (2008).
- [160] Lesage, A., Lelièvre, T., Stoltz, G. & Hénin, J. Smoothed Biasing Forces Yield Unbiased Free Energies With the Extended-System Adaptive Biasing Force Method. *The Journal of Physical Chemistry B* **121**, 3676–3685 (2017).
- [161] Fu, H., Shao, X., Chipot, C. & Cai, W. Extended Adaptive Biasing Force Algorithm. An On-the-Fly Implementation for Accurate Free-Energy Calculations. *Journal of Chemical Theory and Computation* **12**, 3506–3513 (2016).
- [162] Henin, J., Fiorin, G., Chipot, C. & Klein, M. L. Exploring Multidimensional Free Energy Landscapes Using Time-Dependent Biases on Collective Variables. *Journal of Chemical Theory and Computation* **6**, 35–47 (2010).
- [163] Grubmüller, H. Predicting Slow Structural Transitions in Macromolecular Systems: Conformational Flooding. *Physical Review E* **52**, 2893–2906 (1995).
- [164] Barducci, A., Bussi, G. & Parrinello, M. Well-Tempered Metadynamics: A Smoothly Converging and Tunable Free-Energy Method. *Physical Review Letters* **100**, 020603 (2008).
- [165] Dama, J. F., Parrinello, M. & Voth, G. A. Well-Tempered Metadynamics Converges Asymptotically. *Physical Review Letters* **112**, 240602 (2014).
- [166] Sagnella, D. E. & Voth, G. A. Structure and Dynamics of Hydronium in the Ion Channel Gramicidin A. *Biophysical Journal* **70**, 2043–2051 (1996).
- [167] Zapf, C. W., Gerstenberger, B. S., Xing, L., Limburg, D. C., Anderson, D. R., Caspers, N., Han, S., Aulabaugh, A., Kurumbail, R., Shakya, S., Li, X., Spaulding, V., Czerwinski, R. M., Seth, N. & Medley, Q. G. Covalent Inhibitors of Interleukin-2 Inducible T Cell Kinase (ITK) With Nanomolar Potency in a Whole-Blood Assay. *Journal of Medicinal Chemistry* **55**, 10047–10063 (2012).

- [168] Quaytman, S. L. & Schwartz, S. D. Reaction Coordinate of an Enzymatic Reaction Revealed by Transition Path Sampling. *Proceedings of the National Academy of Sciences* **104**, 12253–12258 (2007).
- [169] Schwartz, S. D. & Schramm, V. L. Enzymatic Transition States and Dynamic Motion in Barrier Crossing. *Nature Chemical Biology* **5**, 551–558 (2009).
- [170] Schwartz, S. D. Protein Dynamics and Enzymatic Catalysis. *The Journal of Physical Chemistry B* **127**, 2649–2660 (2023).
- [171] Glasstone, S., Laidler, K. J. & Eyring, H. *Theory of Rate Processes* (McGraw-Hill, New York, 1941).
- [172] Chandler, D. Statistical Mechanics of Isomerization Dynamics in Liquids and the Transition State Approximation. *The Journal of Chemical Physics* **68**, 2959–2970 (1978).
- [173] Pu, J., Gao, J. & Truhlar, D. G. Multidimensional Tunneling, Recrossing, and the Transmission Coefficient for Enzymatic Reactions. *Chemical Reviews* **106**, 3140–3169 (2006).
- [174] Zinovjev, K., a, J. J. & n, I. Toward an Automatic Determination of Enzymatic Reaction Mechanisms and Their Activation Free Energies. *Journal of Chemical Theory and Computation* **9**, 3740–3749 (2013).
- [175] Smoluchowski, M. V. Drei Vorträge Über Diffusion, Brownsche Molekularbewegung und Koagulation von Kolloidteilchen. *Physikalische Zeitschrift* **17**, 557–571 (1916).
- [176] Kramers, H. A. Brownian Motion in a Field of Force and the Diffusion Model of Chemical Reactions. *Physica B+C* **7**, 284–304 (1940).
- [177] Hille, B. *Ionic Channels of Excitable Membranes, 3rd Edition* (Sinauer, Sunderland MA, 2001).
- [178] Lin, Y. L., Meng, Y., Jiang, W. & Roux, B. Explaining Why Gleevec Is a Specific and Potent Inhibitor of Abl Kinase. *Proceedings of the National Academy of Sciences* **110**, 1664–1669 (2013).
- [179] Klaeger, S., Heinzlmeir, S., Wilhelm, M., Polzer, H., Vick, B., Koenig, P. A., 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, S. R., Feuchtinger, A., Wu, Z., Schmidt, T., Rueckert, L., Becker, W., Huenges, J., Garz, A. K., Gohlke, B. O., Zolg, D. P., Kayser, G., Vooder, T., Preissner, R., Hahne, H., nisson, N., Kramer, K., tze, K., Bassermann, F., Schlegl, J., Ehrlich, H. C., Aiche, S., Walch, A., Greif, P. A., Schneider, S., Felder, E. R., Ruland, J., dard, G., Jeremias, I., Spiekermann, K. & Kuster, B. The Target Landscape of Clinical Kinase Drugs. *Science* **358** (2017).

- [180] Angst, D., Gessier, F., Janser, P., Vulpetti, A., Ichli, R., Beerli, C., Littlewood-Evans, A., Dawson, J., Nuesslein-Hildesheim, B., Wieczorek, G., Gutmann, S., Scheufler, C., Hinniger, A., Zimmerlin, A., Funhoff, E. G., Pulz, R. & Cenni, B. Discovery of LOU064 (Remibrutinib), a Potent and Highly Selective Covalent Inhibitor of Bruton's Tyrosine Kinase. *Journal of Medicinal Chemistry* **63**, 5102–5118 (2020).
- [181] Ma, B., Metrick, C. M., Gu, C., Hoemberger, M., Bajrami, B., Bame, E., Huang, J., Mingueneau, M., Murugan, P., Santoro, J. C., Tang, H., Wang, T. & Hopkins, B. T. Optimization of a Novel Piperazinone Series as Potent Selective Peripheral Covalent BTK Inhibitors. *Bioorganic & Medicinal Chemistry Letters* **60**, 128549 (2022).
- [182] Watterson, S. H., Liu, Q., Beaudoin Bertrand, M., Batt, D. G., Li, L., Pattoli, M. A., Skala, S., Cheng, L., Obermeier, M. T., Moore, R., Yang, Z., Vickery, R., Elzinga, P. A., Discenza, L., D'Arienzo, C., Gillooly, K. M., Taylor, T. L., Pulicicchio, C., Zhang, Y., Heimrich, E., McIntyre, K. W., Ruan, Q., Westhouse, R. A., Catlett, I. M., Zheng, N., Chaudhry, C., Dai, J., Galella, M. A., Tebben, A. J., Pokross, M., Li, J., Zhao, R., Smith, D., Rampulla, R., Allentoff, A., Wallace, M. A., Mathur, A., Salter-Cid, L., Macor, J. E., Carter, P. H., Fura, A., Burke, J. R. & Tino, J. A. Discovery of Branebrutinib (BMS-986195): A Strategy for Identifying a Highly Potent and Selective Covalent Inhibitor Providing Rapid In Vivo Inactivation of Bruton's Tyrosine Kinase (BTK). *Journal of Medicinal Chemistry* **62**, 3228–3250 (2019).
- [183] Kaptein, A., de Bruin, G., Emmelot-van Hoek, M., van de Kar, B., de Jong, A., Gulrajani, M., Demont, D., Covey, T., Mittag, D. & Barf, T. Potency and Selectivity of BTK Inhibitors in Clinical Development for B-Cell Malignancies. *Blood* **132** (2018). 60th Annual Meeting of the American-Society-of-Hematology (ASH), San Diego, CA, DEC 01-04, 2018.
- [184] Hopper, M., Gururaja, T., Kinoshita, T., Dean, J. P., Hill, R. J. & Mongan, A. Relative Selectivity of Covalent Inhibitors Requires Assessment of Inactivation Kinetics and Cellular Occupancy: A Case Study of Ibrutinib and Acalabrutinib. *The Journal of Pharmacology and Experimental Therapeutics* **372**, 331–338 (2020).
- [185] Liang, Q., Chen, Y., Yu, K., Chen, C., Zhang, S., Wang, A., Wang, W., Wu, H., Liu, X., Wang, B., Wang, L., Hu, Z., Wang, W., Ren, T., Zhang, S., Liu, Q., Yun, C. H. & Liu, J. Discovery of N-(3-(5-((3-acrylamido-4-(morpholine-4-carbonyl)phenyl)amino)-1-methyl-6-oxo-1,6-dihydropyridin-3-yl)-2-methylphenyl)-4-(tert-butyl)benzamide (CHMFL-BTK-01) as a Highly Selective Irreversible Bruton's Tyrosine Kinase (BTK) Inhibitor. *Eur J Med Chem* **131**, 107–125 (2017).
- [186] Toenjes, S. T., Heydari, B. S., Albright, S. T., Hazin, R., Ortiz, M. A., Piedrafit, F. J. & Gustafson, J. L. Controlling Ibrutinib's Conformations About Its Heterobiaryl Axis to Increase BTK Selectivity. *ACS Medicinal Chemistry Letters* **14**, 305–311 (2023).
- [187] Liepinsh, E. & Otting, G. Proton Exchange Rates From Amino Acid Side Chains—Implications for Image Contrast. *Magnetic Resonance in Medicine* **35**, 30–42 (1996).

- [188] Wallerstein, J., Weininger, U., Khan, M. A., Linse, S. & Akke, M. Site-Specific Protonation Kinetics of Acidic Side Chains in Proteins Determined by pH-dependent Carboxyl (^{13}C) NMR Relaxation. *Journal of the American Chemical Society* **137**, 3093–3101 (2015).
- [189] Chen, J., Yadav, N. N., Stait-Gardner, T., Gupta, A., Price, W. S. & Zheng, G. Thiol-Water Proton Exchange of Glutathione, Cysteine, and N-acetylcysteine: Implications for CEST MRI. *NMR in Biomedicine* **33**, e4188 (2020).
- [190] Branduardi, D., Gervasio, F. L. & Parrinello, M. From A to B in Free Energy Space. *The Journal of Chemical Physics* **126**, 054103 (2007).
- [191] Sattelmeyer, K. W., Tirado-Rives, J. & Jorgensen, W. L. Comparison of SCC-DFTB and NDDO-based Semiempirical Molecular Orbital Methods for Organic Molecules. *The Journal of Physical Chemistry A* **110**, 13551–13559 (2006).
- [192] Mendez-Hernandez, D. D., Gillmore, J. G., Montano, L. A., Gust, D., Moore, T. A., Moore, A. L. & Mujica, V. Building and Testing Correlations for the Estimation of One-Electron Reduction Potentials of a Diverse Set of Organic Molecules. *Journal of Physical Organic Chemistry* **28**, 320–328 (2015).
- [193] Shukla, D., Meng, Y., Roux, B. & Pande, V. S. Activation Pathway of SRC Kinase Reveals Intermediate States as Targets for Drug Design. *Nature Communications* **5**, 3397 (2014).
- [194] Paul, F., Meng, Y. & Roux, B. Identification of Druggable Kinase Target Conformations Using Markov Model Metastable States Analysis of Apo Abl. *Journal of Chemical Theory and Computation* **16**, 1896–1912 (2020).
- [195] Paul, F., Thomas, T. & Roux, B. Diversity of Long-Lived Intermediates Along the Binding Pathway of Imatinib to Abl Kinase Revealed by MD Simulations. *Journal of Chemical Theory and Computation* **16**, 7852–7865 (2020).
- [196] Gao, J., Ma, S., Major, D. T., Nam, K., Pu, J. & Truhlar, D. G. Mechanisms and Free Energies of Enzymatic Reactions. *Chemical Reviews* **106**, 3188–3209 (2006).
- [197] He, Z., Chipot, C. & Roux, B. Committor-Consistent Variational String Method. *The Journal of Physical Chemistry Letters* **13**, 9263–9271 (2022).
- [198] Chakraborty, P., Carlos, A., Thomas, T., Ghaby, K., Fathi, S., Sharma, M., Nguyen, N. K., Farmwald, M., Blachowicz, L., Alcyone, E., Harter, K., Yu, S., Pillai, K. S., Roux, B. & Moellering, R. E. Molecular Dynamics-Guided Design and Chemoproteomic Profiling of Covalent Kinase Activity Probes. *Journal of the American Chemical Society* **148**, 12641–12654 (2026).
- [199] Gujral, T. S., Peshkin, L. & Kirschner, M. W. Exploiting Polypharmacology for Drug Target Deconvolution. *Proceedings of the National Academy of Sciences* **111**, 5048–5053 (2014).

- [200] Ciceri, P., Müller, S., O’Mahony, A., Fedorov, O., Filippakopoulos, P., Hunt, J. P., Lasater, E. A., Pallares, G., Picaud, S., Wells, C., Martin, S., Wodicka, L. M., Shah, N. P., Treiber, D. K. & Knapp, S. Dual Kinase-Bromodomain Inhibitors for Rationally Designed Polypharmacology. *Nature Chemical Biology* **10**, 305–312 (2014).
- [201] Ember, S. W. J., Zhu, J.-Y., Olesen, S. H., Martin, M. P., Becker, A., Berndt, N., Georg, G. I. & Schönbrunn, E. Acetyl-Lysine Binding Site of Bromodomain-Containing Protein 4 (BRD4) Interacts With Diverse Kinase Inhibitors. *ACS Chemical Biology* **9**, 1160–1171 (2014).
- [202] Fabian, M. A., Biggs, W. H., Treiber, D. K., Atteridge, C. E., Azimioara, M. D., Benedetti, M. G., Carter, T. A., Ciceri, P., Edeen, P. T., Floyd, M., Ford, J. M., Galvin, M., Gerlach, J. L., Grotzfeld, R. M., Herrgard, S., Insko, D. E., Insko, M. A., Lai, A. G., Lélias, J.-M., Mehta, S. A., Milanov, Z. V., Velasco, A. M., Wodicka, L. M., Patel, H. K., Zarrinkar, P. P. & Lockhart, D. J. A Small Molecule–Kinase Interaction Map for Clinical Kinase Inhibitors. *Nature Biotechnology* **23**, 329–336 (2005).
- [203] Davis, M. I., Hunt, J. P., Herrgard, S., Ciceri, P., Wodicka, L. M., Pallares, G., Hocker, M., Treiber, D. K. & Zarrinkar, P. P. Comprehensive Analysis of Kinase Inhibitor Selectivity. *Nature Biotechnology* **29**, 1046–1051 (2011).
- [204] Patricelli, M. P., Nomanbhoy, T. K., Wu, J., Brown, H., Zhou, D., Zhang, J., Jagannathan, S., Aban, A., Okerberg, E., Herring, C., Nordin, B., Weissig, H., Yang, Q., Lee, J.-D., Gray, N. S. & Kozarich, J. W. In Situ Kinase Profiling Reveals Functionally Relevant Properties of Native Kinases. *Chemistry & Biology* **18**, 699–710 (2011).
- [205] Serafim, R. A. M., Haarer, L., Pedreira, J. G. B. & Gehring, M. Covalent Chemical Probes for Protein Kinases. *Current Research in Chemical Biology* **3**, 100040 (2023).
- [206] Cohen, P., Cross, D. & Jänne, P. A. Kinase Drug Discovery 20 Years After Imatinib: Progress and Future Directions. *Nature Reviews Drug Discovery* **20**, 551–569 (2021).
- [207] Ostrem, J. M. L. & Shokat, K. M. Targeting KRAS G12C With Covalent Inhibitors. *Annual Review of Cancer Biology* **6**, 49–64 (2022).
- [208] Lanning, B. R., Whitby, L. R., Dix, M. M., Douhan, J., Gilbert, A. M., Hett, E. C., Johnson, T. O., Joslyn, C., Kath, J. C., Niessen, S., Roberts, L. R., Schnute, M. E., Wang, C., Hulce, J. J., Wei, B., Whiteley, L. O., Hayward, M. M. & Cravatt, B. F. A Road Map to Evaluate the Proteome-Wide Selectivity of Covalent Kinase Inhibitors. *Nature Chemical Biology* **10**, 760–767 (2014).
- [209] Zhao, Z. & Bourne, P. E. Systematic Exploration of Privileged Warheads for Covalent Kinase Drug Discovery. *Pharmaceuticals* **15**, 1322 (2022).
- [210] Zhao, Q., Ouyang, X., Wan, X., Gajiwala, K. S., Kath, J. C., Jones, L. H., Burlingame, A. L. & Taunton, J. Broad-Spectrum Kinase Profiling in Live Cells With Lysine-Targeted Sulfonyl Fluoride Probes. *Journal of the American Chemical Society* **139**, 680–685 (2017).

- [211] Tang, G., Wang, W., Zhu, C., Huang, H., Chen, P., Wang, X., Xu, M., Sun, J., Zhang, C.-J., Xiao, Q., Gao, L., Zhang, Z.-M. & Yao, S. Q. Global Reactivity Profiling of the Catalytic Lysine in Human Kinome for Covalent Inhibitor Development. *Angewandte Chemie International Edition* **63**, e202316394 (2024).
- [212] Li, G., Montgomery, J. E., Eckert, M. A., Chang, J. W., Tienda, S. M., Lengyel, E. & Moellering, R. E. An Activity-Dependent Proximity Ligation Platform for Spatially Resolved Quantification of Active Enzymes in Single Cells. *Nature Communications* **8**, 1775 (2017).
- [213] Li, G., Eckert, M. A., Chang, J. W., Montgomery, J. E., Chryplewicz, A., Lengyel, E. & Moellering, R. E. Ultrasensitive, Multiplexed Chemoproteomic Profiling With Soluble Activity-Dependent Proximity Ligation. *Proceedings of the National Academy of Sciences* **116**, 21493–21500 (2019).
- [214] Swenson, C. S., Smitha Pillai, K., Carlos, A. J. & Moellering, R. E. Spatial Chemoproteomics for Mapping the Active Proteome. *Israel Journal of Chemistry* **63**, e202200104 (2023).
- [215] Smitha Pillai, K., Laxton, O., Li, G., Lin, J., Karginova, O., Nanda, R., Olopade, O. I., Tay, S. & Moellering, R. E. Single-Cell Chemoproteomics Identifies Metastatic Activity Signatures in Breast Cancer. *Science Advances* **10**, eadp2622 (2024).
- [216] Li, G. & Moellering, R. E. Chemical Probes for Spatially Resolved Measurement of Active Enzymes in Single Cells. In *Methods in Enzymology*, vol. 628, 243–262 (Elsevier, 2019).
- [217] Christensen, J. G. A Preclinical Review of Sunitinib, a Multitargeted Receptor Tyrosine Kinase Inhibitor With Anti-Angiogenic and Antitumour Activities. *Annals of Oncology* **18**, x3–x10 (2007).
- [218] Shiotsu, Y., Kiyoi, H., Ishikawa, Y., Tanizaki, R., Shimizu, M., Umehara, H., Ishii, K., Mori, Y., Ozeki, K., Minami, Y., Abe, A., Maeda, H., Akiyama, T., Kanda, Y., Sato, Y., Akinaga, S. & Naoe, T. KW-2449, a Novel Multikinase Inhibitor, Suppresses the Growth of Leukemia Cells With FLT3 Mutations or T315I-mutated BCR/ABL Translocation. *Blood* **114**, 1607–1617 (2009).
- [219] Pemovska, T., Johnson, E., Kontro, M., Repasky, G. A., Chen, J., Wells, P., Cronin, C. N., McTigue, M., Kallioniemi, O., Porkka, K., Murray, B. W. & Wennerberg, K. Axitinib Effectively Inhibits BCR-ABL1(T315I) With a Distinct Binding Conformation. *Nature* **519**, 102–105 (2015).
- [220] van Linden, O. P. J., Kooistra, A. J., Leurs, R., de Esch, I. J. P. & de Graaf, C. KLIFS: A Knowledge-Based Structural Database To Navigate Kinase–Ligand Interaction Space. *Journal of Medicinal Chemistry* **57**, 249–277 (2014).

- [221] Gajiwala, K. S., Wu, J. C., Christensen, J., Deshmukh, G. D., Diehl, W., DiNitto, J. P., English, J. M., Greig, M. J., He, Y.-A., Jacques, S. L., Lunney, E. A., McTigue, M., Molina, D., Quenzer, T., Wells, P. A., Yu, X., Zhang, Y., Zou, A., Emmett, M. R., Marshall, A. G., Zhang, H.-M. & Demetri, G. D. KIT Kinase Mutants Show Unique Mechanisms of Drug Resistance to Imatinib and Sunitinib in Gastrointestinal Stromal Tumor Patients. *Proceedings of the National Academy of Sciences* **106**, 1542–1547 (2009).
- [222] Wang, Q., Ye, Q., Xi, X., Cao, X., Wang, X., Zhang, M., Xu, Y., Deng, T., Deng, X., Zhang, G. & Xiao, C. KW2449 Ameliorates Collagen-Induced Arthritis by Inhibiting RIPK1-dependent Necroptosis. *Frontiers in Immunology* **14** (2023).
- [223] Zhang, J., Chen, C., Chen, X., Liao, K., Li, F., Song, X., Liu, C., Su, M.-Y., Sun, H., Hou, T., Tan, C. S. H., Fang, L. & Rao, H. Linker-Free PROTACs Efficiently Induce the Degradation of Oncoproteins. *Nature Communications* **16**, 4794 (2025).
- [224] Liu, H., Ding, X., Liu, L., Mi, Q., Zhao, Q., Shao, Y., Ren, C., Chen, J., Kong, Y., Qiu, X., Elvassore, N., Yang, X., Yin, Q. & Jiang, B. Discovery of Novel BCR-ABL PROTACs Based on the Cereblon E3 Ligase Design, Synthesis, and Biological Evaluation. *European Journal of Medicinal Chemistry* **223**, 113645 (2021).
- [225] Chen, Y., Craven, G. B., Kamber, R. A., Cuesta, A., Zherish, S., Moroz, Y. S., Bassik, M. C. & Taunton, J. Direct Mapping of Ligandable Tyrosines and Lysines in Cells With Chiral Sulfonyl Fluoride Probes. *Nature Chemistry* **15**, 1616–1625 (2023).
- [226] Cortes, J., Roboz, G. J., Kantarjian, H. M., Feldman, E. J., Karp, J. E., Pratz, K. W., Rao, N. S., Akinaga, S. & Levis, M. J. A Phase I Dose Escalation Study of KW-2449, An Oral Multi-Kinase Inhibitor Against FLT3, Abl, FGFR1 and Aurora in Patients With Relapsed/Refractory AML, ALL and MDS or Resistant/Intolerant CML. *Blood* **112**, 2967 (2008).
- [227] Nguyen, T., Dai, Y., Attkisson, E., Kramer, L., Jordan, N., Nguyen, N., Kolluri, N., Muschen, M. & Grant, S. HDAC Inhibitors Potentiate the Activity of the BCR/ABL Kinase Inhibitor KW-2449 in Imatinib-Sensitive or -Resistant BCR/ABL+ Leukemia Cells In Vitro and In Vivo. *Clinical Cancer Research* **17**, 3219–3232 (2011).
- [228] Meng, Y., Gao, C., Clawson, D. K., Atwell, S., Russell, M., Vieth, M. & Roux, B. Predicting the Conformational Variability of Abl Tyrosine Kinase Using Molecular Dynamics Simulations and Markov State Models. *Journal of Chemical Theory and Computation* **14**, 2721–2732 (2018).
- [229] Gilbert, K. E., Vuorinen, A., Aatkar, A., Pogány, P., Pettinger, J., Grant, E. K., Kirkpatrick, J. M., Rittinger, K., House, D., Burley, G. A. & Bush, J. T. Profiling Sulfur(VI) Fluorides as Reactive Functionalities for Chemical Biology Tools and Expansion of the Ligandable Proteome. *ACS Chemical Biology* **18**, 285–295 (2023).

- [230] Brasher, B. B. & Van Etten, R. A. C-Abl Has High Intrinsic Tyrosine Kinase Activity That Is Stimulated by Mutation of the Src Homology 3 Domain and by Autophosphorylation at Two Distinct Regulatory Tyrosines*. *Journal of Biological Chemistry* **275**, 35631–35637 (2000).
- [231] Lamontanara, A. J., Georgeon, S., Tria, G., Svergun, D. I. & Hantschel, O. The SH2 Domain of Abl Kinases Regulates Kinase Autophosphorylation by Controlling Activation Loop Accessibility. *Nature Communications* **5**, 5470 (2014).
- [232] Nagar, B., Hantschel, O., Young, M. A., Scheffzek, K., Veach, D., Bornmann, W., Clarkson, B., Superti-Furga, G. & Kuriyan, J. Structural Basis for the Autoinhibition of C-Abl Tyrosine Kinase. *Cell* **112**, 859–871 (2003).
- [233] van de Schoot, R., Depaoli, S., King, R., Kramer, B., Märten, K., Tadesse, M. G., Vannucci, M., Gelman, A., Veen, D., Willemsen, J. & Yau, C. Bayesian Statistics and Modelling. *Nature Reviews Methods Primers* **1**, 1 (2021).
- [234] Liu, R., Yue, Z., Tsai, C.-C. & Shen, J. Assessing Lysine and Cysteine Reactivities for Designing Targeted Covalent Kinase Inhibitors. *Journal of the American Chemical Society* **141**, 6553–6560 (2019).
- [235] Xie, T., Saleh, T., Rossi, P. & Kalodimos, C. G. Conformational States Dynamically Populated by a Kinase Determine Its Function. *Science* **370**, eabc2754 (2020).
- [236] Becher, I., Savitski, M. M., Savitski, M. F., Hopf, C., Bantscheff, M. & Drewes, G. Affinity Profiling of the Cellular Kinome for the Nucleotide Cofactors ATP, ADP, and GTP. *ACS Chemical Biology* **8**, 599–607 (2013).
- [237] Jones, L. H. Reactive Chemical Probes: Beyond the Kinase Cysteinome. *Angewandte Chemie International Edition* **57**, 9220–9223 (2018).
- [238] Krishnan, S., Miller, R. M., Tian, B., Mullins, R. D., Jacobson, M. P. & Taunton, J. Design of Reversible, Cysteine-Targeted Michael Acceptors Guided by Kinetic and Computational Analysis. *Journal of the American Chemical Society* **136**, 12624–12630 (2014).
- [239] Mihalovits, L. M., Ferenczy, G. G. & Keserü, G. M. The Role of Quantum Chemistry in Covalent Inhibitor Design. *International Journal of Quantum Chemistry* **122**, e26768 (2022).
- [240] Kuzmič, P. DynaFit—A Software Package for Enzymology. In *Methods in Enzymology*, vol. 467, 247–280 (Elsevier, 2009).
- [241] Johnson, K. A. Chapter 23 Fitting Enzyme Kinetic Data With KinTek Global Kinetic Explorer. In *Methods in Enzymology*, vol. 467, 601–626 (Elsevier, 2009).

- [242] Langrish, C. L., Bradshaw, J. M., Francesco, M. R., Owens, T. D., Xing, Y., Shu, J., LaStant, J., Bisconte, A., Outerbridge, C., White, S. D., Hill, R. J., Brameld, K. A., Goldstein, D. M. & Nunn, P. A. Preclinical Efficacy and Anti-Inflammatory Mechanisms of Action of the Bruton Tyrosine Kinase Inhibitor Rilzabrutinib for Immune-Mediated Disease. *The Journal of Immunology* **206**, 1454–1468 (2021).
- [243] Zhao, Z. & Bourne, P. E. Advances in Reversible Covalent Kinase Inhibitors. *Medicinal Research Reviews* **45**, 629–653 (2025).
- [244] Copeland, R. A. The Drug–Target Residence Time Model: A 10-Year Retrospective. *Nature Reviews Drug Discovery* **15**, 87–95 (2016).
- [245] Patel, D., Huma, Z. E. & Duncan, D. Reversible Covalent Inhibition-Desired Covalent Adduct Formation by Mass Action. *ACS Chemical Biology* **19**, 824–838 (2024).
- [246] Krenske, E. H., Petter, R. C. & Houk, K. N. Kinetics and Thermodynamics of Reversible Thiol Additions to Mono- and Deactivated Michael Acceptors: Implications for the Design of Drugs That Bind Covalently to Cysteines. *The Journal of Organic Chemistry* **81**, 11726–11733 (2016).
- [247] Sali, A. & Blundell, T. L. Comparative Protein Modelling by Satisfaction of Spatial Restraints. *Journal of Molecular Biology* **234**, 779–815 (1993).
- [248] Krivov, G. G., Shapovalov, M. V. & Dunbrack Jr., R. L. Improved Prediction of Protein Side-Chain Conformations With SCWRL4. *Proteins: Structure, Function, and Bioinformatics* **77**, 778–795 (2009).
- [249] Schrödinger, L. L. C. The PyMOL Molecular Graphics System, Version 1.8 (2015).
- [250] Huang, J., Rauscher, S., Nawrocki, G., Ran, T., Feig, M., de Groot, B. L., Grubmüller, H. & MacKerell, A. D. CHARMM36m: An Improved Force Field for Folded and Intrinsically Disordered Proteins. *Nature Methods* **14**, 71–73 (2017).
- [251] Frisch, M. J., Trucks, G. W., Schlegel, H. B., Scuseria, G. E., Robb, M. A., Cheeseman, J. R., Scalmani, G., Barone, V., Petersson, G. A., Nakatsuji, H., Li, X., Caricato, M., Marenich, A. V., Bloino, J., Janesko, B. G., Gomperts, R., Mennucci, B., Hratchian, H. P., Ortiz, J. V., Izmaylov, A. F., Sonnenberg, J. L., Williams-Young, D., Ding, F., Lipparini, F., Egidi, F., Goings, J., Peng, B., Petrone, A., Henderson, T., Ranasinghe, D., Zakrzewski, V. G., Gao, J., Rega, N., Zheng, G., Liang, W., Hada, M., Ehara, M., Toyota, K., Fukuda, R., Hasegawa, J., Ishida, M., Nakajima, T., Honda, Y., Kitao, O., Nakai, H., Vreven, T., Throssell, K., Montgomery, J. A., Jr., Peralta, J. E., Ogliaro, F., Bearpark, M. J., Heyd, J. J., Brothers, E. N., Kudin, K. N., Staroverov, V. N., Keith, T. A., Kobayashi, R., Normand, J., Raghavachari, K., Rendell, A. P., Burant, J. C., Iyengar, S. S., Tomasi, J., Cossi, M., Millam, J. M., Klene, M., Adamo, C., Cammi, R., Ochterski, J. W., Martin, R. L., Morokuma, K., Farkas, O., Foresman, J. B. & Fox, D. J. Gaussian~16 Revision C.01 (2016). Gaussian Inc. Wallingford CT.

- [252] Shirts, M. R., Klein, C., Swails, J. M., Yin, J., Gilson, M. K., Mobley, D. L., Case, D. A. & Zhong, E. D. Lessons Learned From Comparing Molecular Dynamics Engines on the SAMPL5 Dataset. *Journal of Computer-Aided Molecular Design* **31**, 147–161 (2017).
- [253] Case, D. A., Cerutti, D. S., Cruzeiro, V. W. D., Darden, T. A., Duke, R. E., Ghazimirsaeed, M., Giambasu, G. M., Giese, T. J., Götz, A. W., Harris, J. A., Kasavajhala, K., Lee, T.-S., Li, Z., Lin, C., Liu, J., Miao, Y., Salomon-Ferrrer, R., Shen, J., Snyder, R., Swails, J., Walker, R. C., Wang, J., Wu, X., Zeng, J., Cheatham III, T. E., Roe, D. R., Roitberg, A., Simmerling, C., York, D. M., Nagan, M. C. & Merz, K. M. J. Recent Developments in Amber Biomolecular Simulations. *Journal of Chemical Information and Modeling* **65**, 7835–7843 (2025).
- [254] Case, D. A., Aktulga, H. M., Belfon, K., Cerutti, D. S., Cisneros, G. A., Cruzeiro, V. W. D., Forouzeshe, N., Giese, T. J., Götz, A. W., Gohlke, H., Izadi, S., Kasavajhala, K., Kaymak, M. C., King, E., Kurtzman, T., Lee, T.-S., Li, P., Liu, J., Luchko, T., Luo, R., Manathunga, M., Machado, M. R., Nguyen, H. M., O’Hearn, K. A., Onufriev, A. V., Pan, F., Pantano, S., Qi, R., Rahnamoun, A., Risheh, A., Schott-Verdugo, S., Shajan, A., Swails, J., Wang, J., Wei, H., Wu, X., Wu, Y., Zhang, S., Zhao, S., Zhu, Q., Cheatham, T. E. I. I. I., Roe, D. R., Roitberg, A., Simmerling, C., York, D. M., Nagan, M. C. & Merz, K. M. J. AmberTools. *Journal of Chemical Information and Modeling* **63**, 6183–6191 (2023).
- [255] Ryckaert, J. P., Ciccotti, G. & Berendsen, H. J. C. Numerical Integration of the Cartesian Equation of Motions of a System With Constraints: Molecular Dynamics of *n*-Alkanes. *Journal of Computational Physics* **23**, 327–341 (1977).
- [256] Stewart, J. J. P. Optimization of Parameters for Semiempirical Methods V: Modification of NDDO Approximations and Application to 70 Elements. *Journal of Molecular Modeling* **13**, 1173–1213 (2007).
- [257] Bonomi, M., Bussi, G., Camilloni, C., Tribello, G. A., Banáš, P., Barducci, A., Bernetti, M., Bolhuis, P. G., Bottaro, S., Branduardi, D., Capelli, R., Carloni, P., Ceriotti, M., Cesari, A., Chen, H., Chen, W., Colizzi, F., De, S., De La Pierre, M., Donadio, D., Drobot, V., Ensing, B., Ferguson, A. L., Filizola, M., Fraser, J. S., Fu, H., Gasparotto, P., Gervasio, F. L., Giberti, F., Gil-Ley, A., Giorgino, T., Heller, G. T., Hocky, G. M., Iannuzzi, M., Invernizzi, M., Jelfs, K. E., Jussupow, A., Kirilin, E., Laio, A., Limongelli, V., Lindorff-Larsen, K., Löhner, T., Marinelli, F., Martin-Samos, L., Masetti, M., Meyer, R., Michaelides, A., Molteni, C., Morishita, T., Nava, M., Paissoni, C., Papaleo, E., Parrinello, M., Pfaendtner, J., Piaggi, P., Piccini, G., Pietropaolo, A., Pietrucci, F., Pipolo, S., Provati, D., Quigley, D., Raiteri, P., Raniolo, S., Rydzewski, J., Salvalaglio, M., Sosso, G. C., Spiwok, V., Šponer, J., Swenson, D. W. H., Tiwary, P., Valsson, O., Vendruscolo, M., Voth, G. A., White, A. & The PLUMED consortium. Promoting Transparency and Reproducibility in Enhanced Molecular Simulations. *Nature Methods* **16**, 670–673 (2019).

- [258] Hummer, G. & Szabo, A. Free Energy Reconstruction From Nonequilibrium Single-Molecule Pulling Experiments. *Proceedings of the National Academy of Sciences* **98**, 3658–3661 (2001).
- [259] Fu, H., Chen, H., Blazhynska, M., Goulard Coderc De Lacam, E., Szczepaniak, F., Pavlova, A., Shao, X., Gumbart, J. C., Dehez, F., Roux, B., Cai, W. & Chipot, C. Accurate Determination of Protein:Ligand Standard Binding Free Energies From Molecular Dynamics Simulations. *Nature Protocols* **17**, 1114–1141 (2022).
- [260] Zwanzig, R. W. High Temperature Equation of State by a Perturbation Method. *The Journal of Chemical Physics* **22**, 1420–1426 (1954).
- [261] Bennett, C. H. Efficient Estimation of Free-Energy Differences From Monte-Carlo Data. *Journal of Computational Physics* **22**, 245–268 (1976).
- [262] Kirkwood, J. G. Statistical Mechanics of Fluid Mixtures. *The Journal of Chemical Physics* **3**, 300–313 (1935).
- [263] Yu, Y. B., Privalov, P. L. & Hodges, R. S. Contribution of Translational and Rotational Motions to Molecular Association in Aqueous Solution. *Biophysical Journal* **81**, 1632–1642 (2001).
- [264] Newville, M., Stensitzki, T., Allen, D. B. & Ingargiola, A. LMFIT: Non-Linear Least-Square Minimization and Curve-Fitting for Python. Zenodo (2014).
- [265] Marquardt, D. W. An Algorithm for Least-Squares Estimation of Nonlinear Parameters. *Journal of the Society for Industrial and Applied Mathematics* **11**, 431–441 (1963).
- [266] Georgi, V., Schiele, F., Berger, B.-T., Steffen, A., Marin Zapata, P. A., Briem, H., Menz, S., Preusse, C., Vasta, J. D., Robers, M. B., Brands, M., Knapp, S. & Fernández-Montalván, A. Binding Kinetics Survey of the Drugged Kinome. *Journal of the American Chemical Society* **140**, 15774–15782 (2018).
- [267] Roux, B. *Computational Modeling and Simulations of Biomolecular Systems* (World Scientific Publishing, Singapore, 2021).
- [268] Kříž, K. & Řezáč, J. Benchmarking of Semiempirical Quantum-Mechanical Methods on Systems Relevant to Computer-Aided Drug Design. *Journal of Chemical Information and Modeling* **60**, 1453–1460 (2020).
- [269] McInnes, L., Healy, J. & Astels, S. Hdbscan: Hierarchical Density Based Clustering. *The Journal of Open Source Software* **2**, 205 (2017).
- [270] Schröder, M., Tan, L., Wang, J., Liang, Y., Gray, N. S., Knapp, S. & Chaikuad, A. Catalytic Domain Plasticity of MKK7 Reveals Structural Mechanisms of Allosteric Activation and Diverse Targeting Opportunities. *Cell Chem Biol* **27**, 1285–1295 (2020).

- [271] Liclican, A., Serafini, L., Xing, W., Czerwieniec, G., Steiner, B., Wang, T., Brendza, K. M., Lutz, J. D., Keegan, K. S., Ray, A. S. *et al.* Biochemical Characterization of Tirabrutinib and Other Irreversible Inhibitors of Bruton’s Tyrosine Kinase Reveals Differences in on-and Off-Target Inhibition. *Biochimica et Biophysica Acta (BBA) - General Subjects* **1864**, 129531 (2020).
- [272] Copeland, R. A., Pompliano, D. L. & Meek, T. D. Drug–Target Residence Time and Its Implications for Lead Optimization. *Nature Reviews Drug Discovery* **5**, 730–739 (2006).
- [273] E, W. & Vanden-Eijnden, E. Towards a Theory of Transition Paths. *Journal of Statistical Physics* **123**, 503–523 (2006).
- [274] Bolhuis, P. G., Dellago, C., Chandler, D. & Geissler, P. Transition Path Sampling: Throwing Ropes Over Mountain Passes, in the Dark. *Ann. Rev. of Phys. Chem.* **53**, 291–318 (2002).
- [275] Prinz, J.-H., Wu, H., Sarich, M., Keller, B., Senne, M., Held, M., Chodera, J. D., Schuette, C. & Noe, F. Markov Models of Molecular Kinetics: Generation and Validation. *The Journal of Chemical Physics* **134** (2011).
- [276] Noé, F., Schütte, C., Vanden-Eijnden, E., Reich, L. & Weikl, T. R. Constructing the Equilibrium Ensemble of Folding Pathways From Short Off-Equilibrium Simulations. *Proceedings of the National Academy of Sciences* **106**, 19011–19016 (2009).
- [277] Omiecinski, C. J., Vanden Heuvel, J. P., Perdew, G. H. & Peters, J. M. Xenobiotic Metabolism, Disposition, and Regulation by Receptors: From Biochemical Phenomenon to Predictors of Major Toxicities. *Toxicological Sciences* **120**, S49–S75 (2011).
- [278] Scheers, E., Leclercq, L., de Jong, J., Bode, N., Bockx, M., Laenen, A., Cuyckens, F., Skee, D., Murphy, J., Sukbuntherng, J. & Mannens, G. Absorption, Metabolism, and Excretion of Oral ¹⁴C Radiolabeled Ibrutinib: An Open-Label, Phase I, Single-Dose Study in Healthy Men. *Drug Metabolism and Disposition* **43**, 289–297 (2015).
- [279] Teorell, T. Kinetics of Distribution of Substances Administered to the Body. I. The Extravascular Modes of Administration. *Archives Internationales de Pharmacodynamie et de Thérapie* **57**, 205–225 (1937).
- [280] Teorell, T. Kinetics of Distribution of Substances Administered to the Body. II. The Intravascular Modes of Administration. *Archives Internationales de Pharmacodynamie et de Thérapie* **57**, 226–240 (1937).
- [281] Turing, A. M. The Chemical Basis of Morphogenesis. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences* **237**, 37–72 (1952).
- [282] Kopelman, R. Fractal Reaction Kinetics. *Science* **241**, 1620–1626 (1988).

- [283] Tkačik, G. & ten Wolde, P. R. Information Processing in Biochemical Networks. *Annual Review of Biophysics* **54**, 249–274 (2025).
- [284] Gillespie, D. T. General Method for Numerically Simulating Stochastic Time Evolution of Coupled Chemical-Reactions. *Journal of Computational Physics* **22**, 403–434 (1976).
- [285] Gillespie, D. T. The Chemical Langevin Equation. *The Journal of Chemical Physics* **113**, 297–306 (2000).
- [286] Glass, D. S., Jin, X. & Riedel-Kruse, I. H. Nonlinear Delay Differential Equations and Their Application to Modeling Biological Network Motifs. *Nature Communications* **12**, 1788 (2021).
- [287] Thiele, I., Jamshidi, N., Fleming, R. M. T. & Palsson, B. Ø. Genome-Scale Reconstruction of Escherichia Coli’s Transcriptional and Translational Machinery: A Knowledge Base, Its Mathematical Formulation, and Its Functional Characterization. *PLoS Computational Biology* **5**, e1000312 (2009).
- [288] Karr, J. R., Sanghvi, J. C., Macklin, D. N., Gutschow, M. V., Jacobs, J. M., Bolival, B., Assad-Garcia, N., Glass, J. I. & Covert, M. W. A Whole-Cell Computational Model Predicts Phenotype From Genotype. *Cell* **150**, 389–401 (2012).
- [289] Gutenkunst, R. N., Waterfall, J. J., Casey, F. P., Brown, K. S., Myers, C. R. & Sethna, J. P. Universally Sloppy Parameter Sensitivities in Systems Biology Models. *PLoS Computational Biology* **3**, e189 (2007).
- [290] Wittig, U., Rey, M., Weidemann, A., Kania, R. & Müller, W. SABIO-RK: An Updated Resource for Manually Curated Biochemical Reaction Kinetics. *Nucleic Acids Research* **46**, D656–D660 (2018).
- [291] Malik-Sheriff, R. S., Glont, M., Nguyen, T. V. N., Tiwari, K., Roberts, M. G., Xavier, A., Vu, M. T., Men, J., Maire, M., Kananathan, S., Fairbanks, E. L., Meyer, J. P., Arankalle, C., Varusai, T. M., Knight-Schrijver, V., Li, L., Dueñas-Roca, C., Dass, G., Keating, S. M., Park, Y. M., Buso, N., Rodriguez, N., Hucka, M. & Hermjakob, H. BioModels—15 Years of Sharing Computational Models in Life Science. *Nucleic Acids Research* **48**, D407–D415 (2020).
- [292] Seeliger, M. A., Young, M., Henderson, M. N., Pellicena, P., King, D. S., Falick, A. M. & Kuriyan, J. High Yield Bacterial Expression of Active C-Abl and c-Src Tyrosine Kinases. *Protein Science* **14**, 3135–3139 (2005).
- [293] Speltz, T. E., Qiao, Z., Swenson, C. S., Shangguan, X., Coukos, J. S., Lee, C. W., Thomas, D. M., Santana, J., Fanning, S. W., Greene, G. L. & Moellering, R. E. Targeting MYC With Modular Synthetic Transcriptional Repressors Derived From bHLH DNA-binding Domains. *Nature Biotechnology* **41**, 541–551 (2023).

- [294] McGibbon, R. T., Beauchamp, K. A., Harrigan, M. P., Klein, C., Swails, J. M., Hernández, C. X., Schwantes, C. R., Wang, L.-P., Lane, T. J. & Pande, V. S. MDTraj: A Modern Open Library for the Analysis of Molecular Dynamics Trajectories. *Biophysical Journal* **109**, 1528–1532 (2015).
- [295] Hunter, J. D. Matplotlib: A 2D Graphics Environment. *Computing in Science & Engineering* **9**, 90–95 (2007).
- [296] Koes, D. R., Baumgartner, M. P. & Camacho, C. J. Lessons Learned in Empirical Scoring With Smina From the CSAR 2011 Benchmarking Exercise. *Journal of Chemical Information and Modeling* **53**, 1893–1904 (2013).
- [297] Morris, G. M., Huey, R., Lindstrom, W., Sanner, M. F., Belew, R. K., Goodsell, D. S. & Olson, A. J. AutoDock4 and AutoDockTools4: Automated Docking With Selective Receptor Flexibility. *Journal of Computational Chemistry* **30**, 2785–2791 (2009).
- [298] Phillips, J. C., Braun, R., Wang, W., Gumbart, J., Tajkhorshid, E., Villa, E., Chipot, C., Skeel, R. D., Kalé, L. & Schulten, K. Scalable Molecular Dynamics With NAMD. *Journal of Computational Chemistry* **26**, 1781–1802 (2005).
- [299] Crowley, M. F., Williamson, M. J. & Walker, R. C. CHAMBER: Comprehensive Support for CHARMM Force Fields Within the AMBER Software. *International Journal of Quantum Chemistry* **109**, 3767–3772 (2009).
- [300] Yu, W., He, X., Vanommeslaeghe, K. & MacKerell Jr., A. D. Extension of the CHARMM General Force Field to Sulfonyl-Containing Compounds and Its Utility in Biomolecular Simulations. *Journal of Computational Chemistry* **33**, 2451–2468 (2012).
- [301] Kumar, A., Yoluk, O. & MacKerell Jr., A. D. FFParm: Standalone Package for CHARMM Additive and Drude Polarizable Force Field Parametrization of Small Molecules. *Journal of Computational Chemistry* **41**, 958–970 (2020).
- [302] Eastman, P., Swails, J., Chodera, J. D., McGibbon, R. T., Zhao, Y., Beauchamp, K. A., Wang, L.-P., Simmonett, A. C., Harrigan, M. P., Stern, C. D., Wiewiora, R. P., Brooks, B. R. & Pande, V. S. OpenMM 7: Rapid Development of High Performance Algorithms for Molecular Dynamics. *PLoS Computational Biology* **13**, e1005659 (2017).
- [303] Branch, M. A., Coleman, T. F. & Li, Y. A Subspace, Interior, and Conjugate Gradient Method for Large-Scale Bound-Constrained Minimization Problems. *SIAM Journal on Scientific Computing* **21**, 1–23 (1999).
- [304] Kumar, R., Carroll, C., Hartikainen, A. & Martin, O. ArviZ a Unified Library for Exploratory Analysis of Bayesian Models in Python. *The Journal of Open Source Software* **4**, 1143 (2019).
- [305] Black, J., Leff, P., Shankley, N. & Wood, J. An Operational Model of Pharmacological Agonism: The Effect of E/[A] Curve Shape on Agonist Dissociation Constant Estimation. *British Journal of Pharmacology* **160** (2010).

- [306] Chen, P., Sun, J., Zhu, C., Tang, G., Wang, W., Xu, M., Xiang, M., Zhang, C.-J., Zhang, Z.-M., Gao, L. & Yao, S. Q. Cell-Active, Reversible, and Irreversible Covalent Inhibitors That Selectively Target the Catalytic Lysine of BCR-ABL Kinase. *Angewandte Chemie International Edition* **61**, e202203878 (2022).
- [307] Gurbani, D., Du, G., Henning, N. J., Rao, S., Bera, A. K., Zhang, T., Gray, N. S. & Westover, K. D. Structure and Characterization of a Covalent Inhibitor of SRC Kinase. *Frontiers in Molecular Biosciences* **7**, 81 (2020).
- [308] Solca, F., Dahl, G., Zoephel, A., Bader, G., Sanderson, M., Klein, C., Kraemer, O., Himmelsbach, F., Haaksma, E. & Adolf, G. R. Target Binding Properties and Cellular Activity of Afatinib (BIBW 2992), an Irreversible ErbB Family Blocker. *The Journal of Pharmacology and Experimental Therapeutics* **343**, 342–350 (2012).
- [309] Gajiwala, K. S., Feng, J., Ferre, R., Ryan, K., Brodsky, O., Weinrich, S., Kath, J. C. & Stewart, A. Insights Into the Aberrant Activity of Mutant EGFR Kinase Domain and Drug Recognition. *Structure* **21**, 209–219 (2013).
- [310] Wang, A., Yan, X. E., Wu, H., Wang, W., Hu, C., Chen, C., Zhao, Z., Zhao, P., Li, X., Wang, L., Wang, B., Ye, Z., Wang, J., Wang, C., Zhang, W., Gray, N. S., Weisberg, E. L., Chen, L., Liu, J., Yun, C. H. & Liu, Q. Ibrutinib Targets Mutant-EGFR Kinase With a Distinct Binding Conformation. *Oncotarget* **7**, 69760–69769 (2016).
- [311] Yan, X. E., Ayaz, P., Zhu, S. J., Zhao, P., Liang, L., Zhang, C. H., Wu, Y. C., Li, J. L., Choi, H. G., Huang, X., Shan, Y., Shaw, D. E. & Yun, C. H. Structural Basis of AZD9291 Selectivity for EGFR T790M. *Journal of Medicinal Chemistry* **63**, 8502–8511 (2020).
- [312] Kung, A., Chen, Y. C., Schimpl, M., Ni, F., Zhu, J., Turner, M., Molina, H., Overman, R. & Zhang, C. Development of Specific, Irreversible Inhibitors for a Receptor Tyrosine Kinase EphB3. *Journal of the American Chemical Society* **138**, 10554–10560 (2016).
- [313] Telliez, J. B., Dowty, M. E., Wang, L., Jussif, J., Lin, T., Li, L., Moy, E., Balbo, P., Li, W., Zhao, Y., Crouse, K., Dickinson, C., Symanowicz, P., Hegen, M., Banker, M. E., Vincent, F., Unwalla, R., Liang, S., Gilbert, A. M., Brown, M. F., Hayward, M., Montgomery, J., Yang, X., Bauman, J., Trujillo, J. I., Casimiro-Garcia, A., Vajdos, F. F., Leung, L., Geoghegan, K. F., Quazi, A., Xuan, D., Jones, L., Hett, E., Wright, K., Clark, J. D. & Thorarensen, A. Discovery of a JAK3-Selective Inhibitor: Functional Differentiation of JAK3-Selective Inhibition Over Pan-JAK or JAK1-Selective Inhibition. *ACS Chemical Biology* **11**, 3442–3451 (2016).
- [314] Wu, J., Powell, F., Larsen, N. A., Lai, Z., Byth, K. F., Read, J., Gu, R. F., Roth, M., Toader, D., Saeh, J. C. & Chen, H. Mechanism and In Vitro Pharmacology of TAK1 Inhibition by (5z)-7-Oxozeaenol. *ACS Chemical Biology* **8**, 643–650 (2013).
- [315] Yang, T., Cuesta, A., Wan, X., Craven, G. B., Hirakawa, B., Khamphavong, P., May, J. R., Kath, J. C., Lapek, J. D., Niessen, S., Burlingame, A. L., Carelli, J. D. &

Taunton, J. Reversible Lysine-Targeted Probes Reveal Residence Time-Based Kinase Selectivity. *Nature Chemical Biology* **18**, 934–941 (2022).